

PARTIAL IDENTIFICATION, DISTRIBUTIONAL PREFERENCES, AND THE WELFARE RANKING OF POLICIES

Maximilian Kasy*

Abstract—We discuss the tension between “what we can get” (identification) and “what we want” (parameters of interest) in models of policy choice (treatment assignment). Our nonstandard empirical object of interest is the ranking of counterfactual policies. Partial identification of treatment effects maps into a partial welfare ranking of treatment assignment policies. We characterize the identified ranking and show how the identifiability of the ranking depends on identifying assumptions, the feasible policy set, and distributional preferences. An application to the project STAR experiment illustrates this dependence. This paper connects the literatures on partial identification, robust statistics, and choice under Knightian uncertainty.

I. Introduction

METHODOLOGICAL difficulties in microeconometrics often arise from the tension between two conflicting objectives: the objective to use models based only on assumptions that seem credible a priori and the objective to estimate parameters that allow evaluating the impact of counterfactual policies. This paper contributes to exploring the frontier in the trade-off between these two objectives. We discuss identification of the ranking of counterfactual policies based on models without functional form assumptions.

Motivated by the twin objectives of credibility and policy relevance, the central contribution of this paper is the consideration of a nonstandard empirical object of interest, the ranking of counterfactual policies.¹ We develop a set of identification results that map distributions of observed data into partial orderings (identified rankings) of policies. A second contribution is that this paper draws connections, both conceptual and formal, between the literature on partial identification and the literature on ambiguity (Bewley, 2002; Stoye, 2011b). There is a close relationship between the identified policy rankings considered here and Knightian uncertainty, as will be discussed below. A third contribution is that the discussion of nonlinear objectives in this paper extends the distributional decomposition literature, which considers distributional effects under conditional exogeneity (Firpo, Fortin, & Lemieux, 2009; Rothe, 2010), by allowing for data with endogeneity.

Received for publication July 2, 2014. Revision accepted for publication December 16, 2014. Editor: Yuriy Gorodnichenko.

* Harvard University and IHS Vienna.

I thank Yuriy Gorodnichenko, Jinyong Hahn, Guido Imbens, James Powell, Jörg Stoye, and Elie Tamer, as well as Susanne Kimm, Jian Li, Zachary Liscow, Alexander Rothenberg, and Sarah Sanga; seminar participants at Harvard University, UC Berkeley, Université de Montreal, UCSD, and Penn State; several anonymous referees; and conference participants at the NAWMES 2012, as well as the Conference on Partial Identification and Decisions at Cornell 2012 for valuable discussions and comments.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00528.

¹The objects considered here—rankings of counterfactual policies—occupy an intermediate place between treatment effects and choice rules mapping data into treatment assignment functions (cf. the analysis of preferences as opposed to choice functions in microeconomic theory: Mas-Colell, Whinston, & Green, 1995). Both analysis of preferences and analysis of choice functions yields important insights.

The following framework is discussed in this paper. The policies h considered allocate a binary treatment D based on covariates X , $h(X) = P(D = 1|X)$, as in Dehejia (2005). The set of feasible policies is possibly subject to constraints. We are interested in identification approaches that partially identify conditional average treatment effects $g(X) = E[Y^1 - Y^0|X]$ or, more generally, the difference between the conditional distributions of the potential outcomes Y^1, Y^0 . The planner’s objective function takes the form $\phi(f)$, where f is the distribution (density) of Y . This setup implies partial identification of the welfare ranking of policies h .² This paper discusses in particular the relationship among policy constraints, objective functions, and data requirements, which are such that it is possible to rank the feasible policy alternatives.

Consider as an example the project STAR experiment, which we discuss in sections II and VI. In this experiment, students were randomly assigned to small and large classrooms, but compliance was imperfect. We can infer bounds on conditional treatment effects from the project STAR data and use these to partially identify the welfare ranking of policies that affect class sizes for different groups of students. Whether a particular set of policies can be ranked in this setting indeed depends on identifying assumptions for treatment effects (monotonicity?), the set of policies considered (subject to a budget constraint?), and distributional preferences (maximize average test scores or test scores of students performing poorly?).

Starting in section III, we develop the theory of identification in this context. In addition to providing a geometric characterization of partial identification of welfare rankings, we develop a series of “possibility results” and “impossibility results.” These results give necessary and sufficient conditions for either full identification of the welfare ranking of a given set of policies, full nonidentification, or partial identification. These results can provide guidance to applied researchers as to when and under what conditions their data contain worthwhile information for policymakers. Additionally, our theoretical results allow us to make a number of interesting connections to other literatures, most notably the literature on ambiguity aversion and the literature on distributional decompositions.

In section III, we assume a linear objective function of the form $\phi = E[Y]$ and provide geometric characterizations of the partial preference ranking of policies h . It is shown that two policies h^a and h^b can be ranked if the policy difference $h^{ab} = h^a - h^b$ lies in the dual cone or the polar cone of

²What is called the identified welfare ranking in this paper has been called dominance, Pareto criterion, or Knightian preferences in different contexts; see, for instance, Manski (2011), Stoye (2011b), and Bewley (2002).

the identified set for g ; h^a and h^b cannot be ranked if h^{ab} is orthogonal to some g in the identified set. Section III also gives necessary and sufficient conditions for sets of feasible policies that are subject to linear budget constraints to be completely ordered and for them to be completely unordered. It is shown that such sets of feasible policies can be completely ordered only if they are one-dimensional. Such sets are completely unordered if and only if we cannot preclude the possibility that treatment effects are proportional to the covariate-specific cost of treatment.

In section IV, we generalize the setup discussed in section III, allowing for general nonlinear but smooth objective functions $\phi(f)$. Examples of such objective functions are quantiles, (quasi) Rawlsian welfare, as well as inequality as measured by the Gini coefficient. This section first studies local policy changes. The effect of such local policy changes on social welfare $\phi(f)$ can be characterized using the functional derivative $\partial\phi/\partial f$. The dual representation of $\partial\phi/\partial f$ is the influence function $IF(Y)$. We find that the effect of local policy changes on welfare ϕ is partially identified, with finite bounds, if and only if ϕ has an influence function that is bounded on the support of Y , that is, if and only if ϕ is a “robust” statistic. Section IV then discusses how the effect of local policy changes on social welfare ϕ relates to welfare differences between policies in a neighborhood of a status quo policy. It is shown that under a condition of continuous differentiability on ϕ , both yield the same conclusions about the welfare ranking of such policies. Based on these results, section IV then generalizes the results of section III, giving necessary and sufficient conditions for identification of the ranking of policies.

In section V, we propose procedures for inference on the partially identified rankings discussed in sections III and IV. We first consider a given pair of policies and discuss tests of the hypotheses: (a) policy a is identified to be preferred to policy b , and (b) policy b is not identified to be preferred to policy a . The latter null hypothesis allows for a nonidentified welfare ranking. We then consider set inference on the set of policies identified to be preferred to a given policy h and on the set of policies for which h is not identified to be preferred to them. The latter set contains policies preferred to h as well as policies for which the welfare ranking is not identified. For either of these two cases, we propose procedures that guarantee uniform coverage of the set of interest.

The online appendixes (appendixes B–D) contain additional reviews and discussions. Appendix B provides a review of partial identification of treatment effects in instrumental variable and in panel data models (based on Manski, 2003, and Chernozhukov, Fernandez-Val, Hahn, et al., 2013), appendix C reviews some concepts from functional analysis and differentiable geometry used in section IV, and appendix D discusses the relationship between this paper and the axiomatic decision theory literature, in particular, Bewley (2002). Adapting results from this literature, we show that given a linear objective function $\phi = E[Y]$, the identified preference order satisfies a property called independence.

Furthermore, any partial order satisfying independence is as if it did arise from some identified set for g .

A. *Connections to the Literature*

This paper contributes to two literatures in particular. First is the literature on partial identification of treatment effects, following Manski (2003), and on treatment choice based on partially identified treatment effects (e.g., Manski, 2011, and Stoye, 2011a). In contrast to those papers, this paper proposes to directly study identification of the welfare ranking of policies instead of treatment effects, which allows us to gain additional insights into the interaction of feasible policy sets, policy objectives, and data requirements. Second, in its discussion of nonlinear objective functions, this paper contributes to the literature on distributional decompositions following the seminal DiNardo, Fortin, and Lemieux (1996) work, building in particular on Firpo et al. (2009). In contrast to this literature, which generally achieves point identification using a conditional independence assumption (see Chernozhukov, Fernandez-Val, & Melly (2013) and the review in Firpo et al. (2011)), this paper allows for endogeneity of treatment relative to potential outcomes without imposing functional form assumptions. Analytically, easily tractable bounds are given for the effect of treatment assignment policies on statistics of the unconditional outcome distribution. From a practitioner’s point of view, the results of this paper suggest new objects of interest, allow for easy analytical calculation of these objects, and provide criteria to check under what conditions a given data set is or is not informative about the ranking of feasible policies.

Several further strands of literature have been important in developing the ideas for this paper. The problem of optimal treatment assignment based on covariates has been analyzed by Manski (2004), Dehejia (2005), Hirano and Porter (2009), Chamberlain (2011), Stoye (2011a), and Bhattacharya and Dupas (2012), among others. These papers focus on the decision-theoretic properties of policy choice based on finite samples from distributions point-identifying treatment effects. Our discussion differs from the one in papers such as Bhattacharya and Dupas (2012) in several respects: (a) We focus on the case of partial identification rather than assuming point identification based on a randomized experiment with perfect compliance. (b) We focus on welfare rankings of policies rather than optimal policy choice. Welfare rankings are in some sense an intermediate object between treatment effects and optimal treatment assignment policies. (c) We focus on issues of identification rather than on (finite sample) decision problems. (d) We allow for nonlinear objective functions of the form $\phi(f)$ rather than requiring linearity.

Choice in the absence of subjective probability distributions (i.e., under ambiguity) has been studied in the decision theory literature since Knight (1921). Formally related to this paper are, in particular, Bewley (2002) and Ryan (2009), building on the classic Anscombe and Aumann

(1963). In economic decision theory, ambiguity is primarily an *ex post* characterization of behavior satisfying certain axioms. In contrast, in the present context, it arises naturally from econometric models in the absence of functional form assumptions. Both stand in contrast to the issue of ambiguity in frequentist statistical decision theory, which does not impose prior distributions on model parameter values (see Stoye, 2011b, and the discussion in supplementary appendix D).

Partial identification of treatment effects has been analyzed in the pioneering work of Manski. The results in this paper apply to any approach yielding partial identification of treatment effects. Two identification approaches are discussed in more detail in supplementary appendix B: non-parametric instrumental variables (see Manski, 2003) and panel data with an assumption about marginal stationarity of unobserved heterogeneity (see Chernozhukov, Fernandez-Val, Hahn, et al. 2013). The trade-off between credibility and relevance mentioned in the beginning has been central in the debates between proponents of causal and those of structural approaches (Deaton, 2010; Imbens, 2010; Angrist & Pischke, 2010; Nevo & Whinston, 2010). In these debates, proponents of causal approaches have emphasized the necessity of credible identification, and proponents of structural approaches have emphasized the importance of estimating parameters that allow an evaluation counterfactual policies. The relationship between relevant policy sets and parameters of interest has been considered by Chetty (2009) and Graham, Imbens, and Ridder (2008). Like this paper, both of these make arguments implying that evaluation of the relative merits of different policies requires knowledge of only some parameters, not full knowledge of the underlying structural relationships. Sen (1995) also raises a related point, arguing that even if there is disagreement over the exact trade-offs among policy objectives, social consensus might be achieved over the ranking of policy alternatives.

II. A Motivating Example

Consider the following scenario. In a policy experiment, children were randomly allocated to small or large classes. Compliance with treatment assignment was incomplete, in particular because a number of students assigned to large classes switched to small classes. We furthermore observe a number of pretreatment characteristics of all children in the participating schools, most notably whether they received free lunch, an indicator of poverty. This scenario in fact describes exactly what happened in the project STAR experiment, to which we return in section VI.

Suppose now that a policymaker wishes to evaluate the relative merits of various alternative policies regarding class sizes. In order to do so, she first has to answer several questions:

1. What assumptions is she willing to impose on the data-generating process—for instance, randomness

of treatment assignment, monotonicity of treatment take-up in assignment, monotonicity of outcomes in treatment?

2. What is the objective function she wishes to maximize—for instance test scores in math or reading or other measurable outcomes—and given the outcome variable of interest, average outcomes, or the outcomes of those doing poorly, or the outcomes of those doing well already? Some weighted average of outcomes?
3. What policy alternatives is she considering? Hiring more teachers so as to reduce class-sizes for everyone? Or is she subject to a budget constraint and can only redistribute across schools or across students—for instance, decreasing class sizes for students on free lunch while increasing class sizes for others?

The answers to all these questions determine whether she will be able to rank the policy alternatives under consideration. The following scenarios are conceivable and in fact describe exactly what we find in our empirical application as discussed in section VI:

1. Assuming only randomness of treatment assignment, alternative policies cannot be ranked. This is due to the partial identification resulting from noncompliance. If additional monotonicity assumptions are imposed, however, then some alternatives can be ranked.
2. If her goal is to maximize the outcomes of good students, that is, increase the upper quantiles of the test score distribution, then the impact of redistributive policies is ambiguous. If her goal is to increase the test scores of poorly performing students, however, then redistributing teachers to students receiving free lunch is unambiguously positive.
3. If the policy alternative under consideration is a class size reduction for all students, then the impact of this alternative is unambiguously positive even when the goal is to maximize the upper quantiles of the score distribution. If the policy alternative under consideration is redistribution, however, the impact of this alternative might be ambiguous.

Before we get to a more detailed discussion of the project STAR data, we will discuss a number of theoretical questions. The next section focuses on the case where the objective function is given by the average $E[Y]$ of an observed outcome Y .³ In this setting, we formally discuss the mapping from bounds on conditional treatment effects (e.g., class size effects for poor and nonpoor students) to a partial ranking of policies (e.g., class size reduction for all, redistribution across students). As it turns out, there

³Note that this allows for a taste for redistribution, for instance, of the form $Y = T^{1/\alpha}$, for test scores T , where larger α corresponds to a greater concern about the outcomes of those at the lower end of the test score distribution.

is a nice geometric interpretation of this mapping, which also opens a tight connection between our setting and the decision-theoretic literature on ambiguity aversion, as discussed in online appendix D. Section IV generalizes to nonlinear objective functions of the form $\phi(f)$, and section V discusses inference focusing on the baseline case of a nonlinear objective function.

III. Welfare Ranking of Policies

Throughout this paper, the following setup is considered. The outcome of interest, Y , is generated by a structural relationship of the form $Y = m(X, D, \epsilon)$, where the treatment D is binary, $D \in \{0, 1\}$. The support of covariates X and unobservables ϵ is left unrestricted. We will denote potential outcomes by $Y^d = m(X, d, \epsilon)$ for $d = 0, 1$. The counterfactual treatment assignment policies considered randomly assign units characterized by X to treatment $D = 1$ with probability $h(X)$: $P(D = 1|X) = h(X)$ and $D \perp (Y^0, Y^1)|X$. The policies h considered include as a special case deterministic policies of the form $h(X) \in \{0, 1\}$, where $D = h(X)$.

We briefly discuss the assumptions that counterfactual treatment assignment is possibly randomized as well as exogenous. Allowing for randomized policies h convexifies the space of policies and has a similar role to the consideration of randomized tests in standard discussions of testing. This convexification allows constraints to be satisfied with equality and leads to a more elegant treatment, even though in some cases, the optimal policy does not involve randomization. All results immediately translate to the set of deterministic policies, since these are contained in the set of randomized policies.

The assumption of exogenous treatment assignment is more restrictive, in particular, since we will consider identification based on data in which compliance to an experimental assignment is incomplete.⁴ How reasonable this assumption is will depend on context; arguably there are many cases in which compliance is more easily guaranteed in the context of a universally implemented policy h than in a local experiment where there are more outside options for participants.

In the context of our motivating example, I believe that this assumption makes sense. In the project STAR experiment, each participating school had both large and small classes, making it fairly easy for parents to transfer their child to a class of different size within the school. Under a counterfactual policy, however, class sizes are most likely determined at the level of schools or even school districts. As a consequence, switching to a class of different size would require

switching to a school with different class sizes, possibly far away, making noncompliance much more costly.

Considering the effect of policies h rather than simple conditional average treatment effects allows us in particular to consider constrained policy sets, for instance, sets subject to budget constraints or political economy constraints. In such constrained sets, there is no trivial mapping from the identified sign of conditional average treatment effects to policy rankings or optimal policy.

In this section we maintain the additional assumption of a linear objective function:

Assumption 1 (*linear objective*). The policy objective is given by $\phi = E[Y]$, and the outcome Y has bounded support, $Y \in [0, 1]$.

All expectations and distributions in this paper are with respect to the target population of the policies under consideration. We later make assumptions about the availability of data that are (partially) informative about the conditional distributions of the potential outcomes in the target population, $P(Y^0|X)$ and $P(Y^1|X)$.

The results of this section immediately generalize to the case of bounded interval support of Y . The assumption of binary treatment D is important, although the identification results discussed in online appendix B.1 and B.2 do generalize to the case of larger, finite support of D . The size of the identified set for conditional average treatment effects in these setups is increasing in the size of the support of D , however. The form of the objective function, $\phi = E[Y]$, is restrictive in that it imposes additive separability of social welfare across units of observation. In an immediate generalization, one might allow for various preferences over inequality or risk, for example, or group-specific costs of treatment. To do so, one could take Y to be a possibly nonlinear transformation of measurable outcomes Y^{obs} , as well as X and D : $Y = Y(Y^{obs}, X, D)$. Section IV will further generalize this setup by allowing for arbitrary objective functions of the form $\phi = \phi(f)$, where ϕ is a (smooth, nonseparable) functional of the outcome distribution $f(Y)$.

This framework is similar to the ones studied by Manski (2004), Dehejia (2005), Bhattacharya and Dupas (2012), and others. Many setups of interest can be subsumed under this framework, where g might reflect a technological or a behavioral relationship. Decision problems that fit into this framework include (a) the assignment of income support programs by a policymaker interested in labor market outcomes, as in Dehejia (2005); (b) the allocation of indivisible capital goods to units of production by profit-maximizing firm owners; (c) the assignment of a medical treatment by doctors maximizing some health outcomes of their patients; (d) the decision of whether to attend college by students' taking into account the economic returns (among other factors), as in Card (2001); or (e) the assignment of students to integrated or segregated classes by an educational policymaker interested in maximizing average rescaled test scores, as in

⁴This approach contrasts with the perspective taken in the literature on marginal policy changes, for instance, in Carneiro, Heckman, and Vytalacil (2010). The latter consider intention-to-treat or "local IV" effects of some exogenous encouragement, where the effects might vary over the support of this encouragement.

Graham et al. (2008).⁵ In all of these applications, we might expect the available data to only partially identify average treatment effects, absent functional form assumptions.

An important limitation of the setup considered here is that it does not allow for incentive compatibility constraints on feasible policy sets. Incentive compatibility introduces a particular form of nonseparability of constraints across treatment units, which complicates the analysis yet is central for many interesting optimal policy problems in mechanism design and optimal taxation, such as those discussed in Chetty (2009). I analyze optimal policy under such constraints in a related paper (Kasy, 2014a).

A. Preliminaries

Partial identification of the potential outcome distributions $P(Y^d|X)$ for $d = 0, 1$ implies a partial ordering of the set of policies h . Consider a policy h^a , where $D^a|X, \epsilon \sim \text{Bernoulli}(h^a(X))$. Define $Y^a := m(X, D^a, \epsilon)$ and $\phi^a := E[Y^a]$. Define Y^b , ϕ^b , and D^b similarly for a policy h^b . Denote the difference of the probability of assignment to treatment between these two policies by $h^{ab} := h^a - h^b$ and let the difference in social welfare achieved by the two policies be given by $\phi^{ab} := \phi^a - \phi^b = E[Y^a] - E[Y^b]$. The social welfare ordering between two policies h^a, h^b is identified if and only if $\text{sign}(\phi^{ab}) = \text{sign}(E[Y^a] - E[Y^b])$ is identified. Note that

$$\begin{aligned} \phi^{ab} &= E[Y^a - Y^b] = E[(D^a - D^b)(Y^1 - Y^0)] \\ &= E[(h^a(X) - h^b(X))(Y^1 - Y^0)] \\ &= E[h^{ab}(X)g(X)], \end{aligned} \quad (1)$$

where $g(X)$ is the conditional average treatment effect defined as

$$g(X) := E[Y^1 - Y^0|X]. \quad (2)$$

Equation (1) holds because of the conditional independence between Y^0, Y^1 and D^a, D^b given X , which follows from the definition of the set of policies h we consider. This paper will use g throughout to denote conditional average treatment effects.

We consider in this section the space of bounded measurable functions of X , equipped with the inner product

$$\langle h, g \rangle := E[h(X)g(X)] \quad (3)$$

and with the norm $\|g\| = \sqrt{\langle g, g \rangle} = \sqrt{E[g^2(X)]}$. In this notation, we can rewrite equation (1) as

$$\begin{aligned} \phi^{ab} &= E[Y^a - Y^b] = E[(h^a(X) - h^b(X))g(X)] \\ &= \langle h^{ab}, g \rangle. \end{aligned} \quad (4)$$

Define the set of policies

$$\mathcal{H} = \{h(\cdot) : 0 \leq h(X) \leq 1\}. \quad (5)$$

The corresponding set of policy differences, $d\mathcal{H} = \mathcal{H} - \mathcal{H} = \{h^{ab} = h^a - h^b : h^a, h^b \in \mathcal{H}\} = \{h : \sup(|h|) \leq 1\}$, is the unit ball of X -measurable functions with respect to the sup norm.

Thus far, we have made no assumption about the available data. In the setup considered here, welfare rankings are a function of the conditional average treatment effects g . We will assume availability of a general identified set for g , denoted \mathcal{G} . As a special case, we will repeatedly consider rectangular sets \mathcal{G} , as in the following assumption:

Assumption 2 (*rectangular identified set for g*). The identified set \mathcal{G} for g is rectangular, that is, \mathcal{G} is of the form

$$\mathcal{G} = \{g(\cdot) : g(X) \in [\underline{g}(X), \bar{g}(X)]\}. \quad (6)$$

Such rectangular identified sets arise, for instance, in non-parametric instrumental variables and panel data models in the absence of functional form assumptions, as follows from the results of Manski (2003) and Chernozhukov, Fernandez-Val, Hahn, et al. (2013). (See online appendix B.1 and B.2 for a review and discussion).⁶ Note that we will mainly invoke this assumption for illustration and in order to obtain more explicit expressions. None of our main results rely on this assumption; it is in particular not invoked in the statement of propositions 1, 3, and 4 in this section or in the statement of theorem 1, proposition 6, and proposition 7 in section IV. Rectangular identified sets are obtained if no restrictions across covariate values are imposed on treatment effects. A leading case where the identified set would not be rectangular would be an assumption of monotonicity of conditional average treatment effects $g(x)$ in x or some component of x .

The identified set for ϕ^{ab} is given by $\{\langle h^{ab}, g \rangle : g \in \mathcal{G}\} = \langle h^{ab}, \mathcal{G} \rangle$. By continuity of the map $g \rightarrow \langle h^{ab}, g \rangle$, if \mathcal{G} is connected, then this set is an interval. If \mathcal{G} is rectangular, this interval is bounded by $\underline{\phi}^{ab} = \langle h^{ab}, g^l \rangle$ and $\bar{\phi}^{ab} = \langle h^{ab}, g^u \rangle$, where

$$\begin{aligned} g^u &= \bar{g} \cdot \mathbf{1}(h^{ab} \geq 0) + \underline{g} \cdot \mathbf{1}(h^{ab} < 0), \\ g^l &= \bar{g} \cdot \mathbf{1}(h^{ab} < 0) + \underline{g} \cdot \mathbf{1}(h^{ab} \geq 0). \end{aligned}$$

Two policies h^a, h^b are strictly ordered if and only if $\text{sign}(\phi^{ab}) = \text{sign}(\underline{\phi}^{ab}) \neq 0$, that is, if either $\underline{\phi}^{ab} > 0$ or $\bar{\phi}^{ab} < 0$.

Definition 1 (*welfare ranking and identified welfare ranking of policies*). Given conditional average treatment effects $g(X)$, the welfare ranking between policies is given by

$$\begin{aligned} h^a \succ^g h^b &:\Leftrightarrow \langle h^{ab}, g \rangle > 0, \\ h^a \succeq^g h^b &:\Leftrightarrow \langle h^{ab}, g \rangle \geq 0. \end{aligned} \quad (7)$$

⁶ It is recommended that readers not familiar with these results read online appendix B before continuing.

⁵ Bhattacharya (2009) condered a similar problem.

Given an identified set \mathcal{G} for conditional average treatment effects g , the identified welfare ranking between policies is given by

$$\begin{aligned} h^a >^{\mathcal{G}} h^b &: \Leftrightarrow \langle h^{ab}, g \rangle > 0 \quad \forall g \in \mathcal{G} \\ h^a \succeq^{\mathcal{G}} h^b &: \Leftrightarrow \langle h^{ab}, g \rangle \geq 0 \quad \forall g \in \mathcal{G}. \end{aligned} \quad (8)$$

Equation (7) defines a complete order of the set of policies \mathcal{H} . The relation $\succeq^{\mathcal{G}}$ is a partial order of \mathcal{H} .⁷ We have

$$g \in \mathcal{G} \Rightarrow (h^a \succeq^{\mathcal{G}} h^b \Rightarrow h^a \succeq^g h^b), \quad (9)$$

and similarly for $>^g, >^{\mathcal{G}}$. We cannot necessarily order all policy pairs based on knowledge of \mathcal{G} . If, however, $h^a \succeq^{\mathcal{G}} h^b$, then we know that h^a is weakly preferred to h^b , $h^a \succeq^g h^b$, even though we have only partially identified the structural relationship g .

B. Geometric Interpretation

The relationship between the set \mathcal{G} and the preference ordering $\succeq^{\mathcal{G}}$ can be given a useful geometric interpretation. The dual cone of a set \mathcal{G} is defined as the set⁸

$$\mathcal{G}^* = \{h : \inf_{g \in \mathcal{G}} \langle h, g \rangle \geq 0\} = \{h : \langle h, \mathcal{G} \rangle \subset [0, \infty)\}. \quad (10)$$

This is the set of all h such that the angle between h and g is no more than 90 degrees for all g in \mathcal{G} . Equivalently, it is the intersection over all $g \in \mathcal{G}$ of the half-spaces $\{h : \langle h, g \rangle \geq 0\}$. The polar cone is given by

$$\begin{aligned} \mathcal{G}^\times &= -\mathcal{G}^* = \{h : \sup_{g \in \mathcal{G}} \langle h, g \rangle \leq 0\} \\ &= \{h : \langle h, \mathcal{G} \rangle \subset (-\infty, 0]\}. \end{aligned} \quad (11)$$

This is the set of all h such that the angle between h and g is at least 90 degrees for all g in \mathcal{G} . We have, by definition, that

$$\begin{aligned} h^a \succeq^{\mathcal{G}} h^b &\Leftrightarrow h^{ab} \in \mathcal{G}^*, \\ h^b \succeq^{\mathcal{G}} h^a &\Leftrightarrow h^{ab} \in \mathcal{G}^\times. \end{aligned} \quad (12)$$

Let $\overline{\mathcal{K}}$ denote the closure of a set \mathcal{K} with respect to the norm $\|\cdot\|$, and \mathcal{K}° the interior of \mathcal{K} .

Proposition 1 (maximal set of ordered policy pairs). *Suppose that assumption 1 holds and that the identified set \mathcal{G} is convex, $0 \notin \mathcal{G}$ and $\operatorname{argmin}_{g \in \mathcal{G}} \|g\|$ exists. Then \mathcal{G} is uninformative about the ordering of h^a, h^b (neither $h^a \succeq^{\mathcal{G}} h^b$ nor $h^b \succeq^{\mathcal{G}} h^a$) if and only if*

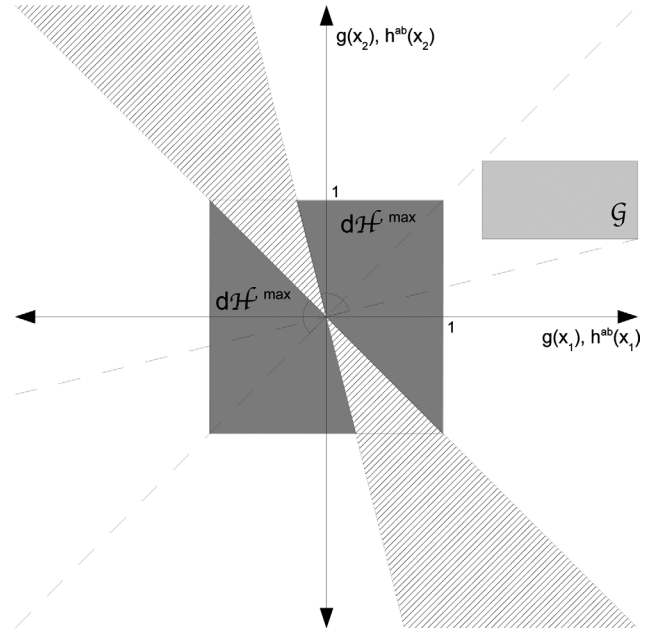
$$h^{ab} \in d\mathcal{H} \setminus (\mathcal{G}^* \cup \mathcal{G}^\times) = d\mathcal{H} \cap \left(\bigcup_{g \in \mathcal{G}} g^\perp \right)^\circ, \quad (13)$$

where $g^\perp = \{h : \langle h, g \rangle = 0\}$ is the orthocomplement of g .

⁷ A partial order satisfies transitivity, $h^a \succeq h^b$ and $h^b \succeq h^c$ implies $h^a \succeq h^c$. A complete order satisfies additionally that for all h^a, h^b , either $h^a \succeq h^b$ or $h^b \succeq h^a$ (see Mas-Colell et al., 1995).

⁸ See, for instance, Bertsekas, Nedić, and Ozdaglar (2003).

FIGURE 1.—IDENTIFICATION REGION OF THE CONDITIONAL AVERAGE TREATMENT EFFECT g AND THE SET OF ORDERED POLICIES IF THE SUPPORT OF X IS $\{x_1, x_2\}$



This figure illustrates the geometry of the relationship between the identification region \mathcal{G} for g and the set $d\mathcal{H}^{\max}$ of policy differences h^{ab} such that h^a, h^b are (weakly) ordered by $\succeq^{\mathcal{G}}$. See the text for a discussion.

The proof of this proposition, and all further proofs, can be found in appendix A. The assumption $0 \notin \mathcal{G}$ simply rules out the trivial case where the data are uninformative about the ranking of any pair of policies. The equality in equation (13) requires the existence of a separating hyperplane between 0 and \mathcal{G} , which follows from the existence of $\operatorname{argmin}_{g \in \mathcal{G}} \|g\|$. This always holds for rectangular \mathcal{G} , as in assumption 2, and if X has finite support. For rectangular \mathcal{G} , an element of $\operatorname{argmin}_{g \in \mathcal{G}} \|g\|$, is given by

$$h' = \begin{cases} \underline{g} & \text{if } |\underline{g}| < |\bar{g}| \\ \bar{g} & \text{else.} \end{cases}$$

The existence of such a separating hyperplane more generally follows from the Hahn-Banach theorem if \mathcal{G}° is not empty.

Define $d\mathcal{H}^{\max}$ to be the maximal set of policy differences h^{ab} such that h^a, h^b are (weakly) ordered by $\succeq^{\mathcal{G}}$, $d\mathcal{H}^{\max} = d\mathcal{H} \cap (\mathcal{G}^* \cup \mathcal{G}^\times)$. Figure 1 illustrates proposition 1 for the case where X has two points of support. The identified region \mathcal{G} for $g(\cdot)$ is rectangular, as in assumption 2. The maximal set of policy differences corresponding to weakly ordered policy pairs, $d\mathcal{H}^{\max}$, is given by the intersection of the set of feasible policy differences (the square in the center), with the dual cone and the polar cone of \mathcal{G} . The dual cone is the set in the upper-right part of the figure; it contains all vectors that have an angle of 90 degrees or less with all vectors in \mathcal{G} . The polar cone is the set in the lower-left part of the figure; it contains all vectors that have an angle of 90 degrees or more with all vectors in \mathcal{G} . The set of policy differences

that cannot be strictly ranked is the set of vectors that are orthogonal to some element of \mathcal{G} .

Proposition 1 characterizes the set of all policy pairs ordered by $\succeq^{\mathcal{G}}$. We are now going to give a sufficient condition for a pair of policies h^a, h^b to be ordered, that is, for identification of $\text{sign}(\phi^{ab})$. Assume the identification region \mathcal{G} is rectangular, and let the function w describe the width of the identified set for $g(X)$, $w(X) = \bar{g}(X) - \underline{g}(X)$. Online appendixes B.1 and B.2 give expressions for w that have the interesting feature that they do not depend on the distribution of Y but are solely a function of the strength of the instrument (in appendix B.1), or the variability of treatment within cross-sectional units over time (in appendix B.2).

Proposition 2 (sufficient condition). *Suppose assumptions 1 and 2 hold. Then, if*

$$\langle |h^{ab}|, w \rangle = \langle h^{ab}, w \cdot \text{sign}(h^{ab}) \rangle < |\langle h^{ab}, g \rangle|, \quad (14)$$

then $\text{sign}(\phi^{ab})$ is identified, and hence either $h^a \succ^{\mathcal{G}} h^b$ or $h^b \succ^{\mathcal{G}} h^a$. If h^a, h^b are deterministic policies, this condition can be rewritten as

$$E[w|\mathcal{X}^{ab}] < |\phi^{ab}| / P(\mathcal{X}^{ab}), \quad (15)$$

where $\mathcal{X}^{ab} = \{x : h^a(x) \neq h^b(x)\}$.

Note that the left-hand side of inequality (14) is the width of the identified set (interval) for ϕ^{ab} , while the right-hand side is equal to $|\phi^{ab}|$.

This proposition can also be interpreted by considering the set

$$\mathcal{G}' = \{g' : g(X) - w(X) \leq g'(X) \leq g(X) + w(X)\}. \quad (16)$$

By definition of w , we have $\mathcal{G} \subset \mathcal{G}'$. This implies $\widehat{\mathcal{G}} \subset \mathcal{G}'$. The set of h^a, h^b for which condition (14) holds is exactly the set of h^a, h^b for which either $h^a \succ^{\mathcal{G}'} h^b$ or $h^b \succ^{\mathcal{G}'} h^a$.

C. Restricted Policy Sets

We might hope that some restricted sets of policies are totally ordered by $\succeq^{\mathcal{G}}$ even if the full set \mathcal{H} is not. Policy sets that lie in affine subspaces intersected with \mathcal{H} are a particularly interesting set of restrictions. Such restrictions might arise, for instance, as a consequence of linear budget constraints of the form $E[h(X)c(X)] = \langle h, c \rangle = C$, where $c(X)$ is the (average) cost of treatment for units characterized by X .

Assumption 3 (affine restrictions on policy set). The set of feasible policies is given by $\mathcal{H}' = \{h(\cdot) : 0 \leq h(x) \leq 1, \langle h, c_i \rangle = C_i, i = 1 \dots k\}$.

The following proposition characterizes when policy sets subject to affine restrictions are totally ordered.

Proposition 3 (affine policy sets that are totally ordered by $\succeq^{\mathcal{G}}$). *Suppose that assumptions 1 and 3 hold, and assume that \mathcal{G} has nonempty interior \mathcal{G}^o . Then if \mathcal{H}' is totally ordered by $\succeq^{\mathcal{G}}$, \mathcal{H}' is at most one-dimensional.*

Proposition 3 requires the assumption that \mathcal{G} has nonempty interior \mathcal{G}^o . In the case of rectangular \mathcal{G} , this is equivalent to assuming that $g(x)$ is not point identified for any x . If $g(x')$ were point identified for some x' , while $g(x)$ is not point identified for all $x \neq x'$, then there might be two-dimensional feasible sets \mathcal{H}' which are totally ordered.

An important example of one-dimensional policy spaces arises if h is restricted to be constant in X or, more generally, if h has to be equal to 0 on a subset of X and constant on the rest of its support. Such restrictions might follow, for instance, if there is a nondiscrimination requirement that all individuals be treated equally or with equal probability. Another example would be the case of binary X , as in figure 1, and a policy space subject to a linear budget constraint.

Proposition 3 characterizes when sets subject to affine restrictions are totally ordered. The following proposition characterizes the opposite extreme: sets subject to linear budget constraints that cannot be ordered at all.

Proposition 4 (affine policy sets for which \mathcal{G} is completely uninformative about $\succeq^{\mathcal{G}}$). *Suppose that assumptions 1 and 3 hold and that \mathcal{G} is convex and \mathcal{G} has nonempty interior \mathcal{G}^o . Then the following two statements are equivalent: (i) There are no $h^a, h^b \in \mathcal{H}'$ such that $h^a \succeq^{\mathcal{G}} h^b$. (ii) $\sum_i \lambda_i c_i$ is an element of \mathcal{G}^o for some $(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$.*

As a special case of proposition 4, consider the case of one linear budget constraint, $k = 1$. In that case, proposition 4 tells us that the data are completely uninformative about the ranking of policies if and only if they are consistent with the hypothesis that treatment effects g are proportional to treatment costs c_1 . If two policies h^a, h^b satisfy such a linear budget constraint, then the policy difference h^{ab} lies in the hyperplane c_1^\perp . If two policies can be ordered, then their difference has to be in the dual cone of \mathcal{G} , or its negative. Proposition 4 reflects the fact that the intersection of the hyperplane c_1^\perp with the dual cone of \mathcal{G} is empty if and only if $\lambda c \in \mathcal{G}$ for some λ .

An interesting example where such a linear budget constraint arises is in the context of reallocations of individuals across groups, which might affect outcome distributions in the presence of social externalities (see Graham et al., 2008). In such a context, the feasible reallocations have to leave the population distribution of individual characteristics constant.

Note that in this section, we have restricted our attention to equality constraints of the form $\langle h, c_i \rangle = C_i$ rather than inequality constraints of the form $\langle h, c_i \rangle \leq C_i$. Inequality constraints do not change any qualitative statements about the identifiability of welfare rankings relative to the unconstrained case. To see this, consider a set \mathcal{H}' subject to inequality constraints that contains an open ball. Such an

open ball is totally ordered by \succeq^g if and only if the same holds for the unconstrained set of policies \mathcal{H} . Similarly, such an open ball is totally unordered by \succeq^g if and only if the same holds for the unconstrained set of policies \mathcal{H} .

IV. Nonlinear Objective Functions and Local Policy Changes

Throughout section III, we assumed a linear objective function of the form $\phi = E[Y]$. This section generalizes the previous results to nonlinear objective functions of the form $\phi(f)$, where f is the unconditional distribution (density) of Y , replacing assumption 1 with the following assumption:

Assumption 4 (nonlinear objective). For $d = 0, 1$ the support of Y^d is contained in $\mathcal{Y} \subset \mathbb{R}$, and the probability distribution of Y^d is absolutely continuous relative to the measure μ on \mathcal{Y} . The planner's objective function is $\phi(f)$, where f is the probability distribution (density) of Y . The support of X is the finite set $\{x^0, \dots, x^n\}$. The outcome distribution f^* of a status quo treatment assignment policy h^* is known.

Otherwise we maintain the same setup as in section III, that is, the outcome of interest, Y , is generated by a structural relationship of the form $Y = m(X, D, \epsilon)$, where the treatment D is binary, $D \in \{0, 1\}$. The counterfactual treatment assignment policies considered randomly assign units characterized by X to treatment $D = 1$ with probability $h(X)$: $P(D = 1|X) = h(X)$ and $D \perp (Y^0, Y^1)|X$. Throughout this section, we denote by f the (conditional) density of Y relative to μ and by f^0 and f^1 the (conditional) densities of the potential outcomes Y^0 and Y^1 . An important example of a status quo policy h^* , as in assumption 4, would be the zero-treatment policy $h^* = 0$. In that case, the potential outcome distribution f^0 is known in principle and any nonidentification of treatment effects arises from ignorance about f^1 . The role of the assumption that the baseline is known is to allow for a local linearization of the nonlinear objective function ϕ around the baseline. The identification results derived for the linear case then translate to the nonlinear one by using this local linearization. A discrete support of X is assumed to streamline the exposition and avoid introducing additional technical complications.

There are many interesting examples of nonlinear objective functions ϕ . A social planner might care about the following approximation to the Rawlsian maximin objective,

$$\phi(f) = \int_0^1 w(\tau)F^{-1}(\tau)d\tau, \quad (17)$$

where $w(\tau)$ are known welfare weights, the mass of which is concentrated on τ close to 0, and $F^{-1}(\tau)$ is the τ th quantile of the distribution corresponding to f . Alternatively, the planner might care about inequality as measured by (1 minus) the Gini coefficient, which can be written as

$$\phi(f) = 2 \cdot \int_0^1 (1 - \tau)F^{-1}(\tau)d\tau / \int_0^1 F^{-1}(\tau)d\tau. \quad (18)$$

Our goal is to give analytical characterizations of the identified welfare ranking and generalize the results of section III to this nonlinear setup. In order to achieve this goal, we will first consider a linearized version of the problem, replacing ϕ by a first-order approximation, which gives the welfare ranking of local policy changes. It will then be shown that all the qualitative conclusions drawn for the linearized problem carry over to the original problem, in a neighborhood of the status quo policy and under a condition of continuous differentiability on ϕ .

These results are easily applied in practice by replacing the outcome Y by the influence function $IF(Y; f^*)$ of ϕ , where the influence function is evaluated at the status quo outcome distribution f^* . For this modified problem, all the previous arguments apply, for instance, those concerning the identified set for treatment effects on welfare and those representing the upper contour set of a given policy as the dual cone of this identified set.

The approach taken here can be understood as a generalization of approaches taken in the recent distributional decomposition literature, in particular, Firpo et al. (2009). Firpo et al. (2009) suggest replacing Y by $IF(Y)$ in Oaxaca-Blinder-type decompositions, thus generalizing such decompositions to general statistics of the outcome distribution, while at the same time preserving the advantages of a linear structure. In contrast to the decomposition literature, which generally assumes conditional independence of potential outcomes from treatment and thus achieves point identification of policy effects (see the review in Firpo et al., 2011), the results presented here are concerned with the implications of endogeneity, absent functional form assumptions, for identification of policy effects on ϕ .

Online appendix C reviews some mathematical preliminaries that are crucial for a concise statement of the following results. The rest of this section is structured as follows. Section IVA characterizes the effects of local policy changes on social welfare ϕ and the identified set for these effects. Section IVB then maps the identified set for these effects into an identified welfare ranking of local policy changes. It is shown that under certain conditions, the welfare ranking of local policy changes is the same as the welfare ranking of corresponding pairs of policies in a neighborhood of the status quo policy. Section IVC finally uses these results in order to generalize the qualitative characterizations derived in section III to the nonlinear case.

A. Effect of Local Policy Changes

We now characterize the identified set for the effect of local policy changes on social welfare ϕ . Our characterizations crucially depend on the differentiability of ϕ . We need to impose one of the following forms of differentiability, which are increasingly restrictive.

Assumption 5 (differentiability of ϕ).

- The function ϕ is Fréchet differentiable at f^* for the L^p norm $\|f\| = (\int |f|^p d\mu)^{1/p}$, for some $p < \infty$.
- The function ϕ is Fréchet differentiable at f^* for the L^1 norm $\|f\| = \int |f| d\mu$.
- The function ϕ is continuously Fréchet differentiable in a neighborhood of f^* for the L^1 norm $\|f\| = \int |f| d\mu$.

In online appendix C, we discuss these forms of differentiability in greater detail.

We can write the effect of a local policy change in various ways, as demonstrated by the following lemma. First, we can approximate the nonlinear functional ϕ by the expectation

$$\phi \approx E[IF(Y; \check{f})]$$

for distributions close to the density \check{f} , where IF is the so-called influence function. This approximation by the expectation of the influence function is valid under L^p differentiability for any p . This linearization of ϕ then leads to a representation of the effect of local policy changes that is reminiscent of our previous representation of the effect of a discrete policy change h^{ab} ,

$$\phi_\theta = \langle h_\theta, g^\phi(\cdot; \check{f}) \rangle,$$

where g^ϕ is the conditional average treatment effect of treatment D on the influence function IF . Compare this to our previous representation,

$$\phi^{ab} = \langle h^{ab}, g \rangle,$$

where g is the conditional average treatment effect of D on Y .

Lemma 1 (dual representations). *Suppose assumptions 4 and 5a hold. Consider the derivative at 0 with respect to θ of $\phi(f(h(\theta)))$, where $h(\cdot)$ is a family of policies as in definition 5, and $f(h)$ is the outcome distribution induced by h . Denote $\check{f} = f(h(0))$. Then there are functions $IF(y; \check{f})$, $g^f(y|x)$, and $g^\phi(x; \check{f})$, independent of the family h , such that the derivatives with respect to θ of ϕ and f are given by*

$$\begin{aligned} \phi_\theta &= \frac{\partial \phi}{\partial f} \cdot f_\theta = \int IF(y; \check{f}) f_\theta(y) d\mu(y), \\ f_\theta(y) &= \frac{\partial f}{\partial h} \cdot h_\theta = \langle h_\theta, g^f(y|\cdot) \rangle, \\ \phi_\theta &= \frac{\partial \phi}{\partial h} \cdot h_\theta = \langle h_\theta, g^\phi(\cdot; \check{f}) \rangle. \end{aligned} \quad (19)$$

Furthermore, $g^f(y|x) = f^1(y|x) - f^0(y|x)$, where f^d denotes the conditional density of Y^d given X , and

$$\begin{aligned} g^\phi(x; \check{f}) &= \int IF(y; \check{f}) g^f(y|x) d\mu(y) \\ &= E[IF(Y^1; \check{f})|X = x] - E[IF(Y^0; \check{f})|X = x]. \end{aligned} \quad (20)$$

If we denote $g^F(y|x) = F^1(y|x) - F^0(y|x)$ and y is continuously distributed on \mathbb{R} , we get $g^f(y|x) = \frac{\partial}{\partial y} g^F(y|x)$ and $g^F(y|x) = E[\mathbf{1}(Y^1 \leq y) - \mathbf{1}(Y^0 \leq y)|X = x]$. Note also how, in equation (19), g^ϕ takes the role that g had in the linear case discussed in section III.

Examples

- Let $\phi = E[Y] = \int y f(y) d\mu(y)$. Then

$$\phi_\theta = \frac{\partial \phi}{\partial f} \cdot f_\theta = \int y f_\theta d\mu,$$

so that $IF(y; \check{f}) = y$ and $g^\phi(x)$ equals $g(x)$ as defined in section III.⁹

- Let $\phi = \text{Var}(Y) = E[Y^2] - E[Y]^2$. Then

$$\begin{aligned} \phi_\theta &= \int y^2 f_\theta d\mu - 2\check{E}[Y] \int y f_\theta d\mu \\ &= \int (y^2 - 2\check{E}[Y] \cdot y) f_\theta d\mu, \end{aligned}$$

that is, $IF(y; \check{f}) = y^2 - 2\check{E}[Y] \cdot y$. Here \check{E} denotes the expectation with respect to the distribution \check{f} . Note that in contrast to the first example, the directional derivative of the variance does depend on the status quo distribution via $\check{E}[Y]$.

- Let ϕ be the τ th quantile of Y , given by $\phi = F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$.¹⁰ By the implicit function theorem, if Y is continuously distributed so that F is differentiable, $\partial \phi / \partial \theta = -F_\theta(\phi) / F_y(\phi)$, where $F_y = f$ is the density of y . Thus, denoting $\check{\phi} = \phi(\check{f})$,

$$\phi_\theta = -\frac{1}{f(\check{\phi})} \int \mathbf{1}(y \leq \check{\phi}) f_\theta d\mu,$$

that is, $IF(y; \check{f}) = -\frac{\mathbf{1}(y \leq \check{\phi})}{f(\check{\phi})}$ if $f(\check{\phi}) > 0$.

The examples of (quasi) Rawlsian welfare, and the Gini coefficient can be analyzed similarly. For ease of exposition, we stick to the simpler examples of mean, variance, and quantiles throughout.

Rectangular identified sets. So far, we have not specified to what extent the conditional potential outcome distributions are identified. The following assumption is the natural generalization of assumption 2, which stated that the identified set for g is rectangular. Lemmas 3 and 5 in online appendix B show that assumption 6 holds in nonparametric instrumental variables and panel data setups. We should

⁹Note that in order to keep notation simple, but contrary to statistical convention, we are not normalizing influence functions to have mean 0 in the examples. This is justified by the fact that $g^f(\cdot|x)d\mu$ and $f_\theta d\mu$ integrate to 0.

¹⁰An axiomatic justification of preferences for maximizing a particular quantile was provided by Rostek (2010); Manski (1988) also considered such preferences.

emphasize again that rectangular identified sets are discussed here only because they are very common, not because they are required for any of our main results.

Assumption 6 (*rectangular identified set for g^f*). The identified set \mathcal{G}^f for g^f , where $g^f(y|x) = f^1(y|x) - f^0(y|x)$, is rectangular over X and has the form

$$\mathcal{G}^f = \{g^f : g^f(\cdot|x) = \tilde{g}^f(\cdot|x) + \gamma^1(x) \cdot f^1(\cdot|x, cf) - \gamma^0(x) \cdot f^0(\cdot|x, cf)\},$$

where $\tilde{g}^f(\cdot|x)$, $\gamma^1(x)$ and $\gamma^0(x)$ are known and $f^1(\cdot|x, cf)$, $f^0(\cdot|x, cf)$ are counterfactual outcome densities (for some subpopulation) ranging over the set of probability densities relative to μ on the support of Y .

Lemma 1 allows constructing a mapping from the identified set for g^f , \mathcal{G}^f , to the identified set $\mathcal{G}^\phi(f^*)$ for $g^\phi(\cdot; f^*)$, where $g^\phi(x; f^*)$ gives the average effect of treatment on welfare ϕ conditional on X , at the status quo distribution f^* . The mapping from g^f to g^ϕ is given by $g^\phi(x; f^*) = \int IF(y; f^*) g^f(y|x) d\mu(y)$. The mapping of identified sets is characterized in the following proposition:

Proposition 5 (bounds on local policy effects and robustness). *Suppose assumptions 4, 5a, and 6 hold. Then the identified set for g^ϕ is rectangular,*

$$\mathcal{G}^\phi(f^*) = \{g^\phi(\cdot; f^*) : \underline{g}^\phi(x; f^*) \leq g^\phi(x; f^*) \leq \overline{g}^\phi(x; f^*)\},$$

and $g^\phi(x; f^*)$ is bounded by¹¹

$$\begin{aligned} \overline{g}^\phi(x; f^*) &= \int IF(y; f^*) \tilde{g}^f(y|x) d\mu(y) \\ &\quad + \gamma^1(x) \cdot \sup_{y \in \mathcal{Y}} IF(y; f^*) \\ &\quad - \gamma^0(x) \cdot \inf_{y \in \mathcal{Y}} IF(y; f^*), \\ \underline{g}^\phi(x; f^*) &= \int IF(y; f^*) \tilde{g}^f(y|x) d\mu(y) \\ &\quad + \gamma^1(x) \cdot \inf_{y \in \mathcal{Y}} IF(y; f^*) \\ &\quad - \gamma^0(x) \cdot \sup_{y \in \mathcal{Y}} IF(y; f^*). \end{aligned} \quad (21)$$

These bounds are finite if and only if assumption 5b holds, that is, iff the influence function IF is bounded on the support of Y .

Examples

1. If $\phi = E[Y]$, we have $IF(y; f^*) = y$. This is unbounded for general outcome distributions; this shows that, in general, assumption 5b does not hold.

¹¹ Note that sup is used here to denote the essential supremum, that is the infimum over all numbers that bound a function with probability 1, and similarly for inf.

The influence function is bounded if and only if \mathcal{Y} is bounded, as we assumed in section III.

2. If $\phi = \text{Var}(Y)$ we have $IF(y; f^*) = (y^2 - 2E^*[Y] \cdot y)$. As in the first example, this is unbounded for general outcome distributions and bounded if and only if \mathcal{Y} is bounded.
3. If $\phi = \inf\{y : F(y) \geq \tau\}$ and F is differentiable, we have $IF(y; f^*) = -\frac{\mathbf{1}(y \leq \phi^*)}{f^*(\phi^*)}$ if $f^*(\phi^*) > 0$. Quantiles thus have bounded influence functions even in the absence of a priori restrictions on the counterfactual outcome distribution; they are ‘‘robust.’’ As a consequence, we get bounded confidence sets for the effect of policy changes on unconditional quantiles of the outcome distribution even for unbounded outcomes.

Proposition 5 characterizes the identified set for $\mathcal{G}^\phi(h^*)$. More generally we will be interested in the set $\mathcal{G}^\phi(h)$. This set maps into the identified set for the effect of local policy changes h_θ , relative to a given policy h , on social welfare ϕ . $\mathcal{G}^\phi(h)$ is given by

$$\mathcal{G}^\phi(h) = \left\{ g^\phi = \int IF(y; f^* + \langle h - h^*, g^f \rangle) g^f d\mu : g^f \in \mathcal{G}^f \right\}.$$

The simple characterization of $\mathcal{G}^\phi(h^*)$ does not hold for general h because of the double dependence of g^ϕ on g^f : through the effect of the policy changes on the distribution of Y and through the distribution f at which $IF(y; f)$ is evaluated. The only exception to this double dependence is the status quo assignment policy h^* , for which we have $\mathcal{G}^\phi(h^*) = \left\{ \int IF(y; f^*) g^f d\mu : g^f \in \mathcal{G}^f \right\}$.

B. Ranking Policies in a Neighborhood of the Status Quo

Section IVA characterized the effect of local policy changes h_θ on social welfare ϕ and the identified set for this effect, $\langle h_\theta, \mathcal{G}^\phi \rangle$. This section relates the difference in welfare for different policies to its first-order approximation, given by the welfare impact of local policy changes. It will be shown that under assumption 5c, the welfare ranking of local policy changes is the same as the welfare ranking of corresponding pairs of policies in a neighborhood of the status quo policy. In order to formally state this and further claims, the following two definitions are useful.

$$\begin{aligned} \underline{\Delta}\phi(h^{ab}; h) &:= \inf_{g^f \in \mathcal{G}^f} (\phi(f^* + \langle h^{ab} + h - h^*, g^f \rangle) \\ &\quad - \phi(f^* + \langle h - h^*, g^f \rangle)), \end{aligned} \quad (22)$$

$$\underline{d}\phi(h_\theta; h) := \inf_{g^f \in \mathcal{G}^f} \frac{\partial}{\partial \theta} \phi(f^* + \langle h(\theta) - h^*, g^f \rangle), \quad (23)$$

where in the latter definition, $h(\theta)$ is such that $\partial h / \partial \theta(0) = h_\theta$ and $h(0) = h$. The difference $\underline{\Delta}\phi(h^{ab}; h)$ gives the lower bound of the identified set for the welfare difference between $h + h^{ab}$ and h . The differential $\underline{d}\phi(h_\theta; h)$ gives the lower bound of the identified set for the marginal welfare effect

of a change in the direction of h_θ , starting from h . Under assumption 5a, as in lemma 1, we can rewrite $\underline{d}\phi(h_\theta; h)$ as

$$\underline{d}\phi(h_\theta; h) = \inf_{g^\phi \in \mathcal{G}^\phi(h)} \langle h_\theta, g^\phi \rangle,$$

and thus the local upper contour set of h is given by¹²

$$\begin{aligned} TU(h) &:= \left\{ h_\theta \in T_h \mathcal{H} : \underline{d}\phi(h_\theta; h) \geq 0 \right\} \\ &= \mathcal{G}^{\phi^*}(h) \cap T_h \mathcal{H}. \end{aligned}$$

Analogously, we can define the upper contour set of h as the set of policies

$$U(h) := \left\{ h^{ab} \in \mathcal{H} : \underline{\Delta}\phi(h^{ab}; h) \geq 0 \right\}.$$

Proposition 5 gave an intuitive and tractable characterization of the set $\mathcal{G}^\phi(h^*)$, which maps into a characterization of the set $TU(h^*)$, the set of welfare increasing local policy changes starting from h^* . Ultimately, however, we are interested in welfare differences between policies h^a and h^b , for which $\underline{\Delta}\phi(h^a; h^b)$ provides the identified lower bound. The following theorem relates the two. This theorem states that the welfare ranking of two policies of difference proportional to h_θ corresponds to the sign of the welfare effect of the local policy change h_θ , at least if the policies compared are close to the baseline h^* .

Theorem 1 (the welfare ranking of local policy changes and of policy differences). *Suppose assumptions 4 and 5c hold. Let h_θ be such that $d\phi(h_\theta; h^*) > 0$. Then there exists a δ such that $\underline{\Delta}\phi(\gamma \cdot h_\theta; h) > 0$ for all h such that $\|h - h^*\| \leq \delta$ and all $0 < \gamma \leq \delta$.*

Examples

We have already discussed L^1 differentiability—boundedness of the influence function. It remains to check continuity of the mapping $f \rightarrow IF(\cdot; f)$ with respect to the L^1 norm for f and the L^∞ norm for IF .

1. If $\phi = E[Y]$, we have $IF(y; f) = y$. Since this does not depend on f , continuity is immediate.
2. If $\phi = \text{Var}(Y)$, we have $IF(y; f) = (y^2 - 2E[Y] \cdot y)$, and thus $IF(y; f^1) - IF(y; f^2) = 2(E^2[Y] - E^1[Y]) \cdot y$. This goes to 0 uniformly in y if y is bounded, as long as $E^2[Y] - E^1[Y]$ goes to 0. The latter in turn follows from $\|f^1 - f^2\| \rightarrow 0$ if, again, Y is bounded.
3. If $\phi = \inf\{y : F(y) \geq \tau\}$ and F is differentiable, we have $IF(y; f) = -\frac{1(y \leq \phi)}{f(\phi)}$ if $f(\phi) > 0$. This implies that ϕ is not continuously L^1 differentiable: closeness in L^1 of f^1, f^2 does not imply closeness pointwise. Thus, the denominator of $-\frac{1(y \leq \phi)}{f(\phi)}$ does not necessarily converge as f converges in L^1 . However, in this case, the implications of theorem 1 still hold even though assumption

5c is violated. This is because the numerator of IF is continuous as a mapping from f to $\mathbf{1}(y \leq \phi(f))$. The denominator does not affect the relative welfare effect of local policy changes, only their scale.

C. Generalization of Results from the Linear Case

We are now ready to generalize to the nonlinear case the sufficient and necessary conditions for complete identification and complete nonidentification of preferences given in propositions 3 and 4. The natural generalization of assumption 3, which imposed linear constraints on the feasible policy set, in a nonlinear context is as follows:

Assumption 7 (nonlinear constraints). The set of feasible policies is given by $\mathcal{H}' = \{h \in [0, 1]^{n+1} : C(h) = 0\}$, where $C : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^k$, $k \leq n$, is differentiable.

Such nonlinear constraints might arise, for instance, if there was an increasing marginal cost of increasing the treatment probability h for a given subgroup of the population. Under assumption 7, the tangent space at the status quo policy h^* is given by

$$T_{h^*} \mathcal{H}' = \left\{ h_\theta : \left\langle h_\theta, \frac{\partial C_i}{\partial h}(h^*) \right\rangle = 0, i = 1 \dots k \right\},$$

that is, $T_{h^*} \mathcal{H}'$ is subject to k linear constraints. Propositions 3 and 4 generalize as follows:

Proposition 6 (restricted policy sets such that $T_{h^*} \mathcal{H}'$ is totally ordered). *Suppose assumptions 4, 5a, and 7 hold, and assume that \mathcal{G}^ϕ has nonempty interior \mathcal{G}^{ϕ^0} . Then if $T_{h^*} \mathcal{H}'$ is totally ordered by $\succeq^{\mathcal{G}^\phi}$, \mathcal{H}' is at most one-dimensional, that is, $k = n$.*

Proposition 7 (restricted policy sets such that $T_{h^*} \mathcal{H}'$ is completely unordered). *Suppose assumptions 4, 5a, and 7 hold, and assume that \mathcal{G}^ϕ is convex and \mathcal{G}^ϕ has nonempty interior \mathcal{G}^{ϕ^0} . Then the following two statements are equivalent: (i) There are no $h_\theta^a, h_\theta^b \in T_{h^*} \mathcal{H}'$ such that $h_\theta^a \succeq^{\mathcal{G}^\phi} h_\theta^b$. (ii) $\sum_i \lambda_i \frac{\partial C_i}{\partial h}(h^*)$ is an element of \mathcal{G}^{ϕ^0} for some $\lambda_i \in \mathbb{R}$.*

How do these two results relate to the identification of the welfare ranking of policies? The following theorem shows that locally full identification, partial identification, or full nonidentification of the welfare ranking of local policy changes imply the same for identification of the ranking of policies in a neighborhood of the status quo.

Theorem 2. *Suppose assumptions 4, 5c, and 7 hold and suppose \mathcal{G}^ϕ is bounded. Then there is a neighborhood N of h^* in \mathcal{H}' such that, for all $h \in N$:*

- i. If $TU(h^*) = \{0\}$, then $U(h) \cap N = \{0\}$.
- ii. If $TU(h^*) \neq \{0\}$, then $U(h) \cap N \neq \{0\}$.

¹² Recall that $\mathcal{G}^{\phi^*}(h)$ denotes the dual cone of $\mathcal{G}^\phi(h)$.

iii. If \mathcal{G}^ϕ has nonempty interior \mathcal{G}^{ϕ^0} and the welfare ranking on $T_h \mathcal{H}^1$ is fully identified, then so is the ranking of N .

Theorems 1 and 2 show that the qualitative conclusions we obtain from the linearized objective function are valid for our original objective function ϕ in some—possibly small—neighborhood. For large discrete policy changes (policy changes that affect a large fraction of the population) or highly nonlinear functionals, we might be worried that our qualitative conclusions are affected by using a linear approximation to the objective function. In that case, our theoretical results are best interpreted again as possibility or impossibility results, providing conditions when alternative policies can be completely ranked, not ranked at all, or partially ranked. In such a setting, exact bounds on welfare effects and the corresponding welfare ranking can still be obtained in many cases by imputing worst-case and best-case counterfactual outcomes to everybody in the population. Such an approach does require monotonicity conditions that allow us to know which the worst-case and best-case outcomes would be, and thus necessarily has to be specific to the chosen objective function ϕ .

When should we be worried whether our theoretical results using first-order approximations yield valid conclusions for discrete, rather than marginal, policy changes? The question is somewhat ill posed. The theoretical results in this section concern the existence of policies that can be ranked and of policies that cannot be ranked. Whether such policies exist in a neighborhood of the status quo is not a function of the magnitude of a given policy change that we might be interested in.

How about the validity of the qualitative conclusions drawn for a specific policy change? The answer depends on the specific objective under consideration. In the application in the next section, we focus on quantiles. The quantiles of a distribution are increased by a policy change if and only if the cumulative distribution function at the corresponding values is decreased. But the cumulative distribution function, which can be written in the form $F(y) = E[\mathbf{1}(Y \leq y)]$, is in fact a linear object. This remarkably implies that the conclusions drawn regarding the sign of the effect of policy changes are always the same based on a linear approximation for quantiles as for the original, nonlinear objective.

For general objectives ϕ , whether a linearization yields valid answers for a given policy change depends on the magnitude of the subpopulation for which the policy is changed. Arguably in many cases, the population of ultimate interest, say, all the schoolchildren in the United States, is much larger than the subpopulation for which a policy is changed, say, half the children in eighty schools in Tennessee. In such a setting, there seems little reason to worry about higher-order properties of the objective function invalidating any qualitative conclusions.

First-order approximations such as the ones we employ are prominently used in many areas of econometrics. They

are a key ingredient to the standard asymptotic theory of m-estimators, the delta method, semiparametric efficiency theory, and so on (van der Vaart, 2000). They also underlie part of the distributional decomposition literature, of which this section could be understood to be an extension to the partially identified case. They finally—and importantly for our setting—underlie much of the normative public finance literature, where social welfare is approximated by a weighted average of individual welfare (see Mirrlees, 1971; Saez & Stantcheva, 2013; the discussion in section 4 of Kasy (2014c)). Such a linearization of social welfare is in fact more than we need for our results, since it requires differentiable influence functions (Kasy, 2014c).

V. Inference on Partially Identified Welfare Rankings

In sections III and IV, we discussed partial identification of the welfare ranking of treatment-assignment policies based on partial identification of treatment effects. In this section, we propose procedures for inference on these partially identified rankings. We first consider a given pair of policies h^a , h^b , and discuss tests of the hypotheses (a) $h^a \succeq^{\mathcal{G}} h^b$ and (b) $h^b \not\succeq^{\mathcal{G}} h^a$. The latter null hypothesis allows for a nonidentified welfare ranking.

We then consider set inference on the set of policies identified to be preferred to a given policy h and on the set of policies for which h is not identified to be preferred to them. The latter set contains policies preferred to h as well as policies for which the welfare ranking is not identified. For either of these two cases, we propose procedures that guarantee uniform coverage of the set of interest.

Throughout this section, we consider the following setup:

- X has finite support (x^0, \dots, x^n) .
- The objective function ϕ is linear, $\phi = E[Y]$.
- The identified set \mathcal{G} is rectangular, $\mathcal{G} = \{g(\cdot) : g(x) \in [g(x), \bar{g}(x)]\}$.
- An asymptotically normal estimator

$$\left(\widehat{g}, \widehat{\underline{g}}\right) = \left(\widehat{g}(x^0), \dots, \widehat{g}(x^n), \widehat{\underline{g}}(x^0), \dots, \widehat{\underline{g}}(x^n)\right)$$

of the bounds defining \mathcal{G} is available, such that

$$\sqrt{n} \left(\widehat{g} - \bar{g}, \widehat{\underline{g}} - \underline{g}\right) \rightarrow^d N(0, \Sigma).$$

- There is a consistent estimator $\widehat{\Sigma}$ of Σ .

Requiring X to be discrete allows us to sidestep a number of technical issues that would arise in the continuous case. Requiring asymptotically normal estimators of bounds is somewhat restrictive. It is, however, easily satisfied for the identification approaches reviewed in online appendix B, where sample analogs of the bounds in equation (40) (instrumental variables) and equation (41) (panel data) are asymptotically normal by direct application of the central

limit theorem. Asymptotic normality might be violated in particular in settings where bounds take the form $\max(T, 0)$ or $\max(T, S)$ for asymptotically normal (T, S) . Bounds of this form can arise when monotonicity restrictions are imposed or in instrumental variables settings without first-stage monotonicity (intersection bounds). In this latter case, asymptotically valid inference might still be conducted by an appropriate modification of our approach, which could be thought of as a parametric bootstrap procedure.

Recall that our object of interest is the preference relationship \succeq^g defined by $h^a \succeq^g h^b \Leftrightarrow \langle h^a - h^b, g \rangle \geq 0$. To this corresponds the identified preference relationship $\succeq^{\mathcal{G}}$ defined by

$$h^a \succeq^{\mathcal{G}} h^b \Leftrightarrow \langle h^a - h^b, g \rangle \geq 0 \quad \forall g \in \mathcal{G}.$$

The relationship $\succeq^{\mathcal{G}}$ is, in general, an incomplete preference relationship, that is, there are policies h^a, h^b such that neither $h^a \succeq^{\mathcal{G}} h^b$ nor $h^b \succeq^{\mathcal{G}} h^a$.

Consider a given pair of policies h^a, h^b . In section VA, we propose tests for the following two null hypotheses:

1. H0,1: $h^a \succeq^{\mathcal{G}} h^b$. This can be rejected if and only if the hypothesis that $h^a \succeq^g h^b$ can be rejected.
2. H0,2: $h^b \not\succeq^{\mathcal{G}} h^a$. This can be rejected if and only if the hypothesis that either (a) $h^a \succeq^g h^b$ or (b) the welfare ranking of h^a, h^b is not identified can be rejected.

Corresponding to these null hypotheses for a given pair of policies h^a, h^b , we might be interested in uniform inference, which is simultaneously valid for arbitrary pairs of policies. In section VB, we discuss (1) how to construct a set estimator \widehat{H}_1 such that

$$P(\forall h^a, h^b : h^a \succeq^{\mathcal{G}} h^b \Rightarrow (h^a - h^b) \in \widehat{H}_1) \geq 1 - \alpha,$$

and (2) how to construct a set estimator \widehat{H}_2 such that

$$P(\forall h^a, h^b : h^b \not\succeq^{\mathcal{G}} h^a \Rightarrow (h^a - h^b) \in \widehat{H}_2) \geq 1 - \alpha.$$

A. Testing Null Hypotheses for a Given Pair of Policies

Using the notation established in section IIIA, we have $h^a \succeq^{\mathcal{G}} h^b$ if and only if

$$\langle h^{ab}, g^l \rangle \geq 0,$$

where $h^{ab} = h^a - h^b$ and

$$g^l = \bar{g} \cdot \mathbf{1}(h^{ab} < 0) + \underline{g} \cdot \mathbf{1}(h^{ab} \geq 0).$$

Let $e^{l,ab}$ be the $n \times 2n$ matrix of 1s and 0s that selects the right components of \bar{g} and \underline{g} such that

$$g^l = e^{l,ab} \cdot (\bar{g}, \underline{g}).$$

$e^{l,ab}$ is a function of $\text{sign}(h^{ab})$, the sign of each component of $h^a - h^b$. We can form an estimator for g^l by

$$\widehat{g}^l := e^{l,ab} \cdot (\widehat{\bar{g}}, \widehat{\underline{g}}).$$

This estimator satisfies

$$\sqrt{n}(\widehat{g}^l - g^l) \rightarrow^d N(0, e^{l,ab} \cdot \Sigma \cdot e^{l,ab}).$$

Similar statements are true for $g^u, e^{u,ab}$, and \widehat{g}^u .

Let

$$T^l := \frac{n \cdot \langle h^{ab}, \widehat{g}^l \rangle}{h^{ab} \cdot e^{l,ab} \cdot \widehat{\Sigma} \cdot e^{l,ab} \cdot h^{ab}}, \quad (24)$$

$$T^u := \frac{n \cdot \langle h^{ab}, \widehat{g}^u \rangle}{h^{ab} \cdot e^{u,ab} \cdot \widehat{\Sigma} \cdot e^{u,ab} \cdot h^{ab}}. \quad (25)$$

When $\langle h^{ab}, g^l \rangle = 0$, we get that $T^l \rightarrow^d N(0, 1)$. When $\langle h^{ab}, g^u \rangle = 0$, we get that $T^u \rightarrow^d N(0, 1)$.

We propose to test the null hypothesis H0,1: $h^a \succeq^{\mathcal{G}} h^b$ using the test statistic T^l , rejecting if and only if

$$T^l < c_\alpha,$$

where c_α is the α quantile of the standard normal distribution. Analogously, we propose to test the null hypothesis H0,2: $h^b \not\succeq^{\mathcal{G}} h^a$ using the test statistic T^u , rejecting if and only if

$$T^u < c_\alpha.$$

It follows from our assumptions and Slutsky's theorem that

$$\lim_{n \rightarrow \infty} P(T^l < c_\alpha) \leq \alpha$$

under H0,1. Similarly

$$\lim_{n \rightarrow \infty} P(T^u < c_\alpha) \leq \alpha$$

under H0,2.

B. Uniform Confidence Sets

Instead of just considering the relative social welfare achievable by two given policies h^a and h^b , we might be interested in the set of all policies h^a that can be identified to be preferred to a given h^b or, equivalently, the set of all h^{ab} such that $\langle h^{ab}, g \rangle \geq 0$ for all $g \in \mathcal{G}$.¹³ This set is the dual cone \mathcal{G}^* of \mathcal{G} .

We might also be interested in the set of all policies h^a such that h^b cannot be identified to be preferred to h^a or,

¹³This section benefited considerably from discussions with Isaiah Andrews. The ‘‘uniform’’ in the title of this section thus refers to uniformity in the set of policies h . Note that this is different from uniformity in the data-generating process P , a central concern of the literature on inference under partial identification. Lack of uniformity in the latter sense often arises as a consequence of kinks in the boundary of sets corresponding to some null hypothesis.

equivalently, the set of all h^{ab} such that $\langle h^{ab}, g \rangle \leq 0$ for all $g \in \mathcal{G}$. This set is the complement of the polar cone \mathcal{G}^\times of \mathcal{G} .

We now construct set estimators \widehat{H}_1 and \widehat{H}_2 containing these two sets with prespecified asymptotic probability $1 - \alpha$. To do so, note first the following:

1. The mapping from \mathcal{G} to its dual cone \mathcal{G}^* is monotonically decreasing in the sense that

$$\mathcal{G}_1 \subset \mathcal{G}_2 \Rightarrow \mathcal{G}_2^* \subset \mathcal{G}_1^*.$$

2. The same holds for the polar cone

$$\mathcal{G}_1 \subset \mathcal{G}_2 \Rightarrow \mathcal{G}_2^\times \subset \mathcal{G}_1^\times,$$

so that

$$\mathcal{G}_1 \subset \mathcal{G}_2 \Rightarrow \mathcal{G}_1^{\times,c} \subset \mathcal{G}_2^{\times,c},$$

where c denotes complements.

Suppose we have set estimators $\widehat{\mathcal{G}}_1$ and $\widehat{\mathcal{G}}_2$ that satisfy that

$$\lim_{n \rightarrow \infty} P(\widehat{\mathcal{G}}_1 \subset \mathcal{G}) \geq 1 - \alpha,$$

$$\lim_{n \rightarrow \infty} P(\widehat{\mathcal{G}}_2 \supset \mathcal{G}) \geq 1 - \alpha.$$

Set

$$\widehat{H}_1 = \widehat{\mathcal{G}}_1^*, \tag{26}$$

$$\widehat{H}_2 = \widehat{\mathcal{G}}_2^{\times,c}. \tag{27}$$

Then, using set monotonicity of the mapping to dual or polar cones, we get

$$\lim_{n \rightarrow \infty} P(\widehat{H}_1 \supset \mathcal{G}^*) \geq 1 - \alpha,$$

$$\lim_{n \rightarrow \infty} P(\widehat{H}_2 \supset \mathcal{G}^{\times,c}) \geq 1 - \alpha.$$

It remains to construct set estimators $\widehat{\mathcal{G}}_1$ and $\widehat{\mathcal{G}}_2$ with the desired coverage properties. Let us first consider $\widehat{\mathcal{G}}_2$. We propose an estimator $\widehat{\mathcal{G}}_2$ that takes the estimated bounds $\widehat{g}, \widehat{\bar{g}}$ and expands them outward using factors $\underline{\sigma}(x) \times \delta$ and $\overline{\sigma}(x) \times \delta$, which are chosen based on $\widehat{\Sigma}$ in such a way to guarantee the required asymptotic coverage. That is, we consider estimators $\widehat{\mathcal{G}}_2$ of the form¹⁴

$$\widehat{\mathcal{G}}_2 = \left\{ g(\cdot) : g(x) \in [\widehat{g}(x) - \underline{\sigma}(x) \cdot \delta, \widehat{\bar{g}}(x) + \overline{\sigma}(x) \cdot \delta] \right\}, \tag{28}$$

where δ is chosen to guarantee the required coverage. The factors $\underline{\sigma}(x)$ and $\overline{\sigma}(x)$ are the asymptotic standard deviations of $\widehat{g}(x), \widehat{\bar{g}}(x)$, that is, the square roots of the appropriate diagonal elements of Σ .

¹⁴ This is a version of the approach proposed by Horowitz and Manski (2000) for inference on identified sets; see also the discussion in Imbens and Manski (2004).

In this setting, δ depends on only the covariance matrix Σ and can be easily calculated numerically to solve

$$\delta(\Sigma) := \min\{\delta : P(Z_i + \overline{\sigma}(x) \cdot \delta \geq 0 \text{ for } i = 1, \dots, n; Z_i - \underline{\sigma}(x) \cdot \delta \leq 0 \text{ for } i = n + 1, \dots, 2n) \geq 1 - \alpha\} \tag{29}$$

for

$$Z \sim N\left(0, \frac{1}{n} \Sigma\right).$$

The relevant probability in this expression can be evaluated using Monte Carlo draws. As the ‘‘critical value’’ δ is one-dimensional and the relevant probability is monotonic in δ , it is straightforward to solve for $\delta(\Sigma)$ using, for instance, a grid search procedure. Using the consistent variance estimator $\widehat{\Sigma}$, define the plug-in estimator

$$\widehat{\mathcal{G}}_2 = \left\{ g(\cdot) : g(x) \in \left[\widehat{g}(x) - \widehat{\underline{\sigma}}(x) \cdot \delta(\widehat{\Sigma}), \widehat{\bar{g}}(x) + \widehat{\overline{\sigma}}(x) \cdot \delta(\widehat{\Sigma}) \right] \right\},$$

where $\widehat{\underline{\sigma}}(x), \widehat{\overline{\sigma}}(x)$ are the estimates of $\underline{\sigma}(x)$ and $\overline{\sigma}(x)$ implied by $\widehat{\Sigma}$. This estimator has the desired asymptotic coverage:

$$\lim_{n \rightarrow \infty} P(\widehat{\mathcal{G}}_2 \supset \mathcal{G}) = 1 - \alpha.$$

Define analogously

$$\epsilon(\Sigma) := \min\{\epsilon : P(Z_i - \overline{\sigma}(x) \cdot \epsilon \geq 0 \text{ for } i = 1, \dots, n; Z_i + \underline{\sigma}(x) \cdot \epsilon \leq 0 \text{ for } i = n + 1, \dots, 2n) \geq 1 - \alpha\}$$

and

$$\widehat{\mathcal{G}}_1 = \left\{ g(\cdot) : g(x) \in \left[\widehat{g}(x) + \widehat{\underline{\sigma}}(x) \cdot \epsilon(\widehat{\Sigma}), \widehat{\bar{g}}(x) - \widehat{\overline{\sigma}}(x) \cdot \epsilon(\widehat{\Sigma}) \right] \right\}.$$

If lower and upper bounds for one of the intervals defining $\widehat{\mathcal{G}}_1$ cross, $\widehat{\mathcal{G}}_1$ is understood to be the empty set. For this definition, we get

$$\lim_{n \rightarrow \infty} P(\widehat{\mathcal{G}}_1 \subset \mathcal{G}) \geq 1 - \alpha.$$

Note that various alternative estimators $\widehat{\mathcal{G}}_1$ and $\widehat{\mathcal{G}}_2$ with the desired coverage properties are easily constructed. In particular, factors of proportionality other than $\underline{\sigma}(x), \overline{\sigma}(x)$ could be used to scale the expansion and shrinkage terms δ, ϵ . Any choice that is based on $\widehat{\Sigma}$ alone (i.e., not based on the estimated bounds $\widehat{g}, \widehat{\bar{g}}$) might be used to construct sets with the desired coverage. Alternative choices might be useful, in particular, if power against particular alternatives is a concern.

VI. Application to Data from the Project STAR Experiment

In this section, the theoretical results developed in this paper are applied to data from the famous project STAR experiment. Data from this experiment have been used by many contributions to the economics of education literature, such as Krueger (1999) and Graham et al. (2008); background information on the experiment can be found in Boyd-Zaharias et al. (2007). The following discussion builds in particular on Krueger (1999).

The project STAR experiment, Student/Teacher Achievement Ratio, was conducted in Tennessee beginning in 1985. In this experiment, kindergarten students and teachers were randomly assigned to one of three groups beginning in the 1985–1986 school year: small classes (13 to 17 students per teacher), regular-size classes (22 to 25 students), and regular classes with a full-time teacher aide. Students were supposed to remain in the same class type for four years. The data set covers students from eighty schools. Each participating school was required to have at least one of each class type, and random assignment took place within schools. Kindergarten attendance was not mandatory in Tennessee during this time. Students entering a participating school in grade 1 or later were randomly assigned to a class on entry. Compliance with experimental assignment was incomplete, as a number of students switched between small and regular classes between grades. Furthermore, a sizable number of students are missing from the data set in years following the initial random assignment.

A number of features of this experiment make it an interesting application for the theoretical results of this paper. First, compliance to experimental assignment was large but incomplete. This implies that conditional average treatment effects are only bound-identified, but the bounds are not too wide so as to be completely uninformative. Second, there appears to be heterogeneity in treatment effects (Krueger, 1999). This heterogeneity implies that reallocations subject to a budget constraint might be potentially welfare improving. Third, there might be disagreement about the objective function, the credibility of various identifying assumptions, and the budget constraint. We demonstrate how the identification of policy rankings depends on the interaction of these three.

For our analysis, we restrict attention to the sample of students who are observed in grades 1 through 3. Table 1 shows a number of summary statistics for this sample. Here “poor” equals 1 for students receiving a free lunch in the first year they are observed in the sample. We pool classes with and without an aide and study only the effect of class size. Assigned treatment Z equals 1 for students assigned to a small class on first entering a project STAR school. The realized treatment variable D equals 1 for students who remained in a small class throughout the study period, except for at most one year. The fact that observations are not balanced across values of Z is not indicative of a violation of

TABLE 1.—SUMMARY STATISTICS FOR STUDENTS ASSIGNED TO SMALL AND REGULAR CLASSES

	Assigned to Regular Class ($Z = 0$)	Assigned to Small Class ($Z = 1$)	All
Poor	0.258 (0.438)	0.324 (0.468)	0.277 (0.447)
Black	0.281 (0.450)	0.264 (0.441)	0.276 (0.447)
Girl	0.499 (0.500)	0.507 (0.500)	0.502 (0.500)
Stayed in small class (D)	0.070 (0.255)	0.942 (0.235)	0.310 (0.463)
Math	−0.058 (0.995)	0.157 (0.998)	0.002 (1.000)
Reading	−0.063 (0.996)	0.187 (0.983)	0.006 (0.999)
Observations	3,090	1,156	4,246

This table shows means and standard deviations of various characteristics for students assigned to small and regular classes. Note that the means are not expected to be equal, since randomization took place at the school level, not in the entire sample.

TABLE 2.—THE JOINT DISTRIBUTION OF ASSIGNED AND REALIZED CLASS SIZE

Z	D		Total
	0	1	
0	2,873	217	3,090
1	74	1,082	1,156
Total	2,947	1,299	4,246

This table shows the joint distribution of assigned treatment Z and realized treatment D in our sample. Assigned treatment equals 1 if students were assigned to a small class on first entering a project STAR school, and realized treatment equals 1 for students who remained in a small class throughout the study period, except for at most one year.

randomization since randomization took place on the school level. Table 2 shows the joint distribution of assigned and realized class size. It can be seen from this table that non-compliance with treatment assignment is an issue. Realized treatment is not equal to assigned treatment for about 7% of students.

Given random assignment of initial class type, a natural assumption in this setting is conditional instrument exogeneity, $Z \perp (Y^0, Y^1) | X$, for potential test scores Y^0, Y^1 that a student would have achieved under a counterfactual class type. The vector X includes controls for *School*, as well as potentially further predetermined variables such as *Poor*. Since compliance is imperfect, $D \neq Z$, this assumption does not allow point-identifying average treatment effects, however. Identified sets under the assumption of instrument exogeneity are discussed in online appendix B. We are also considering identified sets under the additional assumption of monotone treatment response, $Y^1 \geq Y^0$. Identification of potential outcome distributions under this assumption is discussed in Manski (1997).

The outcome variables Y considered are normalized test scores averaged over third- and fourth-grade tests for math and for reading. The objective functions ϕ we look at are various quantiles of the outcome distribution $f(Y)$. Two types of restricted policy sets are studied: policies increasing or decreasing the probability of being assigned to a small class for all students by an equal amount and policies reallocating teachers between nonpoor and poor students,

TABLE 3.—BOUNDS ON $E[g^\phi|\text{POOR}]$, $E[g^\phi|\text{NONPOOR}]$, $E[g^\phi]$, ϕ^{ab} FOR REDISTRIBUTION

Objective ϕ	Poor Students	Nonpoor Students	All Students	Effect of Redistribution
Normalized Average Math Scores in Third and Fourth Grades				
0.3th quantile	[0.141, 0.487]	[−0.027, 0.329]	[0.032, 0.384]	[−0.066, 0.180]
0.5th quantile	[0.081, 0.407]	[−0.011, 0.325]	[0.021, 0.354]	[−0.086, 0.147]
0.7th quantile	[−0.179, 0.218]	[0.113, 0.521]	[0.011, 0.414]	[−0.246, 0.037]
Assuming Instrument Exogeneity and Monotonicity				
0.3th quantile	[0.460, 0.607]	[0.272, 0.410]	[0.336, 0.476]	[0.017, 0.117]
0.5th quantile	[0.445, 0.563]	[0.313, 0.407]	[0.357, 0.458]	[0.013, 0.088]
0.7th quantile	[0.304, 0.404]	[0.502, 0.621]	[0.429, 0.541]	[−0.111, −0.034]
Normalized Average Reading Scores in Third and Fourth Grades				
Assuming Only Instrument Exogeneity				
0.3th quantile	[0.119, 0.461]	[0.000, 0.370]	[0.042, 0.402]	[−0.088, 0.162]
0.5th quantile	[0.048, 0.392]	[0.108, 0.480]	[0.087, 0.449]	[−0.152, 0.100]
0.7th quantile	[0.022, 0.424]	[0.111, 0.546]	[0.080, 0.503]	[−0.184, 0.110]
Assuming Instrument Exogeneity and Monotonicity				
0.3th quantile	[0.457, 0.575]	[0.256, 0.404]	[0.324, 0.461]	[0.019, 0.112]
0.5th quantile	[0.395, 0.502]	[0.398, 0.542]	[0.394, 0.524]	[−0.052, 0.036]
0.7th quantile	[0.428, 0.543]	[0.476, 0.634]	[0.456, 0.598]	[−0.072, 0.024]

This table shows bounds on the effect on various quantiles ϕ of the population distribution of test scores of reducing class size for poor students, nonpoor students, and all students and of redistributing resources toward poor students. See text for additional details.

holding the total number of teachers constant. Table 3 shows estimated bounds on $E[g^\phi|\text{poor}]$ and $E[g^\phi|\text{nonpoor}]$, the “average treatment effects” on ϕ of class size for the groups of students with and without free lunch, for the two outcome variables and various quantiles ϕ .¹⁵ These give the estimated effect of reassigning students from regular to small classes for either group on the unconditional outcome distribution. The final two columns give estimated bounds on the implied effect of two further policies: the first moving all students from regular to small classes, and the second moving all poor students from regular to small classes and compensating by moving an equal number of nonpoor students from small to regular classes. Figure 2 conveys the same information graphically, including additional quantiles.

The results reported in table 3 and figure 2 illustrate nicely the dependence of the identified policy ranking on (a) identifying assumptions, (b) the objective function, and (c) the set of feasible policies:

1. Assuming only instrument exogeneity, we cannot identify the ranking of policies redistributing teachers. If however, we additionally impose monotone treatment effects, the data become informative about the ranking of redistributive policies.
2. Consider the effect of redistributing resources from nonpoor to poor students under the assumptions of exogeneity and monotonicity. The effect of such a redistribution on lower quantiles of the test score distribution is unambiguously positive. The effect on top quantiles is ambiguous, however.
3. If feasible policies are subject to an aggregate budget constraint, then the ranking in terms of top quantiles is not identified. Policies decreasing class sizes for all

students equally can be unambiguously ranked, however, even without the assumption of monotonicity.

It is useful to recall how the theoretical developments in previous sections influenced our empirical approach in this section: First, and most important, is our choice of objects of interest. Identified welfare rankings of policies, while arguably of great applied interest, have not been the subject of previous empirical studies to the best of our knowledge.

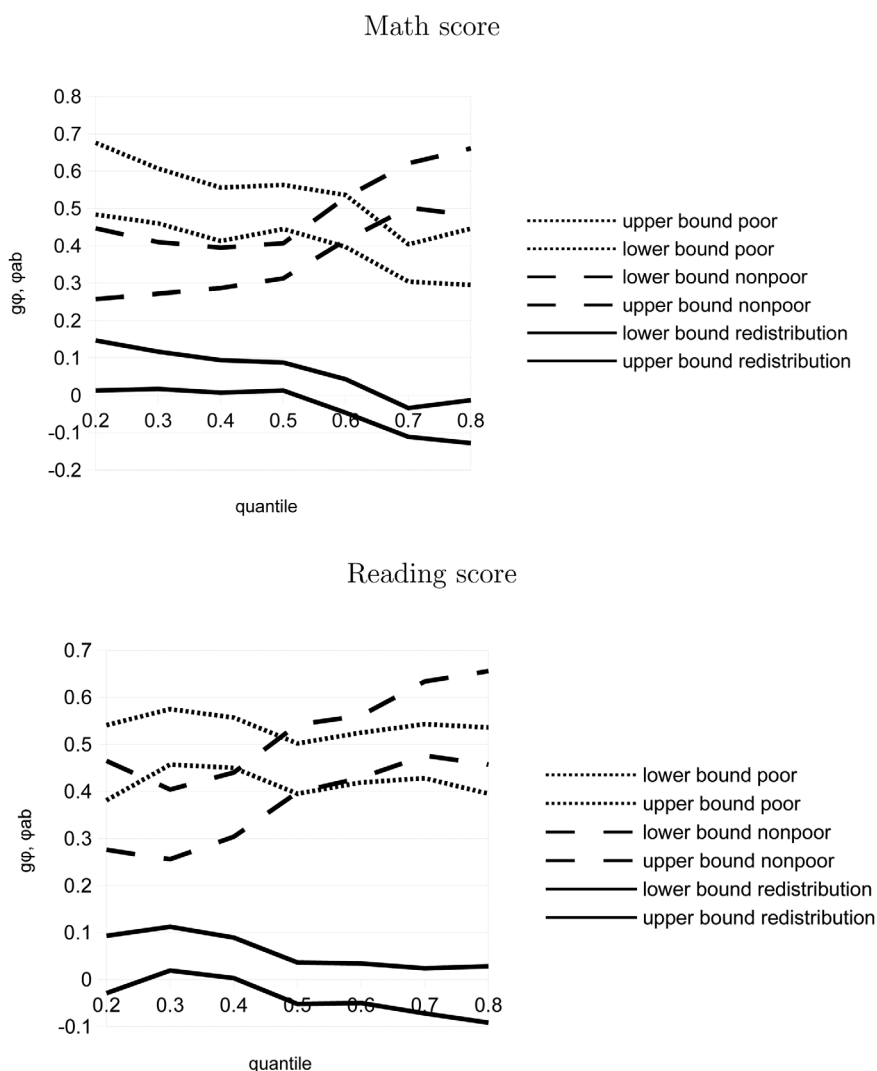
Second, our application demonstrates the interplay between identifying assumptions, objective functions, and constraints on feasible policies in determining the identifiability of policy rankings. From this vantage point, this section is not so much an application of new methods but rather a specific example illustrating our general argument. We elaborated on this interplay in propositions 1 through 4 for the linear case and generalized these propositions to the nonlinear case in propositions 6 and 7.

Third, we focus on quantiles of test score distributions as our objective functions. Bounds on the (approximate) effect of policies on these quantiles are constructed using the results of proposition 5. As discussed at the end of section IV, for the case of quantiles, the fact that we use a linear approximation does not affect our conclusion regarding policy rankings. This is because the sign of policy effects on quantiles is minus the sign of policy effects on the cumulative distribution function at specific points. The cumulative distribution function is in fact linear in the outcome distribution.

Fourth, we perform inference on supersets of the status quo policy in the following section. The inference procedure used is uniform in the set of policies in the sense that we construct sets such that there is an asymptotic probability of 0.95 that all policies in these sets are contained in the true superset. We developed such inference procedures in section VB.

¹⁵ We take the sample distribution of test scores as the status quo distribution, based on which to evaluate the effect of changing treatment assignment policies.

FIGURE 2.—BOUNDS ON EFFECTS ON QUANTILES OF THE TEST SCORE DISTRIBUTION



These figures convey the same information as table 3, depicting effects for additional quantiles.

A. Inference

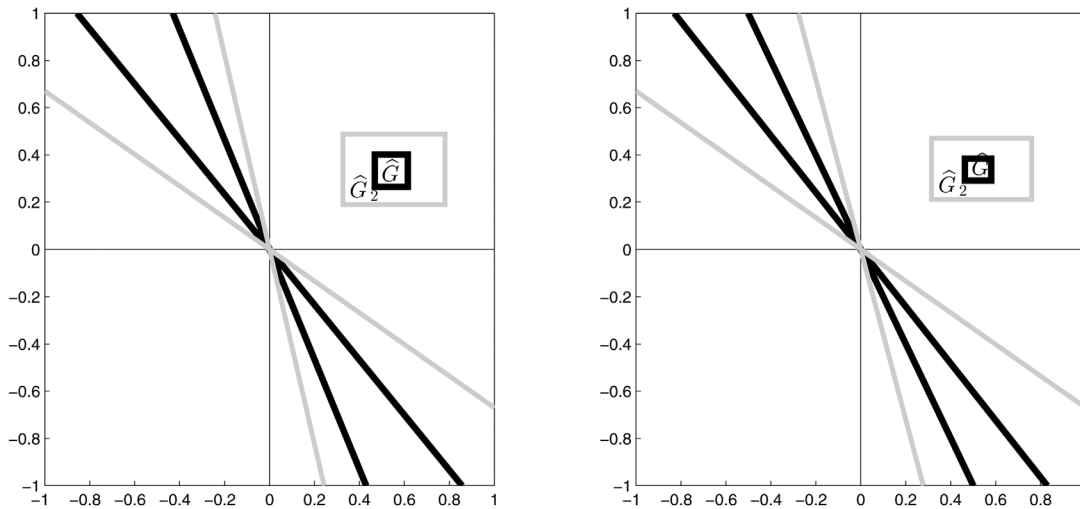
Figure 3 shows the results of implementing the inference procedure proposed in section VB. The black rectangles in this figure correspond to the point estimates of $\hat{\mathcal{G}}$, that is, the point estimates of bounds on conditional quantile treatment effects as displayed in table 3. The larger gray rectangles $\hat{\mathcal{G}}_2$ correspond to confidence sets, which contain the identified set for g^ϕ with probability 0.95. These pictures do not show sets $\hat{\mathcal{G}}_1$ contained in the true set with probability 0.95, as our method yields empty sets for these; the bounds are too imprecisely estimated. Note that the confidence sets $\hat{\mathcal{G}}_2$ are wider than pointwise confidence bands would be, since we require uniform (simultaneous) coverage across all bounds and across poor and nonpoor students (see section VB). Estimates of the covariance matrix of estimated bounds are obtained using the bootstrap.

Figure 3 further displays the dual and polar cones of $\hat{\mathcal{G}}$ and $\hat{\mathcal{G}}_2$ (with respect to the Euclidean inner product). The

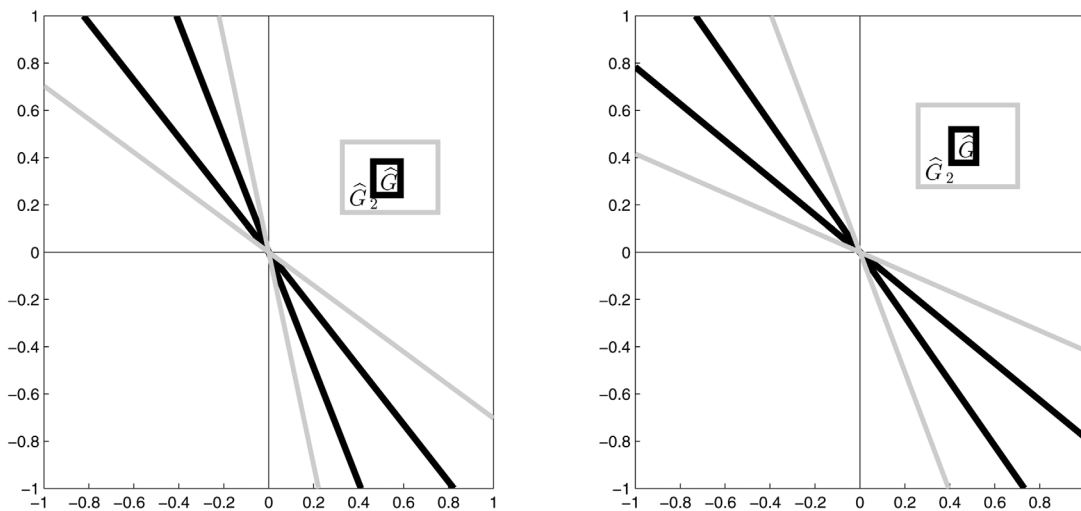
dual cone of $\hat{\mathcal{G}}$ is the set of all points to the northeast of the black lines. This is the set of policy differences h^{ab} (rescaled by population shares of poor and nonpoor students) for which our point estimates indicate an unambiguously positive effect on the respective quantiles of reading and test scores. Purely redistributive policies, subject to a population budget constraint on teachers, correspond to points on the downward-sloping 45 degree line. Consistent with the estimates in table 3, our point estimates suggest that redistribution to poor students has positive effects on the 0.3 quantiles of reading and math scores, as well as on the 0.5 quantile of math scores, while the effect on the 0.5 quantile of reading scores is ambiguous. The dual cone of $\hat{\mathcal{G}}_2$ that is the set of points to the northeast of the gray lines, is a set contained in the dual cone of the identified set $\hat{\mathcal{G}}$ with probability 0.95. The probability of any of the policies in this dual cone not resulting in increases of the objective function is 0.05. As can be seen from the graphs, we cannot conclusively reject the possibility that the effect of

FIGURE 3.—IDENTIFICATION REGIONS FOR THE CONDITIONAL ATE, AND THE SET OF ORDERED POLICIES

Math score, 0.3 quantile and 0.5 quantile



Reading score, 0.3 quantile and 0.5 quantile



These figures are analogous to figure 1. The horizontal axis represents treatment effects for poor (free-lunch) students; the vertical axis represents treatment effects for nonpoor students. The black rectangles $\hat{\mathcal{G}}$ depict point estimates for bounds on the conditional effects g^0 ; the gray rectangles $\hat{\mathcal{G}}_2$ depict confidence sets containing the true \mathcal{G} with probability 0.95. The straight black and gray lines indicate the boundaries of the corresponding dual and polar cones. See the discussion in section VB.

redistribution toward poor students is ambiguous. The effect of a decrease in class sizes for some or all students, however, is unambiguously positive and statistically significantly so, no matter which of the quantiles we choose as our objective.

VII. Conclusion

The goal of this paper is to explore the frontier in the trade-off between the twin goals of recognition of the limits of our knowledge, on the one hand, and the need to give informed policy recommendations, on the other hand. Philosophically speaking, the goal is to reconcile a position of skepticism,

which refrains from making strong a priori assumptions, and of pragmatism, which asserts that relevant empirical parameters are those having implications for policy choice. This paper discusses in particular under what conditions and to what extent the welfare ranking of policies is identified. The answer to this question depends on the interaction of the identified set, the feasible policy set, and the objective function. This paper discusses this question in the context of policies allocating a binary treatment based on observable covariates under partial identification of conditional average treatment effects and with possibly restricted sets of feasible policies. General nonlinear objective functions are allowed for.

APPENDIX A

Proofs

Proof of Proposition 1. The first claim is immediate since either $h^a \succeq^{\mathcal{G}} h^b$ or $h^b \succeq^{\mathcal{G}} h^a$ holds if and only if $h^{ab} \in \mathcal{G}^* \cup \mathcal{G}^\times$.

Next, note that either $h^a \succ^{\mathcal{G}} h^b$ or $h^b \succ^{\mathcal{G}} h^a$ (i.e., a strict ordering of h^a, h^b) holds if and only if $h^{ab} \notin \left(\bigcup_{g \in \mathcal{G}} g^\perp\right)$. Since \mathcal{G} is convex and therefore connected, and since the inner product is continuous, the set $\langle h^{ab}, \mathcal{G} \rangle$ is connected. It therefore contains both positive and nonpositive values only if it contains 0. This is the case if there is a $g \in \mathcal{G}$ such that $\langle h^{ab}, g \rangle = 0$, that is, if there is a $g \in \mathcal{G}$ such that $h^{ab} \in g^\perp$.

The second claim then follows from the fact that the set of h^a, h^b for which either $h^a \succeq^{\mathcal{G}} h^b$ or $h^b \succeq^{\mathcal{G}} h^a$ holds is the closure of the set for which either $h^a \succ^{\mathcal{G}} h^b$ or $h^b \succ^{\mathcal{G}} h^a$ holds. Let $g^1 = \operatorname{argmin}_{g \in \mathcal{G}} \langle h^{ab}, g \rangle$. If h_n is a sequence in $d\mathcal{H} \cap \{h : \min_{g \in \mathcal{G}} \langle h, g \rangle > 0\}$ converging to h^{ab} , then

$$\min_{g \in \mathcal{G}} \langle h^{ab}, g \rangle = \langle h^{ab}, g^1 \rangle = \lim_n \langle h_n, g^1 \rangle \geq \liminf_n \min_{g \in \mathcal{G}} \langle h_n, g \rangle \geq 0.$$

Conversely, for any h^{ab} such that $\min_{g \in \mathcal{G}} \langle h^{ab}, g \rangle \geq 0$, there is an h in any of the neighborhoods of h^{ab} such that $\min_{g \in \mathcal{G}} \langle h, g \rangle > 0$. We can choose, for instance, $h = h^{ab} + \epsilon h'$, where h' is such that $\min_{g \in \mathcal{G}} \langle h', g \rangle > 0$. Such a hyperplane separating 0 from \mathcal{G} exists. Choose $h' = \operatorname{argmin}_{g \in \mathcal{G}} \|g\|$. By $0 \notin \mathcal{G}$, we have $\|h'\| > 0$, $\langle h', g \rangle > 0$ for all $g \in \mathcal{G}$ follows from convexity of \mathcal{G} .

Proof of Proposition 2. Using the definition of g^l, g^u from section IIIA, we have

$$\phi^{ab} - \underline{\phi}^{ab} = \langle h^{ab}, g - g^l \rangle = \langle |h^{ab}|, |g - g^l| \rangle \quad (\text{A1})$$

and

$$\overline{\phi}^{ab} - \phi^{ab} = \langle h^{ab}, g^u - g \rangle = \langle |h^{ab}|, |g^u - g| \rangle. \quad (\text{A2})$$

By definition of w ,

$$|g - g^l| \leq |g^u - g^l| = \bar{g}(X) - \underline{g}(X) = w(X), \quad (\text{A3})$$

similarly for $|g^u - g|$, and therefore

$$\begin{aligned} \phi^{ab} - \underline{\phi}^{ab} &\leq \langle |h^{ab}|, w \rangle, \\ \overline{\phi}^{ab} - \phi^{ab} &\leq \langle |h^{ab}|, w \rangle. \end{aligned} \quad (\text{A4})$$

If h^a, h^b are deterministic policies, then $|h^{ab}| \in \{0, 1\}$, and we can rewrite the right-hand side of inequality (A4) as

$$\langle |h^{ab}|, w \rangle = E[\mathbf{1}(h^{ab} \neq 0) \cdot w] = E[w | \mathcal{X}^{ab}] P(\mathcal{X}^{ab}). \quad (\text{A5})$$

Proof of Proposition 3. Suppose to the contrary that there are three vectors $h^1, h^2, h^3 \in \mathcal{H}'$, such that $h^1 - h^3$ and $h^2 - h^3$ are linearly independent. Take any g from the interior of \mathcal{G} . Define $\alpha^i = \langle h^i - h^3, g \rangle$ for $i = 1, 2$. Since \mathcal{H}' is totally ordered and g is in the interior of \mathcal{G} , $\alpha_i \neq 0$ for $i = 1, 2$. Let

$$h^* = (h^1 - h^3) - \frac{\alpha^1}{\alpha^2} (h^2 - h^3).$$

h^* lies in the linear space spanned by $h^1 - h^3, h^2 - h^3$. Furthermore $\langle h^*, g \rangle = 0$. We can find two policy vectors h^4, h^5 in the triangle spanned by h^1, h^2, h^3 , and hence in the policy set \mathcal{H}' , such that $h^4 - h^5 = \beta h^*$ for some $\beta \neq 0$. Since g is in the interior of \mathcal{G} , we can now find $g^1, g^2 \in \mathcal{G}$, such that $\langle h^4 - h^5, g^1 \rangle < 0 < \langle h^4 - h^5, g^2 \rangle$. Contradiction.

Proof of Proposition 4. We prove the case $k = 1$. The argument immediately generalizes to the case of general k . The subscript 1 is dropped from c, λ . First, we show that statement ii implies statement i. Suppose λc lies in the interior of \mathcal{G} . By definition of \mathcal{H}' , $\langle h^a - h^b, g \rangle = 0$ for all $h^a, h^b \in \mathcal{H}'$

and $g = \lambda c$. Since g lies in the interior of \mathcal{G} , the sign of $\langle h^a - h^b, g' \rangle$ is therefore positive and negative for some $g' \in \mathcal{G}$.

To see that statement i implies statement ii, suppose that λc does not lie in \mathcal{G}^o for any λ . By the Hahn-Banach separating hyperplane theorem, applied to the sets \mathcal{G} and $\{\lambda c : \lambda \in \mathbb{R}\}$, there exists an \tilde{h} such that

$$\sup_{\lambda \in \mathbb{R}} \langle \tilde{h}, \lambda c \rangle \leq \inf_{g \in \mathcal{G}} \langle \tilde{h}, g \rangle.$$

The left-hand side is bounded only if $\langle \tilde{h}, c \rangle = 0$, which thus must hold, and therefore

$$0 \leq \inf_{g \in \mathcal{G}} \langle \tilde{h}, g \rangle.$$

Therefore, any policy pair $h^a, h^b \in \mathcal{H}'$ such that h^{ab} is proportional to \tilde{h} is weakly ordered.

Proof of Lemma 1. Existence of IF follows from assumption 5a and from the general result that the dual of L^p is isomorphic to L^q (see Rudin, 1991). Next, note that $f(y; \theta) - \hat{f}(y) = \langle h(\theta) - h(0), g^f(y|\cdot) \rangle$ by construction, from which the existence of $f_0(y)$ and the second claim immediately follow.

Considering ϕ as a mapping from h to \mathbb{R} , differentiability of this mapping follows from the previous two arguments and the chain rule. The existence of g^ϕ then follows from the finite-dimensional Riesz representation theorem.

Finally, using the chain rule and combining these expressions yields

$$\langle h_0, g^\phi(\cdot; \check{f}) \rangle = \int IF(y; \check{f}) \langle h_0, g^f(y|\cdot) \rangle d\mu(y).$$

Exchanging the order of integration with regard to x and y , justified by Fubini's theorem, gives

$$\langle h_0, g^\phi(\cdot; \check{f}) \rangle = \langle h_0, \int IF(y; \check{f}) g^f(y|\cdot) d\mu(y) \rangle.$$

Since this holds for all h_0 , the last claim follows.

Proof of Proposition 5. By lemma 1 and under assumption 6, we have

$$\begin{aligned} g^\phi(x; f^*) &= \int IF(y; f^*) g^f(y|x) d\mu(y) \\ &= \int IF(y; f^*) (\bar{g}^f(\cdot|x) + \gamma^1(x) \cdot f^1(\cdot|x, cf) \\ &\quad - \gamma^0(x) \cdot f^0(\cdot|x, cf)) d\mu(y) \\ &= \int IF(y; f^*) \bar{g}^f(\cdot|x) d\mu(y) + \gamma^1(x) \cdot E[IF(Y^1; f^*)|x, cf] \\ &\quad - \gamma^0(x) \cdot E[IF(Y^0; f^*)|x, cf], \end{aligned} \quad (\text{A6})$$

where the expectations are taken over the conditional potential outcome distributions f^1, f^0 . These expectations are obviously bounded by $\inf_{y \in \mathcal{Y}} IF(y; f^*)$, $\sup_{y \in \mathcal{Y}} IF(y; f^*)$. Furthermore, these bounds are tight without further restrictions on the potential outcome distributions. The claim follows.

Proof of Theorem 1. First, note that the identified set for $g^f(\cdot|x)$ is necessarily bounded in L^1 (i.e., in total variation), since $\|g^f(\cdot|x)\| = \|f^1(\cdot|x) - f^0(\cdot|x)\| \leq \|f^1(\cdot|x)\| + \|f^0(\cdot|x)\| = 2$. It follows that $\|((h - h^*), g^f)\| \leq 2E[\|(h - h^*)(X)\|]$, and $\|((h - h^*) + \gamma h_0, g^f)\| \leq 2E[\|(h - h^*)(X)\|] + 2\gamma E[\|h_0(X)\|]$.

Now fix any $g^f \in \mathcal{G}^f$. By the mean value theorem,

$$\begin{aligned} \phi(f^* + ((h - h^*) + \gamma h_0, g^f)) - \phi(f^* + ((h - h^*), g^f)) \\ = \frac{\partial \phi}{\partial f}(\check{f}) \cdot \langle \gamma h_0, g^f \rangle, \end{aligned}$$

where \tilde{f} lies between $f^* + \langle (h - h^*) + \gamma h_0, g^f \rangle$ and $f^* + \langle (h - h^*), g^f \rangle$. By the previous argument, we have that $\|\tilde{f} - f^*\| \leq 2E[\langle (h - h^*), X \rangle] + 2\gamma E[\langle h_0, X \rangle]$.

By assumption we have

$$\frac{\partial \phi}{\partial f}(f^*) \cdot \langle h_0, g^f \rangle \geq \inf_{g^{\phi} \in \mathcal{G}^{\phi}(h^*)} \langle h_0, g^{\phi} \rangle = \underline{d\phi}(h_0; h^*) =: \delta_1 > 0.$$

We get the inequality

$$\begin{aligned} \frac{\partial \phi}{\partial f}(\tilde{f}) \cdot \gamma \langle h_0, g^f \rangle &\geq \gamma \delta_1 - \gamma \left\| \frac{\partial \phi}{\partial f}(f^*) - \frac{\partial \phi}{\partial f}(\tilde{f}) \right\| \cdot \|\langle h_0, g^f \rangle\| \\ &\geq \gamma \delta_1 - 2\gamma \left\| \frac{\partial \phi}{\partial f}(f^*) - \frac{\partial \phi}{\partial f}(\tilde{f}) \right\| \cdot E[\langle h_0, X \rangle]. \end{aligned} \quad (A7)$$

By continuity of $\partial \phi / \partial f$, there exists a $\delta_2 > 0$, such that

$$\left\| \frac{\partial \phi}{\partial f}(f^*) - \frac{\partial \phi}{\partial f}(\tilde{f}) \right\| \leq \frac{\delta_1}{4E[\langle h_0, X \rangle]}$$

for $\|\tilde{f} - f^*\| \leq \delta_2$. As shown before, $\|\tilde{f} - f^*\| \leq \delta_2$ if $2E[\langle (h - h^*), X \rangle] + 2E[\langle h_0, X \rangle] \leq \delta_2$. Choosing $\delta = \delta_2 / 4(n + 1)$ yields the claim, as inequality (A7) holds for all g^f .

Proof of Proposition 6. By proposition 3, it is immediate that $T_{h^*} \mathcal{H}'$ is at most one-dimensional if it is totally ordered by $\succeq^{\mathcal{G}^{\phi}}$. But if $T_{h^*} \mathcal{H}'$ is one-dimensional, then so is \mathcal{H}' .

Proof of Proposition 7. This follows immediately from proposition 3 and assumption 7.

Proof of Theorem 2. Note first that $d\phi(h_0; h)$ is Lipschitz continuous in h_0 if $\mathcal{G}^{\phi}(h)$ is bounded, with Lipschitz constant $\sup_x \max(|g^{\phi}(x; h)|, |g^{\phi}(x; h)|)$. Furthermore, by continuous differentiability of ϕ , $d\phi(h_0; h)$ is also continuous in h , since

$$|d\phi(h_0; h^a) - d\phi(h_0; h^b)| \leq \sup_{g^f \in \mathcal{G}^f} 2 \left\| \frac{\partial \phi}{\partial f}(f^a) - \frac{\partial \phi}{\partial f}(f^b) \right\| E[\langle h_0, \cdot \rangle],$$

where $f^a = f^* + \langle h^a - h^*, g^f \rangle$, and similarly for f^b . This follows immediately from the proof of theorem 1 using the boundedness of \mathcal{G}^f . The claim follows since $f^a - f^b \rightarrow 0$ as $h^a - h^b \rightarrow 0$.

i. If $TU(h^*) = \{0\}$, then $d\phi(h_0; h^*) < 0$ for all $0 \neq h_0 \in T_{h^*} \mathcal{H}'$. Since $\dim(\mathcal{H}') < \infty$ the unit ball in \mathcal{H}' is compact. By continuity of $d\phi(h_0; h)$ in both its arguments, this implies that there is an $\epsilon > 0$ such that

$$-\delta := \sup_{h: \|h - h^*\| < \epsilon} \sup_{h_0 \in T_{h^*} \mathcal{H}': \|h_0\| = 1} d\phi(h_0; h) < 0.$$

Finally, the same holds if the supremum over h_0 is taken over some neighborhood of $T_{h^*} \mathcal{H}'$. \mathcal{H}' is in such a neighborhood for $\|h - h^*\|$ small enough by differentiability of the constraints C . The claim now follows, applying the mean value theorem to $\phi(f^* + \langle h^{ab} + h - h^*, g^f \rangle) - \phi(f^* + \langle h - h^*, g^f \rangle)$, given g^f :

$$\begin{aligned} &(\phi(f^* + \langle h^{ab} + h - h^*, g^f \rangle) - \phi(f^* + \langle h - h^*, g^f \rangle)) \\ &= \langle h^{ab}, \partial \phi / \partial f(f) \cdot g^f \rangle, \end{aligned}$$

where f is between $f^* + \langle h^{ab} + h - h^*, g^f \rangle$ and $f^* + \langle h - h^*, g^f \rangle$. Taking the infimum over g^f on both sides yields

$$\underline{\Delta \phi}(h^{ab}; h) = \underline{d\phi}(h^{ab}; \tilde{h})$$

for some \tilde{h} between h and $h + h^{ab}$, and thus $\underline{\Delta \phi}(h^{ab}; h) < -\delta \cdot \|h^{ab}\|$, which proves claim i.

ii. Let h_0^1 be such that $d\phi(h_0^1; h^*) > 0$. By continuity of $d\phi(h_0; h)$, there is a neighborhood of h^* , s.t. $d\phi(h_0; h) > 0$ for all h in this neighborhood, and some h_0 in $T_h \mathcal{H}'$.

By differentiability of the constraints C , for every $\epsilon > 0$, there is a h^{ab} s.t. $h + h^{ab} \in \mathcal{H}'$ such that

$$\|h^{ab} / \|h^{ab}\| - h_0 / \|h_0\| \| < \epsilon.$$

By the mean value theorem, as in the proof of claim i, $\underline{\Delta \phi}(h^{ab}; h) = \underline{d\phi}(h^{ab}; \tilde{h})$ for some \tilde{h} between h and $h + h^{ab}$. As we let $h^{ab} \rightarrow 0$, we get

$$\underline{\Delta \phi}(h^{ab}; h) / \|h^{ab}\| = \underline{d\phi}(h^{ab} / \|h^{ab}\|; \tilde{h}) \rightarrow \underline{d\phi}(h_0 / \|h_0\|; h) > 0,$$

by continuity of $d\phi(h_0; h) > 0$, which proves claim ii.

iii. By the proof of proposition 6, $\dim(\mathcal{H}') = \dim(T_h \mathcal{H}') = 1$. By assumption, there is a $h_0 \in T_{h^*} \mathcal{H}'$ such that $d\phi(h_0; h^*) > 0$. By continuity of $d\phi$, the same holds for all h in some neighborhood of h^* . By the same arguments as in the proof of claim ii, we get $\underline{\Delta \phi}(h^{ab}; h) > 0$ for $\|h^{ab} / \|h^{ab}\| - h_0 / \|h_0\| \|$ small enough, as guaranteed by small enough h^{ab} .

Reversing all signs and exchanging suprema and infima shows that \mathcal{H}' is completely ordered in a neighborhood of h , proving claim 3.

REFERENCES

- Angrist, J. D., and J.-S. Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives* 24 (2010), 3–30.
- Anscombe, F., and R. Aumann, "A Definition of Subjective Probability," *Annals of Mathematical Statistics* 34 (1963), 199–205.
- Bertsekas, D. P., A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization* (Belmont, MA: Athena Scientific, 2003).
- Bewley, T., "Knightian Decision Theory, Part I," *Decisions in Economics and Finance* 25 (2002), 79–110.
- Bhattacharya, D., "Inferring Optimal Peer Assignment from Experimental Data," *Journal of the American Statistical Association* 104 (2009), 486–500.
- Bhattacharya, D., and P. Dupas, "Inferring Welfare Maximizing Treatment Assignment under Budget Constraints," *Journal of Econometrics* 167 (2012), 168–196.
- Boyd-Zaharias, J., J. Finn, R. Fish, and S. Gerber, "Project STAR and Beyond: Database User's Guide," HEROS Inc. and University of New York at Buffalo technical report (2007).
- Card, D., "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica* 69 (2001), 1127–1160.
- Carneiro, P., J. Heckman, and E. Vytlacil, "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica* 78 (2010), 377–394.
- Chamberlain, G., "Bayesian Aspects of Treatment Choice," in John Geweke, Gary Koop, and Herman van Dijk, eds., *Oxford Handbook of Bayesian Econometrics* (New York: Oxford University Press, 2011).
- Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey, "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica* 81 (2013), 535–580.
- Chernozhukov, V., I. Fernandez-Val, and B. Melly, "Inference on Counterfactual Distributions," *Econometrica* 81 (2013), 2205–2269.
- Chetty, R., "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods," *Annual Review of Economics* 1 (2009), 451–488.
- Deaton, A., "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature* 48 (2010), 424–455.
- Dehejia, R., "Program Evaluation as a Decision Problem," *Journal of Econometrics* 125 (2005), 141–173.
- DiNardo, J., N. Fortin, and T. Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica* 64 (1996), 1001–1044.
- Firpo, S., N. Fortin, and T. Lemieux, "Unconditional Quantile Regressions," *Econometrica* 77 (2009), 953–973.
- , "Decomposition Methods in Economics," *Handbook of Labor Economics* 4 (2011), 1–102.
- Graham, B. S., G. Imbens, and G. Ridder, "Measuring the Average Outcome and Inequality Effects of Segregation in the Presence of Social Spillovers," University of California, Berkeley working paper (2008).

- Hirano, K., and J. Porter, "Asymptotics for Statistical Treatment Rules," *Econometrica* 77 (2009), 1683–1701.
- Horowitz, J. L., and C. F. Manski, "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association* 95 (2000), 77–84.
- Imbens, G. W., "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature* 48 (2010), 399–423.
- Imbens, G. W., and C. F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72 (2004), 1845–1857.
- Kasy, M., "Using Data to Inform Policy," Harvard University working paper (2014a).
- "Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment," *Review of Economic Studies* 81 (2014b), 1614–1636.
- "Who Wins, Who Loses? Tools for Distributional Policy Evaluation," working paper (2014c).
- Knight, F. H., *Risk, Uncertainty, and Profit* (Chicago: University of Chicago Press, 1921).
- Krueger, A., "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics* 114 (1999), 497–532.
- Manski, C. F., "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica: Journal of the Econometric Society* (1987), 357–362.
- "Ordinal Utility Models of Decision Making under Uncertainty," *Theory and Decision* 25 (1988), 79–104.
- "Monotone Treatment Response," *Econometrica: Journal of the Econometric Society* 65 (1997), 1311–1334.
- *Partial Identification of Probability Distributions* (New York: Springer-Verlag, 2003).
- "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica* 72 (2004), 1221–1246.
- "Choosing Treatment Policies under Ambiguity," *Annual Review of Economics* 3 (2011), 3.
- Mas-Colell, A., M. Whinston, and J. Green, *Microeconomic Theory* (New York: Oxford University Press, 1995).
- Mirrlees, J., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies* 38 (1971), 175–208.
- Nevo, A., and M. Whinston, "Taking the Dogma Out of Econometrics: Structural Modeling and Credible Inference," *Journal of Economic Perspectives* 24 (2010), 69–81.
- Rostek, M., "Quantile Maximization in Decision Theory," *Review of Economic Studies* 77 (2010), 339–371.
- Rothe, C., "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics* 155 (2010), 56–70.
- Rudin, W., *Functional Analysis* (New York: McGraw-Hill, 1991).
- Ryan, M., "Generalizations of SEU: A Geometric Tour of Some Non-Standard Models," *Oxford Economic Papers* 61 (2009), 327.
- Saez, E., and S. Stantcheva, "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," NBER technical report (2013).
- Sen, A., *Inequality Reexamined* (New York: Oxford University Press, 1995).
- Stoye, J., "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments," *Journal of Econometrics* 166 (2011a), 138–156.
- "Statistical Decisions under Ambiguity," *Theory and Decision* 70 (2011b), 1–20.
- van der Vaart, A., *Asymptotic Statistics* (Cambridge: Cambridge University Press, 2000).