Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters*

Magne Mogstad[†] Andres Santos[‡] Alexander Torgovitsky[§]

May 25, 2018

Abstract

We propose a method for using instrumental variables (IV) to draw inference about causal effects for individuals other than those affected by the instrument at hand. Policy relevance and external validity turn on the ability to do this reliably. Our method exploits the insight that both the IV estimand and many treatment parameters can be expressed as weighted averages of the same underlying marginal treatment effects. Since the weights are identified, knowledge of the IV estimand generally places some restrictions on the unknown marginal treatment effects, and hence on the values of the treatment parameters of interest. We show how to extract information about the treatment parameter of interest from the IV estimand and, more generally, from a class of IV-like estimands that includes the two stage least squares and ordinary least squares estimands, among others. Our method has several applications. First, it can be used to construct nonparametric bounds on the average causal effect of a hypothetical policy change. Second, our method allows the researcher to flexibly incorporate shape restrictions and parametric assumptions, thereby enabling extrapolation of the average effects for compliers to the average effects for different or larger populations. Third, our method can be used to test model specification and hypotheses about behavior, such as no selection bias and/or no selection on gain.

[*] We thank the editor, four anonymous referees, Matt Masten, Derek Neal, Ed Vytacil, Chris Walters, and seminar participants at several universities and conferences for valuable feedback and suggestions. Bradley Setzler provided excellent research assistance. We are grateful to Pascaline Dupas for her help in accessing and understanding the data used in a previous draft of this paper.

[†] Department of Economics, University of Chicago; Statistics Norway; NBER.

[‡] Department of Economics, University of California at Los Angeles. Research supported in part by National Science Foundation grant SES-1426882.

[§] Department of Economics, University of Chicago. Research supported in part by National Science Foundation grant SES-1530538.

1 Introduction

In an influential paper, Imbens and Angrist (1994) provided conditions under which an instrumental variables (IV) estimand can be interpreted as the average causal effect for the subpopulation of compliers, i.e. for those whose treatment status would be affected by an exogenous manipulation of the instrument. In some cases, this local average treatment effect (LATE) is of intrinsic interest, for example if the instrument itself represents an intervention or policy change of interest. On the other hand, in many situations, the causal effect for individuals induced to treatment by the instrument at hand might not be representative of the causal effect for those who would be induced to treatment by a given policy change of interest to the researcher. In these cases, the LATE is not the relevant parameter for evaluating the policy change.

In this paper, we show how to use instrumental variables to draw inference about treatment parameters other than the LATE, thereby learning about causal effects for individuals other than those affected by the instrument at hand. Policy relevance and external validity turn on the ability to do this reliably. Our setting is the canonical program evaluation problem with a binary treatment and a scalar, real-valued outcome.[1] As in Imbens and Angrist (1994), we assume existence of an exogenous instrument that encourages all individuals to participate in treatment. This monotonicity condition is equivalent to a separable selection equation (Vytlacil, 2002), which gives rise to the concept of the marginal treatment effect (MTE) developed by Heckman and Vytlacil (1999, 2005). The MTE can be interpreted as the average effect of treatment for persons on a margin of indifference between participation in treatment and nonparticipation. An important feature of the model is that treatment effects are allowed to vary across individuals with the same observable characteristics in a way that depends on the unobservable determinants of treatment choice.

Our method builds on the observation that both the IV estimand and many treatment parameters can be expressed as weighted averages of the MTE function (Heckman and Vytlacil, 1999, 2001a,b,c, 2005, 2007a,b). Since the weights are identified, this observation implies that knowledge of the IV estimand places some restrictions on the unknown MTE function, and hence on the possible values of treatment parameters other than the LATE. We show how to extract information about these treatment

[1] For discussions of heterogeneous effects IV models with multiple discrete treatments, we refer to Angrist and Imbens (1995), Heckman, Urzua, and Vytlacil (2006), Heckman and Vytlacil (2007b), Heckman and Urzua (2010), Kirkeboen, Leuven, and Mogstad (2016), and Lee and Salanié (2016), among others. Heterogeneous effects IV models with continuous treatments have been considered by Angrist, Graddy, and Imbens (2000), Chesher (2003), Florens, Heckman, Meghir, and Vytlacil (2008), Imbens and Newey (2009), Torgovitsky (2015), Masten (2015), and Masten and Torgovitsky (2016), among others.

parameters from the IV estimand, and, more generally, from a class of "IV–like estimands."

The class of IV-like estimands consists of any cross moment between the outcome and an identified function of the treatment and the instrument. This class is general enough to contain the estimands corresponding to any weighted linear IV estimator, including the two stage least squares (TSLS), optimal generalized method of moments, and ordinary least squares (OLS) estimands. Each IV–like estimand provides a different weighted average of the MTE function, and therefore carries some distinct information about the possible values of the treatment parameter of interest. We show how these IV–like estimands can be chosen systematically so as to provide the tightest possible bounds on the treatment parameter of interest. We propose consistent estimators of these bounds that can be computed using linear programming.

Our method has at least three applications. First, it can be used to construct nonparametric bounds on the average causal effect of a hypothetical policy change. Second, our method enables extrapolation of the average effects for compliers to the average effects for different or larger populations. Third, our method can be used to perform tests of model specification and of individual behavior, such as testing the null hypotheses of no selection bias and/or no selection on gains. In all of these applications, our method provides a researcher the option to be fully nonparametric, or to impose shape and/or parametric restrictions, if desired.

Our paper contributes to several literatures. A large body of work is concerned with using instrumental variables to draw inference about treatment parameters other than the LATE. Nonparametric point identification of these parameters generally requires a continuous instrument with large support (Heckman and Vytlacil, 2005). In practice, however, instruments have limited support and are often discrete or even binary. For these situations, many common target parameters of interest, such as the average treatment effect, are not nonparametrically point identified. Analytic expressions for sharp bounds on the average treatment effect have been derived by Manski (1989, 1990, 1994, 1997, 2003), Balke and Pearl (1997), Heckman and Vytlacil (2001b) and Kitagawa (2009), among others.[2]

Analytic expressions for bounds are useful because they provide intuition on the source and strength of identification. However, it can be difficult to derive analytic bounds for more complicated parameters, such as the policy relevant treatment effects

[2] Note that Manski's analyses did not impose the monotonicity condition of Imbens and Angrist (1994); see Heckman and Vytlacil (2001b) and Kitagawa (2009) for further discussion. Also related is work by Shaikh and Vytlacil (2011), Bhattacharya, Shaikh, and Vytlacil (2012), and Mourifié (2015), who augment this monotonicity condition with a similar assumption for the potential outcomes.

(PRTEs) studied by Heckman and Vytlacil (2001a, 2005) and Carneiro, Heckman, and Vytlacil (2010, 2011). Our methodology is particularly useful in such settings. In addition, our method provides a unified framework for imposing shape restrictions such as monotonicity, concavity, monotone treatment selection (Manski, 1997; Manski and Pepper, 2000, 2009) and separability between observed and unobserved factors in the MTE function (Brinch, Mogstad, and Wiswall, 2017). It can be especially difficult to derive analytic bounds for treatment parameters that incorporate these types of assumptions in flexible combinations. In contrast, our general computational approach allows one to flexibly adjust the parameter of interest, as well as the maintained assumptions, without requiring additional identification analysis.

In addition, our paper is related to recent work that considers extrapolation in instrumental variables models under additional assumptions. While our method delivers bound on the target parameter in general, these bounds nest important point identification results as special cases. For example, our method nests existing approaches that extrapolate by assuming no unobserved heterogeneity in the treatment effect (Heckman and Robb, 1985; Angrist and Fernandez-Val, 2013), and those that parameterize this unobserved heterogeneity (Heckman, Tobias, and Vytlacil, 2003; Brinch et al., 2017). One attractive feature of our method is that the constraints in the linear programs being solved require an MTE function to also yield the usual, nonparametrically point identified LATE. Hence, our method allows for extrapolation to other parameters of interest without sacrificing the internal validity of the LATE.

Our paper also relates to a literature on specification tests in settings with instrumental variables. This is a consequence of our method, since if there are no MTE functions that are consistent with a set of IV–like estimands, then the model is misspecified. Balke and Pearl (1997) and Imbens and Rubin (1997) noted that the assumptions used to point identify the LATE have testable implications, while Machado, Shaikh, and Vytlacil (2013), Huber and Mellace (2014), Kitagawa (2015), and Mourifié and Wan (2016) have developed this observation into formal statistical tests. Our method builds on the work of these authors by allowing the researcher to maintain additional assumptions, such as parametric and/or shape restrictions. In addition to testing whether the model is misspecified, our method can also be used to test null hypotheses such as no selection bias and/or no selection on gains.

The remainder of the paper is organized as follows. In Section 2, we present the model and develop our method for bounding a target parameter of interest while potentially maintaining additional shape constraints. We discuss implementation of our approach in Section 3. In Section 4, we discuss key applications of our method, which we illustrate in Section 5 with a numerical example. In Section 6, we provide a concluding discussion on common empirical settings in which our approach is particularly useful. Our working paper (Mogstad, Santos, and Torgovitsky, 2017) contains a method for hypothesis testing and constructing confidence intervals, as well as an application to Dupas's (2014) study of the effects of price subsidies on the adoption and usage of an antimalarial bed net in Kenya.

2 A General Framework for Identification in the IV Model

2.1 Model

Our model is the canonical program evaluation problem with a binary treatment $D \in \{0,1\}$ and a scalar, real-valued outcome, Y. Corresponding to the two treatment arms are unobservable potential outcomes, Y_0 and Y_1. These represent the realization of Y that would have been experienced by an individual had their treatment status been exogenously set to 0 or 1. The relationship between observed and potential outcomes is given by

$$Y = DY_1 + (1 - D)Y_0. \tag{1}$$

Following Heckman and Vytlacil (1999, 2005), we assume that treatment is determined by the weakly separable selection or choice equation

$$D = \mathbb{1}[\nu(Z) - U \geq 0], \tag{2}$$

where ν is an unknown function, U is a continuously distributed random variable, and Z is a vector of observable covariates.

The observable variables in the model are the outcome $Y \in \mathbf{R}$, the binary treatment $D \in \{0,1\}$, and the covariates $Z \in \mathbf{R}^{d_z}$. We decompose Z into $Z = (X, Z_0)$, where $X \in \mathbf{R}^{d_x}$ are used as control variables and $Z_0 \in \mathbf{R}^{d_{z_0}}$ will be assumed to be exogenous instruments. Our focus in this section is on identification, so we assume throughout that the researcher knows the joint distribution of (Y, D, Z). The unobservables are the potential outcomes (Y_0, Y_1), and the variable U in the selection equation, which represents unobservable factors that affect treatment choice. We maintain the following standard assumptions throughout the paper.

Assumptions I

I.1 *$U \perp\!\!\!\perp Z_0 | X$, where $\perp\!\!\!\perp$ denotes (conditional) statistical independence.*

I.2 *$E[Y_d|Z,U] = E[Y_d|X,U]$ and $E[Y_d^2] < \infty$ for $d \in \{0,1\}$.*

I.3 *U is continuously distributed, conditional on X.*

Assumptions I.1 and I.2 require Z_0 to be exogenous with respect to both the selection and outcome processes. Vytlacil (2002) showed that, given I.1, the assumption that the index of the selection equation is additively separable as in (2) is equivalent to the assumption that Z_0 affects D monotonically in the sense introduced by Imbens and Angrist (1994). Hence, I.1 combined with (2) imposes substantive restrictions on choice behavior: Conditional on X, changes in Z_0 either weakly encourage or weakly discourage all agents to choose $D = 1$. Assumption I.2 imposes an exclusion restriction that the conditional means of Y_0 and Y_1 depend on $Z = (Z_0, X)$ only through the covariates X.

Assumption I.3 is a weak regularity condition that enables us to impose a standard normalization. As is well known, equation (2) may be rewritten as

$$D = \mathbb{1}\left[F_{U|X}(U|X) \leq F_{U|X}(\nu(Z)|X)\right] \equiv \mathbb{1}[\widetilde{U} \leq \widetilde{\nu}(Z)], \tag{3}$$

where we are using the notation $F_{U|X}(u|x) \equiv P(U \leq u|X = x)$ and we have defined $\widetilde{U} \equiv F_{U|X}(U|X)$ and $\widetilde{\nu}(Z) \equiv F_{U|X}(\nu(Z)|X)$. Under Assumptions I.1 and I.3, $\widetilde{U}$ is uniformly distributed on $[0, 1]$, conditional on $Z = (Z_0, X)$. Working with this normalized model simplifies the analysis and does not affect its empirical content. Hence, we drop the tilde and maintain throughout the paper the normalization that U itself is distributed uniformly over $[0, 1]$ conditional on Z. A consequence of this normalization is that

$$p(z) \equiv P(D = 1|Z = z) = F_{U|Z}(\nu(z)|z) = \nu(z), \tag{4}$$

where $p(z)$ is the propensity score.

It is important to observe what is *not* being assumed under Assumptions I. First, we do not impose any conditions on the support of Z: Both the control (X) and exogenous (Z_0) components of Z may be either continuous, discrete and ordered, categorical, or binary. Second, the IV model as specified here allows for rich forms of observed and unobserved heterogeneity. In particular, it allows $Y_1 - Y_0$ to vary not only across individuals with different values of X, but also among individuals with the same X. The treatment D may be statistically dependent with Y_0 (indicating selection bias), or $Y_1 - Y_0$ (indicating selection on the gain), or both, even conditional on X. Third, the model does not specify why individuals make the treatment choice that they do, in contrast to a stylized Roy model in which $D = \mathbb{1}[Y_1 > Y_0]$. However, the model also does not preclude the possibility that individuals choose treatment with full or partial knowledge of the potential outcomes (Y_0, Y_1). Any such knowledge will be reflected through dependence between the potential outcomes, (Y_0, Y_1), and the unobserved component treatment choice, U. Assumption I does not restrict this dependence.

As observed by Heckman and Vytlacil (1999, 2005), a wide range of treatment parameters can be written as weighted averages of the underlying MTE function, which is defined as

$$\text{MTE}(u, x) \equiv E\left[Y_1 - Y_0 \mid U = u, X = x\right].$$

We use a slight generalization of their observation. Instead of working with the MTE function directly, we consider treatment parameters that can be expressed as functions of the two marginal treatment response (MTR) functions, defined as

$$m_0(u, x) \equiv E\left[Y_0 \mid U = u, X = x\right] \quad \text{and} \quad m_1(u, x) \equiv E\left[Y_1 \mid U = u, X = x\right]. \tag{5}$$

Of course, each pair $m \equiv (m_0, m_1)$ of MTR functions generates an associated MTE function $m_1 - m_0$. One benefit of working with MTR functions instead of MTE functions is that it allows us to consider parameters that weight m_0 and m_1 asymmetrically. Another benefit is that it allows the researcher to impose assumptions on m_0 and m_1 separately.

Our method allows the specification of MTR functions to be completely unrestricted, but it also allows the researcher the ability to impose restrictions. For example, it is possible to maintain nonparametric shape restrictions, as discussed in Section 2.6. It is also possible to parameterize the MTR functions. Polynomials are widely used as a parameterization in empirical work using MTEs (Moffitt, 2008; French and Song, 2014; Brinch et al., 2017), but our method allows for other parameterizations as well. The uniform normalization is convenient for interpreting these parameterizations, since it allows $U = u$ to be interpreted as a quantile of the distribution of latent willingness to pay for treatment. An MTR specification that is, for example, linear in u, is interpretable as imposing a linear relationship between average outcomes and the quantile of unobserved willingness to pay. A higher order polynomial allows for a flexible but smooth relationship between average outcomes and these quantiles.

2.2 What We Want to Know: The Target Parameter

We assume that the researcher is interested in a target parameter $\beta^\star$ that can be written for any candidate pair of MTR functions $m \equiv (m_0, m_1)$ as

$$\beta^\star \equiv E\left[\int_0^1 m_0(u, X)\omega_0^\star(u, Z)\, d\mu^\star(u)\right] + E\left[\int_0^1 m_1(u, X)\omega_1^\star(u, Z)\, d\mu^\star(u)\right], \tag{6}$$

where $\omega_0^\star$ and $\omega_1^\star$ are identified weighting functions, and $\mu^\star$ is an integrating measure that is chosen by the researcher and usually taken to be the Lebesgue measure on $[0,1]$. For example, to set $\beta^\star$ to be the average treatment effect (ATE), observe that

$$E[Y_1 - Y_0] = E[m_1(U,X) - m_0(U,X)] = E\left[\int_0^1 m_1(u,X)du\right] - E\left[\int_0^1 m_0(u,X)du\right],$$

take $\omega_1^\star(u,z) = 1$, $\omega_0^\star(u,z) = -1$, and let $\mu^\star$ be the Lebesgue measure on $[0,1]$. Similarly, to set $\beta^\star$ to be the ATE conditional on X lying in some known set $\mathcal{X}^\star$, take

$$\omega_1^\star(u,z) \equiv \omega_1^\star(u,x,z_0) = \frac{\mathbb{1}[x \in \mathcal{X}^\star]}{P(X \in \mathcal{X}^\star)},$$

$\omega_0^\star(u,z) = -\omega_1^\star(u,z)$, and let $\mu^\star$ be as before. The resulting target parameter is then the population average treatment effect for individuals with covariates $x \in \mathcal{X}^\star$.

In Table 1, we provide formulas for the weights $\omega_0^\star$ and $\omega_1^\star$ that correspond to a variety of different treatment parameters. Any of these can be taken to be the target parameter $\beta^\star$. Examples include (i) the average treatment effect for the treated (ATT), i.e. the average impact of treatment for individuals who actually take the treatment; (ii) the average treatment effect for the untreated (ATU), i.e. the average impact of treatment for individuals who do not take treatment; (iii) LATE$[\underline{u},\overline{u}]$, i.e. the average treatment effect for individuals who would take the treatment if their realization of the instrument yielded $p(z) = \overline{u}$, but not if $p(z) = \underline{u}$; and (iv) the policy relevant treatment effect (PRTE), i.e. the average impact on Y (either gross or per net individual affected) due to a change from the baseline policy to some alternative policy.

For most of the parameters in Table 1, the integrating measure $\mu^\star$ is Lebesgue measure on $[0,1]$. However, researchers are sometimes interested in the MTE function itself, e.g. Carneiro et al. (2011) and Maestas, Mullen, and Strand (2013). Our specification of $\beta^\star$ accommodates this by replacing $\mu^\star$ with the Dirac measure (i.e., a point mass) at some specified point $\overline{u}$ and taking $\omega_0^\star(u,z) = -\omega_1^\star(u,z) = -1$. The resulting target parameter is the MTE function averaged over X, i.e. $E[m_1(\overline{u},X) - m_0(\overline{u},X)]$.

2.3 What We Know: IV–Like Estimands

A key point for our method is that a set of identified quantities can also be written in a form similar to (6). Consider, for example, the IV estimand that results from using Z as an instrument for D in a linear instrumental variables regression that includes a constant term, but which does not include any other covariates X. Assuming

$\text{Cov}(D, Z) \neq 0$, this estimand is given by

$$\beta_{\text{IV}} \equiv \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)}. \tag{7}$$

For example, if $Z \in \{0, 1\}$ is binary, then β_{IV} reduces to the standard Wald estimand

$$\beta_{\text{IV}} = \frac{E\left[Y \mid Z = 1\right] - E\left[Y \mid Z = 0\right]}{E\left[D \mid Z = 1\right] - E\left[D \mid Z = 0\right]}. \tag{8}$$

Heckman and Vytlacil (2005) show that β_{IV} can also be written in the form (6) as a weighted average of the MTE function. This observation suggests that useful information about $\beta^\star$ can be extracted from knowledge of β_{IV}. The next proposition shows that, more generally, any cross moment of Y with an identified (or known) function of $(D, Z) \equiv (D, X, Z_0)$ can also be expressed as the weighted sum of the two MTR functions, m_0 and m_1. We refer to such cross moments as IV–like estimands.

Proposition 1. *Suppose that $s : \{0, 1\} \times \mathbf{R}^{d_z} \to \mathbf{R}$ is an identified (or known) function that is measurable and has a finite second moment. We refer to such a function s as an* IV–like specification *and to $\beta_s \equiv E[s(D, Z)Y]$ as an* IV–like estimand. *If (Y, D) are generated according to* (1) *and* (2) *under Assumptions I, then*

$$\beta_s = E\left[\int_0^1 m_0(u, X)\omega_{0s}(u, Z)\, du\right] + E\left[\int_0^1 m_1(u, X)\omega_{1s}(u, Z)\, du\right], \tag{9}$$

$$\begin{aligned} \textit{where} \quad & \omega_{0s}(u, z) \equiv s(0, z)\mathbb{1}[u > p(z)] \\ \textit{and} \quad & \omega_{1s}(u, z) \equiv s(1, z)\mathbb{1}[u \leq p(z)]. \end{aligned}$$

The weights in Proposition 1 can be shown to reduce to the weights for β_{IV} derived by Heckman and Vytlacil (2005) by taking

$$s(d, z) \equiv s(d, x, z_0) = \frac{z_0 - E[Z_0]}{\text{Cov}(D, Z_0)}, \tag{10}$$

which is an identified function of d (trivially) and z.[3] As we elaborate further in Supplemental Appendix S1, Proposition 1 applies more broadly to include any well-defined weighted linear IV estimand that uses some function of D and Z as included and excluded instruments for a set of endogenous variables also constructed from D

[3] An implication of (9) is that the IV-like specifications $s(d, z) = (z_0 - E[Z_0])/\operatorname{Cov}(D, Z_0)$ and $\tilde{s}(d, z) = (z_0 - E[Z_0])$ differ by only a multiplicative constant, e.g. $\text{Cov}(D, Z_0)$, and so yield the same restrictions on the MTR functions. As a result, if β_{IV} is weakly identified, one might prefer to use the IV-like specification $\tilde{s}(d, z) = (z_0 - E[Z_0])$ instead of (10).

and Z.[4] For example, the OLS estimand corresponds to taking s to be

$$s(d,z) = \frac{d - E[D]}{\mathrm{Var}(D)}.$$

More generally, any subvector of the TSLS estimand can also be written as an IV–like estimand. Table 2 contains expressions for some notable IV–like estimands.

2.4 From What We Know to What We Want to Know

We now show how to extract information about the target parameter $\beta^\star$ from the general class of IV-like estimands introduced in Section 2.3. Let $\mathcal{S}$ denote some collection of IV–like specifications (i.e. functions $s : \{0,1\} \times \mathbf{R}^{d_z} \to \mathbf{R}$) chosen by the researcher, that each satisfy the conditions set out in Proposition 1. Corresponding to each $s \in \mathcal{S}$ is an IV–like estimand $\beta_s \equiv E[s(D,Z)Y]$. We assume that the researcher has restricted the pair of MTR functions $m \equiv (m_0, m_1)$ to lie in some *parameter space* $\mathcal{M}$ contained in a vector space. The parameter space incorporates assumptions that the researcher wishes to maintain about m, such as parametric or shape restrictions. Our goal is to characterize bounds on the values of the target parameter $\beta^\star$ that could have been generated by MTR functions $m \in \mathcal{M}$ that also could have delivered the collection of identified IV–estimands through (9).

To this end, we denote the weighting expression in Proposition 1 as a linear map $\Gamma_s : \mathcal{M} \to \mathbf{R}$, defined for any IV–like specification $s \in \mathcal{S}$ as

$$\Gamma_s(m) \equiv E\left[\int_0^1 m_0(u,X)\omega_{0s}(u,Z)\,du\right] + E\left[\int_0^1 m_1(u,X)\omega_{1s}(u,Z)\,du\right], \tag{11}$$

where we recall that $\omega_{0s}(u,z) \equiv s(0,z)\mathbb{1}[u > p(z)]$ and $\omega_{1s}(u,z) \equiv s(1,z)\mathbb{1}[u \leq p(z)]$. By Proposition 1, if (Y,D) are generated according to (1) and (2) under Assumptions I, then the MTR functions $m \equiv (m_0, m_1)$ must satisfy $\Gamma_s(m) = \beta_s$ for every IV–like specification $s \in \mathcal{S}$. As a result, m must lie in the set

$$\mathcal{M}_{\mathcal{S}} \equiv \{m \in \mathcal{M} : \Gamma_s(m) = \beta_s \text{ for all } s \in \mathcal{S}\}.$$

The target parameter, $\beta^\star$, can also be expressed as an identified linear map of the MTR functions. From (6), we define this map as $\Gamma^\star : \mathcal{M} \to \mathbf{R}$, with

$$\Gamma^\star(m) \equiv E\left[\int_0^1 m_0(u,X)\omega_0^\star(u,Z)d\mu^\star(u)\right] + E\left[\int_0^1 m_1(u,X)\omega_1^\star(u,Z)d\mu^\star(u)\right]. \tag{12}$$

[4] The phrases "included" and "excluded" instrument are meant in the sense typically introduced in textbook treatments of the linear IV model without heterogeneity.

It follows that if (Y, D) is generated according to (1) and (2) under Assumptions I, then the target parameter must belong to the identified set

$$\mathcal{B}^\star_\mathcal{S} \equiv \{b \in \mathbf{R} : b = \Gamma^\star(m) \text{ for some } m \in \mathcal{M}_\mathcal{S}\}.$$

Intuitively, $\mathcal{B}^\star_\mathcal{S}$ is the set of values for the target parameter that could have been generated by MTR functions that satisfy the researcher's assumptions and are also consistent with the IV–like estimands. The next result shows that if $\mathcal{M}$ is convex, then $\mathcal{B}^\star_\mathcal{S}$ is an interval that can be characterized by solving two convex optimization problems.

Proposition 2. *Suppose that $\mathcal{M}$ is convex. Then either $\mathcal{M}_\mathcal{S}$ is empty, and hence $\mathcal{B}^\star_\mathcal{S}$ is empty, or else the closure of $\mathcal{B}^\star_\mathcal{S}$ (in $\mathbf{R}$) is equal to the interval $[\underline{\beta}^\star, \overline{\beta}^\star]$, where*

$$\underline{\beta}^\star \equiv \inf_{m \in \mathcal{M}_\mathcal{S}} \Gamma^\star(m) \qquad \textit{and} \qquad \overline{\beta}^\star \equiv \sup_{m \in \mathcal{M}_\mathcal{S}} \Gamma^\star(m). \tag{13}$$

2.5 Information and Point Identification

The set $\mathcal{M}_\mathcal{S}$ consists of all MTR functions in $\mathcal{M}$ that are consistent with the IV–like estimands chosen by the researcher. However, $\mathcal{M}_\mathcal{S}$ may not necessarily exhaust all of the information available in the data. In particular, $\mathcal{M}_\mathcal{S}$ could contain MTR functions that would be ruled out if $\mathcal{S}$ were expanded to include additional IV–like specifications. If this is the case, then incorporating these further specifications can shrink $\mathcal{B}^\star_\mathcal{S}$.

We examine this issue by considering the conditional means of Y that would be generated through (1) and (2) under Assumptions I by a given MTR pair $m = (m_0, m_1)$. Whenever $0 < p(Z) < 1$, these conditional means can be written as

$$E[Y|D = 0, Z] = E[Y_0|U > p(Z), Z] = \frac{1}{(1 - p(Z))} \int_{p(Z)}^1 m_0(u, X)\, du, \tag{14}$$

$$\text{and} \qquad E[Y|D = 1, Z] = E[Y_1|U \leq p(Z), Z] = \frac{1}{p(Z)} \int_0^{p(Z)} m_1(u, X)\, du. \tag{15}$$

MTR pairs that (almost surely) satisfy (14)–(15) are compatible with the observed conditional means of Y. Our next result shows that any such MTR pair will be in $\mathcal{M}_\mathcal{S}$ for any choice of $\mathcal{S}$. Moreover, we show that if $\mathcal{S}$ is chosen correctly, then $\mathcal{M}_\mathcal{S}$ will contain *only* MTR pairs that are compatible with the observed conditional means of Y.

Proposition 3. *Suppose that every $m \in \mathcal{M}$ satisfies $E[\int_0^1 m_d(u, X)^2 du] < \infty$ for*

$d \in \{0, 1\}$. *If $\mathcal{S}$ contains functions that satisfy the conditions of Proposition 1, then*

$$\{m \in \mathcal{M} : m \text{ satisfies (14) and (15) almost surely}\} \subseteq \mathcal{M}_{\mathcal{S}}. \tag{16}$$

Moreover, suppose that $\mathcal{S} \equiv \{s(d, z) = \mathbb{1}[d = d']f(z)$ for $(d', f) \in \{0, 1\} \times \mathcal{F}\}$, where $\mathcal{F}$ is a collection of functions. If the linear span of $\mathcal{F}$ is norm dense in $\mathbf{L}^2(Z) \equiv \{f : \mathbf{R}^{d_z} \to \mathbf{R}$ s.t. $E[f(Z)^2] < \infty\}$, then

$$\{m \in \mathcal{M} : m \text{ satisfies (14) and (15) almost surely}\} = \mathcal{M}_{\mathcal{S}}. \tag{17}$$

Proposition 3 shows that if $\mathcal{S}$ is a sufficiently rich class of functions, then $\mathcal{M}_{\mathcal{S}}$ exhausts the available data in the sense of being the smallest subset of $\mathcal{M}$ that is compatible with the observed conditional means of Y. It follows that $\mathcal{B}^{\star}_{\mathcal{S}}$ is also the smallest set of values for the target parameter that are consistent with both the conditional means of Y and the assumptions of the model. For example, if $D \in \{0, 1\}$ and $Z \in \{0, 1\}$, then (17) holds if we take $\mathcal{F} = \{\mathbb{1}[z = 0], \mathbb{1}[z = 1]\}$, so that

$$\mathcal{S} = \{\mathbb{1}[d = 0, z = 0],\ \mathbb{1}[d = 0, z = 1],\ \mathbb{1}[d = 1, z = 0],\ \mathbb{1}[d = 1, z = 1]\}.$$

The information contained in the corresponding IV–like estimands is the same as that contained in the coefficients of a saturated regression of Y on D and Z. More generally, if Z is continuous, then (17) can be satisfied by taking $\mathcal{F}$ to be certain parametric families of functions of Z. For example, if $Z \in \mathbf{R}^{d_z}$, then one such family is the set of half spaces, $\mathcal{F} = \{\mathbb{1}[z \leq z'] : z' \in \mathbf{R}^{d_z}\}$. Other examples can be found in e.g. Bierens (1990) and Stinchcombe and White (1998).

While we view partial identification as the standard case, we emphasize that our analysis does not preclude point identification. Letting $|\mathcal{S}|$ denote the cardinality of $\mathcal{S}$, notice that the restrictions

$$\Gamma_s(m) = \beta_s \text{ for all } s \in \mathcal{S} \tag{18}$$

constitute a linear system of $|\mathcal{S}|$ equations in terms of m. As such, there are three possibilities for the cardinality of the solution set: Either there is a unique solution, there are an infinite number of solutions, or there is no solution. These cases correspond respectively to point identification of m, partial identification of m, and model misspecification. We discuss the latter case in Section 4.3. In the case that m is point identified, note that then any target parameter $\beta^{\star}$ is also point identified.

These observations about point identification are implicit in the work of Brinch et al.

(2017). Those authors show that if $\mathcal{M}$ is restricted to be a set of polynomials, then point identification of the MTR functions can be established by considering regressions of Y on $p(Z)$ and D. In terms of (18), this just means that $\mathcal{S}$ creates at least as many non-redundant equations as the dimension of m. Their results allow for Z to be discrete, but require the specification of $\mathcal{M}$ to be no richer than the support of $p(Z)$. For example, if Z is binary, their results require $\mathcal{M}$ to only contain MTR pairs that are linear in u.[5] In contrast, our results allow the researcher to specify $\mathcal{M}$ independently of the data.

In some situations, point identification of the target parameter, $\beta^\star$, can also be established when $|\mathcal{S}|$ is infinite and $\mathcal{M}$ is infinite dimensional. Indeed, relationships similar to (14) and (15) have been used previously to establish point identification of the MTE function. For example, Heckman and Vytlacil (1999, 2001c, 2005) and Carneiro et al. (2010, 2011) show that if Z_0 is continuously distributed, then the MTE is point identified over the support of the propensity score. As a consequence, any target parameter $\beta^\star$ whose weights have support contained within the interior of the support of the propensity score will also be point identified. Proposition 3 implies that the same is true in our framework if $\mathcal{S}$ is chosen to be a sufficiently rich collection of functions.

2.6 Shape Restrictions

Our method allows researchers to easily impose shape restrictions either on the MTR functions $m = (m_0, m_1)$ or directly on the MTE function $m_1 - m_0$. For example, to impose the monotone treatment response assumption considered by Manski (1997), i.e. that $Y_1 \geq Y_0$ with probability 1, the set $\mathcal{M}$ should be specified to only contain MTR pairs for which $m_1 - m_0$ is non-negative. Similarly, one could assume that $m_1(\cdot, x) - m_0(\cdot, x)$ is weakly decreasing for every x. This restriction would reflect the assumption that those more likely to select into treatment (those with small realizations of U) are also more likely to have larger gains from treatment, which is like the monotone treatment selection assumption of Manski and Pepper (2000). Maintaining combinations of assumptions simultaneously (e.g. both monotone treatment response and monotone treatment selection) is simply a matter of imposing both restrictions on $\mathcal{M}$ at the same time.

Another type of nonparametric shape restriction that can be used to tighten the bounds is separability between the observed (X) and unobserved (U) components.

[5] Recently, Kowalski (2016) has applied the linear case, which was studied in depth by Brinch et al. (2012, 2017), to analyze the Oregon Health Insurance Experiment.

Although restrictive, separability of this sort is standard (often implicit) in applied work using instrumental variables.[6] In our framework, separability between X and U can be imposed by restricting $\mathcal{M}$ to only contain MTR pairs (m_0, m_1) that can be decomposed as

$$m_d(u, x) = m_d^U(u) + m_d^X(x) \quad \text{for } d = 0, 1, \tag{19}$$

where m_d^U and m_d^X are functions that can themselves satisfy some shape restrictions. This type of separability implies that the slopes of the MTR functions with respect to u do not vary with x. Alternatively, if complete separability is viewed as too strong of a restriction, it is straightforward to interact u with only certain components of x. Specifications like (19) can also be used to mitigate the curse of dimensionality, for example by specifying $m_d^X(x)$ to be a linear function of x.

3 Implementation

In this section, we show how to implement Proposition 2. First, we discuss a tractable computational approach for solving the optimization problems defined in (13). Then, in Section 3.2, we propose a consistent estimator of the identified set for the target parameter, i.e. $\mathcal{B}^{\star}_{\text{id}} \equiv [\underline{\beta}^{\star}, \overline{\beta}^{\star}]$.

3.1 Computing the Bounds

Our approach for implementing Proposition 2 is to replace the possibly infinite dimensional parameter space of functions $\mathcal{M}$ by a finite dimensional subset $\mathcal{M}_{\text{fd}} \subseteq \mathcal{M}$.[7] The upper bound for the target parameter with this finite dimensional subset is given by

$$\overline{\beta}^{\star}_{\text{fd}} \equiv \sup_{m \in \mathcal{M}_{\text{fd}}} \Gamma^{\star}(m) \quad \text{s.t.} \quad \Gamma_s(m) = \beta_s \text{ for all } s \in \mathcal{S}, \tag{20}$$

while $\underline{\beta}^{\star}_{\text{fd}}$ is defined as the analogous infimum. Suppose that we specify $\mathcal{M}_{\text{fd}}$ as the finite linear basis

$$\mathcal{M}_{\text{fd}} \equiv \left\{ (m_0, m_1) \in \mathcal{M} : m_d(u, x) = \sum_{k=1}^{K_d} \theta_{dk} b_{dk}(u, x) \text{ for some } \{\theta_{dk}\}_{k=1}^{K_d}, d = 0, 1 \right\}, \tag{21}$$

[6] In some settings, additive separability can be motivated by assumptions about primitives such as technology or preferences. See Brinch et al. (2017) for examples.

[7] This is as in sieve estimation methods (Chen, 2007), but our focus here is on computation rather than regularization.

where $\{b_{dk}\}_{k=1}^{K_d}$ are known basis functions and $\theta \equiv (\theta_0', \theta_1')'$ parameterizes functions in $\mathcal{M}_{\text{fd}}$ with $\theta_d \equiv (\theta_{d1}, \ldots, \theta_{dK_d})'$. The parameter space $\mathcal{M}$ generates a parameter space

$$\Theta \equiv \left\{ (\theta_0, \theta_1) \in \mathbf{R}^{K_0} \times \mathbf{R}^{K_1} : \left(\sum_{k=1}^{K_0} \theta_{0k} b_{0k}, \sum_{k=1}^{K_1} \theta_{k1} b_{1k} \right) \in \mathcal{M} \right\},$$

for the finite dimensional parameter θ. Using the linearity of the mappings $\Gamma^\star$ and Γ_s defined in (11) and (12), we write (20) as

$$\begin{aligned} \overline{\beta}^\star_{\text{fd}} \equiv \sup_{(\theta_0,\theta_1)\in\Theta} & \sum_{k=1}^{K_0} \theta_{0k} \Gamma_0^\star(b_{0k}) + \sum_{k=1}^{K_1} \theta_{1k} \Gamma_1^\star(b_{1k}) \\ \text{s.t.} & \sum_{k=1}^{K_0} \theta_{0k} \Gamma_{0s}(b_{0k}) + \sum_{k=1}^{K_1} \theta_{1k} \Gamma_{1s}(b_{1k}) = \beta_s \text{ for all } s \in \mathcal{S}, \end{aligned} \tag{22}$$

where we have decomposed $\Gamma^\star(m) \equiv \Gamma_0^\star(m_0) + \Gamma_1^\star(m_1)$ with

$$\Gamma_d^\star(m_d) \equiv E\left[\int_0^1 m_d(u, X) \omega_d^\star(u, Z) d\mu^\star(u) \right] \quad \text{for } d = 0, 1, \tag{23}$$

and similarly for $\Gamma_s(m) = \Gamma_{0s}(m_0) + \Gamma_{1s}(m_1)$ with

$$\Gamma_{ds}(m_d) \equiv E\left[\int_0^1 m_d(u, X) \omega_{ds}(u, Z) d\mu^\star(u) \right] \quad \text{for } d = 0, 1. \tag{24}$$

If Θ is a polyhedral set, and $\mathcal{S}$ is a finite set, then (22) is a finite-dimensional linear program. Linear programs are used routinely in empirical work involving quantile regressions, e.g. Abadie, Angrist, and Imbens (2002), in part because they can be solved quickly and reliably. Whether a given shape restriction on $\mathcal{M}$ translates into Θ being polyhedral depends on the basis functions. In Supplemental Appendix S2, we discuss the Bernstein polynomial basis, which is particularly attractive in this regard. We also observe there that for certain choices of bases, including the Bernstein polynomial basis, one can compute the integrals in the definitions of $\Gamma_d^\star(b_{dk})$ and $\Gamma_{ds}(b_{dk})$ analytically. For other cases, these one-dimensional integrals can be computed numerically.

In some situations, $\mathcal{M}$ can be replaced by a finite dimensional set $\mathcal{M}_{\text{fd}}$ without affecting the bounds on the target parameter, i.e. while ensuring $\overline{\beta}^\star_{\text{fd}} = \overline{\beta}^\star$. This can be interpreted as an *exact* computational approach for determining nonparametric bounds on the target parameter. Suppose that Z has discrete support and that the weight functions $(u, z) \mapsto \omega_d^\star(u, z)$ for the target parameter are piecewise constant in u. Then define a partition $\{\mathcal{U}_j\}_{j=1}^J$ of $[0, 1]$ such that $\omega_d^\star(u, z)$ and $\mathbb{1}[u \leq p(z)]$ are constant (as functions of u) on each $\mathcal{U}_j$.[8] Let $\{x_1, \dots x_L\}$ denote the support of X and then use

$$b_{jl}(u, x) \equiv \mathbb{1}[u \in \mathcal{U}_j, x = x_l] \text{ for } 1 \leq j \leq J \text{ and } 1 \leq l \leq L \tag{25}$$

as the basis functions employed in the construction of $\mathcal{M}_{\text{fd}}$ in (21), with $K_d = JL$.

A linear combination of the basis functions (25) forms a constant spline as a function of u over $[0, 1]$ for each x. The constant spline that provides the best mean squared error approximation to a given function $m_d(u, x)$ can be shown to be

$$\Pi m_d(u, x) \equiv \sum_{j=1}^{J} \sum_{l=1}^{L} E[m_d(U, x_l) | U \in \mathcal{U}_j, X = x_l] b_{jl}(u, x). \tag{26}$$

This corresponds to taking $\theta_{d(j,l)} = E[m_d(U, x_l) | U \in \mathcal{U}_j, X = x_l]$ for an element of (21), with the slight abuse of notation that $k = (j, l)$. The next proposition uses (26) to show that constant splines can reproduce the nonparametric bounds on the target parameter.

Proposition 4. *Suppose that Z has discrete support and that $\omega^\star_d(u, z)$ are piecewise constant in u. Let $\{\mathcal{U}_j\}_{j=1}^{J}$ be a partition of* $[0, 1]$ *such that $\omega^\star_d(u, z)$ and $\mathbb{1}[u \leq p(z)]$ are constant on $u \in \mathcal{U}_j$ for any z. Suppose that $\mathcal{M}_{fd}$ is constructed using basis functions* (25) *and that* $(\Pi m_0, \Pi m_1) \in \mathcal{M}$ *for every* $(m_0, m_1) \in \mathcal{M}$. *Then $\overline{\beta}^\star_{fd} = \overline{\beta}^\star$ and $\underline{\beta}^\star_{fd} = \underline{\beta}^\star$.*

Proposition 4 shows that one can solve the infinite dimensional problems defining $\underline{\beta}^\star$ and $\overline{\beta}^\star$ *exactly* by solving (20) with $\mathcal{M}_{\text{fd}}$ consisting of constant splines. Besides requiring Z to have discrete support, the result also assumes $(\Pi m_0, \Pi m_1) \in \mathcal{M}$ for every $m \in \mathcal{M}$.[9] The second condition requires that the constant spline generated from a MTR pair $m \in \mathcal{M}$ through (26) still belong to the parameter space $\mathcal{M}$. For certain shape restrictions, such as boundedness or monotonicity, (m_0, m_1) satisfying the restriction implies the projection $(\Pi m_0, \Pi m_1)$ will satisfy them as well. As a result, in such applications, the assumption $(\Pi m_0, \Pi m_1) \in \mathcal{M}$ is satisfied whenever $\mathcal{M}$ includes piecewise constant functions. We observe that the latter condition can be relaxed to

[8] For example, if $u \mapsto \omega^\star_d(u, z)$ is left-continuous, then we may set $\mathcal{U}_1 \equiv [u_0, u_1]$ and $\mathcal{U}_j \equiv (u_{j-1}, u_j]$ for $2 \leq j \leq J$, where $\{u_j\}_{j=0}^{J}$ are the ordered unique elements of the union of $\{0, 1\}$, $\operatorname{supp} p(Z)$, and the discontinuity points of $\{\omega^\star_d(\cdot, z) : d \in \{0, 1\}, z \in \operatorname{supp} Z\}$. Moreover, if $\mu^\star$ is absolutely continuous with respect to Lebesgue measure, then it suffices that $\omega^\star_d(U, z)$ and $\mathbb{1}[U \leq p(z)]$ be almost surely constant on $\mathcal{U}_j$.

[9] The assumption that $\omega^\star_d(u, z)$ is piecewise constant in u for fixed z is not restrictive for common target parameters; see Table 1. The requirement that Z is discretely distributed can be relaxed, since when Z_0 has a continuous component, the MTR functions are point identified over the interior of the support of the propensity score. (This can be seen by slightly generalizing the local instrument variable argument of Heckman and Vytlacil (1999) to (14) and (15).) One can then apply Proposition 4 to the portion of the MTR functions that are not point identified by the continuous variation.

instead assuming that $\mathcal{M}$ includes functions that can approximate $(\Pi m_0, \Pi m_1)$ suitably well.[10]

3.2 Estimation

So far, we have assumed that the researcher knows the population joint distribution of the observed data (Y, D, Z). In practice, researchers typically model the observed data as a finite sample drawn from the population. Features of the distribution of observables must then be viewed as estimators subject to statistical error. In this section, we propose a modification of our procedure that allows for such error. In Supplemental Appendix S3, we show that our proposed set estimator is consistent for the population identified set under weak conditions. In our working paper (Mogstad et al., 2017), we develop a procedure that can be used for hypothesis testing and for constructing confidence intervals.

Suppose that $\hat{\Gamma}^\star$, $\hat{\Gamma}_s$ and $\hat{\beta}_s$ are consistent estimators of $\Gamma^\star, \Gamma_s$ and β_s, respectively, and that $\mathcal{S}$ has a finite number of elements. We discuss construction of these estimators in Supplemental Appendix S4. Using these estimators, we estimate $\overline{\beta}^\star$ with

$$\hat{\overline{\beta}}^\star \equiv \sup_{m \in \mathcal{M}} \hat{\Gamma}^\star(m) \quad \text{s.t.} \quad \sum_{s \in \mathcal{S}} |\hat{\Gamma}_s(m) - \hat{\beta}_s| \leq \inf_{m' \in \mathcal{M}} \sum_{s \in \mathcal{S}} |\hat{\Gamma}_s(m') - \hat{\beta}_s| + \kappa_n, \tag{27}$$

where $\kappa_n > 0$ is a tuning parameter that converges to zero at an appropriate rate with the sample size, n. The analogous minimization problem provides an estimator $\hat{\underline{\beta}}^\star$ for $\underline{\beta}^\star$. In addition to replacing $\Gamma^\star$, Γ_s and β_s by estimators, these programs also differ from those in Proposition 2 in the way they handle the set $\mathcal{M}_{\mathcal{S}}$ of observationally equivalent MTR functions. Instead of constraining m to lie in this set as in (13), the constraint in (27) restricts attention to the subset of $m \in \mathcal{M}$ that come closest to satisfying these constraints, up to a tolerance κ_n. This modification means that (27) will always have a feasible solution, so that our estimators $\hat{\overline{\beta}}^\star$ and $\hat{\underline{\beta}}^\star$ always exist.

Our conditions for consistency in Supplemental Appendix S3 are quite weak. In particular, they allow for $|\mathcal{S}|$ to be infinite and $\mathcal{M}$ infinite dimensional. To handle the former case, we generalize the constraint in (27) to allow for loss functions other than a (finite) sum of absolute deviations. However, the absolute deviations loss in (27) is attractive from a computational standpoint. Computing $\hat{\overline{\beta}}^\star$ in this case requires two simple steps. First, one solves the program in the right-hand side of the constraint in (27). This can be reformulated as a linear program by adding slack variables for

[10] Formally, if for every $m = (m_0, m_1) \in \mathcal{M}_{\mathcal{S}}$ and $\Pi m = (\Pi m_0, \Pi m_1)$ there is a sequence $m_n \in \mathcal{M}_{\mathcal{S}}$ such that $\Gamma^\star(m_n) \to \Gamma^\star(\Pi m)$, then Proposition 4 continues to hold. The assumption that $(\Pi m_0, \Pi m_1) \in \mathcal{M}$ for every $m \in \mathcal{M}$ is simply a sufficient condition for this more general requirement.

the positive and negative parts of $\hat{\Gamma}_s(m') - \hat{\beta}_s$. Second, one solves (27) and the analogous problem for $\underline{\hat{\beta}}^\star$, both of which can similarly be reformulated as linear programs. Since $\kappa_n > 0$, these programs will always be feasible. In practice, one would typically take $\mathcal{M} = \mathcal{M}_{\text{fd}}$, which renders these programs finite-dimensional through the same argument as in (22) and (23).

4 Applications of the Method

4.1 Partial Identification of Policy Relevant Treatment Effects

The policy relevant treatment effect (PRTE) is the average causal effect on Y of changing from a baseline policy to an alternative policy that provides different incentives to participate in treatment (Heckman and Vytlacil, 1999, 2005). In many situations, this policy comparison does not directly correspond to the variation in treatment induced by the instrument, so the PRTE is not point identified. In this section, we discuss how researchers can use our method to construct bounds on the PRTE in such cases.

Consider a policy a that operates by changing factors that affect an agent's treatment decision. We follow Heckman and Vytlacil (1999, 2005) in assuming that a has no direct effect on the potential outcomes (Y_0, Y_1), and in particular that it does not affect the parameter space $\mathcal{M}$ of admissible MTR functions. This assumption is similar to the exclusion restriction. The policy can then be summarized by a propensity score and instrument pair (p^a, Z^a). Treatment choice under policy a is given by

$$D^a \equiv \mathbb{1}[U \leq p^a(Z^a)],$$

where U is the same unobservable term as in the selection equation for the status quo policy, D. The outcome of Y that would be observed under the new policy is therefore

$$Y^a = D^a Y_1 + (1 - D^a) Y_0.$$

The PRTE of policy a_1 relative to another policy a_0 is defined as

$$\text{PRTE} \equiv \frac{E[Y^{a_1}] - E[Y^{a_0}]}{E[D^{a_1}] - E[D^{a_0}]} \tag{28}$$

where we assume that $E[D^{a_1}] \neq E[D^{a_0}]$.[11]

In Table 1, we provide weights $\omega_0^\star$ and $\omega_1^\star$ that can be used to express the PRTE

[11] If this assumption is a concern, one can also define the PRTE as $E[Y^{a_1}] - E[Y^{a_0}]$, see Heckman and Vytlacil (2001a) or pp. 380–381 of Carneiro et al. (2010). Our approach directly applies to this definition as well.

as a target parameter $\beta^\star$ with the form given in (6) for different policies a_0 and a_1. After choosing the weights that correspond to the policy comparison of interest, our method can be used to estimate bounds for the PRTE. These bounds can be fully nonparametric, but they can also incorporate parametric or shape restrictions if desired.

The way in which different policy comparisons translate into different weights is illustrated in Table 1 through the three specific examples considered by Carneiro et al. (2011). Each of these comparisons is between a hypothetical policy a_1 and the status quo policy a_0, the latter of which is characterized by the pair $(p^{a_0}, Z^{a_0}) = (p, Z)$ observed in the data. The comparisons are: (i) an additive α change in the propensity score, i.e. $p^{a_1} = p + \alpha$; (ii) a proportional $(1 + \alpha)$ change in the propensity score, i.e. $p^{a_1} = (1 + \alpha)p$; and (iii) an additive α shift in the distribution the jth component of Z, i.e. $Z^{a_1} = Z + \alpha e_j$, where e_j is the jth unit vector. The first and second of these represent policies that increase (or decrease) participation in the treatment by a given amount α or a proportional amount $(1 + \alpha)$. The third policy represents the effect of shifting the distribution of a variable that impacts treatment choice.

In all of these definitions, α is a quantity that could either be hypothesized by the researcher, estimated from some auxiliary choice model, or extrapolated from the estimated propensity score under parametric assumptions. For example, in the application in our working paper (Mogstad et al., 2017), we estimate the value of α by parametrically extrapolating a demand curve fit with experimentally varied prices. Since α is interpretable in terms of the change of treatment participation probability, a simpler approach is to just specify a value of that represents an empirically interesting change in the probability of choosing treatment.

4.2 Extrapolation of Local Average Treatment Effects

Imbens and Angrist (1994) showed that the LATE is point identified under the assumptions considered in this paper. As argued by Imbens (2010, pp. 414–415), it is desirable to report both the LATE, with its high internal validity, but possibly limited external validity, and extrapolations of the LATE to larger or different populations. We now show how to use our method to perform this type of extrapolation, thereby allowing researchers to assess the sensitivity of a given LATE estimate to an expansion (or contraction) of the complier subpopulation.

For doing this, it is useful to connect the LATE parameter to the PRTE. To see the relationship, suppose that there are no covariates X, i.e. $Z = Z_0$, and suppose that Z_0 is binary. Consider the PRTE that results from comparing a policy a_1 under which every agent receives $Z = 1$ against a policy a_0 under which every agent receives $Z = 0$.

Choices under these policies are

$$D^{a_0} \equiv \mathbb{1}[U \leq p(0)] \quad \text{and} \quad D^{a_1} \equiv \mathbb{1}[U \leq p(1)],$$

where $p(1) > p(0)$ are the propensity score values in the observed data. The PRTE for this policy comparison is

$$\frac{E[Y^{a_1}] - E[Y^{a_0}]}{E[D^{a_1}] - E[D^{a_0}]} = \frac{E\left[(D^{a_1} - D^{a_0})(Y_1 - Y_0)\right]}{p(1) - p(0)} = E\left[Y_1 - Y_0 \mid U \in (p(0), p(1)]\right], \quad (29)$$

where we used $D^{a_1} - D^{a_0} = \mathbb{1}[U \in (p(0), p(1)]]$. The right-hand side of (29) is precisely the LATE as defined by Imbens and Angrist (1994).

Extrapolation of the LATE amounts to changing $p(0)$, $p(1)$, or both. For example, suppose that the researcher wants to examine the sensitivity of the LATE to an expansion of the complier subpopulation that includes individuals with lower willingness to pay for treatment. This sensitivity check corresponds to shifting $p(1)$ to $p(1) + \alpha$ for $\alpha > 0$. Arguing as in (29), the extrapolated LATE can be shown to be

$$\text{PRTE}(\alpha) = E\left[Y_1 - Y_0 \mid U \in (p(0), p(1) + \alpha]\right]. \quad (30)$$

PRTE(α) is still a LATE as defined by Heckman and Vytlacil (2005), but one that is not point identified by the IV estimand unless $\alpha = 0$. As we discussed in the previous section, we can use our method to bound PRTE(α) for $\alpha > 0$.

To gain some intuition on the determinants of these bounds, we write PRTE(α) as

$$\begin{aligned}&\text{PRTE}(\alpha)\\ &\quad = \left(\frac{p(1) - p(0)}{\alpha + p(1) - p(0)}\right)\text{LATE} + \left(\frac{1}{\alpha + p(1) - p(0)}\right)\int_{p(1)}^{p(1)+\alpha} m_1(u) - m_0(u)\, du,\end{aligned}$$

where LATE is the usual point identified LATE in (29). This decomposition shows that the conclusions that can be drawn about PRTE(α) depend on what can be said about $m_1(u) - m_0(u)$ for $u \in [p(1), p(1) + \alpha]$. Therefore, restrictions on the parameter space $\mathcal{M}$ of admissible MTR pairs translate directly into restrictions on the possible values of PRTE(α). For example, if we know bounds on the support of Y, which is always the case for example if Y is binary, then even nonparametric bounds can be informative about PRTE(α).

Our method allows a researcher to formally and transparently balance their desire for robust assumptions against their desire for broader extrapolation. Stronger assumptions are reflected through a more restrictive parameter space $\mathcal{M}$ of admissible MTR pairs. Less ambitious extrapolations are reflected through smaller values of α. For a given α, more restrictive specifications of $\mathcal{M}$ yield smaller bounds, while for a given specification of $\mathcal{M}$, smaller values of α also yield smaller bounds. Both margins can be smoothly adjusted, with point identification obtained as a limiting case as $\alpha \to 0$. We illustrate this tradeoff in our numerical example in Section 5.

4.3 Testable Implications

If the set $\mathcal{M}_{\mathcal{S}}$ is empty, then the model is misspecified: There does not exist a pair of MTR functions m that is in the parameter space ($m \in \mathcal{M}$), and which could have generated the observed data. If $\mathcal{M}$ is an unrestricted class of functions, then this is attributable to a falsification of selection equation (2) together with Assumptions I. The testable implications of these assumptions for the IV model are well-known, see e.g. Balke and Pearl (1997), Imbens and Rubin (1997), and Kitagawa (2015). On the other hand, if other restrictions have been placed on $\mathcal{M}$, then misspecification could be due either to the failure of (2) with Assumptions I, or the specification of $\mathcal{M}$, or both.

In our framework, this reasoning can be used more generally to provide testable implications about properties of the underlying MTR functions. For example, Table 1 reports the weights that correspond to the quantity $E[Y_0|D = 1] - E[Y_0|D = 0]$. This quantity is often described as a measure of selection bias, since it captures the extent to which average untreated outcomes differ solely on the basis of treatment status. The weights provide a linear mapping $\Gamma^{\star}_{\text{sel}}$ such that $\Gamma^{\star}_{\text{sel}}(m)$ indicates the amount of selection bias under an MTR pair m. Suppose that we restrict $\mathcal{M}$ to contain only MTR pairs m for which $\Gamma^{\star}_{\text{sel}}(m) = 0$. Then rejecting the null hypothesis that $\mathcal{M}_{\mathcal{S}}$ is nonempty can be interpreted as rejecting the hypothesis that there is no selection bias, as long as (2), Assumptions I, and any other restrictions on $\mathcal{M}$ are not deemed suspect.

Alternatively, one might be interested in testing the joint hypothesis that there is no selection on unobservables; that is, no selection bias and no selection on the gain. Weights for a typical measure of selection on the gain are provided in Table 1. These too provide a linear mapping $\Gamma^{\star}_{\text{gain}}$ on the set of MTR pairs $\mathcal{M}$. The hypothesis that there is no selection on unobservables would be rejected if we were to reject the null hypothesis that $\mathcal{M}_{\mathcal{S}}$ is nonempty when $\mathcal{M}$ is restricted to contain only MTR pairs m for which $\Gamma^{\star}_{\text{sel}}(m) = \Gamma^{\star}_{\text{gain}}(m) = 0$.

5 Numerical Illustration

In this section, we illustrate how to use our method to construct nonparametric bounds on treatment parameters of interest, and how shape restrictions and parametric assumptions can be used to tighten these bounds.

5.1 The Data Generating Process

We consider a simple example with a trinary instrument, $Z \in \{0, 1, 2\}$, with $P(Z = 0) = .5$, $P(Z = 1) = .4$, and $P(Z = 2) = .1$. The propensity score is specified as $p(0) = .35, p(1) = .6$, and $p(2) = .7$. We take the outcome $Y \in \{0, 1\}$ to be binary and restrict $\mathcal{M}$ to contain only MTR pairs that are bounded between 0 and 1. The data is generated using the MTR functions

$$\begin{aligned} & m_0(u) = .6b_0^2(u) + .4b_1^2(u) + .3b_2^2(u) \\ \text{and} \quad & m_1(u) = .75b_0^2(u) + .5b_1^2(u) + .25b_2^2(u), \end{aligned} \tag{31}$$

where b_k^2 is the kth Bernstein basis polynomial of degree 2.[12]

5.2 IV Estimand, Weights, and Target Parameter

Figure 1 contains two plots with two vertical axes. The left plot is for $d = 0$, while the right plot is for $d = 1$, and both vertical axes apply to both plots. The left axis measures weight functions, which are indicated with colored curves and, for the sake of clarity, are not drawn over regions where they are zero. The blue weights correspond to ω_{ds} when $s(D, Z)$ is taken to be (10) so that β_s is the IV slope coefficient estimand (7) from using Z as an instrument for D. These weights are positive between the smallest and largest values of the propensity score, i.e. from $p(0) = .35$ to $p(2) = .7$, and they change value at $p(1) = .6$.

As shown by Imbens and Angrist (1994), three LATEs are nonparametrically point identified in this setting: $\text{LATE}(.35, .6), \text{LATE}(.35, .7)$ and $\text{LATE}(.6, .7)$. Suppose that the researcher wants to examine the sensitivity of these average causal effects to an expansion of the complier subpopulation. Then they might be interested in the target parameter

$$\text{LATE}(.35, .9) \equiv E\left[Y_1 - Y_0 | U \in (.35, .9]\right].$$

[12] Supplemental Appendix S2 contains the definition of the Bernstein polynomials, along with a discussion of some useful properties of the Bernstein polynomial basis.

The weights for this parameter are drawn in red in Figure 1. As shown in Table 1, the weights are constant over $[.35, .9]$ with magnitude $(.9 - .35)^{-1} \approx 1.81$.

The right vertical axis in Figure 1 measures MTR functions for both the $d = 0$ and $d = 1$ plots. The MTR functions that were used to generate the data, i.e. (31), are plotted in black. These MTR functions imply a value of approximately .074 for the IV slope coefficient. By Proposition 1, this value is equal to the integral of the product of the black and blue curves, summed over the $d = 0$ and $d = 1$ plots. Similarly, these MTR functions imply a value of approximately .046 for LATE$(.35, .9)$ through an analogous sum of integrals using the red curves.

5.3 Bounds on the Target Parameter

Figure 2 is like Figure 1, except the MTR functions that are plotted yield the nonparametric upper bound on LATE$(.35, .9)$, computed using a constant spline as discussed in Section 3.1.[13] The pair $m \equiv (m_0, m_1)$ in this plot is generated by trying to make m_1 as large as possible, and m_0 as a small as possible, on the support of the red weights, while still yielding a value of .074 for the IV slope coefficient determined by the blue weights. These MTR functions imply a value of .5 for the target parameter LATE$(.35, .9)$, which is the largest value that is consistent with the IV slope estimand. There are multiple pairs of MTR functions with this property. In particular, notice that neither weights are positive over the region $[0, .35]$, so that MTR pairs may be freely adjusted on this region without changing the implied values of the IV slope estimand or LATE$(.35, .9)$. The lower bound of $-.421$ indicated in Figure 2 is obtained through an analogous minimization problem that follows the same logic.

Figure 3 repeats this exercise while including a second IV–like estimand. The second estimand is the OLS slope coefficient, whose weights are drawn in light blue. Notice that, whereas the blue and red weights are symmetric between $d = 0$ and $d = 1$ in the sense of having the same magnitude but different signs, the light blue weights for the OLS slope coefficient are asymmetric. A maximizing or minimizing MTR pair must yield the implied values for both the IV slope coefficient and the OLS slope coefficient, which is approximately .253. In this DGP, the additional constraint from the OLS slope coefficient has no effect on the upper bound of LATE$(.35, .9)$, but does tighten the lower bound slightly from $-.421$ to $-.411$.

In Figure 4, instead of using Z as a single instrument, we split Z into two binary indicators, $\mathbb{1}[Z = 2]$ and $\mathbb{1}[Z = 3]$, to create two IV slope estimands. This tightens both

[13] This amounts to solving two linear programs. We wrote these programs in AMPL (Fourer, Gay, and Kernighan, 2002) and solved them with Gurobi (Gurobi Optimization, 2015).

bounds on LATE$(.35, .9)$ considerably. The tightest possible nonparametric bounds are obtained in Figure 5, which includes a collection of six IV–like specifications that is rich enough to satisfy the conditions of Proposition 3. The resulting bounds of $[-.138, .407]$ are the sharp nonparametric bounds for LATE$(.35, .9)$.

The bounds in Figure 5 can be tightened by imposing nonparametric shape restrictions. For example, in Figure 6, the MTR functions are restricted to be decreasing like the DGP MTR functions shown in Figure 1. This tightens the sharp identified set for LATE$(.35, .9)$ by ruling out non-decreasing MTR pairs like the one shown in Figure 5. Even tighter bounds can be obtained by also requiring the MTR functions to be smooth. This may be a desirable a priori assumption if one believes that the jump-discontinuous MTR pairs in Figure 6 are sufficiently poorly behaved as to be an unlikely description of the relationship between selection unobservables, U, and potential outcomes, Y_0 and Y_1. For example, in Figure 7, the MTR functions are restricted to be decreasing and characterizable by a polynomial of order 10 or lower. This eliminates the possibility of non-smooth MTR functions like those in Figure 6, and in this example reduces the bounds to $[0, 0.067]$.

5.4 Tradeoffs between Tightness and Extrapolation

Figure 8 illustrates how the bounds change as the parameter of interest changes. In particular, instead of LATE$(.35, .9)$, we construct bounds on

$$\text{LATE}(.35, \overline{u}) \equiv E\left[Y_1 - Y_0 | U \in (.35, \overline{u}]\right],$$

for different values of $\overline{u}$, using the same specification as in Figure 7. Sharp lower and upper bounds on this parameter are given by the blue and red curves, respectively. As evident from Figure 8, the bounds collapse to a point for $\overline{u} = .6$ and $.7$, i.e the two other points of support for the propensity score. For these values of $\overline{u}$, LATE$(.35, \overline{u})$ is a point identified LATE as in Imbens and Angrist (1994). For other values of $\overline{u}$ this parameter is not point identified, such as for $\overline{u} = .9$, which is indicated by the dotted vertical line. For $\overline{u}$ between .6 and .7 the bounds are very tight, as shown in the magnified region. As $\overline{u}$ decreases from .6 or increases above .7, the bounds widen.

This property reflects the increasing difficulty of drawing inference about a parameter the more dissimilar it is from what was observed in the data, i.e. the more extrapolation it requires. Given a desired tightness of the bounds, a more ambitious extrapolation can be obtained only by imposing stronger assumptions. Given a set of assumptions, tighter bounds can be obtained only by less ambitious extrapolations. Our framework allows the researcher to achieve bounds that are as narrow as they desire, while requiring them to honestly acknowledge the strength of their assumptions and the degree of extrapolation involved in their counterfactual.

6 Discussion: When to Use Our Method

We have proposed a method for using instrumental variables to draw inference about a wide range of treatment parameters other than the LATE. This enables researchers to learn about causal effects for a broad range of individuals, not just those who are affected by the instrument observed in the data. The ability to do this is critical to ensuring that estimates obtained through IV strategies are both externally valid and relevant to policy. An important aspect of our method is that it allows the researcher a large amount of flexibility in choosing a parameter of interest, and in choosing auxiliary identifying assumptions that can be used to help tighten their empirical conclusions.

Our method can be used to analyze conventional treatment parameters such as the ATE, the ATT, and the ATU. These parameters are relevant for policy counterfactuals that hypothesize mandating a choice of treatment. However, many policy discussions focus on interventions that change the costs or benefits of choosing certain activities, while still allowing individuals to freely select into these activities. Our method is especially useful for these settings, since it allows researchers to draw inference about target parameters in the class of PRTEs. These parameters are useful for analyzing interventions that influence (but may not fully determine) an individual's treatment choice, for example by changing the costs associated with the treatment alternatives. These types of interventions will often involve extrapolation to new environments, which makes our focus on partial identification natural.

Partial identification approaches are sometimes criticized for yielding empirical conclusions that are insufficiently informative for practitioners. Computational methods, such as the one proposed in this paper, are useful tools for answering this criticism. The flexibility of our framework means that a researcher can smoothly adjust their policy question (target parameter), or the assumptions they are willing to maintain, in a way that approaches point identification as a special case. As a result, the tightness of the bounds they report is at their discretion, while still being disciplined by the reality that stronger conclusions require stronger assumptions.

One common empirical setting for which our method is particularly well suited is when the instrument is based on the lack of pattern or predictability in a natural event that cannot be shifted by policy, such as the weather (Angrist et al., 2000; Miguel, Satyanath, and Sergenti, 2004), or the gender composition of children (Angrist and Evans, 1998; Black, Devereux, and Salvanes, 2005). Instruments derived from such “natural experiments” are popular in empirical work due to their apparent exogeneity. However, the associated LATEs can have low external validity, since the group of individuals whose behavior is affected by a natural experiment is likely to be different from the group of individuals who would be affected under an interesting policy counterfactual. Our method can be used to rigorously address this critique by analyzing the sensitivity of the empirical conclusions to the size or composition of the group affected by the natural experiment.

Another common application for our method is when the instrument represents a past policy change, but the researcher is interested in the effect of expanding or contracting the policy. In this case, standard IV methods already identify the causal effect for individuals whose treatment choice was affected by the prior policy change. However, using the previous change to inform about a new expansion requires extrapolation. Our method can be used to conduct this extrapolation rigorously. For example, in our working paper (Mogstad et al., 2017), we apply our method to analyze how actual and counterfactual price subsidies affect the adoption and usage of antimalarial bed nets in Kenya. Our results there suggest that generous subsidy regimes encourage usage among individuals who would otherwise not use a bed net, albeit at a high cost of subsidizing other inframarginal individuals.

A Proofs

Proof of Proposition 1. Using equation (1), we first note that

$$\beta_s = E[s(D,Z)DY_1] + E[s(D,Z)(1-D)Y_0]. \tag{32}$$

Using equation (2) with Assumptions I.1 and I.2, observe that the first term of (32) can be written as

$$\begin{aligned} E[s(D,Z)DY_1] &= E[s(D,Z)\mathbb{1}[U \leq p(Z)]E[Y_1|U,Z]] \\ &\equiv E[s(1,Z)\mathbb{1}[U \leq p(Z)]m_1(U,X)], \end{aligned} \tag{33}$$

where the first equality follows because $s(D,Z)D$ is a deterministic function of (U,Z), and the second equality uses the definition of m_1 and I.2, together with the identity that

$$s(D,Z)\mathbb{1}[U \leq p(Z)] \equiv s\left(\mathbb{1}[U \leq p(Z)], Z\right)\mathbb{1}[U \leq p(Z)] = s(1,Z)\mathbb{1}[U \leq p(Z)].$$

Using the normalization that the distribution of U conditional on Z is uniformly distributed on $[0,1]$ for any realization of Z, it follows from (33) that

$$\begin{aligned} E[s(D,Z)DY_1] &= E\left[E\left[s(1,Z)\mathbb{1}[U \leq p(Z)]m_1(U,X)|Z\right]\right] \\ &= E\left[\int_0^1 s(1,Z)\mathbb{1}[u \leq p(Z)]m_1(u,X)\,du\right] \\ &\equiv E\left[\int_0^1 \omega_{1s}(u,Z)m_1(u,X)\,du\right]. \end{aligned}$$

The claimed result follows after applying a symmetric argument to the second term on the right hand side of equation (32). *Q.E.D.*

Proof of Proposition 2. Since $\Gamma_s : \mathcal{M} \to \mathbf{R}$ is linear for every $s \in \mathcal{S}$, it follows from convexity of $\mathcal{M}$ that $\mathcal{M}_{\mathcal{S}}$ is convex as well. (Note that the empty set is trivially convex.) If $\mathcal{M}_{\mathcal{S}}$ is empty, then by definition we also have $\mathcal{B}^\star_{\mathcal{S}} = \emptyset$. On the other hand, if $\mathcal{M}_{\mathcal{S}} \neq \emptyset$, then the linearity of $\Gamma^\star : \mathcal{M} \to \mathbf{R}$ implies that $\mathcal{B}^\star_{\mathcal{S}} \equiv \Gamma^\star(\mathcal{M}_{\mathcal{S}}) \subseteq \mathbf{R}$ is a convex set, and so its closure is $[\inf_{m \in \mathcal{M}_{\mathcal{S}}} \Gamma^\star(m), \sup_{m \in \mathcal{M}_{\mathcal{S}}} \Gamma^\star(m)] \equiv [\underline{\beta}^\star, \overline{\beta}^\star]$. *Q.E.D.*

Proof of Proposition 3. For notational simplicity, we define the set

$$\mathcal{M}_{\text{id}} \equiv \{m \in \mathcal{M} : m \text{ satisfies (14) and (15) almost surely}\}.$$

For any $m \equiv (m_0, m_1) \in \mathcal{M}_{\text{id}}$ and $s \in \mathcal{S}$ we obtain from the definition of β_s that

$$\beta_s = E[s(D,Z)E[Y|D,Z]] = \sum_{d \in \{0,1\}} E[\mathbb{1}[D=d]s(d,Z)E[Y|D=d,Z]]. \tag{34}$$

Examining the $d = 0$ term in the summation, we obtain

$$\begin{aligned} E[\mathbb{1}[D=0]s(0,Z)E[Y|D=0,Z]] &= E\left[\mathbb{1}[D=0]s(0,Z)\frac{1}{(1-p(Z))}\int_{p(Z)}^1 m_0(u,X)du\right] \\ &= E\left[\mathbb{1}[D=0]\frac{1}{(1-p(Z))}\int_0^1 m_0(u,X)\omega_{0s}(u,Z)du\right] \\ &= E\left[\int_0^1 m_0(u,X)\omega_{0s}(u,Z)du\right], \end{aligned} \tag{35}$$

where the first equality follows from $m \in \mathcal{M}_{\text{id}}$ satisfying (14), the second equality uses the definition $\omega_{0s}(u,z) \equiv s(0,z)\mathbb{1}[u > p(z)]$, and the final equality is implied by $P(D=0|Z) = 1 - p(Z)$. By analogous arguments, we also obtain

$$E[\mathbb{1}[D=1]s(1,Z)E[Y|D=1,Z]] = E\left[\int_0^1 m_1(u,X)\omega_{1s}(u,Z)du\right]. \tag{36}$$

Together, (34), (35), and (36) imply that $\Gamma_s(m) = \beta_s$. In particular, since $s \in \mathcal{S}$ and $m \in \mathcal{M}_{\text{id}}$ were arbitrary, we conclude that $\mathcal{M}_{\text{id}} \subseteq \mathcal{M}_{\mathcal{S}}$, as claimed.

Next, suppose $\mathcal{S} = \{s(d,z) = \mathbb{1}[d = d']f(z) : (d', f) \in \{0,1\} \times \mathcal{F}\}$ and that the closed linear span of $\mathcal{F}$ is equal to $\mathbf{L}^2(Z)$. Then note that for any $m \in \mathcal{M}_{\mathcal{S}}$ and $s \in \mathcal{S}$ with the structure $s(d,z) = \mathbb{1}[d=0]f(z)$ we obtain by definition of β_s and Γ_s that

$$E[Y\mathbb{1}[D=0]f(Z)] \equiv \beta_s = \Gamma_s(m) = E\left[\int_{p(Z)}^1 m_0(u,X)du \times f(Z)\right] \tag{37}$$

where the second equality follows from $m \in \mathcal{M}_{\mathcal{S}}$ and the final equality is due to $\omega_{0s}(u,z) \equiv \mathbb{1}[u > p(z)]s(0,z)$ and $s(0,z) = \mathbb{1}[0=0]f(z)$. Furthermore, define

$$\Delta(Z) \equiv E[Y\mathbb{1}[D=0]|Z] - \int_{p(Z)}^1 m_0(u,X)du \tag{38}$$

and note that (37) implies that $E[\Delta(Z)f(Z)] = 0$ for all $f \in \mathcal{F}$. Since $E[Y_0^2] < \infty$ by Assumption I.2 and $E[\int m_d^2(u,X)du] < \infty$, Jensen's inequality implies that $\Delta \in \mathbf{L}^2(Z)$. Thus, since $E[\Delta(Z)f(Z)] = 0$ for all $f \in \mathcal{F}$ and the closed linear span of $\mathcal{F}$ equals $\mathbf{L}^2(Z)$, we conclude that $\Delta(Z) = 0$ almost surely. Equivalently, since $P(D = 0|Z) =$

$1 - p(Z)$ by definition of $p(Z)$, we obtain whenever $1 - p(Z) > 0$ that

$$E[Y|D = 0, Z] = \frac{1}{1 - p(Z)} \int_{p(Z)}^{1} m_0(u, X) du \tag{39}$$

almost surely, i.e. m_0 satisfies (14). Analogous arguments imply that m_1 satisfies (15). Since $(m_0, m_1) = m \in \mathcal{M}_{\mathcal{S}}$ was arbitrary, we conclude that $\mathcal{M}_{\mathcal{S}} \subseteq \mathcal{M}_{\text{id}}$, which together with (16) establishes (17). *Q.E.D.*

Proof of Proposition 4. We prove the proposition for the upper bound of the target parameter. The proof for the lower bound follows by identical arguments.

Observe that since $\mathcal{M}_{\text{fd}} \subseteq \mathcal{M}$ by definition of $\mathcal{M}_{\text{fd}}$, we can immediately conclude that $\overline{\beta}^{\star}_{\text{fd}} \leq \overline{\beta}^{\star}$. To establish the opposite inequality, we show that $\Gamma_{ds}(\Pi m_d) = \Gamma_{ds}(m_d)$ for any $(m_0, m_1) \in \mathcal{M}$, where Πm_d is defined as in (26). As shorthand, let $\mu_{jl} \equiv E[m_d(U, x_l)|U \in \mathcal{U}_j, X = x_l]$ in the definition of Πm_d. Then

$$\begin{aligned}
\Gamma_{ds}(\Pi m_d) &\equiv E\left[\int \left(\sum_{j=1}^{J}\sum_{l=1}^{L} \mu_{jl} b_{jl}(u, X)\right) \omega_{ds}(u, Z)\, du\right] \\
&= E\left[\sum_{l=1}^{L} \mathbb{1}[X = x_l] \sum_{j=1}^{J} \int_{\mathcal{U}_j} \mu_{jl}\omega_{ds}(u, Z)\, du\right] \\
&= E\left[\sum_{l=1}^{L}\sum_{k=1}^{K_l} \mathbb{1}[X = x_l, Z = z_{lk}] \sum_{j=1}^{J} \int_{\mathcal{U}_j} \mu_{jl}\omega_{ds}(u, z_{lk})\, du\right],
\end{aligned} \tag{40}$$

where the second equality uses the definition of b_{jl} in (25), and the third equality follows after enumerating the support of Z, conditional on $X = x_l$, as $\{z_{lk}\}_{k=1}^{K_l}$. Since $\mathcal{U}_j$ is assumed to be constructed so that $\mathbb{1}[u \leq p(z)]$, and hence $\omega_{ds}(u, z)$ are constant over $\mathcal{U}_j$, it follows that

$$\begin{aligned}
\sum_{j=1}^{J} \int_{\mathcal{U}_j} \mu_{jl}\omega_{ds}(u, z_{lk})\, du &\equiv \sum_{j=1}^{J} \int_{\mathcal{U}_j} E\left[m_d(U, x_l)|U \in \mathcal{U}_j, X = x_l\right] \omega_{ds}(u, z_{lk})\, du \\
&= \sum_{j=1}^{J} E\left[m_d(U, x_l)\omega_{ds}(U, z_{lk})|U \in \mathcal{U}_j, X = x_l\right] P[U \in \mathcal{U}_j] \\
&= \sum_{j=1}^{J} \int_{\mathcal{U}_j} m_d(u, x_l)\omega_{ds}(u, z_{lk})\, du \\
&= \int_0^1 m_d(u, x_l)\omega_{ds}(u, z_{lk})\, du,
\end{aligned} \tag{41}$$

where the second and third equalities used the normalization that U is uniformly distributed over $[0,1]$, conditional on X. Substituting (41) into (40), we have

$$\begin{aligned}\Gamma_{ds}(\Pi m_d) &= E\left[\sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbb{1}[X = x_l, Z = z_{lk}]\int_0^1 m_d(u, x_l)\omega_{ds}(u, z_{lk})\, du\right] \\ &= E\left[\int_0^1 m_d(u, X)\omega_{ds}(u, Z)\, du\right] \equiv \Gamma_{ds}(m_d),\end{aligned}$$

as claimed. An identical argument yields $\Gamma^\star_d(m_d) = \Gamma^\star_d(\Pi m_d)$. Given the assumption that $\Pi m \equiv (\Pi m_0, \Pi m_1) \in \mathcal{M}_{\text{fd}}$, it follows that

$$\begin{aligned}\overline{\beta}^\star &\equiv \sup_{m \in \mathcal{M}} \{\Gamma^\star(m) \text{ s.t. } \Gamma_s(m) = \beta_s \text{ for all } s \in \mathcal{S}\} \\ &= \sup_{m \in \mathcal{M}} \{\Gamma^\star(\Pi m) \text{ s.t. } \Gamma_s(\Pi m) = \beta_s \text{ for all } s \in \mathcal{S}\} \\ &\leq \sup_{m \in \mathcal{M}_{\text{fd}}} \{\Gamma^\star(m) \text{ s.t. } \Gamma_s(m) = \beta_s \text{ for all } s \in \mathcal{S}\} \equiv \overline{\beta}^\star_{\text{fd}}. \end{aligned} \tag{42}$$

We conclude that $\overline{\beta}^\star_{\text{fd}} = \overline{\beta}^\star$. *Q.E.D.*

References

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117. 14

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *The American Economic Review*, 88, 450–477. 24

Angrist, J. D. and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External Validity and," in *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, Cambridge University Press, vol. 51, 401–. 3

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *The Review of Economic Studies*, 67, 499–527. 1, 24

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442. 1

Balke, A. and J. Pearl (1997): "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 2, 3, 20

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2012): "Treatment effect bounds: An application to SwanGanz catheterization," *Journal of Econometrics*, 168, 223–243. 2

Bierens, H. J. (1990): "A consistent conditional moment test of functional form," *Econometrica: Journal of the Econometric Society*, 1443–1458. 11

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005): "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education," *The Quarterly Journal of Economics*, 120, 669–700. 24

Brinch, C. N., M. Mogstad, and M. Wiswall (2012): "Beyond LATE with a Discrete Instrument," *Working paper.* 12

——— (2017): "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, 125, 985–1039. 3, 6, 11, 12, 13

Carneiro, P., J. J. Heckman, and E. Vytlacil (2010): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica*, 78, 377–394. 3, 12, 17

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–81. 3, 7, 12, 18

Chen, X. (2007): "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. Volume 6, Part 2, 5549–5632. 13

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. 1

Dupas, P. (2014): "ShortRun Subsidies and LongRun Adoption of New Health Products: Evidence From a Field Experiment," *Econometrica*, 82, 197–228. 4

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008): "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206. 1

Fourer, R., D. M. Gay, and B. W. Kernighan (2002): *AMPL: A Modeling Language for Mathematical Programming*, Cengage Learning. 22

French, E. and J. Song (2014): "The Effect of Disability Insurance Receipt on Labor Supply," *American Economic Journal: Economic Policy*, 6, 291–337. 6

Gurobi Optimization, I. (2015): "Gurobi Optimizer Reference Manual," . 22

Heckman, J., J. L. Tobias, and E. Vytlacil (2003): "Simple Estimators for Treatment Parameters in a Latent-Variable Framework," *Review of Economics and Statistics*, 85, 748–755. 3

Heckman, J. J. and R. J. Robb (1985): "Alternative methods for evaluating the impact of interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press. 3

Heckman, J. J. and S. Urzua (2010): "Comparing IV with structural models: What simple IV can and cannot identify," *Journal of Econometrics*, 156, 27–37. 1

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432. 1

Heckman, J. J. and E. Vytlacil (2001a): "Policy-Relevant Treatment Effects," *The American Economic Review*, 91, 107–111. 1, 3, 17

——— (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. 1, 2, 3, 4, 6, 8, 12, 17, 19

Heckman, J. J. and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences of the United States of America*, 96, 4730–4734. 1, 4, 6, 12, 15, 17

——— (2001b): "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," in *Econometric Evaluations of Active Labor Market Policies in Europe*, ed. by M. Lechner and F. Pfeiffer, Heidelberg and Berlin: Physica. 1, 2

——— (2001c): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by K. M. C Hsiao and J. Powell, Cambridge University Press. 1, 12

——— (2007a): "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. Volume 6, Part 2, 4779–4874. 1

——— (2007b): "Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. Volume 6, Part 2, 4875–5143. 1

Huber, M. and G. Mellace (2014): "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints," *Review of Economics and Statistics*, 97, 398–411. 3

Imbens, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423. 18

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 1, 2, 5, 18, 19, 21, 23

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512. 1

Imbens, G. W. and D. B. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 64, 555–574. 3, 20

Kirkeboen, L., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings and Self-Selection," *The Quarterly Journal of Economics*, 131, 1057–1111. 1

Kitagawa, T. (2009): "Identification Region of the Potential Outcome Distributions under Instrument Independence," *Cemmap working paper.* 2

——— (2015): "A Test for Instrument Validity," *Econometrica*, 83, 2043–2063. 3, 20

Kowalski, A. (2016): "Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Examine Treatment Effect Heterogeneity in Experiments," *NBER Working paper 22363*. 12

Lee, S. and B. Salanié (2016): "Identifying Effects of Multivalued Treatments," *Working paper*. 1

Machado, C., A. M. Shaikh, and E. J. Vytlacil (2013): "Instrumental Variables and the Sign of the Average Treatment Effect," *Working paper*. 3

Maestas, N., K. J. Mullen, and A. Strand (2013): "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *The American Economic Review*, 103, 1797–1829. 7

Manski, C. (1994): "The selection problem," in *Advances in Econometrics, Sixth World Congress*, vol. 1, 143–70. 2

Manski, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24, 343–360. 2

——— (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, 319–323. 2

——— (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334. 2, 3, 12

——— (2003): *Partial identification of probability distributions*, Springer. 2

Manski, C. F. and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. 3, 12

——— (2009): "More on monotone instrumental variables," *Econometrics Journal*, 12, S200–S216. 3

Masten, M. A. (2015): "Random Coefficients on Endogenous Variables in Simultaneous Equations Models," *cemmap working paper 25/15*. 1

Masten, M. A. and A. Torgovitsky (2016): "Identification of Instrumental Variable Correlated Random Coefficients Models," *The Review of Economics and Statistics*, forthcoming. 1

Miguel, E., S. Satyanath, and E. Sergenti (2004): "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 112, 725–753. 24

Moffitt, R. (2008): "Estimating Marginal Treatment Effects in Heterogeneous Populations," *Annales d'conomie et de Statistique*, 239–261. 6

Mogstad, M., A. Santos, and A. Torgovitsky (2017): "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters," *NBER Working Paper*. 4, 16, 18, 25

Mourifié, I. (2015): "Sharp bounds on treatment effects in a binary triangular system," *Journal of Econometrics*, 187, 74–81. 2

Mourifié, I. and Y. Wan (2016): "Testing Local Average Treatment Effect Assumptions," *The Review of Economics and Statistics*, 99, 305–313. 3

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955. 2

Stinchcombe, M. B. and H. White (1998): "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric theory*, 14, 295–325. 11

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197. 1

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. 1, 5

Table 1: Weights for a Variety of Target Parameters

		Weights		
Target Parameter	**Expression**	$\boldsymbol{\omega_0(u,z) \equiv \omega_0(u,x,z_0)}$	$\boldsymbol{\omega_1(u,z) \equiv \omega_1(u,x,z_0)}$	**Measure** $\boldsymbol{\mu^\star}$
Average Untreated Outcome	$E[Y_0]$	1	0	Leb.$[0,1]$
Average Treated Outcome	$E[Y_1]$	0	1	Leb.$[0,1]$
Average Treatment Effect (ATE)	$E[Y_1 - Y_0]$	-1	1	Leb.$[0,1]$
Average Treatment Effect (ATE) given $X \in \mathcal{X}^\star$	$E[Y_1 - Y_0 \mid X \in \mathcal{X}^\star]$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[x \in \mathcal{X}^\star]}{P(X \in \mathcal{X}^\star)}$	Leb.$[0,1]$
Average Treatment on the Treated (ATT)	$E[Y_1 - Y_0 \mid D = 1]$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \le p(z)]}{P(D=1)}$	Leb.$[0,1]$
Average Treatment on the Untreated (ATU)	$E[Y_1 - Y_0 \mid D = 0]$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u > p(z)]}{P(D=0)}$	Leb.$[0,1]$
Marginal Treatment Effect at $\overline{u}$	$E[Y_1 - Y_0 \mid U = \overline{u}]$	-1	1	Dirac$(\overline{u})$
Local Average Treatment Effect for $U \in [\underline{u}, \overline{u}]$ (LATE$(\underline{u}, \overline{u})$)	$E[Y_1 - Y_0 \mid U \in [\underline{u}, \overline{u}]]$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \in [\underline{u}, \overline{u}]]}{(\overline{u} - \underline{u})}$	Leb.$[0,1]$
Policy Relevant Treatment Effect (PRTE) for new policy $(p^\star, Z^\star)$	$\dfrac{E[Y^\star] - E[Y]}{E[D^\star] - E[D]}$	$-\omega_1^\star(u,z)$	$\dfrac{P\left[u \le p^\star(Z^\star)\right] - P\left[u \le p(Z)\right]}{E[p^\star(Z^\star)] - E[p(Z)]}$	Leb.$[0,1]$
Additive PRTE with magnitude α	PRTE with $Z^\star = Z$ and $p^\star(z) = p(z) + \alpha$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \le p(z) + \alpha] - \mathbb{1}[u \le p(z)]}{\alpha}$	Leb.$[0,1]$
Proportional PRTE with magnitude α	PRTE with $Z^\star = Z$ and $p^\star(z) = (1+\alpha)p(z)$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \le (1+\alpha)p(z)] - \mathbb{1}[u \le p(z)]}{\alpha E[p(Z)]}$	Leb.$[0,1]$
PRTE for an additive α shift of the j^{th} component of Z	PRTE with $Z^\star = Z + \alpha e_j$ and $p^\star(z) = p(z)$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \le p(z + \alpha e_j)] - \mathbb{1}[u \le p(z)]}{E[p(Z + \alpha e_j)] - E[p(Z)]}$	Leb.$[0,1]$
Average Selection Bias	$E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0]$	$\dfrac{\mathbb{1}[u \le p(z)]}{P(D=1)} - \dfrac{\mathbb{1}[u > p(z)]}{P(D=0)}$	0	Leb.$[0,1]$
Average Selection on the Gain	$E[Y_1 - Y_0 \mid D = 1] - E[Y_1 - Y_0 \mid D = 0]$	$-\omega_1^\star(u,z)$	$\dfrac{\mathbb{1}[u \le p(z)]}{P(D=1)} - \dfrac{\mathbb{1}[u > p(z)]}{P(D=0)}$	Leb.$[0,1]$
Sum of two quantities $\beta_A^\star$, $\beta_B^\star$ with common measure $\mu^\star$	$\beta_A^\star + \beta_B^\star$	$\omega_{A,0}^\star(u,z) + \omega_{B,0}^\star(u,z)$	$\omega_{A,1}^\star(u,z) + \omega_{B,1}^\star(u,z)$	Common $\mu^\star$

Table 2: Notable IV–Like Estimands

Estimand	$\boldsymbol{\beta_s}$	$\boldsymbol{s(d,z) \equiv s(d,x,z_0)}$	**Notes**
IV slope	$\frac{\text{Cov}(Y,Z_0)}{\text{Cov}(D,Z_0)}$	$\frac{z_0 - E[Z_0]}{\text{Cov}(D,Z_0)}$	Z_0 scalar
IV slope using $p(Z)$	$\frac{\text{Cov}(Y,p(Z))}{\text{Cov}(D,p(Z))}$	$\frac{p(z) - E[p(Z)]}{\text{Cov}(D,p(Z))}$	—
IV (jth component)	$e_j' E[\widetilde{Z}\widetilde{X}']^{-1} E[\widetilde{Z}Y]$	$e_j' E[\widetilde{Z}\widetilde{X}']^{-1} \widetilde{z}$	$\widetilde{X} \equiv [1, D, X']'$ $\widetilde{Z} \equiv [1, Z, X']'$ Z scalar e_j the jth unit vector
TSLS (jth component)	$e_j' \left(\Pi E[\widetilde{Z}\widetilde{X}']\right)^{-1} \left(\Pi E[\widetilde{Z}Y]\right)$	$e_j' (\Pi E[\widetilde{Z}\widetilde{X}'])^{-1} \Pi \widetilde{z}$	$\Pi \equiv E[\widetilde{X}\widetilde{Z}'] E[\widetilde{Z}\widetilde{Z}']^{-1}$ Z vector
OLS slope	$\frac{\text{Cov}(Y,D)}{\text{Var}(D)}$	$\frac{d - E[D]}{\text{Var}(D)}$	—
OLS (jth component)	$e_j' E[\widetilde{X}\widetilde{X}']^{-1} E[\widetilde{X}Y]$	$e_j' E[\widetilde{X}\widetilde{X}']^{-1} \widetilde{x}$	$\widetilde{X} \equiv [1, D, X']'$ e_j the jth unit vector

Figure 1: MTRs Used in the Data Generating Process (DGP)

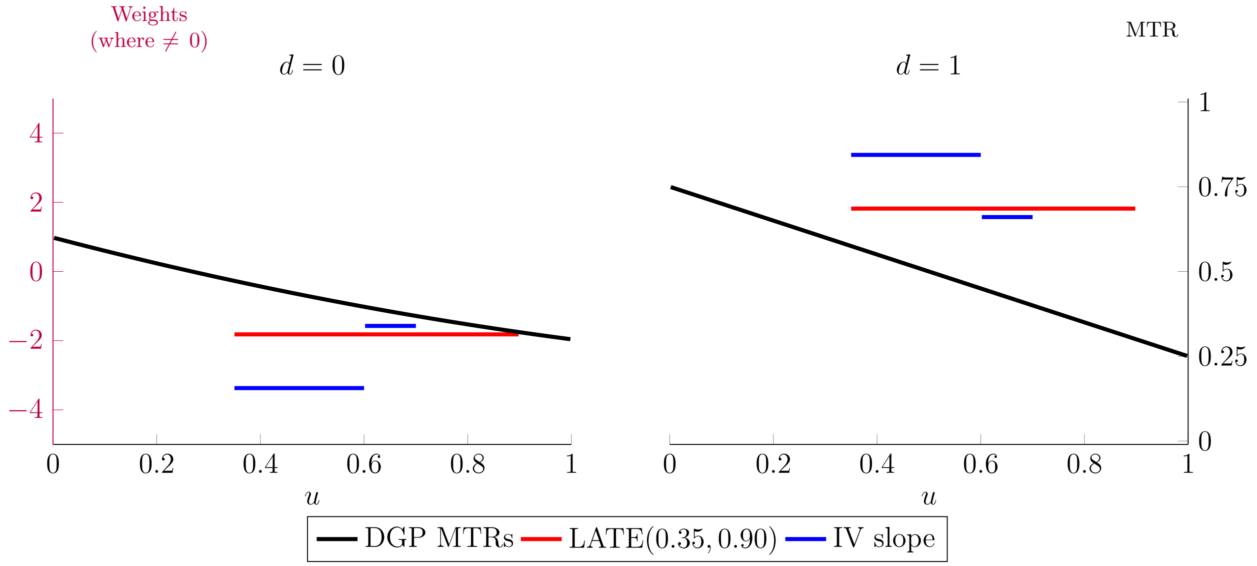


Figure 2: Maximizing MTRs When Using Only the IV Slope Coefficient

Nonparametric bounds: [-0.421,0.500]

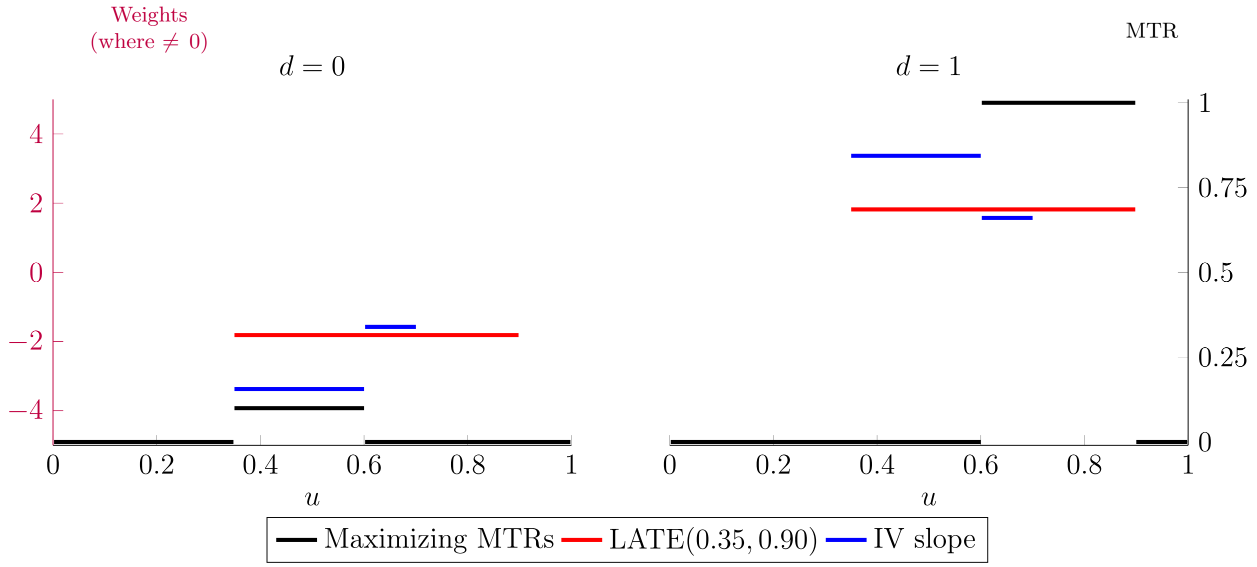


Figure 3: Maximizing MTRs When Using Both the IV and OLS Slope Coefficients

Nonparametric bounds: [-0.411,0.500]

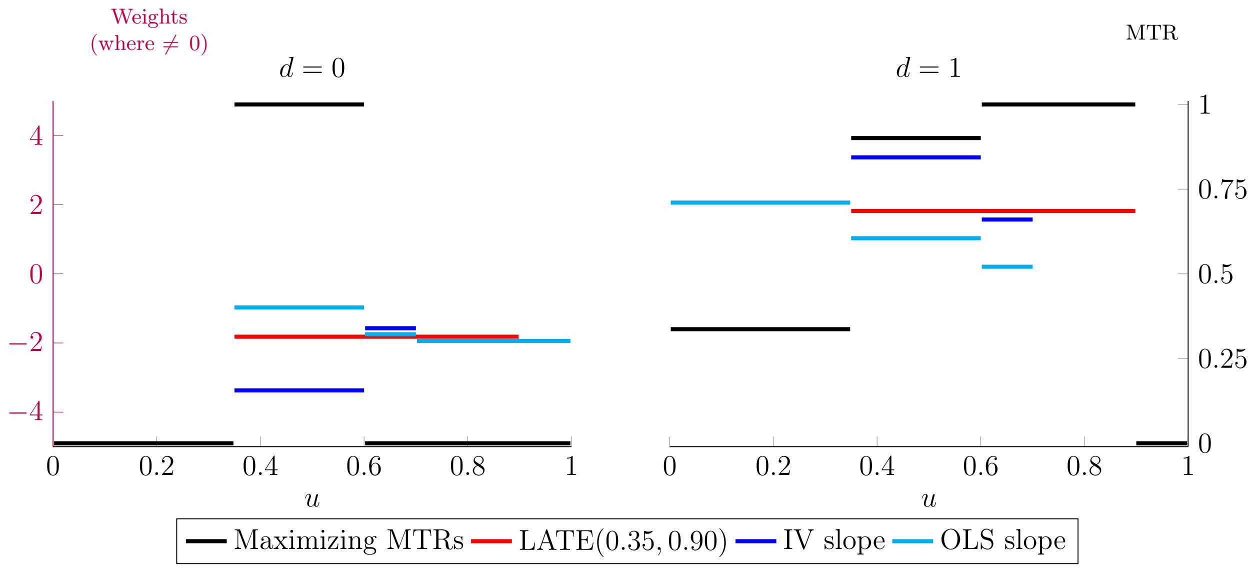


Figure 4: Maximizing MTRs When Breaking the IV Slope into Two Components

Nonparametric bounds: [-0.320,0.407]

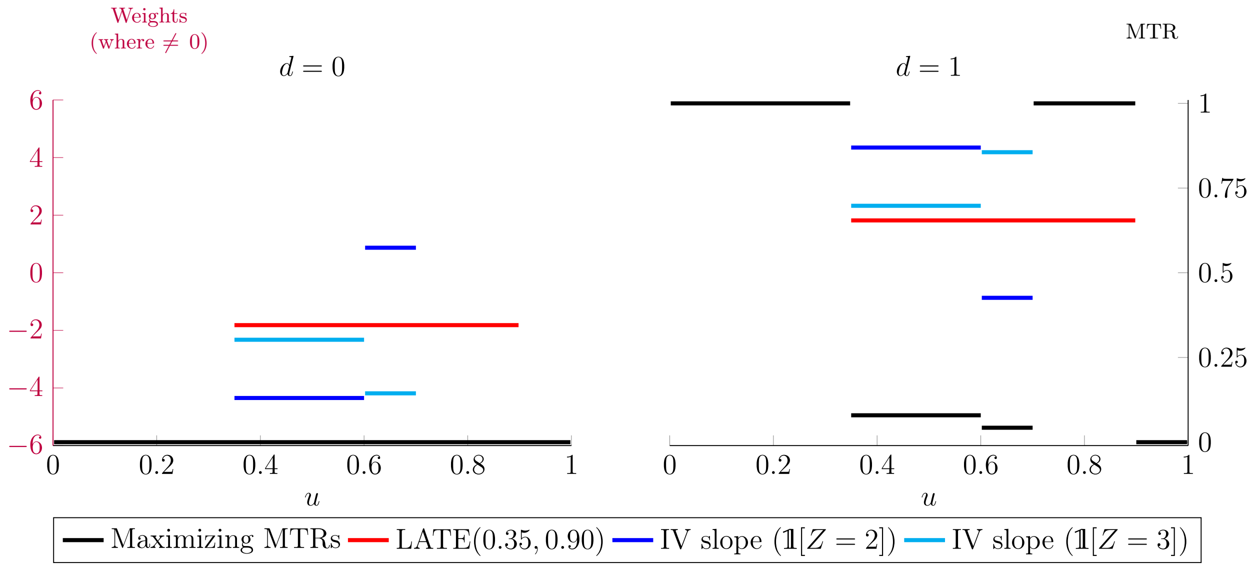


Figure 5: Maximizing MTRs When Using All IV–like Estimands (Sharp Bounds)

Nonparametric bounds: [-0.138,0.407]

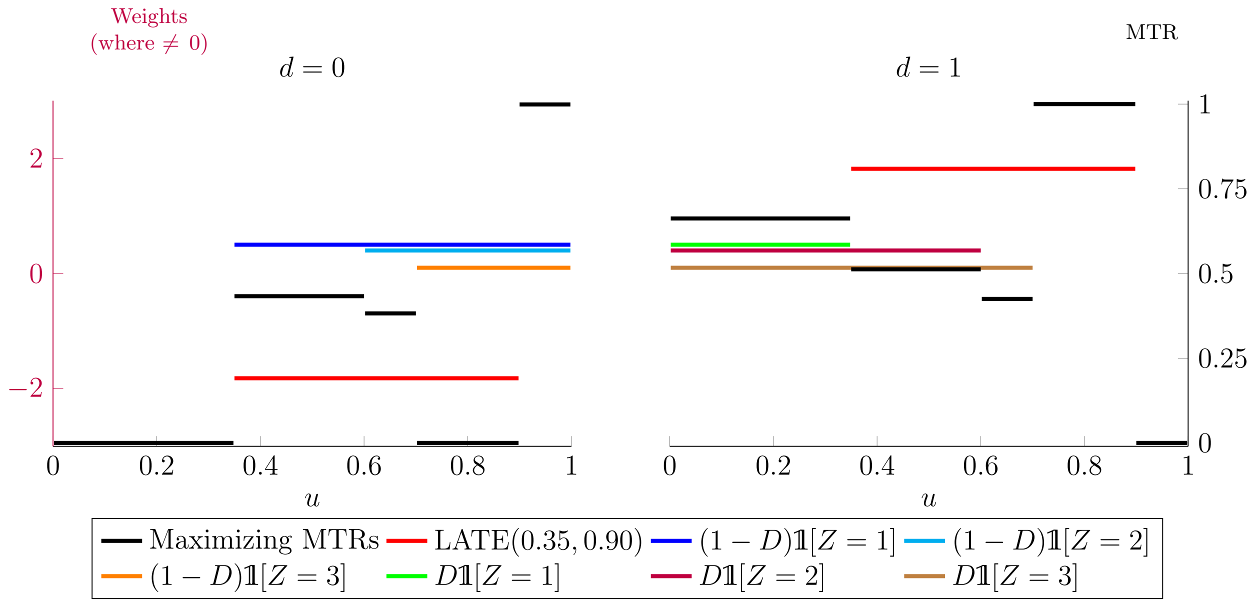


Figure 6: Maximizing MTRs When Restricted to be Decreasing

Nonparametric bounds, MTRs decreasing: [-0.095,0.077]

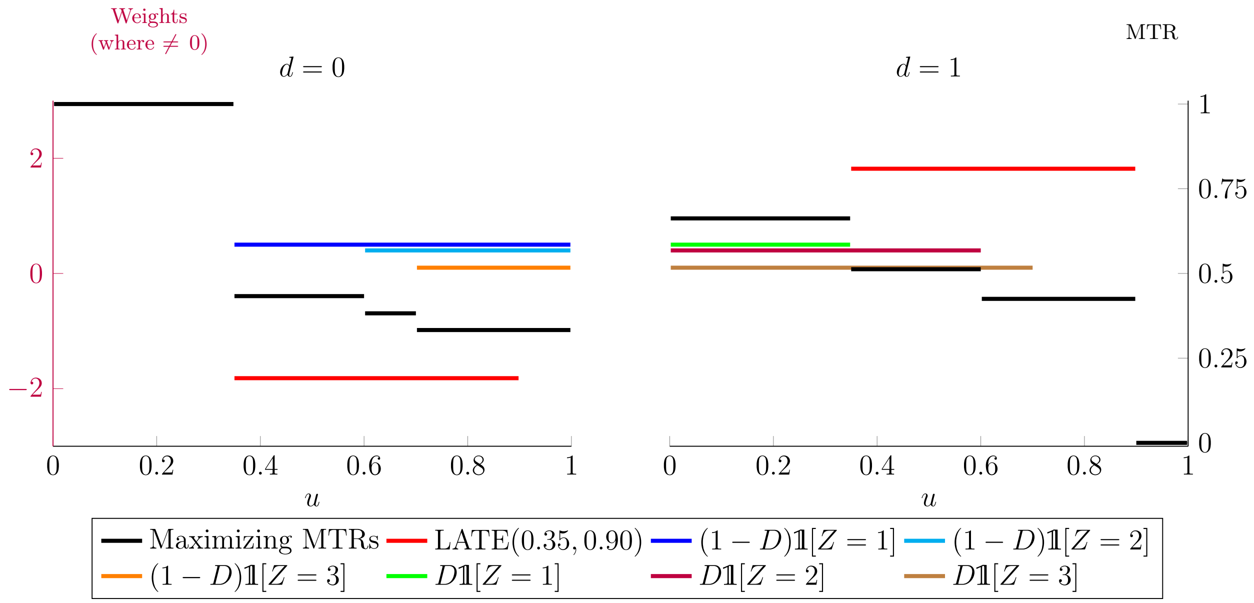


Figure 7: Maximizing MTRs When Further Restricted to be a 10th Order Polynomial

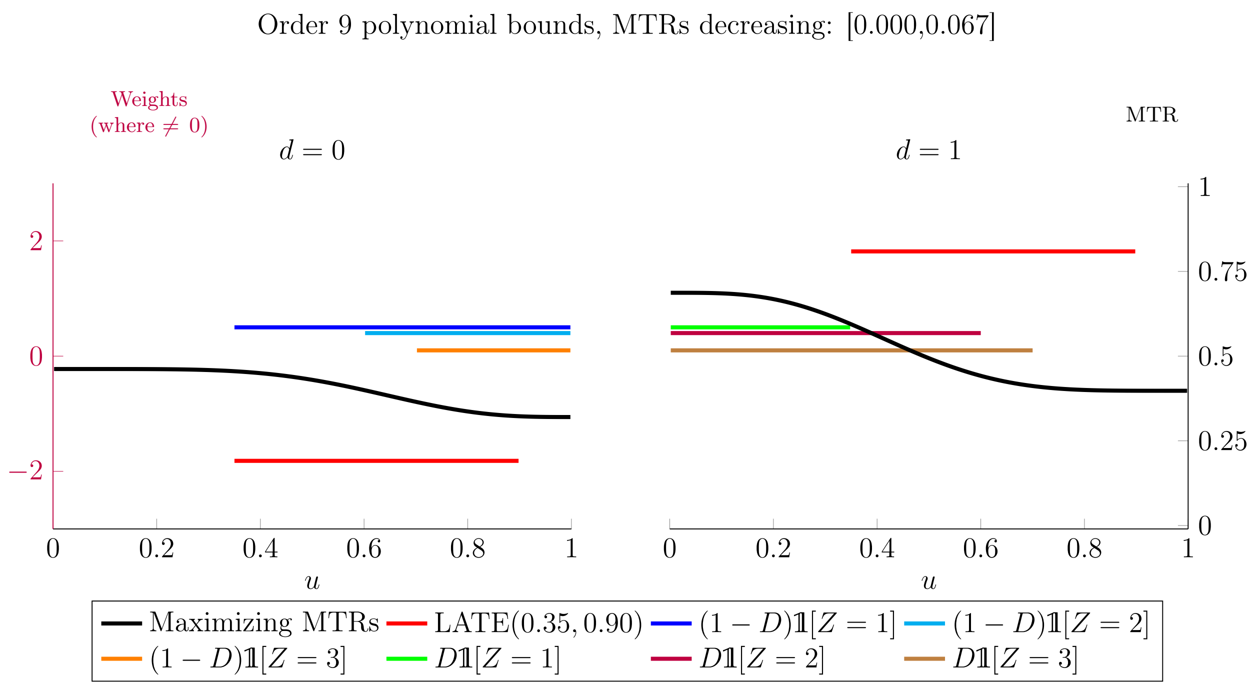
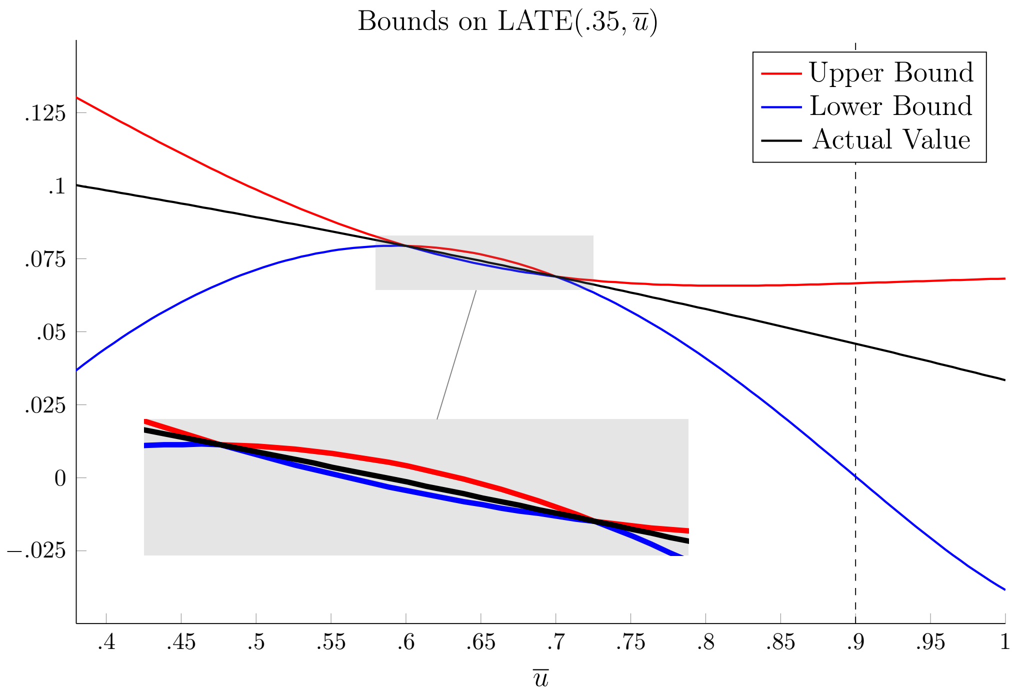


Figure 8: Bounds on a Family of PRTEs


Bounds on LATE(.35, $\overline{u}$)
Upper Bound
Lower Bound
Actual Value
$\overline{u}$