Orthogonal Policy Learning Under Ambiguity*

Riccardo D'Adamo†

November 23, 2022

Job Market Paper

[Click here for the latest version]

Abstract

This paper studies the problem of estimating individualized treatment rules when treatment effects are partially identified, as it is often the case with observational data. By drawing connections between the treatment assignment problem and classical decision theory, we characterize several notions of optimal treatment policies in the presence of partial identification. The proposed framework allows to incorporate user-defined constraints on the policies, such as restrictions for transparency or interpretability, while also ensuring computational feasibility. We show that partial identification leads to a novel statistical learning problem with risk directionally – but not fully – differentiable with respect to an infinite-dimensional nuisance component. We propose an estimation procedure that ensures Neyman-orthogonality with respect to the nuisance component and provide statistical guarantees that depend on the amount of concentration around the points of non-differentiability in the data-generating process. The proposed method is illustrated using data from the Job Partnership Training Act study.

*I am grateful to my advisors Martin Weidner and Toru Kitagawa for their help and support throughout my graduate studies. I would also like to thank Timothy Christensen, Sukjin Han, Whitney Newey, Tomasz Olma, Joris Pinkse, Andrei Zeleneev, as well as participants to the NeurIPS EconML Workshop, IAAE 2022 Conference, Bristol Econometric Study Group for helpful comments and discussions.

†Department of Economics, University College London. E-mail: uctpdad@ucl.ac.uk

1 Introduction

The problem of choosing an optimal treatment assignment based on data is ubiquitous in economics and other fields, including medicine and marketing. Individuals often display heterogeneous responses to the same treatment. Decision-makers in policy and industry are therefore interested in leveraging the growing availability of rich granular data to tailor treatment assignment to individuals based on their characteristics. As a result, a fast-growing literature has emerged focused on developing procedures for estimation of individualized treatment rules. While a variety of approaches have been recently established, these typically assume that the available data allow to provide credible point estimates for the effect of the treatment, that is treatment effects are point identified. While of important stylized value, this assumption is often hard to justify in many empirical settings. For example, economists have long been aware that popular quasi-experimental and observational research designs, such as instrumental variables (IV), allow to point identify treatment effects only for specific sub-populations (Imbens and Angrist, 1994). Even in randomized control trials, point identification of the treatment effects is often precluded due to non-random attrition, e.g. when participants dropout from a program or the researcher is denied information on the outcome variable (Lee, 2009). In such settings, the data may only provide partial knowledge about the treatment response in the form of credible bounds, i.e. the treatment effects are *partially identified*. As a result, the decision-maker may have ambiguous evidence on whether a candidate policy should be preferred to another, so that only a partial ordering of policies can be deduced in general. While informative from a scientific perspective, a partial ordering of policies is unsatisfying when the ultimate goal of the analysis is to select a single policy to be implemented in the real world. In this scenario, a decision-maker has to confront two sources of ambiguity. The first source concerns ambiguous knowledge of the treatment response τ conditional on knowledge of distribution of the data P, due to partial identification. The second source is the lack of knowledge of the distribution P, which must be estimated from the data.

In this paper, we develop methods to handle both sources of ambiguity within the framework of "empirical welfare maximization" (Kitagawa and Tetenov, 2018), also referred to as "policy learning" (Athey and Wager, 2021). This approach considers treatment policies that are exogenously constrained to have low complexity in terms of Vapnik-Chervonenkis (VC) dimension. This encompasses many practical settings of interest, as policies often have to satisfy requirements imposed for institutional or practical reasons, such as fairness, budget or interpretability. The empirical welfare maximization (EWM) method selects the optimal policy as the maximizer of the empirical analogue of the population welfare, formulated as the average of the individual outcomes in the target population. The EWM estimation procedure has the convenient structure of an empirical risk minimization problem, which is exploited by Kitagawa and Tetenov (2018) and Athey and Wager (2021) to study its statistical properties.

We extend the EWM framework to settings with partial identification by making several contributions. First, we study the problem of assigning treatment under partial identification at the population level (i.e. where the distribution of data P is known) from a general perspective. In particular, we show how classic optimality criteria for decision under ambiguity, such as minimax risk and minimax regret, can be applied in the context of welfare maximization. Our unified framework accommodates different attitudes towards ambiguity and a wide range of popular identification assumptions, including Manski (1990) and Manski and Pepper (2000) bounds. Our analysis delivers several notions of optimal treatment policies, which we refer to as *ambiguity-robust*: they are "robust" in the sense that each of them delivers a notion of single optimal policy in the presence of partial identification, while they all reduce to the same optimal treatment assignment in the special case of point-identification. As part of this analysis, we establish general conditions on the identification sets under which the treatment assignment problem can be expressed in a simplified form, leading to computationally tractable sample analogues. In particular, we show that all ambiguity-robust policies can be represented as maximizers of a "surrogate" welfare, in which identification bounds are combined to form a proxy for the partially identified CATE. The surrogate welfare depends on several nuisance components, and its specific form is determined by the identification assumptions and attitude towards ambiguity held by the decision-maker.

We then propose an algorithm for computing the estimated ambiguity-robust policy and provide statistical guarantees on its performance in terms of the regret convergence of the surrogate welfare. Similarly to Athey and Wager (2021) and Foster and Syrgkanis (2019), our procedure leverages insights from the literature on double/de-biased machine learning (Chernozhukov et al., 2022) by making use of Neyman-orthogonalized estimates of the surrogate welfare. This, coupled with sample-splitting, allows us to guarantee rates of convergence for the estimated ambiguity-robust policy while imposing minimal requirements on the estimation of the nuisance components. One unique feature of the partially identified setting studied in this paper is the restricted degree of smoothness enjoyed by the welfare criterion. In particular, we show that popular choices of identification assumptions and optimality criteria for choice under ambiguity lead to surrogate welfare criteria that are only *directionally differentiable* with respect to the data-generating process. We highlight the importance of this feature for the problem at hand and develop new theoretical results showing how the extent of non-differentiability in the data-generating process affects the statistical properties of the learning procedure. To the best of our knowledge, we are the first to investigate the role of non-differentiabilities in the context of semiparametric statistical learning problems. Our results are therefore of independent interest and may be relevant beyond the treatment assignment problem of this paper.

The results of this paper contribute to the recent literature on EWM methods, e.g. Kitagawa and Tetenov (2018), Athey and Wager (2021), Mbakop and Tabord-Meehan (2021), and more broadly to the literature studying statistical treatment choice (see Hirano and Porter, 2020, and references therein). Kitagawa and Tetenov (2018) introduced the EWM method and provided theoretical results showing its optimality when implemented with experimental data. Athey and Wager (2021) leverage insights from the recent literature on orthogonal machine learning (Chernozhukov et al., 2022) and propose doubly-robust estimation of the treatment effect which leads to optimal learning rates even with observational data. We build on their work by adopting Neyman-orthogonal estimates while we relax the fundamental assumption that treatment effects are point identified. Cui and Tchetgen (2021) also develop procedures for learning optimal treatments rules with instrumental variables but consider unconstrained policy classes. Similarly to Athey and Wager (2021), they ensure point-identification of treatment response by restricting their analysis to the effect on compliers.

Kasy (2016), Han (2019) and Byambadalai (2022) provide methods for comparing policies in the presence of covariates and partially identified treatment effects. The focus of their work is on characterizing the partial ordering of policies in terms of their associated welfare rather than resolving the ambiguity and estimating an optimal treatment rule.

In a series of papers, Manski (2009, 2010, 2011) studies the problem of a social planner who must choose treatment for a population under partial knowledge of the treatment response in the absence of covariates. He shows that when the sign of the treatment effect is ambiguous, the minimax regret criterion leads to policies that randomize treatment in the population. While our study of the population problem is inspired by Manski's work in this area, the focus of our paper is on deterministic rules assigning individualized treatment, i.e. based on (potentially continuous) covariates. Stoye (2012), Ishihara and Kitagawa (2021) and Yata (2021) consider treatment assignment under partial identification from a finite-sample minimax perspective, while Christensen et al. (2022) adopt a local-asymptotic approach. However, these works do not consider individualization of the treatment assignment.

More closely related to our work is Kallus and Zhou (2018), who extend the EWM framework to learn an optimal policy in the presence of partially identified treatment effects under violations of unconfoundedness. In particular, they target welfare improvement with respect to a baseline pre-existing policy and consider partial-identification of the welfare criterion through Rosenbaum's sensitivity model (Rosenbaum, 1987). Adjaho and Christensen (2022) and Kido (2022) examine policies with minimax welfare guarantees when the target population lies in a Wasserstein neighborhood of the experimental population. The identification assumptions (and associated estimation procedures) considered in these papers are distinct and do not nest those covered by our framework. As a result, our contributions are complementary to these works. In independent work, Pu and Zhang (2021) study policy learning under ambiguity from a classification perspective and derive an optimal policy which coincides with one notion of ambiguity-robust policy studied in this paper. However, our estimation procedure crucially differs from theirs for the use of Neyman-orthogonalization which, combined with a refined proof-strategy that accounts for the lack of full-differentiability in the welfare criterion, allows us to guarantee considerably faster rates of convergence. In this sense, our results extend and improve on those in Pu and Zhang (2021).

Finally, we contribute to a body of literature dealing with estimation and inference for directionally-differentiable functionals. Hirano and Porter (2012) show that if a target estimand is not differentiable in the parameters of the data distributions, then no asymptotically unbiased or regular estimator exists. Ponomarev (2022) studies efficient estimation of directionally differentiable functionals from a local minimax perspective. Fang and Santos (2018) and Kitagawa et al. (2020) provide inference results for directionally differentiable functions from a frequentist and Bayesian perspective, respectively. Also motivated by partial identification, Christensen et al. (2022) consider estimation of treatment rules when the welfare criterion is only directionally differentiable with respect to a finite-dimensional nuisance component. Our framework instead involves infinite-dimensional nuisance components and therefore our analysis must account for the lack of differentiability with novel theoretical results that substantially differ from Christensen et al. (2022).

The rest of this paper is organized as follows. Section 2 introduces the setup. Section 3 presents several notions of ambiguity-robust optimal policy. Section 4 presents the proposed estimation procedure for the ambiguity-robust optimal policy. Section 5 provides statistical guarantees for the estimated optimal policy. Section 6 presents an empirical illustration based on the Job Training Partnership Act Study. Section 7 concludes the paper. Proofs and extensions are given in the Appendix.

Notation. Throughout the paper, for $d \in \mathbb{N}$, let $\mathbb{R}^d$ denote the Euclidean space, with $\|\cdot\|_p$ and $\langle \cdot, \cdot \rangle$ being the usual ℓ_p-norm and inner product, respectively. For two vectors $x \in \mathbb{R}^p$ and $y \in \mathbb{R}^q$, $x \subset y$ means that x is a sub-vector of y. For a symmetric matrix A, $\lambda_{\max}(A)$ denotes its largest eigenvalue. Unless otherwise stated, the expectation $\mathbb{E}[\cdot]$, probability $\mathbb{P}(\cdot)$, and variance $\mathrm{Var}(\cdot)$ operators will be taken with respect to the underlying distribution of observables P. Given a random variable $Z \in \mathcal{Z}$ with $\mathcal{Z} \subseteq \mathbb{R}^d$, the associated probability measure P_Z, and a function $f : \mathcal{Z} \to \mathcal{W}$ with $\mathcal{W} \subseteq \mathbb{R}^q$, we define $\|f\|_{Lp(P_Z)} = \left(\mathbb{E}_{P_Z}\left[\|f(Z)\|_p^p\right]\right)^{1/p}$ for $p \in (0\, \infty)$. We extend this definition to $p = \infty$ in the natural way. For a sequence of real numbers x_n and y_n, $x_n = o(y_n)$ and $x_n = O(y_n)$ mean, respectively, that $x_n/y_n \to 0$ and $x_n \leq C y_n$ for some constant C as $n \to \infty$. For real numbers a, b, $a \lesssim b$ means that there exists a constant C such that $a \leq Cb$. For a positive real number a, $\lfloor a \rfloor$ denotes its nearest smallest integer. The notation $\to_p$ denotes convergence in probability.

2 Setup

Let $Y_i \in \mathbb{R}$ be an outcome measuring utility, $D_i \in \{0,1\}$ a binary treatment, $X_i \in \mathcal{X} \subseteq \mathbb{R}^{k_x}$ a set of pre-treatment covariates for an individual i from an i.i.d. population of interest. We use standard notation to define the potential outcomes $Y_i(0), Y_i(1)$. The conditional average treatment effect (CATE) $\tau : \mathcal{X} \to \mathbb{R}$ is then defined as

$$\tau(x) = y_1(x) - y_0(x), \quad y_d(x) = \mathbb{E}[Y_i(d)|X_i = x],\, d = 0, 1,$$

where the expectation is taken with respect to the distribution of the population, and we will henceforth suppress the i-subscript for convenience. The decision-maker (DM) is interested in choosing a deterministic treatment assignment rule (or policy) $\pi : \mathcal{X} \to \{0,1\}$, which maps from the space of individual covariates to the binary decision "treat" ($\pi(x) = 1$) or "do not treat" ($\pi(x) = 0$). Following Manski (2004), we define the utilitarian social welfare associated with a policy π and a given configuration of the expected potential outcomes $y_0(\cdot), y_1(\cdot)$ as

$$\begin{aligned} W_{y_0,y_1}(\pi) &= \mathbb{E}_{P_X}\left[y_1(X) \cdot \pi(X) + y_0(X) \cdot (1 - \pi(X))\right] \\ &= \underbrace{\mathbb{E}_{P_X}\left[\pi(X) \cdot \tau(X)\right]}_{=:I_\tau(\pi)} + \mathbb{E}_{P_X}\left[y_0(X)\right], \end{aligned} \tag{1}$$

where $I_\tau(\pi)$ represents the average impact of policy π. The optimal policy for a given configuration of the CATE function is the one that maximizes the associated welfare:

$$\pi^* = \operatorname*{argmax}_{\pi \in \Pi} W_{y_0,y_1}(\pi) = \operatorname*{argmax}_{\pi \in \Pi} I_\tau(\pi), \tag{2}$$

where Π is a family of candidate policies.[1] The DM has knowledge of the CATE through the distribution $P \in \mathcal{P}$ of observable random variables W, where $(Y, D, X) \subseteq W$. In particular, we denote $\mathcal{T}(P)$ the set of plausible CATE functions associated with a certain distribution of observables. When the DM has perfect knowledge of P and $\mathcal{T}(P)$ is a singleton, i.e. τ is point-identified, she can obtain π^* by solving (2).

Suppose now that $\mathcal{T}(P)$ is a non-singleton set, i.e. τ is partially identified. In that case, even under perfect knowledge of P, there exists a set of plausible values for the impact $I_\tau(\pi)$ of a candidate policy π. Notice that partial identification of the CATE does not necessarily

[1] Throughout the paper, we will assume that the maximization problem in (2) has at least one solution. If multiple solutions exist, the DM is assumed to arbitrarily pick π^* from the set of maximizers.

imply that the DM cannot obtain the optimal policy π^*. In particular, it is easy to see that under point-identification of the CATE one has $\pi^*(x) = \mathbb{1}\{\tau(x) \geq 0\}$ when the class of candidate policies Π is unrestricted, so that identification of the sign of the CATE is sufficient to obtain the optimal policy.[2] However, the unrestricted policy class has limited relevance in many practical settings. For example, the policy space Π may be exogenously constrained for institutional reasons, e.g. as policies may be required to satisfy specific requirements for budget, fairness or interpretability. While the DM may still hope that his specification of Π contains the first-best policy $\mathbb{1}\{\tau(X) \geq 0\}$, it is useful to interpret π^* as the "best-in-class" policy for the chosen class Π, when this does not contain the first-best. When Π is constrained, the DM is not able to obtain π^* in general without full knowledge of the CATE, although a partial ordering of policies can still be deduced (see Kasy, 2016; Han, 2019; Byambadalai, 2022).

Under partial identification, the DM therefore faces two sources of ambiguity. First, she does not know the distribution P. However, we assume that she has access to a sample $(W_i)_{i=1,\dots,n}$ from which she can learn about P. Second, she does not have knowledge about τ within the identified-set $\mathcal{T}(P)$, even under perfect knowledge of P. The broad objective of this paper is to provide a framework that allows the DM to handle both sources of ambiguity. We will approach the problem in two steps. First, we will study the decision problem faced by the DM under perfect knowledge of P. In particular, we will handle the ambiguity arising from partial identification of τ using well-known optimality criteria for decision under ambiguity. Each of the optimality criteria we consider will deliver a corresponding notion of optimal policy, which we call "ambiguity-robust". The ambiguity-robust optimal policy is a unique treatment assignment rule that is preferred to all other policies in Π according to preferences of the DM, and that coincides with the usual notion of optimal policy π^* in (2) in the special case of point identification of the CATE. In the next section, we study several notions of ambiguity-robust optimal policy.

In the second part of our analysis, we study how to handle the ambiguity in P by showing how a sample $(W_i)_{i=1,\dots,n}$ can be used to obtain an estimate $\widehat{\pi}_n$ for the ambiguity-robust optimal policy. The estimation procedure and the associated statistical guarantees are presented in Section 4 and 5, respectively.

Remark 1. *Unrestricted policy classes may also be precluded for practical reasons related to the estimation of the optimal policy. For example, the researcher may need to condition on a large number of covariates X for identification of the treatment effects, but only be interested in assigning treatment based on a restricted set of the covariates $\widetilde{X} \subset X$ (e.g. because she may not observe the full set of covariates when assigning treatment to new individuals from the*

[2] Cui and Tchetgen (2021) study a case in which sole point-identification of the sign of the CATE via an instrumental variable allows to obtain the optimal policy.

population). In that case, a practical way to side-step computation of an estimate for the lower-dimensional CATE, $\mathbb{E}[Y_i(1)|\widetilde{X}_i = \widetilde{x}] - \mathbb{E}[Y_i(0)|\widetilde{X}_i = \widetilde{x}]$*, is to impose restrictions directly on the policy class* Π *and estimate the optimal policy based on the estimated higher-dimensional CATE via the sample analogue of* (2)*.*

3 Ambiguity-robust optimal policies

The study of decision under ambiguity has a long tradition in decision theory and has received considerable attention in the context of treatment assignment problems (see Manski, 2011, for a review). In this section we review some classical optimality criteria for decision under ambiguity and study how they can be applied in the context of the treatment assignment problem at hand, leading to several notions of ambiguity-robust optimal policy.

A well-known optimality criterion for decision under ambiguity is minimax risk (see, e.g. Wald, 1950). In the context of our treatment assignment problem we can interpret welfare as negative risk, and this criterion leads to the optimal *maximin welfare* policy

$$\pi^*_{\mathtt{MMW}} = \operatorname*{argmax}_{\pi\in\Pi} \min_{(y_0,y_1)\in\mathcal{Y}(P)} W_{y_0,y_1}(\pi), \tag{3}$$

where $\mathcal{Y}(P)$ is the ambiguity set for $(y_0(\cdot), y_1(\cdot))$ identified from the distribution P of observables random variables. The optimal maximin welfare policy maximizes the lowest possible welfare under any configuration of the expected potential outcome functions in the identified set $\mathcal{Y}(P)$. An alternative application of minimax risk optimality in the context of treatment assignment is *maximin impact*, leading to the optimal policy

$$\pi^*_{\mathtt{MMI}} = \operatorname*{argmax}_{\pi\in\Pi} \min_{\tau\in\mathcal{T}(P)} I_\tau(\pi), \tag{4}$$

where $\mathcal{T}(P)$ denotes the ambiguity set for the CATE function. The optimal maximin impact policy maximizes the lowest possible impact under any configuration of the CATE in the identified set $\mathcal{T}(P)$. Notice that the minimax welfare criterion reflects an extreme degree of pessimism with regards to outcomes associated with both treatment and non-treatment scenarios; on the other hand, the minimax impact criterion reflects an extreme degree of pessimism with regards to the impact of the policy, thus directly raising the threshold for treatment.[3] Despite its intuitive appeal, minimax optimality has been criticised for being too conservative and often delivering decisions that are especially sensitive to changes in the ambiguity set.[4]

[3]In the empirical application of Section 6, both minimax welfare and minimax impact criteria result in $\pi(x) = 0$ for the entire population.

[4]In his classic textbook, Berger goes as far as saying that "In actually making decisions, the use of the minimax principle is definitely suspect." (Berger, 1985).

An alternative criterion that alleviates some of these concerns is *minimax regret*, with corresponding optimal policy

$$\begin{aligned}\pi^*_{\texttt{MMR}} &= \underset{\pi\in\Pi}{\operatorname{argmin}} \max_{(y_0,y_1)\in\mathcal{Y}(P)} \left[\left(\max_{\pi\,:\,\mathcal{X}\to\{0,1\}} W_{y_0,y_1}(\pi)\right) - W_{y_0,y_1}(\pi)\right] \\ &= \underset{\pi\in\Pi}{\operatorname{argmin}} \max_{\tau\in\mathcal{T}(P)} \left[\left(\max_{\pi\,:\,\mathcal{X}\to\{0,1\}} I_\tau(\pi)\right) - I_\tau(\pi)\right],\end{aligned} \tag{5}$$

The minimax regret criterion delivers a policy that minimizes the largest possible distance between attained welfare and the highest level of welfare attainable by the "oracle" treatment rule $\pi^* = \mathbb{I}\{\tau(x)\geq 0\}$ that has knowledge of the true τ. Minimax regret optimality has been advocated by Manski (2004) for its balanced consideration of the possible states of nature and for delivering more "reasonable" decisions rules in practice, compared to minimax risk approaches.

Remark 2. *An alternative version of the minimax regret criterion is minimax regret with respect to the welfare attained by the best-in-class policy in* Π, *resulting in the objective*

$$\pi^*_{\mathit{MMR2}} = \underset{\pi\in\Pi}{\operatorname{argmin}} \max_{\tau\in\mathcal{T}(P)} \left[\left(\max_{\pi\in\Pi} I_\tau(\pi)\right) - I_\tau(\pi)\right]. \tag{6}$$

While these two versions can be expected to enjoy similar properties, the first version we have considered is considerably more tractable. In fact, the innermost maximization in (5) has the closed-form solution $\max_{\pi\,:\,\mathcal{X}\to\{0,1\}} W_\tau(\pi) = \mathbb{E}_{P_X}\left[\max\{\tau(X),0\}\right]$. *As we show in Proposition 2 below, this allows to more explicitly characterize the properties of the optimization problem and the resulting optimal policy, as well as reduce the computational burden in solving the empirical analogue of the problem. For this reason we will focus on the version in (5) of the criterion. We also note that whenever the class* Π *is "well-specified", in the sense that* $\mathbb{I}\{\tau(x)\geq 0\}\in\Pi$ *for all* $\tau\in\mathcal{T}(P)$, *the two optimality criteria are equivalent.*

One critical drawback in the application of the optimality criteria just presented to the treatment assignment problem of this paper is that the optimal policies cannot be obtained in closed form. This is due to the form of (3), (4) and (5) involving several nested optimizations whose solutions cannot be easily characterized at the current level of generality when X includes continuously distributed covariates and Π may be arbitrarily restricted, which are both primary cases of interest of this paper. To make progress, we impose the following restrictions on the ambiguity sets for the expected potential outcomes and CATE.

Assumption 3.1 (Rectangular identified set for (y_0, y_1))**.** *The identified set for* (y_0, y_1) *is rectangular, that is,* $\mathcal{Y}$ *is of the form*

$$\mathcal{Y} = \{(y_0(\cdot), y_1(\cdot)) : (y_0(x), y_1(x)) \in \mathcal{Y}(x)\},$$

where $\mathcal{Y}(x)$ *is a compact subset of* $\mathbb{R}^2$.

Assumption 3.2 (Rectangular identified set for τ)**.** *The identified set for τ is rectangular, that is, $\mathcal{T}$ is of the form*

$$\mathcal{T} = \{\tau(\cdot) : \tau(x) \in [\underline{\tau}(x), \overline{\tau}(x)]\},$$

where $|\overline{\tau}(x)| < \infty$, $|\underline{\tau}(x)| < \infty$ for all $x \in \mathcal{X}$.

Assumptions 3.1 and 3.2 impose separation of the identified sets for the expected potential outcomes and CATE across the support of the covariates $\mathcal{X}$.[5] They are typically satisfied by identification schemes that do not impose shape restrictions on counterfactual outcomes with respect to the covariates X_i. These assumptions are widely adopted in the partial identification literature, and we refer the reader to Appendix B in Kasy (2016) for an extensive review of identification schemes that result in rectangular identified sets. Below we present three examples of identification schemes for the CATE that satisfy this assumption.

Example 3.1 (Manski bounds)**.** *Suppose there exists a binary instrument $Z_i \in \{0,1\}$ that satisfies the well know exogeneity and exclusion restrictions $Y_i(0), Y_i(1), D_i(0), D_i(1) \perp Z_i|X_i$, where $Y_i(d)$ and $D_i(z)$ denote the counterfactual outcome and treatment functions, respectively. If the instrument Z_i also satisfies the overlap condition*

$$\eta \leq \mathbb{P}(Z_i = 1|X_i) \leq 1 - \eta, \quad \eta > 0,$$

and the monotonicity condition (also known as no-defiers condition):

$$\mathbb{P}\big(D_i(1) \leq D_i(0)|X_i\big) = 1 \quad \text{or} \quad \mathbb{P}\big(D_i(1) \geq D_i(0)|X_i\big) = 1,$$

then seminal work by Imbens and Angrist (1994) shows point-identification of the conditional local average treatment effect (LATE):

$$\mathbb{E}[Y_i(1) - Y_i(0) \,|\, D_i(1) \neq D_i(0),\ X_i = x].$$

Let us now assume that $Y \in [Y_L, Y_U]$, i.e. the outcome is bounded, and define

$$\begin{aligned}
&h(z,x) = \mathbb{E}[Y_i|Z_i = z, X_i = x],\\
&m(d,z,x) = \mathbb{E}[Y_i|D_i = d, Z_i = z, X_i = x],\\
&p(z,x) = \mathbb{P}(D_i = 1|Z_i = z, X_i = x),\\
&z(x) = \mathbb{P}(Z_i = 1|X_i = x).
\end{aligned}$$

[5] Notice that Assumption 3.1 implies Assumption 3.2, but not viceversa.

The identified sets for the expected potential outcomes $y_0(x)$ and $y_1(x)$ are contained within the bounds

$$\overline{y}_0(x) = \min_{z\in\{0,1\}} \big\{m(0,z,x)\cdot(1-p(z,x)) + Y_U \cdot p(z,x)\big\},$$
$$\underline{y}_0(x) = \max_{z\in\{0,1\}} \big\{m(0,z,x)\cdot(1-p(z,x)) + Y_L \cdot p(z,x)\big\},$$

and

$$\overline{y}_1(x) = \min_{z\in\{0,1\}} \big\{m(1,z,x)\cdot p(z,x) + Y_U \cdot (1-p(z,x))\big\},$$
$$\underline{y}_1(x) = \max_{z\in\{0,1\}} \big\{m(1,z,x)\cdot p(z,x) + Y_L \cdot (1-p(z,x))\big\}.$$

The identified set for the CATE is then contained within the bounds

$$\overline{\tau}(x) = \overline{y}_1(x) - \underline{y}_0(x),$$
$$\underline{\tau}(x) = \underline{y}_1(x) - \overline{y}_0(x).$$

If no further functional form assumption on the distribution of potential outcomes is made, these bounds are sharp (Heckman and Vytlacil, 2001) and the sharp identified sets for the average potential outcomes and CATE respectively satisfy Assumption 3.1 and Assumption 3.2.

Example 3.2 (Balke-Pearl)**.** *Suppose that the same assumptions as in Example 3.1 hold, and additionally the monoticity assumption is strengthened to*

$$\mathbb{P}\big(D_i(1) \geq D_i(0)|X_i\big) = 1,$$

that is, the direction of the monotonicity is known and positive. The bounds for the potential outcomes simplify to

$$\overline{y}_0(x) = m(0,0,x)\cdot(1-p(0,x)) + Y_U\cdot p(0,x),$$
$$\underline{y}_0(x) = m(0,0,x)\cdot(1-p(0,x)) + Y_L\cdot p(0,x),$$
$$\overline{y}_1(x) = m(1,1,x)\cdot p(1,x) + Y_U\cdot(1-p(1,x),$$
$$\underline{y}_1(x) = m(1,0,x)\cdot p(0,x) + Y_L\cdot(1-p(0,x)),$$

and the CATE is contained within the bounds

$$\overline{\tau}(x) = h(1,x) - h(0,x) + p(0,x)\cdot\big(m(1,0,x) - Y_L\big) + (1-p(1,x))\cdot\big(Y_U - m(0,1,x)\big),$$
$$\underline{\tau}(x) = h(1,x) - h(0,x) + p(0,x)\cdot\big(m(1,0,x) - Y_U\big) + (1-p(1,x))\cdot\big(Y_L - m(0,1,x)\big).$$

where $p(0,x)$ and $1-p(1,x)$ identify the proportions of always-takers and never-takers at $X_i = x$, respectively. If no further functional form assumption on the distribution of outcomes for non-compliant populations is made, these bounds are sharp (Balke and Pearl, 1997) and the sharp *identified sets for the average potential outcomes and CATE respectively satisfy Assumption 3.1 and Assumption 3.2.*

Example 3.3 (Manski-Pepper bounds)**.** *Suppose that instead of full exogeneity, the instrumental variable Z_i satisfies the weaker "monotone IV" condition*

$$\mathbb{E}[Y_i(d)|Z_i = 0, X_i] \leq \mathbb{E}[Y_i(d)|Z_i = 1, X_i], \quad d = 0, 1. \tag{7}$$

Manski and Pepper (2000) show that when the outcome is bounded one has

$$\begin{aligned}
&\sum_{z=0,1} \mathbb{P}\left(Z_i = z|X_i\right) \cdot \max_{z_1 \leq z} \left\{m(d, z_1, X_i) \cdot \mathbb{P}(D_i = d|Z_i = z_1, X_i) + Y_L \cdot \mathbb{P}(D_i = 1 - d|Z_i = z_1, X_i)\right\} \\
&\qquad\qquad\qquad\qquad \leq \ \mathbb{E}[Y_i(d)|X_i] \ \leq \\
&\sum_{z=0,1} \mathbb{P}\left(Z_i = z|X_i\right) \cdot \min_{z_2 \geq z} \left\{m(d, z_2, X_i) \cdot \mathbb{P}(D_i = d|Z_i = z_2, X_i) + Y_L \cdot \mathbb{P}(D_i = 1 - d|Z_i = z_2, X_i)\right\}.
\end{aligned}$$

Upper (lower) bounds for the CATE are obtained by combining upper (lower) bounds for $\mathbb{E}[Y_i(1)|X_i = x]$ with the lower (upper) bound for $\mathbb{E}[Y_i(0)|X_i = x]$:

$$\begin{aligned}
\overline{\tau}(x) &= z(x) \cdot \psi_{1,1}(x; Y_U) + (1 - z(x)) \cdot \min\left\{\psi_{0,1}(x; Y_U), \psi_{1,1}(x; Y_U)\right\} \\
&\quad - z(x) \cdot \max\left\{\psi_{0,0}(x; Y_L), \psi_{1,0}(x; Y_L)\right\} - (1 - z(x)) \cdot \psi_{0,0}(x; Y_L), \\
\underline{\tau}(x) &= z(x) \cdot \max\left\{\psi_{0,1}(x; Y_L), \psi_{1,1}(x; Y_L)\right\} + (1 - z(x)) \cdot \psi_{0,1}(x; Y_L) \\
&\quad - z(x) \cdot \psi_{1,1}(x; Y_U) - (1 - z(x)) \cdot \min\left\{\psi_{0,0}(x; Y_U), \psi_{1,0}(x; Y_U)\right\}, .
\end{aligned}$$

where

$$\psi_{z,d}\left(x; Y_{(\cdot)}\right) = m(d, z, x) \cdot (d \cdot p(z, x) + (1 - d) \cdot (1 - p(z, x)) + Y_{(\cdot)} \cdot (d \cdot (1 - p(z, x)) + (1 - d) \cdot p(z, x)).$$

Under no further assumption on the distribution of potential outcomes, these bounds are sharp (Manski and Pepper, 2000) and satisfy Assumptions 3.1 and 3.2.

Having restricted the identified sets $\mathcal{Y}$ and $\mathcal{T}$ as in Assumptions 3.1-3.2, we are now able to provide a simpler characterization of the maximin welfare and maximin impact policies.

Proposition 1. *Define $\underline{y}_d(x) = \min_{y_d(x) \in \mathcal{Y}(x)} y_d(x)$ and $\overline{y}_d(x) = \max_{y_d(x) \in \mathcal{Y}(x)} y_d(x)$. Under Assumption 3.1 the optimal maximin welfare policy is*

$$\pi^*_{\mathtt{MMW}} = \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_{P_X}\left[(2\pi(X) - 1) \cdot (\underline{y}_1(X) - \underline{y}_0(X))\right]. \tag{8}$$

Furthermore, under Assumption 3.2 the optimal maximin impact policy is

$$\pi^*_{\mathtt{MMI}} = \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_{P_X}\left[(2\pi(X) - 1) \cdot \underline{\tau}(X)\right]. \tag{9}$$

Proposition 1 shows that the optimal maximin welfare and maximin impact policies maximize surrogate versions of welfare substituting the unidentified CATE with the difference in the lower bounds of potential outcomes $\underline{y}_1(x) - \underline{y}_0(x)$ and the lower bound for CATE $\underline{\tau}(x)$ at every point in the covariate space, respectively. Notice that when $\mathcal{Y}(x)$ is rectangular with respect to the two potential outcomes, i.e. $y_0(x) \in \mathcal{Y}_0(x), y_1(x) \in \mathcal{Y}_1(x)$ and $\mathcal{Y}(x) = \mathcal{Y}_0(x) \times \mathcal{Y}_1(x)$, we have $\underline{\tau}(x) = \underline{y}_1(x) - \overline{y}_0(x)$, thus highlighting the "pessimistic" nature of the maximin impact criterion.

Remark 3. *The maximin welfare and maximin impact optimal policies coincide when $y_0(\cdot)$ is point-identified. This case is relevant when $y_0(x)$ represents the (conditional) average outcome under the status-quo in the entire population and is typically point-identified from observational data.*

The simplification of these two maximin problems into single maximisation problems has important benefits for the study of the optimal policies and their estimation from the data. In fact, the sample analogues of optimizations (8) and (9) are amenable to standard computation procedures for a variety of policy classes Π. Furthermore, their solution can be studied using tools for empirical risk minimisation problems, as discussed in Section 4.

Despite the involvement of an additional maximization problem compared to maximin welfare and maximin impact, Assumption 3.2 allows to provide a simpler characterization also for the minimax regret optimal policy.

Proposition 2. *Under Assumption 3.2 the optimal minimax welfare regret policy is*

$$\pi^*_{\mathtt{MMR}} = \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_{P_X}\left[(2\pi(X) - 1) \cdot \widetilde{\tau}(X) \right] \tag{10}$$

where

$$\widetilde{\tau}(x) = \overline{\tau}(x) \cdot \mathbb{1}\big\{\overline{\tau}(x) \geq 0\big\} + \underline{\tau}(x) \cdot \mathbb{1}\big\{\underline{\tau}(x) \leq 0\big\} \tag{11}$$

This simpler characterization of the minimax regret problem as a single maximization sheds light on the properties of its associated optimal policy. In particular, we see that the objective function symmetrically treats individuals whose expected treatment effect sign is identified by assigning as surrogate for the CATE their outer bound, i.e. the CATE upper (lower) bound for individuals with identified positive (negative) sign for CATE. Individuals for which the sign of the treatment effect is ambiguous are assigned an intermediate point within their respective CATE bounds, which depends on the extent to which the identified set lies in the positive or negative region. Intuitively, the criterion prioritizes correct treatment allocation to individuals who unambiguously benefit from (or are harmed by) the treatment and down-weights the importance of individuals for which the sign of the treatment response is ambiguous in the treatment allocation problem. As an extreme case, individuals with CATE bounds exactly symmetric around 0 (i.e. $\overline{\tau}(x) = -\underline{\tau}(x)$) are given no consideration in the solution of the treatment allocation problem. This intuition can be further supported by noticing that the original welfare maximization under point-identification in (2) can be re-casted as the weighted classification problem

$$\pi^* = \operatorname*{argmin}_{\pi\in\Pi} \mathbb{E}_{P_X}\left[\mathbb{1}\big\{(2\pi(X)-1) \neq \operatorname{sign}(\tau(X))\big\} \cdot |\tau(X)|\right],$$

of which the minimax welfare regret optimal policy in (11) turns out to solve the minimax version under Assumption 3.2:

$$\pi^*_{\texttt{MMR}} = \operatorname*{argmin}_{\pi\in\Pi} \max_{\tau\in\mathcal{T}} \mathbb{E}_{P_X}\left[\mathbb{1}\big\{(2\pi(X)-1) \neq \operatorname{sign}(\tau(X))\big\} \cdot |\tau(X)|\right].$$

It is from this minimax classification risk perspective that Pu and Zhang (2021) obtain and study the minimax regret policy, which they call the "IV-optimal policy".

An alternative version of minimax regret optimality which has been used in the context of treatment choice is minimax regret with respect to a baseline policy. Kallus and Zhou (2018) assume the existence of a fixed policy π_{B} from which the DM does not want to unnecessarily deviate. They define the optimal policy as minimizing regret with respect to this baseline policy:

$$\begin{aligned}\pi^*_{\texttt{MMRB}} &= \operatorname*{argmin}_{\pi\in\Pi} \max_{\tau\in\mathcal{T}} \left(I_\tau(\pi_B) - I_\tau(\pi)\right)\\ &= \operatorname*{argmax}_{\pi\in\Pi} \mathbb{E}_{P_X}\left[(2\pi(X)-1)\cdot\left(\overline{\tau}(X)\cdot\mathbb{1}\left\{\pi_B(X)\geq 0\right\} + \underline{\tau}(X)\cdot\mathbb{1}\left\{\pi_B(X)<0\right\}\right)\right],\end{aligned}$$

where the second equality uses Assumption 3.2. While potentially appealing in certain settings, e.g. when π_B represents the existing standard of care in a medical setting, this optimality criterion suffers the potential drawback of requiring the DM to specify (and motivate) the baseline policy for it to be operational. Adopting the never-treat baseline policy, i.e. $\pi_B(x) = 0,\ \forall x \in \mathcal{X}$, could be seen as an appealing "agnostic" choice, which however makes this criterion default to maximin impact and thus inherit its potentially undesirable properties.

The last notion of ambiguity-robust optimal policy that we present in this section is based on the Hurwicz criterion (Hurwicz, 1951), arguably one of the most widely used in decision-making under ambiguity. In the context of the treatment assignment problem at hand, the Hurwicz criterion leads to the ambiguity-robust policy

$$\pi^*_{\texttt{HurW},\delta_0,\delta_1} =$$

$$\operatorname*{argmax}_{\pi\in\Pi} \mathbb{E}_{P_X}\left[(2\pi(X)-1)\cdot\Big(\{\delta_1\cdot\overline{y}_1(X) + (1-\delta_1)\cdot\underline{y}_1(X)\} - \{\delta_0\cdot\overline{y}_0(X) + (1-\delta_0)\cdot\underline{y}_0(X)\}\Big)\right],$$

where $\delta_1 \in [0,1]$ and $\delta_0 \in [0,1]$ are user-defined weights reflecting the degree of optimism with respect to the outcomes under treatment and non-treatment, respectively. It is easy to see that the maximin welfare criterion in (3) corresponds to the choice $\delta_1 = 0, \delta_0 = 0$. An analogous notion of optimality focused on impact rather than welfare, leads to the optimal policy

$$\pi^*_{\mathtt{HurI},\delta} = \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_{P_X}\left[(2\pi(X) - 1) \cdot \big(\delta \cdot \overline{\tau}(X) + (1-\delta) \cdot \underline{\tau}(X)\big) \right],$$

where $\delta \in [0,1]$ controls the degree of optimism with respect to the effect of treatment, with the maximin impact optimal policy corresponding to the choice $\delta = 0$. Interestingly, when $\delta = 1/2$ and Π is well-specified, in the sense that it contains the first-best assignment $\mathbb{1}\{\overline{\tau}(x) + \underline{\tau}(x) \geq 0\}$, we have

$$\pi^*_{\mathtt{MMR}} = \mathbb{1}\{\widetilde{\tau}(x) \geq 0\} = \mathbb{1}\{\overline{\tau}(x) + \underline{\tau}(x) \geq 0\} = \pi^*_{\mathtt{HurI},\frac{1}{2}}.$$

Therefore the minimax regret and Hurwicz impact optimal policies coincide, as they assign treatment based on the middle point between the upper and lower CATE bounds. When Π is not well-specified, however, minimax regret optimality is not nested into any of the Hurwicz-type criteria just presented, thus highlighting the radically different attitude towards ambiguity implied by minimax regret compared to maximin welfare/impact. In particular, minimax regret is the only criterion of those presented (along with Hurwicz impact under $\delta = 1/2$) that treats symmetrically individuals with CATE bounds symmetric around 0, in the sense that $\Gamma(P_1; x) = -\Gamma(P_2, x)$ whenever $\overline{\tau}_1(x) = -\underline{\tau}_2(x)$ and $\underline{\tau}_1(x) = -\overline{\tau}_2(x)$. For this reason, minimax regret does not reflect an optimistic/pessimistic attitude towards ambiguity but rather an "opportunistic" one, in light of its prioritization of correct treatment assignment to individuals whose CATE sign is unambiguously identified.

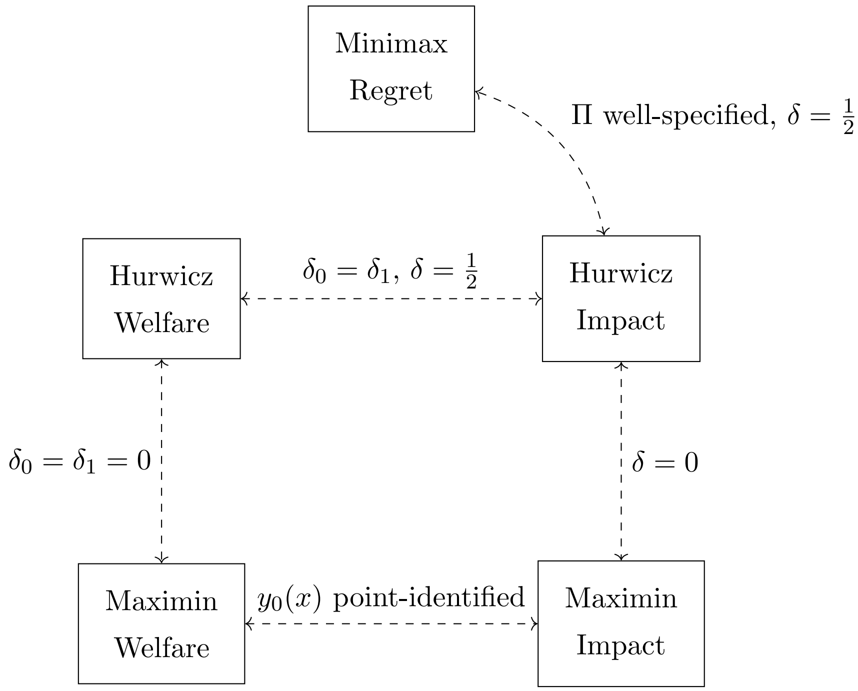


Figure 1: Relationship between ambiguity-robust optimal policies

3.1 A common framework

While accommodating a wide range of attitudes towards ambiguity, the notions of optimality presented in Section 5.2 share a common structure. In fact, by virtue of Assumptions 3.1 and 3.2, the corresponding optimal policies can all be written as

$$\pi^*(P) = \operatorname*{argmax}_{\pi \in \Pi} Q(P;\pi), \qquad Q(P;\pi) := \mathbb{E}_{P_X}\Big[(2\pi(X)-1)\cdot\Gamma(P;X)\Big], \tag{12}$$

for a specific score function[6] $\Gamma(P;\cdot)$, where we have highlighted the dependence of the score on the distribution P. The specific dependence on P is determined by the optimality criterion (as summarized in Table 1) as well as the identification assumptions (e.g. Balke-Pearl, Manski-Pepper etc.). This common structure also nests the point identified setting as the special case $\Gamma(P;X) = \tau(X)$ and thus suggests that existing estimation procedures for this special case can be extended to the partially identified setting.

However, one peculiar feature of the partially identified setting is the restricted degree of smoothness enjoyed by the objective function, in particular the differentiability of the scores with respect to P. Under point-identification of the CATE via standard unconfoundedness assumptions, one has $\Gamma(P;x) = \mathbb{E}[Y|D=1, X=x] - \mathbb{E}[Y|D=1, X=x]$ and the full differentiability of the score with respect to the expectation $\mathbb{E}[Y|D,X]$ is immediately apparent. However, for the

[6]The term 'score function' is borrowed from Athey and Wager (2021).

minimax regret criterion we notice that the score is *directionally differentiable*[7] with respect to P at $\overline{\tau}(x) = 0$ or $\underline{\tau}(x) = 0$. Even when $\Gamma(P; x)$ depends smoothly on expected outcomes/CATE bounds, lack of full differentiability of the scores can arise through a lack of differentiability of the expected outcomes and CATE bounds themselves. In fact, many popular identification assumptions, including the Manski and Manski-Pepper bounds from Examples 3.1 and 3.3, deliver bounds that are only directionally differentiable with respect to identified parameters due to the presence of min / max operators (see Chernozhukov et al., 2013, and examples therein). Whether a consequence of the optimality criterion or the identification assumptions, lack of full differentiability of the scores is a unique and pervasive feature of the treatment assignment problem under partial identification, one that has not been explicitly acknowledged in the most recent contributions in this area.[8] A major contribution of this paper is to account for the role played by the lack of full differentiability when we establish procedures for estimating ambiguity-robust optimal policies in Section 4.

Table 1: Optimality criteria and associated scores

Optimality criterion	$\Gamma(P; x)$
Maximin Welfare	$\underline{y}_1(x) - \underline{y}_0(x)$
Maximin Impact	$\underline{\tau}(x)$
Minimax Regret (oracle)	$\overline{\tau}(x) \cdot \mathbb{1}\big\{\overline{\tau}(x) \geq 0\big\} + \underline{\tau}(x) \cdot \mathbb{1}\big\{\underline{\tau}(x) \leq 0\big\}$
Minimax Regret (baseline)	$\overline{\tau}(x) \cdot \mathbb{1}\big\{\pi_{\mathtt{B}}(x) = 1\big\} + \underline{\tau}(x) \cdot \big\{\pi_{\mathtt{B}}(x) = 0\big\}$
Hurwicz (welfare)	$\big(\{\delta_1 \cdot \overline{y}_1(x) + (1-\delta_1) \cdot \underline{y}_1(x)\} - \{\delta_0 \cdot \overline{y}_0(x) + (1-\delta_0) \cdot \underline{y}_0(x)\}\big)$ $\delta_1, \delta_0 \in [0, 1]$
Hurwicz (impact)	$\delta \cdot \overline{\tau}(x) + (1-\delta) \cdot \underline{\tau}(x),\ \delta \in [0, 1]$

[7] Let $P \in \mathcal{P}$ be a probability distribution on which the function $f : \mathcal{P} \to \mathbb{R}$ depends. We say that f is directionally differentiable at P_0 if the limit

$$\lim_{t \downarrow 0} \frac{f(P_0 + t(h - P_0)) - f(P_0)}{t} = \dot{f}_{P_0}[h]$$

exists for every $h \in \mathcal{P}$, in which case $\dot{f}_{P_0}[\cdot]$ denotes the directional derivative of f at P_0. If it exists, the directional derivative $\dot{f}_{P_0}[\cdot]$ is positively homogeneous of degree one but not necessarily linear. If $\dot{f}_{P_0}[\cdot]$ is linear then f is fully differentiable at P_0.

[8] The only exception is Christensen et al. (2022), who deal with estimation of optimal treatment decisions in the absence of individualization.

4 Estimation

In this section we present the statistical framework underlying the problem of estimation of optimal treatment rules under partial identification. We will discuss heuristics underlying several features of the estimation problem, and then present our proposed estimation procedure.

We work in a learning setting where the estimand $\pi^*(P)$ is as in (12), and we observe an i.i.d. sample $(W_i)_{i=1,\dots,n}$ of size n from the unknown distribution P for the observed random variables $W \in \mathcal{W}$, $X \subset W$. To retain generality of the framework, we do not specify the exact dependence of the functional $\Gamma(P;x)$ on P, which will depend on the choice of optimality criterion for the resolution of ambiguity (maximin welfare, minimax regret etc.) and identification assumptions determining the sets $\mathcal{Y}(P), \mathcal{T}(P)$. However, we will assume that the scores depend on P only through a vector of nuisance functions $g : \mathcal{V} \to \mathbb{R}^J$ specified by the moment equations

$$\mathbb{E}[U - g(V) \mid V] = 0, \tag{13}$$

where U and V are random vectors with $U \subseteq W$ and $X \subseteq V \subset W$. Furthermore, we will stipulate that the dependence of $\Gamma(g;x)$ on the nuisance functions g from the possibly infinite-dimensional space $\mathcal{G}$ can be reduced as

$$\Gamma(g;x) = \Gamma(\theta(x), x),$$

where, for a fixed x, the parameter $\theta(x) \in \Theta_x \subseteq \mathbb{R}^M$ is a finite-dimensional vector of conditional moments of U deduced from g. This latter restriction rules out scores $\Gamma(g;X)$ that at a single point in the covariate space depend on exhaustive evaluations of the nuisance functions g over continuous supports. This is the case, for example, in versions of the CATE bounds from Examples 3.1-3.3 featuring instruments with continuous support $\mathcal{Z}$. In those settings, the CATE bounds depend on objects such as $\sup_{z\in\mathcal{Z}} \mathbb{E}[Y \mid Z = z, X = x]$, and are therefore not covered by the results of this paper. Finally, we will assume that that $\Gamma(\theta(x);x)$ can be expressed as

$$\Gamma(\theta(x);x) = \varphi_0(\theta(x);x) + \sum_{\ell=1}^{L} a_\ell \cdot \varphi_\ell(\theta(x);x) \cdot \mathbb{1}\left\{\varphi_\ell(\theta(x);x) \geq 0\right\}, \quad a_\ell \in \{-1, 1\}, \tag{14}$$

where the functions $\varphi_\ell(\theta(x);x) : \Theta_x \times \mathcal{X} \to \mathbb{R}$ are fully differentiable with respect to $\theta(x)$ for all $x \in \mathcal{X}$. While seemingly ad-hoc, this restriction is sufficiently general to accommodate a wide range of popular partial identification assumptions for the CATE as well as optimality criteria for the resolution of ambiguity. In particular, formulation (14) accommodates linear combinations of min/max operators, which typically feature in many identification bounds for the CATE with discrete instruments. In fact, our framework can be shown to be applicable to any combination of the optimality criteria discussed in Section 3 and the identification schemes contained in the recent survey paper by Swanson et al. (2018).[9]

Example 3.2 (Continued)**.** *Under the identification assumptions of the Balke-Pearl bounds and resolution of ambiguity via Minimax Regret, we have*

$$\begin{aligned} g &= (h, m, p), \\ \theta(x) &= (h(1,x), h(0,x), m(1,0,x), m(0,1,x), p(1,x), p(0,x)), \end{aligned}$$

and

$$\Gamma(g;x) = \varphi_1(\theta(x);x) \cdot \mathbb{1}\left\{\varphi_1(\theta(x);x) \geq 0\right\} - \varphi_2(\theta(x);x) \cdot \mathbb{1}\left\{\varphi_2(\theta(x);x) \geq 0\right\},$$

where $\varphi_1(\theta(x);x) = \overline{\tau}(\theta(x);x)$, $\varphi_2(\theta(x);x) = -\underline{\tau}(\theta(x);x)$ are differentiable with respect to $\theta(x)$.

In this framework, a natural approach for estimation is via the so-called "empirical risk minimisation" (ERM) principle (Vapnik, 1998), in which the estimate for the optimal policy is obtained as the maximiser of a sample analogue of the population objective Q:

$$\widehat{\pi}_n = \operatorname{argmax}\left\{\widehat{Q}_n(\pi) \,:\, \pi \in \Pi\right\}, \qquad \widehat{Q}_n(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1) \cdot \widehat{\Gamma}_i \tag{15}$$

where $\widehat{\Gamma}_i$ is some suitable estimate for $\Gamma(g;X_i)$. The ERM approach is a cornerstone of statistical learning theory and is at the foundation of many traditional and modern estimation methods in statistics, econometrics and machine learning. The ERM principle has also guided much of the recent literature on individualized treatment rules, where different variations have been applied under the names of "outcome-weighted learning" (Zhao et al., 2012) and "empirical welfare maximization" (Kitagawa and Tetenov, 2018). A major challenge in the implementation of (15) comes from the presence of the nuisance functions g, which are typically unknown and thus need to be estimated. Assuming that we have access to appropriate algorithms/nonparametric procedures for estimation the nuisance functions, one simple approach would be to use the sample $(W_i)_{i=1,\ldots,n}$ to obtain the estimates $\widehat{g}$ and then form plug-in estimates for the score as $\widehat{\Gamma}_i = \Gamma(\widehat{g};X_i)$. While seemingly natural, this "naive plug-in" approach has undesirable properties. In particular, policies estimated via the naive plug-in approach can typically only be shown to converge at sub-optimal rates to their population counterparts, unless very restrictive assumptions are imposed on first-stage estimators for the nuisance components (see, e.g., Foster and Syrgkanis, 2019).

[9] Even though not accommodated by formulation (14), our framework and theoretical results also apply to scores that feature a finite number of nested linear combinations of min / max operators. We discuss this extension in Appendix A.

One crucial reason underlying the undesirable statistical properties of the naive plug-in approach is that the resulting objective function estimate $\widehat{Q}_n$ is overtly sensitive to error in estimating the nuisance functions g. In order to gain intuition, it is useful to consider the following expansion of the population objective function $Q(g;\pi) = \mathbb{E}_{P_X}\left[(2\pi(X)-1)\cdot\Gamma(g;X)\right]$,

$$Q(\widetilde{g};\pi) - Q(g;\pi) = \left.\frac{\partial Q(g + t(\widetilde{g}-g);\pi)}{\partial t}\right|_{t=0} + \Delta(\widetilde{g},g;\pi) + O\left(\|\widetilde{g}-g\|^2_{L_2(P)}\right) \tag{16}$$

where

$$\begin{aligned}&\Delta(\widetilde{g},g;\pi) = \\ &\mathbb{E}_{P_X}\left[(2\pi(X)-1)\cdot\left(\sum_{\ell=1}^{L} a_\ell\cdot\varphi_\ell(g;X)\cdot\left(\mathbb{1}\left\{\varphi_\ell(\widetilde{g};X)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X)\geq 0\right\}\right)\right)\right].\end{aligned}$$

Von Mises expansions like the above are at the heart of the theory of orthogonal machine learning (Chernozhukov et al., 2022). In our setting, it allows to describe the impact of a small deviation from g in the direction $\widetilde{g}-g$ as consisting of three terms. The first term is the so-called "pathwise derivative" of $Q(g;\pi)$ and typically scales with $\|\widetilde{g}-g\|_{L_1(P)}$. The second term $\Delta(\widetilde{g},g;\pi)$ is due the presence of the type of non-differentiabilities arising under partial identification, and is unique to the framework of this paper. This term accounts for the bias that arises from misclassifying whether the component functions φ_ℓ are above or below 0, as we move away from g in the direction $\widetilde{g}-g$. The third term is a second-order remainder scaling with the mean-square distance between $\widetilde{g}$ and g. A central feature of our proposed estimation procedure is the construction of a new objective function, called Neyman-orthogonal, with reduced sensitivity to local perturbations away from g. For this purpose, we will assume that there exists functionals $\alpha_\ell(\{g,f\};V)$ such that for every $\widetilde{g}\in\mathcal{G}$

$$\varphi_\ell(\widetilde{g};x) = \mathbb{E}[\langle\alpha_\ell(\{\widetilde{g},f\};V),\widetilde{g}(V)\rangle \mid X = x], \qquad \ell = 1,\ldots,L,$$

where $f\in\mathcal{F}$ is a vector of additional nuisance functions defined analogously to g.[10] We then construct Neyman-orthogonal formulations for the component functions as

$$\varphi_\ell^{\texttt{NO}}(\{g,f\};w) = \varphi_\ell(\theta;x) + \phi_\ell(\{g,f\};w), \quad \phi_\ell(\{g,f\};w) = \langle\alpha_\ell(\{g,f\},v), u - g(v)\rangle.$$

The functionals α_ℓ are the Riesz-representers of φ_ℓ, while the functionals ϕ_ℓ are their so-called influence function adjustments. We refer the reader to Ichimura and Newey (2022) for their

[10] We will also assume that for the j-th entry of the Riesz-representer we have $\alpha_\ell^{(j)}(\{(\widetilde{g}_{-j},\widetilde{g}_j),\widetilde{f}\},x) = \alpha_\ell^{(j)}(\{\widetilde{g}_{-j},\widetilde{f}\},x)$, where $\widetilde{g}_{-j}$ denotes the exclusion of the j-th entry $\widetilde{g}_j$ from the vector of nuisance functions $\widetilde{g}$. This restriction is sufficiently general to accommodate component functions $\varphi_\ell(\theta(x);x)$ that feature linear combinations of products of the parameters $\theta(x)$, thus encompassing all the discussed identification schemes, including Examples 3.2-3.3.

properties and general methods for their calculation, while we provide below their specific form for the Balke-Pearl CATE bounds of Example 3.2.[11]

Example 3.2 (Continued)**.** *Following Ichimura and Newey (2022), we compute the influence function adjustment $\phi_U(\{g,f\},W_i)$ for the CATE upper bound by taking the Gateaux derivative of $\overline{\tau}(g;X)$, which yields*

$$\begin{aligned}\phi_U(\{g,f\};W_i) = &\underbrace{\left[\frac{Z_i}{z(X_i)} - \frac{1-Z_i}{1-z(X_i)}\right]}_{\alpha_U^{(1)}(\{g,f\},V_i)} \cdot \big(Y_i - h(Z_i,X_i)\big) \\ &+ \underbrace{\left[\frac{D_i(1-Z_i)}{1-z(X_i)} + \frac{(1-D_i)Z_i}{z(X_i)}\right]}_{\alpha_U^{(2)}(\{g,f\},V_i)} \cdot \big(Y_i - m(D_i,Z_i,X_i)\big) \\ &+ \underbrace{\left[(m(1,0,X_i) - Y_L) \cdot \frac{1-Z_i}{1-z(X_i)} - (Y_U - m(0,1,X_i)) \cdot \frac{Z_i}{z(X_i)}\right]}_{\alpha_U^{(3)}(\{g,f\},V_i)} \cdot \big(D_i - p(Z_i,X_i)\big),\end{aligned}$$

where the associated Riesz-representer is $\alpha_U(\{g,f\},V_i) = (\alpha_U^{(1)},\alpha_U^{(2)},\alpha_U^{(3)})'$ with $f = z(x)$ and $V_i = (D_i,Z_i,X_i)'$. The influence function and Riesz-representer for the CATE lower bound are obtained by interchanging Y_U and Y_L in the expressions above.

Finally we form Neyman-orthogonal formulations for the scores as

$$\Gamma^{\mathtt{NO}}(\{g,f\};w) = \varphi_0^{\mathtt{NO}}(\{g,f\};w) + \sum_{\ell=1}^{L} a_\ell \cdot \varphi_\ell^{\mathtt{NO}}(\{g,f\};w) \cdot \mathbb{1}\left\{\varphi_\ell(g;x) \geq 0\right\},$$

which are then used to form the Neyman-orthogonal objective function

$$Q^{\mathtt{NO}}(\{\cdot,\cdot\};\pi) = \mathbb{E}_{P_W}\Big[(2\pi(X)-1)\cdot\Gamma^{\mathtt{NO}}(\{\cdot,\cdot\};W)\Big].$$

Our construction of Neyman-orthogonal scores features the addition of the influence function adjustments ϕ_ℓ to the component functions φ_ℓ outside the indicators, but crucially not inside. Heuristically, the influence function adjustments serve the purpose of reducing the bias induced by the evaluation of the component functions $\varphi_\ell(\cdot;x)$ away from g. Since the indicators vary discontinuously with g it is not possible to linearly approximate the dependence of the indicators on the nuisance functions at the point of discontinuity $\varphi_\ell(g;x) = 0$. As a result, it is not possible to reduce the bias induced by the presence of the indicators (represented by the term $\Delta(\widetilde{g},g,\pi)$ in (16)) by means of influence function adjustments, whose de-biasing properties implicitly rely on

[11] See also Kennedy (2022) for a user-friendly discussion of methods for computation influence function adjustments.

the validity of such linear approximation.[12] Notice that $Q^{\texttt{NO}}(\{g, f\}; \pi) = Q(g; \pi)$ by the mean-zero property of the influence function adjustments, so that orthogonalization of the objective does not change the notion of optimal policy $\pi^*(P)$. Nonetheless, for the orthogonalized objective it can be shown that

$$Q^{\texttt{NO}}(\{\widetilde{g}, \widetilde{f}\}; \pi) - Q^{\texttt{NO}}(\{g, f\}; \pi) = \Delta(\widetilde{g}, g; \pi) + O\left(\|\widetilde{g} - g\|^2_{L_2(P)} + \|\widetilde{f} - f\|^2_{L_2(P)}\right). \tag{17}$$

Comparing the above with (16), we see that the von Mises expansion for the orthogonalized objective does not feature the pathwise derivative term, implying that $Q^{\texttt{NO}}(\,\cdot\,; \pi)$ is less sensitive to deviations away from g compared to the original objective $Q(\,\cdot\,; \pi)$. As shown in Section 5, this property will generally translate in improved statistical guarantees for the estimated policy when the nuisance functions have to be learned from the data. It should however be noticed that the term $\Delta(\widetilde{g}, g; \pi)$ still appears in the relevant expansion after orthogonalization. The contribution of this term is quantified in Section 5, and it is shown to be of first-order importance for the statistical properties of the estimation procedures.

The second key component of our approach is the use of sample-splitting, which is a commonly employed method in semiparametric inference (Chernozhukov et al., 2022) and statistical learning (Foster and Syrgkanis, 2019). The main purpose of sample-splitting is to reduce the risk of overfitting that generally arises from using the same data to estimate the nuisance functions as well as the optimal policy, as in the naive plug-in approach. Similarly to Athey and Wager (2021), we employ a particular form of sample-splitting known as K-fold cross-fitting (described below). This procedure ensures that in the estimate for $\Gamma^{\texttt{NO}}(\{g, f\}; W_i)$, the estimates for the nuisance functions $\{g, f\}$ are independent from the data-point W_i for that same unit. This independence property is crucial for the theoretical guarantees of our proposed method.

Our proposed estimation procedure is therefore as follows. We first randomly split the data into K evenly-sized folds and for each fold $k = 1, \ldots, K$ we obtain estimates $\{\widehat{g}^{(-k)}, \widehat{f}^{(-k)}\}$ using the remaining $K-1$ folds. These are then used to form cross-fitted Neyman-orthogonal estimates for the scores

$$\widehat{\Gamma}_i = \widehat{\Gamma}^{\texttt{NO}}\left(\{\hat{g}^{-k(i)}, \hat{f}^{-k(i)}\}; W_i\right), \qquad i = 1, \ldots, n, \tag{18}$$

where $k(i) \in \{1, \ldots, K\}$ denotes the fold containing the i-th observation. Finally, the estimated optimal policy rule $\widehat{\pi}_n$ is obtained via the optimization problem (15).

[12] On the contrary, naively adding the influence function adjustments inside the indicators would lead to a bias increase, rather than a reduction.

5 Statistical guarantees for the estimated policy

Let $\widehat{\pi}_n$ be the estimated treatment policy defined in (15), with estimated scores as in (18). Following Manski (2004), we assess the performance of the estimated treatment policy in terms of (statistical) regret with respect to population optimal policy. Let the population ambiguity-robust optimal policy be $\pi_n^*(P) \in \operatorname{argmax}_{\pi \in \Pi_n} Q(P; \pi)$, where we have included the n-subscript to the policy class Π_n to allow this to depend on the sample size for generality. The statistical regret of an estimated policy $\widehat{\pi}_n$ is defined as

$$R_n(P; \widehat{\pi}_n) = \mathbb{E}_{P_n}\big[Q(P; \pi_n^*) - Q(P; \widehat{\pi}_n)\big] \geq 0, \tag{19}$$

where $\mathbb{E}_{P_n}$ is the expectation with respect to the i.i.d. sample of observable random variables $(W_i)_{i=1,\ldots,n}$ used to estimate $\widehat{\pi}_n$. The next few subsections build up to a final result providing asymptotic convergence guarantees for $\widehat{\pi}_n$ to π_n^* in terms of statistical regret.

5.1 Assumptions

We make the following assumptions.

Assumption 5.1 (VC-class)**.** *There exists constants* $0 \leq \nu < 1/2$ *and* $N \geq 1$ *such that* $\mathrm{VC}(\Pi_n) \lesssim n^{\nu}$ *for all* $n \geq N$*.*

Assumption 5.1 restricts the policy class to have finite VC-dimension, which is a standard requirement for controlling the complexity of a policy class in the classification literature. The VC-dimension of the policy-class Π is defined as the largest interger m such that there exist points $x_1, \ldots, x_m$ that are shattered by Π, i.e. where the policy values $\pi(x_1), \ldots, \pi(x_m)$ can take on all 2^m possible combinations in $\{0,1\}^m$ (for more on VC-dimension, see Wainwright, 2019). Several practically relevant classes of treatment rules satisfy this requirement, including the linear-index and quadrant rules used in the empirical application of Section 6. Our assumption allows the VC-dimension of the policy class to grow moderately with the sample size, thus allowing the treatment rule to depend on high-dimensional covariates.

Assumption 5.2 (Regularity conditions for data-generating process)**.**

(i) There exist constants $\mathcal{C}_{1,\varphi}, \mathcal{C}_{1,\alpha}$ *such that for all* $\{\widetilde{g}, \widetilde{f}\} \in \mathcal{G} \times \mathcal{F}$

$$\|\varphi_\ell(\widetilde{g}; X) - \varphi_\ell(g; X)\|_{L_\infty(P_X)} \leq \mathcal{C}_{1,\varphi} \cdot \|\widetilde{g} - g\|_{L_\infty(P_V)},$$
$$\left\|\alpha_\ell(\{\widetilde{g}, \widetilde{f}\}; V) - \alpha_\ell(\{g, f\}; V)\right\|_{L_\infty(P_V)} \leq \mathcal{C}_{1,\alpha} \cdot \left(\|\widetilde{g} - g\|_{L_\infty(P_V)} + \left\|\widetilde{f} - f\right\|_{L_\infty(P_V)}\right),$$

for $\ell = 0, \ldots, L$*.*

(ii) *There exist constants* $\mathcal{C}_{2,\varphi}, \mathcal{C}_{2,\alpha}$ *such that for all* $\{\widetilde{g}, \widetilde{f}\} \in \mathcal{G} \times \mathcal{F}$

$$\|\varphi_\ell(\widetilde{g}; X) - \varphi_\ell(g; X)\|_{L_2(P_V)} \leq \mathcal{C}_{2,\varphi} \cdot \|\widetilde{g} - g\|_{L_2(P_V)},$$
$$\left\|\alpha_\ell(\{\widetilde{g}, \widetilde{f}\}; V) - \alpha_\ell(\{g, f\}; V)\right\|_{L_2(P_V)} \leq \mathcal{C}_{2,\alpha} \cdot \left(\|\widetilde{g} - g\|_{L_2(P_V)} + \left\|\widetilde{f} - f\right\|_{L_2(P_V)}\right),$$

for $\ell = 0, \ldots, L$.

(iii) *There exist constants* $\mathcal{C}_{3,\varphi}, \mathcal{C}_{3,\alpha}$ *such that for all* $\{\widetilde{g}, \widetilde{f}\} \in \mathcal{G} \times \mathcal{F}$

$$\|\varphi_\ell(\widetilde{g}; X)\|_{L_\infty(P_X)} \leq \mathcal{C}_{3,\varphi},$$
$$\left\|\alpha_\ell(\{\widetilde{f}, \widetilde{g}\}; V)\right\|_{L_\infty(P_V)} \leq \mathcal{C}_{3,\alpha},$$

for $\ell = 0, \ldots, L$.

(iv) *The irreducible noise* $\varepsilon_i := U_i - g(V_i)$ *is a sub-Gaussian vector conditional on* V_i*, with conditional variance* $\mathrm{Var}(\varepsilon_i \mid V_i) = \Sigma(V_i)$ *satisfying* $\|\lambda_{\max}(\Sigma(V))\|_{L_\infty(P_V)} \leq \overline{\lambda} < \infty$.

Assumptions 5(i) and 5(ii) impose Lipschitz continuity of the component functions and Riesz-representers with respect to the nuisance component in the L_∞ and L_2-norm, respectively. These requirements are typically met under mild conditions within the framework of this paper. For the Balke-Pearl bounds of Example 3.2, these assumptions hold under the overlap condition whenever $\mathcal{G}$ and $\mathcal{F}$ are subsets of the space of bounded functions[13], which is automatically satisfied since $U_i = (Y_i, D_i, Z_i)'$ is a vector of random variables with bounded support. Assumption 5.2(iii) is a uniform bound on the component functions and Riesz-representers, the former implying uniform boundedness of the scores $\Gamma(g; \cdot)$. Assumption 5.2(iv) is a standard requirement in statistical learning theory restricting the tail behaviour of statistical noise. It is automatically satisfied when U_i has bounded support, as in the Balke-Pearl bounds, but also allows for outcomes with unbounded support whose conditional distribution has sufficiently thin tails. Together with Assumption 5.2(iii), this assumption implies sub-gaussianity of $\Gamma^{\mathtt{NO}}(\{g, f\}, W_i)$.

The next two assumptions impose requirements on the estimators for the nuisance components.

Assumption 5.3 (Regularity conditions for fist-step estimators)**.**

(i) *The estimators of the nuisance functions* $\{\widehat{g}_n, \widehat{f}_n\}$ *belong to the function classes* $\mathcal{G} \times \mathcal{F}$ *with probability 1.*

[13] That is, there exists a constant $B > 0$ such that $\|\{\widetilde{g}, \widetilde{f}\}\|_{L_\infty(P_V)} \leq B$, $\forall \{\widetilde{g}, \widetilde{f}\} \in \mathcal{G} \times \mathcal{F}$

(ii) There exists a constant $\mathcal{C}_4 > 0$ such that

$$\|\widehat{g}_n - g\|_{L_\infty(P_V)} \leq \mathcal{C}_4,$$
$$\left\|\widehat{f}_n - f\right\|_{L_\infty(P_V)} \leq \mathcal{C}_4,$$

with probability approaching 1 as $n \to \infty$.

Part (i) of Assumption 5.3 is needed to ensure the validity of the Lipschitz continuity requirements of Assumption 5.2 for the component functions and Riesz-representers when evaluated at the first-stage estimates. In the context of the Balke-Pearl bounds, it is satisfied when $\widehat{g}_n, \widehat{f}_n$ are uniformly bounded and the estimated propensity score $\widehat{z}(X_i)$ is uniformly bounded away from 0 and 1, with probability one. The first condition is satisfied by virtually any estimation procedure when the outcomes U_i have bounded support. The second requirement can be guaranteed under appropriate trimming of the estimated propensities. Part (ii) requires that estimation errors for the nuisance components are uniformly bounded, which is satisfied under Assumption 5.3(i) when $\mathcal{G} \times \mathcal{F}$ is a subset of the space of bounded functions. When U_i has unbounded support and $\mathcal{G} \times \mathcal{F}$ includes unbounded functions, a more primitive condition for (ii) would be uniform consistency of the first stage estimates, that is $\|\{\widehat{g}_n, \widehat{f}_n\} - \{g, f\}\|_{L_\infty(P_V)} \to_p 0$.[14]

Assumption 5.4 (L_2 convergence rates)**.** *The estimators of the nuisance functions satisfy*

$$\mathbb{E}_{P_n}\left[\|\widehat{g}_n - g\|^2_{L_2(P_V)}\right] \leq \frac{r_n}{n^{1/2}},$$
$$\mathbb{E}_{P_n}\left[\left\|\widehat{f}_n - f\right\|^2_{L_2(P_V)}\right] \leq \frac{r_n}{n^{1/2}},$$

for some sequence $r_n = o(1)$.

The above requirement on the L_2-convergence rates for the learners of the nuisance functions is a standard assumption in the semiparametric inference literature (see, e.g., Farrell, 2015, and Chernozhukov et al., 2022). It can be shown to provably hold for traditional nonparametric estimation methods such as sieve methods (Chen, 2007) as well as modern black-box machine learning algorithms including Lasso (see, e.g., Farrell, 2015), deep neural networks (Farrell et al., 2021), boosting and others, for which stronger guarantees such as Donsker-type properties are typically not available. The ability to invoke a mild L_2-convergence requirement is a virtue of the combined use of Neyman-orthogonalization and sample-splitting, a key insight brought forward by Chernozhukov et al. (2022) for semiparametric GMM inference, and subsequently leveraged

[14] However, it should be noted that the uniform consistency requirement is not completely innocuous when $\{\widehat{g}_n, \widehat{f}_n\}$ are machine learning estimators (Farrell et al., 2021).

by Athey and Wager (2021) and Foster and Syrgkanis (2019) in the context of statistical learning problems.[15]

Finally, we present an assumption that concerns the distribution of the component functions φ_ℓ at the population level.

Assumption 5.5 (Margin). *There exist constants $\mathcal{C}_m > 0$ and $\gamma \geq 0$ such that*

$$\mathbb{P}_X\Big(0 < |\varphi_\ell(g; X)| \leq t\Big) \leq \mathcal{C}_m t^\gamma, \quad \forall\, t > 0.$$

for $\ell = 1, \dots, L$.

The above assumption restricts the extent to which the distribution of the component functions $\varphi_\ell(g; X)$ can concentrate around the point of non-differentiability 0 and it is a form of "margin assumption", first introduced by Mammen and Tsybakov (1999). Such an assumption has been widely used in statistics to obtain fast learning rates in classification problems (see, e.g., Arlot and Bartlett, 2011). Notice that the above formulation for the margin assumption restricts the concentration of probability for the distribution of the components functions in a neighbourhood of 0, but still allows for arbitrary probability mass at 0.

Example 5.1 ($\gamma = 1$). *Suppose X contains an absolutely continuous covariate $\tilde{\mathrm{x}}$ and $\varphi_\ell(g; X) \cdot \mathbb{1}\{\tilde{\mathrm{x}} \neq 0\}$ is absolutely continuous with density bounded above by $\overline{f}$ for $\ell = 1, \dots, L$. Then Assumption (5.5) holds with $\gamma = 1$ and $\mathcal{C}_m = 2\overline{f}$.*

Example 5.2 ($\gamma = \infty$). *Suppose there exists a $t_0 > 0$ such that $\mathbb{P}_X\left(0 < |\varphi_\ell(X)| < t_0\right) = 0$ for $\ell = 1, \dots, L$. Then Assumption (5.5) holds with $\gamma = \infty$ and some $\mathcal{C}_m > 0$.*

In the context of the Balke-Pearl bounds from Example 3.2 with resolution of ambiguity via Minimax Regret, Assumption 5.5 restricts the extent to which the CATE bounds $\overline{\tau}, \underline{\tau}$ can concentrate around 0 in the data-generating process. Under $\gamma = \infty$ the support of each CATE bound is required to be fully separated from 0, while $\gamma = 1$ requires that each CATE bound has bounded density in a neighborhood of 0.

In the next section we present our theoretical results based on Assumptions 5.1-5.5.

5.2 Regret convergence rates

In this section, we provide asymptotic rates of convergence for the regret of the estimated policy $R_n(P; \widehat{\pi}_n)$ as defined in (19). In line with the existing literature, we study *uniform* regret

[15] Unlike Athey and Wager (2021), our assumptions do not allow to trade-off accuracy in the estimation across the different nuisance functions. This is because our framework allows for $\varphi_\ell(g; x)$ to be a potentially non-linear functional of the nuisance functions g, as is the case in Examples 3.1-3.3, thus precluding such double-robustness property.

bounds that are valid for all distributions $P \in \mathcal{P}$ satisfying Assumptions 5.1-5.5. All results in this section are thus intended to hold uniformly in the above sense, and we will drop the dependence on P for notational convenience.

We begin by noticing that controlling the convergence of $\widehat{\pi}_n$ to the best-in-class policy π_n^* intuitively requires accounting for: 1) the estimation error in the component functions and influence function adjustments due to estimation of the nuisance components $\{g, f\}$, 2) the difference between the population Neyman-orthogonal score and true score[16], and 3) the fact that we estimate our policy using a sample from the distribution of the covariates X_i rather than their true distribution. We define the following quantities:

$$\widehat{Q}_n^{\mathtt{NO}}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\big(2\pi(X_i)-1\big)\cdot\Gamma^{\mathtt{NO}}(\{\widehat{g},\widehat{f}\};W_i),$$
$$Q_n^{\mathtt{NO}}(\pi) = \frac{1}{n}\sum_{i=1}^{n}\big(2\pi(X_i)-1\big)\cdot\Gamma^{\mathtt{NO}}(\{g,f\},W_i),$$

and formalize this intuition in the next proposition.

Proposition 3. *The regret of $\widehat{\pi}_n$ obeys the following bound:*

$$R_n(\widehat{\pi}_n) \leq 2\mathbb{E}\left[\sup_{\pi\in\Pi_n}\left|\widehat{Q}_n^{\mathtt{NO}}(\pi)-Q_n^{\mathtt{NO}}(\pi)\right|\right] + \mathbb{E}\left[\sup_{\pi\in\Pi_n}\left|Q_n^{\mathtt{NO}}(\pi)-Q(\pi)\right|\right]. \tag{20}$$

The second term in the above bound accounts for points 2) and 3). $Q_n(\pi) - Q(\pi)$ is a centred (mean-zero) empirical process and therefore its uniform expectation can be shown to be $O\left(\sqrt{\mathrm{VC}(\Pi_n)/n}\right)$ using symmetrization and chaining arguments (see, e.g., Wainwright, 2019, Ch. 5.3). Controlling the first term, which accounts for point 1), is particularly challenging and requires tailored arguments that deal with the particular form of the population scores in (14), in particular their lack of full differentiability.

Lemma 1. *Suppose that Assumptions 5.1-5.5 hold and define $\kappa_n = \lfloor n(1-1/K)\rfloor$. Then we have*

$$\mathbb{E}_{P_n}\left[\sup_{\pi\in\Pi_n}\left|\widehat{Q}_n^{\mathtt{NO}}(\pi)-Q_n^{\mathtt{NO}}(\pi)\right|\right] = O\left(\frac{r_{\kappa_n}}{\sqrt{n}} + \sqrt{\frac{\mathrm{VC}(\Pi_n)}{n}} + \left(\frac{r_{\kappa_n}}{\sqrt{n}}\right)^{\frac{\gamma+1}{\gamma+2}}\right).$$

Lemma 1 is the central result of this paper. It provides an asymptotic rate of convergence to zero of the empirical process $\left|\widehat{Q}_n^{\mathtt{NO}}(\pi)-Q_n^{\mathtt{NO}}(\pi)\right|$ uniformly over the policy class Π_n, which depends on the VC-dimension of the class and the degree of concentration of the component

[16]That is, we need to account for the fact that we have added the influence function adjustments to the component functions.

functions $\varphi_\ell(g;X)$ around 0, as indexed by γ. In order to convey intuition on this result we provide a brief outline of the proof, which is based on the decomposition

$$\begin{aligned}\widehat{Q}_n^{\texttt{NO}}(\pi) - Q_n^{\texttt{NO}}(\pi) &= \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\left[\widehat{\Gamma}^{\texttt{NO}}\left(\{\widehat{g}^{-k(i)},\widehat{f}^{-k(i)}\},W_i\right) - \Gamma^{\texttt{NO}}\left(\{g,f\},W_i\right)\right] \\ &= A_0(\pi) + \sum_{\ell=1}^{L} a_\ell\cdot\left[A_{1,\ell}(\pi)+A_{2,\ell}(\pi)+A_{3,\ell}(\pi)\right],\end{aligned} \tag{21}$$

where

$$A_0(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\left[\varphi_0^{\texttt{NO}}(\{\widehat{g}^{-k(i)},\widehat{f}^{-k(i)}\},W_i) - \varphi_0^{\texttt{NO}}(\{g,f\},W_i)\right],$$

$$A_{1,\ell}(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\left[\varphi_\ell^{\texttt{NO}}(\{\widehat{g}^{-k(i)},\widehat{f}^{-k(i)}\},W_i) - \varphi_\ell^{\texttt{NO}}(\{g,f\},W_i)\right]\cdot\mathbb{1}\left\{\varphi_\ell\left(\widehat{g}^{-k(i)};X_i\right)>0\right\},$$

$$A_{2,\ell}(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\phi_\ell(\{g,f\};W_i)\cdot\left[\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right],$$

$$A_{3,\ell}(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\varphi_\ell(g;X_i)\cdot\left[\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right].$$

Terms $A_0(\pi)$ and $A_{1,\ell}(\pi)$ can be controlled using similar arguments to Athey and Wager (2021) and are responsible for the $O\left(r_{\kappa_n}/\sqrt{n}\right)$ term in the bound of Lemma 1. The de-biasing properties of Neyman-orthogonalization combined with sample-splitting play a crucial role in this context, as they ensure that the error in estimating $\varphi_\ell(g;x)$ only has a second-order contribution. As a result, term $A_{1,\ell}(\pi)$ scales with the mean-squared estimation error in the nuisance functions and, under Assumption 5.4, its expectation decays faster than $1/\sqrt{n}$ uniformly over Π_n.[17] If plug-in (non-orthogonalized) estimates for φ_ℓ are instead used to form the score estimates $\widehat{\Gamma}_i$, the estimation error in the nuisance functions has a first-order impact on term $A_{1,\ell}(\pi)$. As a result, its uniform expectation would scale with the L_1 estimation error which, under Assumption 5.4, implies the much slower convergence $\mathbb{E}[\sup_{\pi\in\Pi_n}|A_{1,\ell}(\pi)|] = o(n^{1/4})$.

For term $A_{2,\ell}(\pi)$, the mean-zero property of the influence function adjustments together with sample-splitting ensures that this term is a centred empirical process and thus it is responsible for a $O\left(\sqrt{\text{VC}(\Pi_n)/n}\right)$ contribution again by symmetrization and chaining arguments.

Finally, for term $A_{3,\ell}(\pi)$ we show that

$$\mathbb{E}\left[\sup_{\pi\in\Pi_n} A_{3,\ell}(\pi)\right] \leq \mathbb{E}\left[\left|\varphi_\ell(g;X_i)\cdot\left(\mathbb{1}\left\{\varphi_\ell^{-k(i)}(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\right|\right],$$

[17]Notice that, by virtue of sample-splitting, the presence of the indicator $\mathbb{1}\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\}$ is immaterial when controlling the expectation of $A_{1,\ell}(\pi)$ uniformly over Π_n.

where the RHS can be recognized to be the classification loss of an estimator for the sign of $\varphi_\ell(g;x)$ based on thresholding $\varphi_\ell(\widehat{g}^{-k(i)};x)$. Rates of convergence in binary classification problems intuitively depend on the degree of separation of the true regression function from 0, as indexed by γ. We thus leverage results from the literature on classification (Audibert and Tsybakov, 2007) to quantify the contribution of $A_{3,\ell}$ in the bound of Lemma 1 in terms of γ.

We are now ready to combine the rates of convergence for the three terms in Proposition 3 to obtain a final regret bound for our proposed estimation procedure.

Theorem 2. *Suppose Assumptions 5.2-5.5 hold. Then the regret obeys*

$$R_n(\widehat{\pi}_n) = O\left(\sqrt{\frac{\mathrm{VC}(\Pi_n)}{n}} \vee \left(\frac{r_{\kappa_n}}{\sqrt{n}}\right)^{\frac{\gamma+1}{\gamma+2}}\right).$$

We see that regret convergence for our policy learning procedure happens at a rate corresponding to whichever is the leading term in the asymptotic expansion of Lemma 1, which depends on ν and γ. When the policy class Π_n has fixed VC-dimension ($\nu = 0$), regret convergence happens at rates ranging from $o(n^{1/4})$ in the least favourable case ($\gamma = 0$) to $O(\sqrt{\mathrm{VC}(\Pi)/n})$ in the most favourable case ($\gamma = \infty$). The latter case is in line with existing results for policy learning with point identified CATE, in which full-differentiability of the scores leads to $\sqrt{\mathrm{VC}(\Pi_n)/n}$ learning rates (see Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Foster and Syrgkanis, 2019). For the intermediate case $\gamma = 1$ of Example 5.1 our procedure guarantees regret convergence at rate $o(n^{1/3})$.

It is useful to compare the performance guarantees in this paper with Pu and Zhang (2021), whose procedure involves the use of non-orthogonalized estimates for the scores with sample-splitting. They show that the regret of a policy estimated via the maximization (15) based on cross-fitted non-orthogonalized scores is upper bounded by the L_1-norm of the estimation error in the nuisance functions. Under Assumption 5.4, this implies $o(n^{1/4})$ convergence for the regret, which is strictly slower than our rates *for all values* of $\gamma > 0$. The faster speed of convergence guaranteed by our procedure is not just due to a refined proof strategy but crucially depends on the use of Neyman-orthogonalization, as elucidated by our discussion of Lemma 1.

Remark 4. *The procedure of Pu and Zhang (2021) also differs from ours in its final implementation, which in their case is carried out via support vector machines (SVM) with Π_n assumed to be a reproducing kernel Hilbert space. While the use of surrogate losses (such as the hinge loss in SVM) to convexify problem (15) can bring considerable computational benefits in terms of speed and scalability, it comes at the cost of even slower convergence guarantees than the $o(n^{-1/4})$ discussed above. We stress that our insights regarding the benefits of Neyman-orthogonalization in terms of faster learning rates apply irrespective of the final implementation. Notice also that the use of surrogate loss functions does not guarantee convergence of the estimated optimal policy* *to the best-in-class π_n^* in general when the policy class Π_n does not contain the "first-best" policy $\mathbb{1}\{\Gamma(g;x) \geq 0\}$, as shown by the recent work of Kitagawa et al. (2021).*

6 Empirical application

In this section we apply the methods discussed in this paper to data from the National Job Training Partnership Act (JTPA) Study. This study randomly selected applicants to receive various training and services, including job-search assistance, for a period of 18 months. The study collected background information on applicants before random assignment and then recorded their earnings in the 30-month period following treatment assignment. Kitagawa and Tetenov (2018) apply their EWM method to a sample of 9,223 adult JTPA applicants to estimate the optimal allocation of *eligibility* into the programme that maximizes individual earnings across the population. In particular, they take total individual earnings in the 30 months after assignment as the welfare outcome measure Y_i, and consider policies that allocate eligibility in the programme based on the individual's observable characteristics. Kitagawa and Tetenov's analysis is from an *intent-to-treat* perspective as they focus on the problem of deciding who should be given eligibility to participate in the programme. Since eligibility in the JTPA study is randomly assigned, the effect of eligibility on earnings is point identified from the data and methods for policy learning under point-identification can be applied in this setting. We depart from Kitagawa and Tetenov (2018) and instead consider optimal assignment of *actual participation* in the training. This analysis would be of interest to a policy-maker that expects to achieve (close to) perfect compliance to her treatment decision, e.g. when participation is made a condition for receipt of a sufficiently generous unemployment benefit. Compliance in the JTPA study is imperfect as roughly 23% of applicants' participation status $D_i = 0, 1$ deviates from their assigned eligibility status $Z_i = 0, 1$, as shown in Table 2. As a result, random assignment of the eligibility instrument Z_i is not sufficient to point-identify the effect of participation in the training, motivating the use of the methods of this paper.

For partial identification of the CATE we consider the Balke-Pearl scheme of Example 3.2, where bounds for the 30-month post-treatment earnings are $Y_L = \$0$ and $Y_U = \$59,640$.[18] We compare this with point-identification of the CATE as the conditional local average treatment effect (LATE), predicated under the assumption of no unobserved heterogeneity. We subtract \$1200 from both the CATE bounds and the conditional LATE; this is the average cost of services

[18]The outcome upper bound corresponds to the 97.5th percentile of the earnings distribution rather than highest recorded value of $\$155,760$. Outcome bounds in Balke-Pearl bounds effectively impute unidentified expected earnings for never-takers and always-takers. Restricting expected earnings to be below such high quantile is in effect a mild requirement which brings considerable identification power.

Table 2: Joint distribution of eligibility and participation, JTPA study

	Eligibility (Z_i)		
Participation (D_i)	0	1	Total
0	3047	2118	5165
1	43	4015	4058
Total	3090	6133	9233

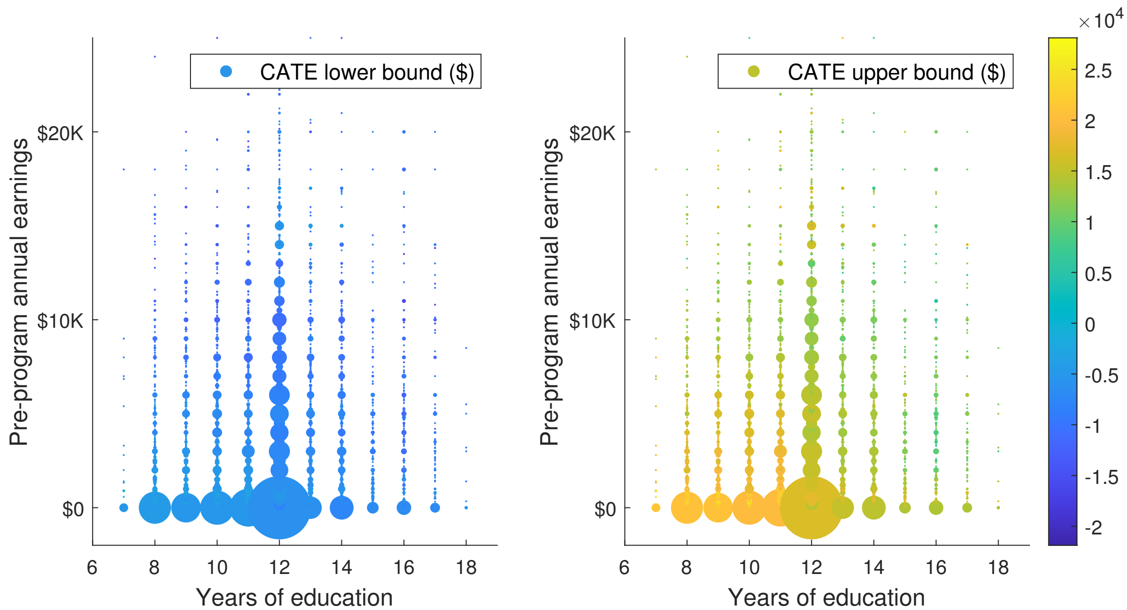
Data source: Kitagawa and Tetenov (2018) and Abadie, Angrist, and Imbens (2002).

per actual treatment, estimated from Table 5 in Bloom et al. (1997). Following Kitagawa and Tetenov (2018), we condition treatment assignment on two pre-treatment variables: the individual's years of education and earnings in the year prior to assignment. Estimation of the optimal policy follows the procedure described in Section 4, with $K = 10$ evenly-sized data folds used to form cross-fitted Neyman-orthogonal estimates for the CATE bounds and conditional LATE functions. The nuisance functions are estimated via boosted regression trees, performed by the MATLAB function `fitrensenmble`.[19] Figure 1 demonstrates cross-fitted plug-in estimates for the CATE bounds (a) and LATE/MMR scores (b), where the size of the dots indicates the number of individuals with different covariate values. We notice that the estimated CATE lower bounds are negative for the whole sample, and thus the maximin impact policy never assigns treatment in this application. We will therefore focus our analysis on minimax regret (MMR).[20]

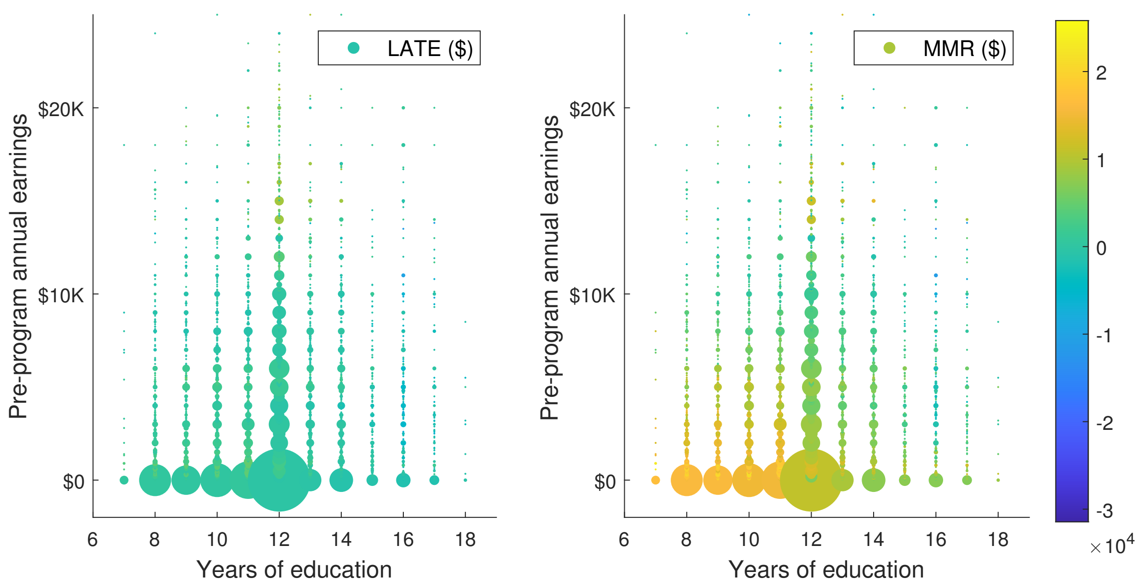
We consider two alternative choices for the candidate policy class Π. The first is the class of quadrant treatment policies. To be assigned to treatment according to this policy, an individual's education and pre-program earnings have to be above (or below) some specific threshold. Figure 3 illustrates the quadrant treatment policies selected by our proposed method, where the colored shaded areas indicate individuals assigned to treatment by the two respective policies. The optimal MMR policy assigns treatment to individuals with education below 15 years and pre-treatment earnings below \$23,551. The optimal LATE policy features the same threshold for education, but selects individuals with pre-treatment earnings above \$200 for treatment.

[19]Tuning parameters have been chosen via cross-validation within each data-fold. For further details on the estimation procedure we refer to the MATLAB documentation for the command.

[20]Maximin welfare also results in no treatment for the whole population in this application.



(a) CATE bounds



(b) Scores

Figure 2: JTPA - Plug-in cross-fitted estimates (net of $1200)

Table 3: Treatment proportions of alternative treatment assignment policies

	Share of Population to be treated	Share of Population receiving same treament under MMR and LATE
Quadrant Rule		0.68
Minimax Regret	0.96	–
LATE	0.64	–
Linear Index Rule		0.70
Minimax Regret	0.95	–
LATE	0.69	–

The rows labeled "Minimax Regret" give information on estimated optimal minimax regret policies based on the scores in Equation (11) with Balke-Pearl CATE bounds. The rows labeled "LATE" give information on optimal policies for scores that identify the conditional LATE.

While the two policies appear similar, they substantially differ in the proportion of population assigned to treatment (96% by the MMR policy versus 64% by the LATE policy). This is due to the large concentration of individuals with pre-treatment earnings close to (or equal) zero. As a result, 32% of individuals receive a different treatment assignment across the two policies. Second, we consider the class of linear treatment policies. This class consists of policies that assign treatment to an individual according to whether a linear index in his observable characteristics is above a certain threshold. Figure 4 illustrates how the direction of treatment assignment as a function of prior earnings differs between the MMR and LATE policy in a similar fashion to the quadrant rules; contrary to the LATE policy, MMR prioritizes treatment assignment to individuals with lower pre-program earnings. Nonetheless, 70% of the population still receives the same treatment under the two different policies, in light of the relatively low concentration of individuals in the areas of the covariate space where the two policies differ. Similarly to the quadrant policy rule, the MMR policy assigns treatment to a larger share of the population (95%) compared to the LATE policy (69%).

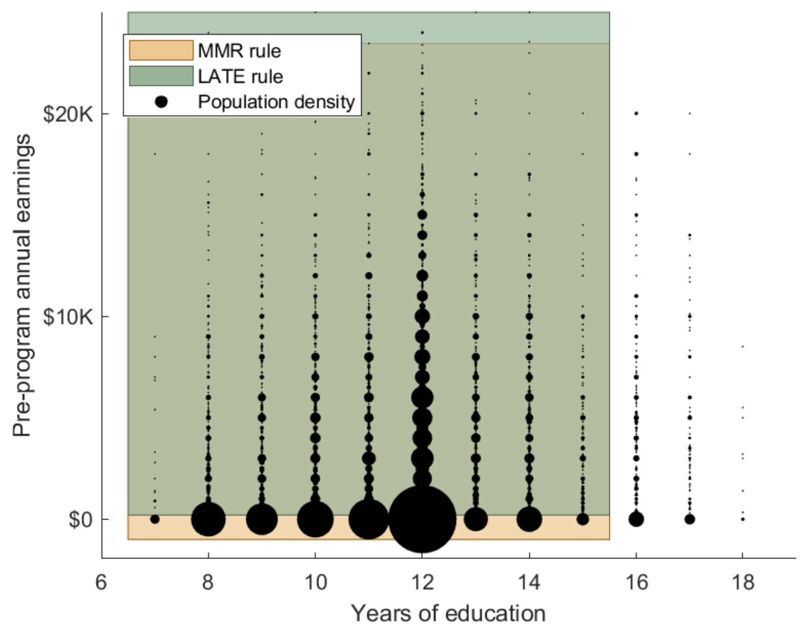


Figure 3: Estimated optimal policies from the quadrant policy class

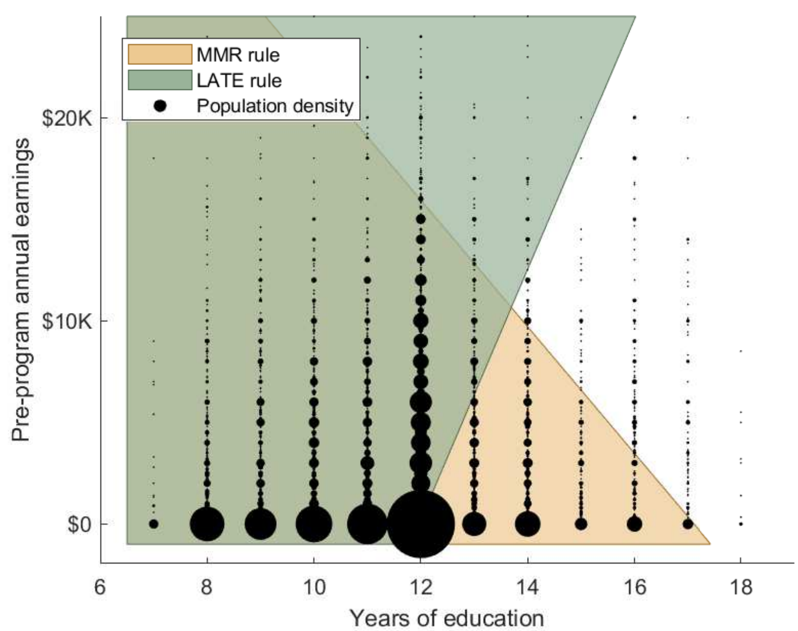


Figure 4: Estimated optimal policies from the linear-index policy class

7 Conclusion

This paper develops a general policy learning framework for estimation of individualized treatment rules when treatment effects are partially identified. By drawing connections between the treatment assignment problem and classical decision theory, we have characterized several notions of optimal treatment policies in the presence of partial identification. We have shown how partial identification leads to a new policy learning problem where the risk is only directionally-differentiable with respect to a nuisance infinite-dimensional component. We have proposed an estimation procedure that ensures Neyman-orthogonality with respect to the nuisance components and we have provide statistical guarantees that depend on the amount of concentration around the points of non-differentiability in the data-generating process. Our proposed methods are illustrated with an application to the Job Training Partnership Act study, where we have shown that allowing for partial identification delivers substantially different programme participation policies compared to existing methods that assume point-identification.

There are several avenues for future research. First, it would be interesting to extend the theory of this paper to partial identification via instrumental variables with continuous support. Second, it would be useful to extend the methods to more general identification sets that incorporate smoothness restrictions on unobserved counterfactual quantities, such as those considered in Kim et al. (2018). Finally, it would be interesting to assess the optimality of our proposed estimation procedure by deriving minimax lower bound rates for semiparametric statistical learning problems with directionally-differentiable risk.

Appendices

A Extension to nested min/max operators

In this section we describe how the proposed estimation procedure and the theoretical results of Section 4 can be extended to scores $\Gamma(g;X)$ that feature nested linear combinations of min / max operators. This extension comprises min / max operators over multiple components, since $\max\{a,b,c\} = \max\{\max\{a,b\},c\}$.

We begin by noticing that our proposed estimation described in Section 4 can be also defined as follows. First, for each $\min\{a(g;x), b(g;x)\}$ (or max) operator contained in $\Gamma(g;x)$, one substitutes the operator with $a(g;x)$ or $b(g;x)$ based on their cross-fitted plug-in (non-orthogonalized) estimates $\widehat{a}_i := a(\widehat{g}^{-k(i)};X_i)$ and $\widehat{b}_i := b(\widehat{g}^{-k(i)};X_i)$. Then, the selected component is estimated ($a(g;x)$ or $b(g;x)$) is estimated by their cross-fitted Neyman-orthogonal analogue ($\widehat{a}_i^{\mathtt{NO}}$ or $\widehat{b}_i^{\mathtt{NO}}$). In the presence of nested min / max operators, our estimation is generalized as follows. First, in succession from the most inner to the most outer min / max, each operator is substituted with their smallest/largest argument based on cross-fitted non-orthogonalized estimates. Then, the selected components are estimated by their cross fitted Neyman-orthogonal analogue. As an illustration, consider the hypothetical score

$$
\begin{aligned}
\Gamma(g;x) &= d(g;x) + \max\{c(g;x) + \min\{a(g;x), b(g;x)\}, 0\} \\
&= d(g;x) + \max\{\underbrace{c(g;x) + b(g;x) + (a(g;x) - b(g;x)) \cdot \mathbb{1}\{a(g;x) - b(g;x) \leq 0\}}_{\varrho(g;x)}, 0\} \\
&= d(g;x) + \varrho(g;x) \cdot \mathbb{1}\left\{\varrho(g;x) \geq 0\right\}.
\end{aligned}
$$

Applying the above procedure to this example gives the following expression for the estimated Neyman-orthogonal score:

$$
\Gamma^{\mathtt{NO}}(\{\widehat{g}^{-k(i)}, \widehat{f}^{-k(i)}\}, W_i) = \widehat{d}_i^{\mathtt{NO}} + \left[\widehat{c}_i^{\mathtt{NO}} + \widehat{b}_i^{\mathtt{NO}} + (\widehat{a}_i^{\mathtt{NO}} - \widehat{b}_i^{\mathtt{NO}}) \cdot \mathbb{1}\{\widehat{a}_i - \widehat{b}_i \leq 0\}\right] \cdot \mathbb{1}\left\{\widehat{\varrho}_i \geq 0\right\},
$$

where

$$
\widehat{\varrho}_i = \widehat{c}_i + \widehat{b}_i + (\widehat{a}_i - \widehat{b}_i) \cdot \mathbb{1}\{\widehat{a}_i - \widehat{b}_i \leq 0\}.
$$

We will now show how the theoretical results of Section 5 can be generalized to this example.[21]

[21] The extension to general scores containing an arbitrary finite number of nested min / max operators follows immediately from our discussion of this example.

Following the arguments of Section 5.2, we have

$$\begin{aligned}
\widehat{Q}_n^{\texttt{NO}}(\pi) - Q_n^{\texttt{NO}}(\pi) &= \sum_{i=1}^{n}(2\pi(X_i)-1)\cdot(\widehat{d}_i^{\texttt{NO}} - d_i^{\texttt{NO}}) \\
&+ \sum_{i=1}^{n}(2\pi(X_i)-1)\cdot(\widehat{c}_i^{\texttt{NO}} + \widehat{b}_i^{\texttt{NO}} - c_i^{\texttt{NO}} - b_i^{\texttt{NO}})\cdot \mathbb{1}\{\widehat{\varrho}_i \geq 0\} \\
&+ \sum_{i=1}^{n}(2\pi(X_i)-1)\cdot(\widehat{a}_i^{\texttt{NO}} - \widehat{b}_i^{\texttt{NO}})\cdot \mathbb{1}\{\widehat{a}_i - \widehat{b}_i \leq 0\}\cdot \mathbb{1}\{\widehat{\varrho}_i \geq 0\} \\
&+ \sum_{i=1}^{n}(2\pi(x)-1)\cdot(a_i^{\texttt{NO}} - b_i^{\texttt{NO}})\cdot \Big[\mathbb{1}\{\widehat{a}_i - \widehat{b}_i \leq 0\} - \mathbb{1}\{a_i - b_i \leq 0\}\Big]\cdot \mathbb{1}\{\widehat{\varrho}_i \geq 0\} \\
&+ \sum_{i=1}^{n}(2\pi(x)-1)\cdot \varrho_i^{\texttt{NO}}\cdot[\mathbb{1}\{\widehat{\varrho}_i \geq 0\} - \mathbb{1}\{\varrho_i \geq 0\}].
\end{aligned}$$

The first term in the expansion has the same structure as $A_{0,\ell}$ and thus obeys the same bound. The second and third term obey the same bound as $A_{1,\ell}$ since, by virtue of sample-splitting, the indicators $\mathbb{1}\{\widehat{\varrho}_i \geq 0\}$ and $\mathbb{1}\{\widehat{a}_i - \widehat{b}_i \geq 0\}$ are immaterial when controlling the expectation of these term uniformly over Π_n (see arguments in the Proof of Lemma 1). The fourth term has the same structure as $A_{2,\ell} + A_{3,\ell}$ except for the presence of the indicator $\mathbb{1}\{\widehat{\varrho}_i \geq 0\}$, which again can be shown to be immaterial for controlling the $A_{2,\ell}$-like term by virtue of sample-splitting. For the $A_{3,\ell}$-like term we instead have the bound

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\pi\in\Pi_n}\frac{1}{n}\sum_{i=1}^{n}(2\pi(x)-1)\cdot(c_i-b_i)\cdot\Big(\mathbb{1}\{\widehat{c}_i - \widehat{b}_i \leq 0\} - \mathbb{1}\{c_i - b_i \leq 0\}\Big)\cdot \mathbb{1}\{\widehat{\varrho}_i \geq 0\}\right] \\
&\leq \mathbb{E}\left[\left|(c_i-b_i)\cdot\Big(\mathbb{1}\{\widehat{c}_i - \widehat{b}_i \leq 0\} - \mathbb{1}\{c_i - b_i \leq 0\}\Big)\right|\cdot|\mathbb{1}\{\widehat{\varrho}_i \geq 0\}|\right] \\
&\leq \mathbb{E}\left[\left|(c_i-b_i)\cdot\Big(\mathbb{1}\{\widehat{c}_i - \widehat{b}_i \leq 0\} - \mathbb{1}\{c_i - b_i \leq 0\}\Big)\right|\right],
\end{aligned}$$

which can be bounded in the same fashion as $A_{3,\ell}$ under a margin assumption on $a(g;X) - b(g;X)$. Finally, the fifth term also has the same structure as $A_{2,\ell} + A_{2,\ell}$, and can be controlled using arguments from Section 5.2 under a margin assumption on $\rho(g;X)$.

B Proofs

B.1 Proof of Proposition 1

For the maximin welfare policy we have

$$\begin{aligned}
\min_{(y_0,y_1)\in\mathcal{Y}}\mathbb{E}_{P_X}\left[\pi(X)\cdot y_{\pi(X)}(X)\right] &= \mathbb{E}_{P_X}\left[\min_{(y_0(x),y_1(x))\in\mathcal{Y}(x)}\pi(X)\cdot y_{\pi(X)}(X)\right] \\
&= \mathbb{E}_{P_X}\left[\pi(X)\cdot \underline{y}_1(X) + (1-\pi(X))\cdot \underline{y}_0(X)\right],
\end{aligned}$$

where the first equality is justified by 3.1. Thus we have

$$\underset{\pi\in\Pi}{\operatorname{argmax}} \min_{(y_0,y_1)\in\mathcal{Y}} W_\tau(\pi) = \underset{\pi\in\Pi}{\operatorname{argmax}} \mathbb{E}_{P_X}\left[\pi(X)\cdot \underline{y}_1(X) + (1-\pi(X))\cdot \underline{y}_0(X)\right].$$

For the maximin impact policy we have

$$\min_{\tau\in\mathcal{T}} \mathbb{E}_{P_X}\left[\pi(X)\cdot\tau(X)\right] = \mathbb{E}_{P_X}\left[\min_{\tau\in\mathcal{T}} \pi(X)\cdot\tau(X)\right] = \mathbb{E}_{P_X}\left[\pi(X)\cdot\underline{\tau}(X)\right].$$

where the first equality is justified by Assumption 3.2. Thus we have

$$\underset{\pi\in\Pi}{\operatorname{argmax}} \min_{\tau\in\mathcal{T}} W_\tau(\pi) = \underset{\pi\in\Pi}{\operatorname{argmax}} \mathbb{E}_{P_X}\left[\pi(X)\cdot\underline{\tau}(X)\right].$$

The statement of the proposition again follows from the invariance of the maximizer to positive affine transformations of the objective function.

B.2 Proof of Proposition 2

We notice that

$$\begin{aligned}
&\max_{\tau\in\mathcal{T}}\left(\max_{\pi:\mathcal{X}\to\{0,1\}} W_\tau(\pi) - W_\tau(\pi)\right)\\
&= \max_{\tau\in\mathcal{T}} \mathbb{E}_{P_X}\left[\left(\frac{1}{2}+\frac{1}{2}\operatorname{sgn}(\tau(X)) - \pi(X)\right)\cdot\tau(X)\right]\\
&= \mathbb{E}_{P_X}\left[\max_{\tau\in\mathcal{T}}\left(\frac{1}{2}+\frac{1}{2}\operatorname{sgn}(\tau(X)) - \pi(X)\right)\cdot\tau(X)\right]\\
&= \mathbb{E}_{P_X}\Bigg[\underbrace{\max_{\tau\in\mathcal{T}} \mathbb{1}\big\{\underline{\tau}(x)\geq 0\big\}\cdot(1-\pi(X))\cdot\tau(X)}_{=(1-\pi)\cdot\mathbb{1}\{\underline{\tau}(x)\geq 0\}\cdot\overline{\tau}(X)}\Bigg] + \mathbb{E}_{P_X}\Bigg[\underbrace{\max_{\tau\in\mathcal{T}} \mathbb{1}\big\{\overline{\tau}(x)\leq 0\big\}\cdot -\pi(X)\cdot\tau(X)}_{=-\pi\cdot\mathbb{1}\{\underline{\tau}(x)\geq 0\}\cdot\underline{\tau}(X)}\Bigg]\\
&\qquad + \mathbb{E}_{P_X}\Bigg[\underbrace{\max_{\tau\in\mathcal{T}} \mathbb{1}\big\{\underline{\tau}(x)<0<\overline{\tau}(x)\big\}\cdot\left(\frac{1}{2}+\frac{1}{2}\operatorname{sgn}(\tau(X)) - \pi(X)\right)\cdot\tau(X)}_{=-\pi\cdot\mathbb{1}\{\underline{\tau}(X)<0<\overline{\tau}(X)\}(\overline{\tau}(X)-\underline{\tau}(X))+\mathbb{1}\{\underline{\tau}(X)<0<\overline{\tau}(X)\}\cdot\overline{\tau}(X)}\Bigg]\\
&= -\mathbb{E}_{P_X}\left[\pi\cdot\left(\mathbb{1}\big\{\underline{\tau}(X)\geq 0\big\}\cdot\overline{\tau}(X) + \mathbb{1}\big\{\overline{\tau}(X)\leq 0\big\}\cdot\underline{\tau}(X) + \mathbb{1}\big\{\underline{\tau}(X)<0<\overline{\tau}(X)\big\}\cdot\big(\overline{\tau}(X)-\underline{\tau}(X)\big)\right)\right]\\
&\qquad + \mathbb{E}_{P_X}\left[\overline{\tau}(X)\cdot\mathbb{1}\big\{\underline{\tau}(X)\geq 0\big\} + \mathbb{1}\left\{\underline{\tau}(X)<0<\overline{\tau}(X)\right\}\cdot\overline{\tau}(X)\right],
\end{aligned}$$

where the first equality uses the fact $\operatorname{argmax}_{\pi:\mathcal{X}\to\{0,1\}} W_\tau(\pi) = \frac{1}{2}+\frac{1}{2}\operatorname{sgn}(\tau(X))$, and the second equality uses Assumption 3.2. The statement of the proposition then follows from the invariance of the maximizer to positive affine transformations of the objective function.

B.3 Proof of Proposition 3

We being by decomposing regret as follows:

$$Q(\pi_n^*) - Q(\widehat{\pi}_n) = \left[Q(\pi_n^*) - Q_n^{\texttt{NO}}(\pi_n^*)\right] + \left[Q_n^{\texttt{NO}}(\pi_n^*) - \widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n)\right] + \left[\widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n) - Q(\widehat{\pi}_n)\right]. \tag{22}$$

The first term is zero in expectation. The second term can be upper bounded as

$$\left[Q_n^{\texttt{NO}}(\pi_n^*) - \widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n)\right] \leq \left[Q_n^{\texttt{NO}}(\pi_n^*) - \widehat{Q}_n^{\texttt{NO}}(\pi_n^*)\right] + \left[\widehat{Q}_n^{\texttt{NO}}(\pi_n^*) - \widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n)\right] \leq \sup_{\pi \in \Pi_n} \left|Q_n^{\texttt{NO}}(\pi) - \widehat{Q}_n^{\texttt{NO}}(\pi)\right|,$$

where we have used that $\widehat{Q}_n^{\texttt{NO}}(\pi^*) - \widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n) \leq 0$, which follows from $\widehat{\pi}_n$ being the maximizer of $\widehat{Q}_n^{\texttt{NO}}(\cdot)$. The third term can be further expanded and upper bounded as follows

$$\widehat{Q}_n^{\texttt{NO}}(\widehat{\pi}_n) - Q(\widehat{\pi}_n) \leq \sup_{\pi \in \Pi_n} \left|\widehat{Q}_n^{\texttt{NO}}(\pi) - Q_n^{\texttt{NO}}(\pi)\right| + \sup_{\pi \in \Pi_n} \left|Q_n^{\texttt{NO}}(\pi) - Q(\pi)\right|.$$

Using the last two displays and taking expectations in (22) yields the desired conclusion.

B.4 Proof of Lemma 1

We will establish each of the following bounds in turn:

$$\begin{aligned}
\mathbb{E}\left[\sup_{\pi \in \Pi_n} |A_0(\pi)|\right] &= O\left(\sqrt{\mathrm{VC}(\Pi_n) \cdot \frac{r_{\kappa_n}}{n^{3/2}}} + \frac{r_{\kappa_n}}{n^{1/2}}\right), \\
\mathbb{E}\left[\sup_{\pi \in \Pi_n} |A_{1,\ell}(\pi)|\right] &= O\left(\sqrt{\mathrm{VC}(\Pi_n) \cdot \frac{r_{\kappa_n}}{n^{3/2}}} + \frac{r_{\kappa_n}}{n^{1/2}}\right), \\
\mathbb{E}\left[\sup_{\pi \in \Pi_n} |A_{2,\ell}(\pi)|\right] &= O\left(\sqrt{\frac{\mathrm{VC}(\Pi_n)}{n}}\right) \\
\mathbb{E}\left[\sup_{\pi \in \Pi_n} |A_{3,\ell}(\pi)|\right] &= O\left(\left(\frac{r_{\kappa_n}}{n}\right)^{\frac{\gamma+1}{\gamma+2}}\right)
\end{aligned}$$

Combining the above through decomposition (21) gives the desired final bound.

Bound for A_0 and $A_{1,\ell}$

We prove a bound for $\sup_{\pi \in \Pi_n} |A_{1,\ell}(\pi)|$; it will be immediate that $A_0(\pi)$ obeys the same bound. We begin with the following decomposition

$$\begin{aligned}
A_{1,\ell}(\pi) = &\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1) \cdot \langle \widehat{\alpha}_i - \alpha_i, U_i - g(V_i)\rangle, && (= B_1(\pi)), \\
&+ \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1) \cdot \langle \widehat{\alpha}_i - \alpha_i, g_i - \widehat{g}_i\rangle, && (= B_2(\pi)), \\
&+ \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1) \cdot \left(\varphi_\ell(\widehat{g}^{-k(i)}; X_i) - \varphi_\ell(g; X_i) + \langle \alpha_i, g_i - \widehat{g}_i\rangle\right), && (= B_3(\pi)),
\end{aligned}$$

where we have used the shorthand notation $\widehat{g}_i := \widehat{g}^{-k(i)}(V_i)$, $g_i = g(V_i)$, $\widehat{\alpha}_i := (\{\widehat{g}^{-k(i)}, f^{-k(i)}\}; V_i)$, $\alpha_i := (\{g, f\}; V_i)$. Starting with $B_1(\pi)$, the contribution of the k-th fold is

$$B_1^{(k)}(\pi) = \frac{1}{n}\sum_{i:k(i)=k}(2\pi(X_i) - 1) \cdot \langle \widehat{\alpha}_i - \alpha_i, \varepsilon_i\rangle \cdot \mathbb{1}\{\widehat{\varphi}_i \geq 0\}.$$

The sample-splitting procedure guarantees that $\{\widehat{g}^{-k(i)}, \widehat{f}^{-k(i)}\}$ only depend on data from the remaining $K-1$ folds, and thus conditioning on these estimates for the nuisance components makes $B_1^{(k)}(\pi)$ a sum of independent mean-zero terms, in light of

$$\mathbb{E}\left[U_i - g(V_i) \mid V_i, \widehat{g}^{-k(i)}, \widehat{f}^{-k(i)}\right] = 0.$$

Furthermore, the terms are also sub-Gaussian since it is a linear combination of sub-Gaussian random variables with bounded weights w.p.a 1, in light of

$$\|\mathbb{1}\{\widehat{\varphi}_i \geq 0\} \cdot (\widehat{\alpha}_i - \alpha_i)\|_{L_\infty(P_V)} \leq \left\|\widehat{g}^{-k}(V) - g(V)\right\|_{L_\infty(P_V)} + \left\|\widehat{f}^{-k}(V) - f(V)\right\|_{L_\infty(P_V)} \leq 2 \cdot \mathcal{C}_2 \cdot \mathcal{C}_3,$$

w.p.a 1, where the first inequality uses Assumption 5.2(i) and the second inequality uses Assumption 5.2(iii). Having computed the variance of $B_1^{(k)}(\pi)$ conditional on $\left(\widehat{g}^{-k}, \widehat{f}^{-k}\right)$

$$V_n(k) = \mathbb{E}\left[(\widehat{\alpha}_i^{-k} - \alpha_i)' \, \Sigma(V_i) \, (\widehat{\alpha}_i^{-k} - \alpha_i) \cdot \mathbb{1}\left\{\widehat{\varphi}_i \geq 0\right\} \mid \widehat{g}^{-k}, \widehat{f}^{-k}\right],$$

we can apply Corollary 3 in Athey and Wager (2021) to establish the bound

$$\frac{n}{n_k} \mathbb{E}\left[\sup_{\pi \in \Pi} \left|B_1^{(k)}(\pi)\right| \mid \hat{g}^{-k}\right] = O\left(\sqrt{V_n(k) \frac{\mathrm{VC}(\Pi_n)}{n_k}}\right), \tag{23}$$

where n_k denotes the number of observations in the k-th fold. Using Assumptions 5.2(ii) and 5.4, we have

$$\begin{aligned}
\mathbb{E}\left[V_n(k)\right] &\leq \mathbb{E}_{P_n}\left[\overline{\lambda} \cdot \|\widehat{\alpha}_i - \alpha_i\|^2_{L_2(P_V)}\right] \\
&\leq 2 \cdot \overline{\lambda} \cdot \mathcal{C}^2_{2,\alpha} \cdot \mathbb{E}_{P_n}\left[\left\|\widehat{g}^{-k(i)} - g\right\|^2_{L_2(P_V)} + \left\|\widehat{f}^{-k(i)} - f\right\|^2_{L_2(P_V)}\right] \\
&= O\left(\frac{r_{\kappa_n}}{\sqrt{n}}\right).
\end{aligned} \tag{24}$$

Finally, we apply (23) repeatedly for each of the K data-folds and using Jensen's Inequality and (24) and obtain the final bound

$$\mathbb{E}\left[\sup_{\pi \in \Pi} |B_1(\pi)|\right] = O\left(\sqrt{\mathrm{VC}(\Pi_n) \cdot \frac{r_{\kappa_n}}{n^{3/2}}}\right). \tag{25}$$

We now turn to $B_2(\pi)$, for which we have

$$
\begin{aligned}
B_2(\pi) &= \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\langle\widehat{\alpha}_i-\alpha_i,\widehat{g}_i-g_i\rangle\cdot\mathbb{1}\{\widehat{\varphi}_i\geq 0\}\\
&= \sum_{j=1}^{J}\left[\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot(\widehat{\alpha}_i^{(j)}-\alpha_i^{(j)})\cdot(\widehat{g}_i^{(j)}-g_i)\cdot\mathbb{1}\{\widehat{\varphi}_i\geq 0\}\right]\\
&\leq \sum_{j=1}^{J}\left[\frac{1}{n}\sum_{i=1}^{n}|\widehat{\alpha}_i^{(j)}-\alpha_i^{(j)}|\cdot|\widehat{g}_i^{(j)}-g_i|\right]\\
&\leq \sum_{j=1}^{J}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\alpha}_i^{(j)}-\alpha_i^{(j)}\right)^2}\times\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(g_i^{(j)}-g_i\right)^2},
\end{aligned}
$$

where the last inequality uses Cauchy-Schwarz inequality. This bound does not depend on π and thus holds uniformly over Π_n. We then apply Cauchy-Schwarz again and use Asuumption 5.4 to verify that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\pi\in\Pi}|B_2(\pi)|\right] &\leq \sum_{j=1}^{J}\mathbb{E}_{P_n}\left[\left\|\widehat{\alpha}_i^{(j)}-\alpha_i^{(j)}\right\|_{L_2(P_V)}^2\right]^{1/2}\times\mathbb{E}_{P_n}\left[\left\|\widehat{g}_i^{(j)}-g_i^{(j)}\right\|_{L_2(P_V)}^2\right]^{1/2},\\
&\leq J\cdot\mathbb{E}_{P_n}\left[\|\widehat{\alpha}_i-\alpha_i\|_{L_2(P_V)}^2\right]^{1/2}\times\mathbb{E}_{P_n}\left[\|\widehat{g}_i-g_i\|_{L_2(P_V)}^2\right]^{1/2}\\
&\lesssim J\cdot\mathbb{E}_{P_n}\left[\left\|\widehat{f}^{-k(i)}-f\right\|_{L_2(P_V)}^2+\left\|\widehat{g}^{-k(i)}-g\right\|_{L_2(P_V)}^2\right]^{1/2}\times\mathbb{E}_{P_n}\left[\left\|\widehat{g}^{-k(i)}-g\right\|_{L_2(P_V)}^2\right]^{1/2}\\
&= O\left(\frac{r_{\kappa_n}}{\sqrt{n}}\right).
\end{aligned}
$$

We now turn to $B_3(\pi)$. We begin by considering the following telescoping

$$
\widehat{g}_i-g_i=\sum_{j=1}^{J}\left[(g_i^{(\bullet:j-1)},\widehat{g}_i^{(j:\bullet)})-(g_i^{(\bullet:j)},\widehat{g}_i^{(j+1:\bullet)})\right]=\sum_{j=1}^{J}\left(0,0,\ldots,\widehat{g}_i^{(j)}-g_i^{(j)},0,\ldots,0\right),
$$

where $g^{(\bullet:j)}$ and $g^{(j:\bullet)}$ denote, respectively, the first and last j entries of $g(V_i)$, where we adopt the convention $g^{(\bullet:0)}=g^{(J+1:\bullet)}=\emptyset$. We can therefore decompose $B_3(\pi)$ as follows:

$$
\begin{aligned}
&B_3(\pi)=\sum_{j=1}^{J}\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\\
&\times\left(\varphi((g^{(\bullet:j-1)},\widehat{g}_i^{(j:\bullet)});X_i)-\varphi((g_i^{(\bullet:j)},\widehat{g}_i^{(j+1:\bullet)});X_i)-\alpha^{(j)}(\{(g^{(\bullet:j-1)},\widehat{g}^{(j:\bullet)}),f\})\cdot\left(\widehat{g}_i^{(j)}-g_i^{(j)}\right)\right)\cdot\mathbb{1}\{\widehat{\varphi}_i\geq 0\}\\
&-\sum_{j=1}^{J}\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i)-1)\cdot\left[\left(\alpha_i^{(j)}-\alpha^{(j)}(\{(g^{(\bullet:j-1},\widehat{g}^{(j:\bullet)}),f\})\right)\cdot\left(\widehat{g}_i^{(j)}-g_i^{(j)}\right)\right]\cdot\mathbb{1}\{\widehat{\varphi}_i\geq 0\}.
\end{aligned}
$$

By the definition of the Riesz-representer and cross-fitting we have

$$
\begin{aligned}
\mathbb{E}\Big[&\varphi((g^{(\bullet:j-1)},\widehat{g}_i^{(j:\bullet)});X_i)-\varphi((g_i^{(\bullet:j)},\widehat{g}_i^{(j+1:\bullet)});X_i)\\
&-\alpha^{(j)}(\{(g^{(\bullet:j-1)},\widehat{g}^{(j:\bullet)}),f\},W_i)\cdot\left(\widehat{g}_i^{(j)}-g_i^{(j)}\right)\mid V_i,\widehat{g}^{-k(i)},\widehat{f}^{-k(i)}\Big]=0,
\end{aligned}
$$

where we have used the property $\alpha_\ell^{(j)}(\{(\widetilde{g}_{-j}, \widetilde{g}_j), \widetilde{f}\}, x) = \alpha_\ell^{(j)}(\{\widetilde{g}_{-j}, \widetilde{f}\}, x)$. Furthermore, the term within the expectation operator is sub-Gaussian since uniformly bounded by Assumption 5.2(iii). Therefore the first term in the expansion of $B_3(\pi)$ can be controlled uniformly using identical arguments as for $B_1(\pi)$ and obeys the same bound. The second term in the expansion of $B_3(\pi)$ can be bounded with identical arguments as for $B_2(\pi)$ and obeys the same bound. We therefore conclude that

$$\mathbb{E}\left[\sup_{\pi\in\Pi}|B_3(\pi)|\right] = O\left(\sqrt{\mathrm{VC}(\Pi_n)\cdot\frac{r_{\kappa_n}}{n^{3/2}}} + \frac{r_{\kappa_n}}{\sqrt{n}}\right).$$

Combining the bounds for $B_1(\pi)$,$B_2(\pi)$ and $B_3(\pi)$ via the triangle inequality finally gives the desired bound for $\mathbb{E}\left[\sup_{\pi\in\Pi_n}|A_{1,\ell}(\pi)|\right]$.

Bound for $A_{2,\ell}$

We first notice that

$$\mathbb{E}\left[\phi_\ell(\{g,f\};W_i)\cdot\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right) \mid V_i, \widehat{g}^{-k(i)}\right] = 0, \tag{26}$$

by the mean-zero property of the influence function adjustments ϕ_ℓ and cross-fitting. Furthermore, the term inside expectation is sub-Gaussian by uniform boundedness of the Riesz-representer, guaranteed by Assumption 5.2(iii). Thus we can use similar arguments to those used for $B_1(\pi)$ to show

$$\mathbb{E}\left[\sup_{\pi\in\Pi_n}|A_{2,\ell}(\pi)|\right] = O\left(\sqrt{\frac{\mathrm{VC}(\Pi_n)}{n}}\right).$$

Bound for $A_{3,\ell}$

We begin by noticing that $\varphi_\ell(g;X_i)\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\leq 0$ and thus, since the "never treat" policy belongs to any policy class Π for which $\mathrm{VC}(\Pi)\geq 1$, we have

$$\sup_{\pi\in\Pi_n} A_{3,\ell}(\pi) = \frac{1}{2n}\sum_{i=1}^n\left|\varphi_\ell(g;X_i)\cdot\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\right|,$$

and thus we obtain the uniform bound[22]

$$\mathbb{E}\left[\sup_{\pi\in\Pi_n} A_{3,\ell}(\pi)\right] = \frac{1}{2}\mathbb{E}\left[\left|\varphi_\ell(g;X_i)\cdot\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\} - \mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\right|\right]. \tag{27}$$

[22] For a policy class of zero VC-dimension, (27) holds as an inequality.

For the RHS in (27), we closely follow Lemma 5.2 in Audibert and Tsybakov (2007), but we report the steps of the proof for completeness. For $\gamma > 0$ and any $t > 0$ we have

$$
\begin{aligned}
&\mathbb{E}\left[\left|\varphi_\ell(g;X_i)\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\}-\mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\right|\right]\\
&\leq \mathbb{E}\left[|\varphi_\ell(g;X_i)|\cdot\mathbb{1}\left\{\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\geq|\varphi_\ell(g;X_i)|\right\}\right]\\
&\leq \mathbb{E}\left[|\varphi_\ell(g;X_i)|\cdot\mathbb{1}\left\{\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\geq|\varphi_\ell(g;X_i)|\right\}\cdot\mathbb{1}\left\{0<|\varphi_\ell(g;X_i)|\leq t\right\}\right]\\
&\qquad+\mathbb{E}\left[|\varphi_\ell(g;X_i)|\cdot\mathbb{1}\left\{\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\geq|\varphi_\ell(g;X_i)|\right\}\cdot\mathbb{1}\left\{|\varphi_\ell(g;X_i)|>t\right\}\right]\\
&\leq \mathbb{E}\left[\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\cdot\mathbb{1}\left\{0<|\varphi_\ell(g;X_i)|\leq t\right\}\right]\\
&\qquad+\mathbb{E}\left[\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\cdot\mathbb{1}\left\{\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|>t\right\}\right]\\
&\leq \mathbb{E}\left[\left(\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right)^2\right]^{1/2}\cdot\mathbb{P}\big(0<|\varphi_\ell(g;X_i)|\leq t\big)^{1/2}+\frac{\mathbb{E}\left[\left(\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right)^2\right]}{t}\\
&\leq C_0^{1/2}\mathbb{E}\left[\left(\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right)^2\right]^{1/2}t^{\gamma/2}+\frac{\mathbb{E}\left[\left(\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right)^2\right]}{t},
\end{aligned}
$$

where the penultimate inequality uses Cauchy-Schwarz and Markov inequalities, and the last inequality uses the Margin Assumption. Minimizing the last display over t gives

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\pi\in\Pi_n}A_{3,\ell}(\pi)\right]&\leq(\gamma+2)\cdot\left(\frac{2}{\gamma}\right)^{\gamma/(\gamma+2)}\cdot\mathcal{C}_m^{1/(\gamma+2)}\cdot\mathbb{E}\left[\left(\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right)^2\right]^{\frac{\gamma+1}{\gamma+2}}\\
&\leq(\gamma+2)\cdot\left(\frac{2}{\gamma}\right)^{\gamma/(\gamma+2)}\cdot\mathcal{C}_m^{1/(\gamma+2)}\cdot\mathcal{C}_{2,\varphi}^{\frac{2(\gamma+1)}{\gamma+2}}\cdot\mathbb{E}_{P_n}\left[\left\|\widehat{g}^{-k}-g\right\|^2_{L_2(P_X)}\right]^{\frac{\gamma+1}{\gamma+2}}
\end{aligned}
$$

For $\gamma = 0$, a similar argument gives

$$
\begin{aligned}
&\mathbb{E}\left[\left|\varphi_\ell(g;X_i)\left(\mathbb{1}\left\{\varphi_\ell(\widehat{g}^{-k(i)};X_i)\geq 0\right\}-\mathbb{1}\left\{\varphi_\ell(g;X_i)\geq 0\right\}\right)\right|\right]\\
&\leq \mathbb{E}\left[\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\cdot\mathbb{1}\left\{0<|\varphi_\ell(g;X_i)|\leq t\right\}\right]\\
&\qquad\qquad+\mathbb{E}\left[\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|\cdot\mathbb{1}\left\{\left|\varphi_\ell(\widehat{g}^{-k(i)};X_i)-\varphi_\ell(g;X_i)\right|>t\right\}\right]\\
&\leq 2\mathbb{E}_{P_n}\left[\left\|\varphi_\ell(\widehat{g}^{-k};X)-\varphi_\ell(g;X)\right\|^2_{L_2(P_X)}\right]^{1/2}\\
&\leq 2\cdot\mathcal{C}_{2,\varphi}^2\cdot\mathbb{E}_{P_n}\left[\left\|\widehat{g}^{-k}-g\right\|^2_{L_2(P_X)}\right]^{1/2}.
\end{aligned}
$$

Combining the cases $\gamma > 0$ and $\gamma = 0$, and using the L_2-risk bounds for $\widehat{g}^{-k}$ from Assumption 5.4 we finally get

$$\mathbb{E}\left[\sup_{\pi\in\Pi_n} A_{3,\ell}(\pi)\right] = O\left(\left(\frac{r_{\kappa_n}}{\sqrt{n}}\right)^{\frac{\gamma+1}{\gamma+2}}\right).$$

B.5 Proof of Theorem 2

The Neyman-orthogonalized score $\Gamma^{\mathtt{NO}}(\{g, f\}; W_i)$ satisfies the assumptions of Corollary 3 in Athey and Wager (2021), and thus it can be applied verbatim to show that

$$\mathbb{E}\left[\sup_{\pi\in\Pi_n} \left|Q_n^{\mathtt{NO}}(\pi) - Q(\pi)\right|\right] = O\left(\sqrt{\frac{\mathrm{VC}(\Pi_n)}{n}}\right).$$

Combining the above bound with Lemma 1 via Proposition 3 gives the statement of the theorem.

References

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Adjaho, C. and T. Christensen (2022). Externally valid treatment choice.

Arlot, S. and P. L. Bartlett (2011). Margin-adaptive model selection in statistical learning. *Bernoulli 17*(2), 687 – 713.

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

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

Balke, A. and J. Pearl (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association 92*(439), 1171–1176.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer Series in Statistics, 2nd edition.

Byambadalai, U. (2022). Identification and inference for welfare gains without unconfoundedness.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Volume 6 of *Handbook of Econometrics*, pp. 5549–5632. Elsevier.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022). Locally robust semiparametric estimation. *Econometrica 90*(4), 1501–1535.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.

Christensen, T., H. R. Moon, and F. Schorfheide (2022). Optimal discrete decisions when payoffs are partially identified.

Cui, Y. and E. T. Tchetgen (2021). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *Journal of the American Statistical Association 116*(533), 162–173. PMID: 33994604.

Fang, Z. and A. Santos (2018, 09). Inference on Directionally Differentiable Functions. *The Review of Economic Studies 86*(1), 377–412.

Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics 189*(1), 1–23.

Farrell, M. H., T. Liang, and S. Misra (2021). Deep neural networks for estimation and inference. *Econometrica 89*(1), 181–213.

Foster, D. J. and V. Syrgkanis (2019). Orthogonal statistical learning.

Han, S. (2019). Optimal dynamic treatment regimes and partial welfare ordering.

Heckman, J. J. and E. J. Vytlacil (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluation of Labour Market Policies*, pp. 1–15. Springer.

Hirano, K. and J. R. Porter (2012). Impossibility results for nondifferentiable functionals. *Econometrica 80*(4), 1769–1790.

Hirano, K. and J. R. Porter (2020). Asymptotic analysis of statistical decision rules in econometrics. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin (Eds.), *Handbook of Econometrics, Volume 7A*, Volume 7 of *Handbook of Econometrics*, pp. 283–354. Elsevier.

Hurwicz, L. (1951). The generalised bayes-minimax principle: A criterion for decision-making under uncertainty.

Ichimura, H. and W. K. Newey (2022). The influence function of semiparametric estimators. *Quantitative Economics 13*(1), 29–61.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Ishihara, T. and T. Kitagawa (2021). Evidence aggregation for treatment choice.

Kallus, N. and A. Zhou (2018). Confounding-robust policy improvement.

Kasy, M. (2016, 03). Partial Identification, Distributional Preferences, and the Welfare Ranking of Policies. *The Review of Economics and Statistics 98*(1), 111–131.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review.

Kido, D. (2022). Distributionally robust policy learning with wasserstein distance.

Kim, W., K. Kwon, S. Kwon, and S. Lee (2018). The identification power of smoothness assumptions in models with counterfactual outcomes. *Quantitative Economics 9*(2), 617–642.

Kitagawa, T., J. L. Montiel Olea, J. Payne, and A. Velez (2020). Posterior distribution of nondifferentiable functions. *Journal of Econometrics 217*(1), 161–175.

Kitagawa, T., S. Sakaguchi, and A. Tetenov (2021). Constrained classification and policy learning.

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

Lee, D. S. (2009, 07). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102.

Mammen, E. and A. B. Tsybakov (1999). Smooth discrimination analysis. *The Annals of*

Statistics 27(6), 1808 – 1829.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review 80*(2), 319–323.

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

Manski, C. F. (2009). Diversified treatment under ambiguity. *International Economic Review 50*(4), 1013–1041.

Manski, C. F. (2010). Vaccination with partial knowledge of external effectiveness. *Proceedings of the National Academy of Sciences 107*(9), 3953–3960.

Manski, C. F. (2011). Choosing treatment policies under ambiguity. *Annual Review of Economics 3*(1), 25–49.

Manski, C. F. and J. V. Pepper (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica 68*(4), 997–1010.

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

Ponomarev, K. (2022). Efficient estimation of directionally differentiable functionals.

Pu, H. and B. Zhang (2021, mar). Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 83*(2), 318–345.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika 74*(1), 13–26.

Stoye, J. (2012). Minimax regret treatment choice with covariates or with limited validity of experiments. *Journal of Econometrics 166*(1), 138–156. Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday.

Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: Wiley.

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wald, A. (1950). *Statistical Decision Functions*. Wiley.

Yata, K. (2021). Optimal decision rules under partial identification.

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association 107*(499), 1106–1118.