STRUCTURAL EQUATIONS, TREATMENT EFFECTS AND ECONOMETRIC POLICY EVALUATION[1]

By James J. Heckman and Edward Vytlacil

First Draft, August 2000; Revised, June 2001

Final Version, September 25, 2003.

[1]This paper was presented by Heckman as the Fisher-Schultz Lecture at the Eighth World Meetings of the Econometric Society, Seattle, Washington, August 13, 2000. This paper was also presented at the seminar on Applied Price Theory, at the Graduate School of Business, University of Chicago in October 2000 and at the Montreal Econometrics Seminar, September, 2003. We thank Jaap Abbring, Richard Blundell, Pedro Carneiro, Guido Imbens and two anonymous referees for very helpful comments. Pedro Carneiro also made several analytical suggestions. We benefited from the close reading of Ricardo Avelino, Fernanda Ruiz, Sergio Urzua and Weerachart Kilenthong on this draft. Sergio Urzua provided valuable research assistance in programming the simulations reported in this paper and was assisted by Hanna Lee. This research was supported by NSF 97-09-873, NSF-00-99195 and NICHD-40-403-000-85-261 and the American Bar Foundation.

Abstract

This paper uses the marginal treatment effect (MTE) to unify the nonparametric literature on treatment effects with the econometric literature on structural estimation using a nonparametric analog of a policy invariant parameter; to generate a variety of treatment effects from a common baseline functional form; to organize the literature on instrumental variable estimation, matching and control function estimation; to explore what policy questions commonly used estimators in the treatment effect literature answer; and to define and identify the policy relevant treatment effects for a class of interventions that affect treatment choice but not potential outcomes. A method for estimating the marginal treatment effect is developed. Two methods for estimating policy relevant treatment effects are presented. The analysis is extended to consider multiple treatments from an ordered choice model.

JEL:C1

Evaluating the impacts of public policies, forecasting their effects in new environments and predicting the effects of policies never tried are three central tasks of economics. The structural approach and the treatment effect approach are two competing paradigms of policy evaluation.

The structural approach emphasizes clearly articulated economic models that can be used to accomplish all three tasks under the exogeneity and policy invariance assumptions presented in that literature. (See Hansen and Sargent, 1981, Hendry, 1995.) Economic theory is used to guide the construction of models and to suggest included and excluded variables. The functional form and exogeneity assumptions invoked in this literature are sometimes controversial (see e.g. Angrist and Krueger, 1999) and the sources of identification of parameters of these models are often not clearly articulated.

The treatment effect literature as currently developed focuses on the first task-evaluating the impact of a policy in place in the special case when there is a "treatment group" and a "control group," i.e. a group of nonparticipants. In the language of that literature, "internal validity" is the primary goal and issues of forecasting out of sample or of evaluating new policies receive little attention.[2] Because of its more limited goals, fewer explicit functional form and exogeneity assumptions are invoked. The literature on treatment effects has given rise to a new language of economic policy analysis where the link to economic theory is often obscure and the economic policy questions being addressed are not always clearly stated. Different instruments—natural or unnatural—answer different economic questions that are not clearly stated. Relationships among the policy parameters implicitly defined by alternative choices of instruments are not articulated.

This paper unites the two approaches to policy evaluation under the assumption that analysts have access to treatment and control groups using the Marginal Treatment Effect (MTE). The MTE is the mean response of persons to treatment at a margin that is precisely defined in this paper. It is a willingness to pay measure when outcomes are values under alternative treatment regimes.

Under the conditions specified in this paper, the MTE can be used to construct and compare alternative conventional treatment effects, a new class of policy relevant treatment effects and the estimands produced from instrumental variable estimators and matching estimators. Using the MTE, this paper unites the selection (control function) approach, defined in a nonparametric setting, with the recent literature on instrumental variables.[3]

A major focus in the recent microeconomic policy evaluation literature, and a major theme of this paper, is on constructing and estimating models with heterogeneity in responses to treatment among

[2]Internal validity means that a treatment parameter defined in a specified environment is free of selection bias. It is defined more precisely below.

[3]Thus, as a byproduct, we clear up the confusion in the applied literature that often inappropriately contrasts these two estimation strategies. (See, e.g., Angrist and Krueger, 1999).

otherwise observationally identical people. This literature emphasizes that responses to treatment vary among observationally identical people, and crucially, that agents select (or are selected) into treatment at least in part based on their own idiosyncratic response to it. This emphasis is in marked contrast to the emphasis in the conventional representative-agent macro time series literature that ignores such heterogeneity despite the ample microeconometric evidence on it.[4]

Entire classes of econometric evaluation estimators can be organized by whether or not they allow for the possibility of selection based on unobserved components of heterogeneous responses to treatment. In the presence of such heterogeneity, a variety of different mean treatment effects can be defined for different instruments and conditioning sets. In the absence of such heterogeneity, these different treatment effects collapse to the same parameter.[5]

The dependence of estimated treatment parameters on instruments is an important and not widely understood feature of models with heterogeneous responses on which people act.[6] Instrument-dependent parameters arise in this class of models, something excluded by assumption in conventional structural econometric models that emphasize the estimation of invariant parameters. Two economists analyzing the same data set but using different valid instruments will estimate different parameters that have different economic interpretations. Even more remarkably, two economists using the same instrument but with different notions about what variables belong in choice equations will interpret the output of an instrumental variable analysis differently. Intuitions about "identifying strategies" acquired from analyzing conventional models where responses to treatment do not vary among persons are no longer valid in the more general setting analyzed in this paper. The choice of an instrument *defines* the treatment parameter being estimated. The relevant question regarding the choice of instrumental variables in the general class of models studied in this paper is "What parameter is being identified by the instrument?" not the traditional question of "What is the efficient combination of instruments for a fixed parameter?" - the question that has traditionally occupied the attention of econometricians who study instrumental variables.

The plan of this paper is as follows. Section 1 presents a prototypical microeconometric structural model as a benchmark against which to define and motivate the various treatment parameters used in the literature and to compare and contrast structural estimation approaches with those used in the literature on treatment effects. Section 2 extends the treatment effect literature by introducing choice theory into it and by using a weaker set of assumptions than those used in the structural literature to define and identify the marginal treatment effect (MTE). We show that the MTE can be used to generate and unify the

[4]Heckman (2001b) summarizes the evidence on heterogeneity in responses to treatment on which agents select into treatment.

[5]See Heckman (1997); Heckman and Robb (1985, 1986 reprinted 2000); Heckman and Vytlacil (1999).

[6]This dependence was first noted by Heckman and Robb (1985, p. 196). See also Angrist, Graddy and Imbens (2000).

various treatment parameters advocated in the recent literature and provides an economic foundation for the treatment effect literature. Section 3 uses the *MTE* to define policy relevant parameters that answer well posed economic questions. Evaluation of different policies requires different weights for the *MTE*. The *MTE* plays the role of a policy invariant structural parameter in conventional econometrics for a class of policy interventions defined in this paper.

Section 4 organizes entire classes of econometric estimators on the basis of what they assume about the role of unobservables in the *MTE* function, conditional on X. We focus on instrumental variables in this paper but we also consider matching.[7]

Section 5 returns to the policy evaluation problem. The treatment effect literature can be used to answer certain narrowly focused questions under weaker assumptions than are required to recover conventional structural parameters that answer a broad range of questions. When we attempt to address the broader set of questions entertained in the structural econometrics literature, additional conditions are required to extrapolate existing policies to new environments and to provide accurate forecasts of new policies never previously experienced. The weaker identifying assumptions invoked in the treatment effect literature are possible because of the narrower set of questions addressed by that literature. In the language of the treatment effect literature, internal validity (absence of selection bias) does not imply external validity (the ability to generalize). When the same policy forecasting questions addressed by the structural literature are asked of the treatment effect literature, the assumption sets used in the two literatures look very similar, especially for nonparametric versions of structural models. External validity requires stronger conditions. All of the analysis in the first five sections is for a two potential outcome model. We present models with multiple outcomes in Section 6. Section 7 concludes.

1 A Latent Variable Framework

The treatment effect literature investigates a class of policies that have partial coverage at a point in time so there is a "treatment" group and a "control" group. It is not helpful in evaluating policies that have universal participation.[8] Throughout this paper we follow the conventional practice in the literature and ignore general equilibrium effects.[9]

In order to link our discussion to the literature on structural econometrics, it is fruitful to compare how the two different approaches analyze a Generalized Roy Model for two potential outcomes (Y_0, Y_1). This

[7] These and other estimators are analyzed in greater depth in Heckman and Vytlacil (2004).

[8] The structural econometric literature substitutes functional form and support conditions for control groups. See Heckman and Vytlacil (2004).

[9] See, however, the studies by Heckman, Lochner and Taber (1998a,b; 1999) and Heckman (2001a,b), who demonstrate the empirical importance of investigating such effects in the context of evaluating the returns to schooling.

model is widely used in applied econometrics.[10] We extend our methods to multiple potential outcome models in Section 6.[11]

Write potential outcomes (Y_0, Y_1) for conditioning variables X as[12]

$$Y_0 = \mu_0(X) + U_0 \tag{1a}$$

and

$$Y_1 = \mu_1(X) + U_1. \tag{1b}$$

Let $D = 1$ denote the event that Y_1 is observed, while $D = 0$ denotes the event that Y_0 is observed, so the outcome Y is

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

Let

$$C = \mu_C(Z) + U_C \tag{1d}$$

denote the cost of receiving treatment. Net utility is

$$\begin{aligned} D^* &= Y_1 - Y_0 - \mu_C(Z) - U_C \\ D &= \mathbf{1}[D^* \geq 0)]. \end{aligned} \tag{2}$$

when $\mu_C(Z) = 0$ and $U_C = 0$. The Roy Model (1951) is the case. In a model of educational attainment, Y_1 is the present value of college earnings. Y_0 is the present value of earnings in the benchmark no-treatment state (e.g. high school). In a model of labor supply,[13] Y_1 is the market wage and Y_0 is the reservation wage. In this setting, Y_0 is usually not observed. More generally, if $\mathcal{I}$ is an information set, the decision to participate is based on $\mathcal{I}$ and $D = \mathbf{1}[D^* > 0]$ where D^* is some random variable measurable with respect to $\mathcal{I}$.[14] Tuition and family income operate through direct costs $\mu_C(Z)$ to determine college attendance; fixed costs and taxes affect the decision to work.

Conventional approaches used in the structural econometrics literature assume that $(X, Z) \perp\!\!\!\perp (U_0, U_1, U_C)$, where "$\perp\!\!\!\perp$" denotes independence. In addition, they adopt parametric assumptions about the distributions

[10] See Amemiya (1985).

[11] Florens, Heckman, Meghir and Vytlacil (2001) and Heckman and Vytlacil (2004) develop multiple outcome versions of the analysis of this paper more extensively.

[12] Throughout this paper, we denote random variables/random vectors by capital letters and potential realizations by the corresponding lower case letter. For example, X denotes the random vector, and x denotes a potential realization of the random vector X.

[13] See Heckman (1974).

[14] For example, $D^* = E(Y_1 - Y_0 - C|\mathcal{I})$.

of the error terms and functional forms of the estimating equations, and identify the full model that can then be used to construct a variety of policy counterfactuals. The most commonly used specification of this model writes $\mu_0(X) = X\beta_0$, $\mu_1(X) = X\beta_1$, $\mu_C(Z) = Z\beta_C$ and assumes $(U_0, U_1, U_C) \sim N(0, \Sigma)$. This is the normal selection model. (Heckman, 1976).

The parametric normal framework can be used to answer all three policy evaluation questions. It can be used to evaluate existing policies by asking how policy-induced changes in X or Z affect (Y, D). It can be used to extrapolate old policies to new environments by computing outcomes for the values of X, Z that characterize the new environment. Linearity and distributional assumptions make extrapolation straightforward. This framework can be used to evaluate new policies if they can be expressed as some known functions of (X, Z). For example, consider the effect of charging tuition in an environment where tuition has never before been charged. If tuition can be put on the same footing as (made comparable with) another measure of cost that is measured and varies, or with returns that can be measured and vary, then we can use the estimated response to the variation in observed costs or returns to estimate the response to the new tuition policy.[15]

This paper relaxes the functional form and distributional assumptions used in the structural literature and still identifies an economically interpretable model that can be used for policy analysis. Recent semiparametric approaches relax both distributional and functional form assumptions of selection models, but typically maintain exogeneity of X assumptions (see, e.g., Powell, 1994) and do not estimate treatment effects except through limit arguments (Heckman, 1990; Andrews and Schafgans, 1998).[16] The treatment effect literature seeks to bypass the *ad hoc* assumptions used in the structural literature and estimate treatment effects under weaker conditions. The goal of this literature is to examine the effects of policies in place (i.e. to produce internally valid estimators) rather than to forecast new policies or old policies on new populations.

2 Treatment Effects

The model of treatment effects developed in Heckman and Vytlacil (1999, 2000), relaxes most of the controversial assumptions discussed in Section 1. It is a nonparametric selection model with testable restrictions that can be used to unify the treatment effect literature, identify different treatment effects, link the literature on treatment effects to the literature in structural econometrics and interpret the implicit

[15] In a present value income maximizing model of schooling, costs and returns are on the same footing so knowledge of how schooling responds to returns is enough to determine how schooling responds to costs. See Section 5.2.

[16] A large part of the literature is concerned with estimation of slope coefficients (e.g., Ahn and Powell, 1993) and not the counterfactuals needed for policy analysis. Heckman (1990) discusses the more demanding conditions required to identify counterfactuals.

economic assumptions underlying instrumental variables and matching methods.

We use the general framework of Section 1, equations (1a)–(1d), and define Y as the measured outcome variable. We do not impose any assumption on the support of the distribution of Y. We use the more general nonlinear and nonseparable outcome model

$$Y_1 = \mu_1(X, U_1) \tag{3a}$$

$$Y_0 = \mu_0(X, U_0).^{17} \tag{3b}$$

The individual treatment effect associated with moving an otherwise identical person from "0" to "1" is $Y_1 - Y_0 = \Delta$ and is defined as the effect on Y of a *ceteris paribus* move from "0" to "1". These *ceteris paribus* effects are called "causal effects." To link this framework to the literature on structural econometrics, we characterize the decision rule for program participation by an index model:

$$D^* = \mu_D(Z) - U_D\,; \qquad D = 1 \quad \text{if} \quad D^* \geq 0\,; \qquad D = 0 \quad \text{otherwise}, \tag{4}$$

where (Z, X) is observed and (U_1, U_0, U_D) is unobserved.[18] U_D may be a function of (U_0, U_1).[19] Without loss of generality, Z includes all of the elements of X. However, our analysis requires that Z contain at least one element not in X. The following assumptions are weaker than those used in the literature on structural econometrics or the recent literature on semiparametric selection models (see Powell, 1994) and at the same time can be used to both define and identify different treatment parameters.[20] The assumptions are:

(A-1) $\mu_D(Z)$ is a nondegenerate random variable conditional on X;

(A-2) (U_1, U_D) and (U_0, U_D) are independent of Z conditional on X;[21]

(A-3) The distribution of U_D is absolutely continuous with respect to Lebesgue measure;

(A-4) $E|Y_1|$ and $E|Y_0| < \infty$; and

(A-5) $1 > \Pr(D = 1 \mid X) > 0$.

[17] Examples include conventional latent variable models: $Y_i = 1$ if $Y_i^* = \mu_i(X) + U_i \geq 0$ and $Y_i = 0$ otherwise; $i = 0, 1$. Notice that in the general case, $E(Y_i \mid X) - \mu_i(X, U_i) \neq U_i, i = 0, 1$ so even if the μ_i are structural, the $E(Y_i \mid X)$ are not.

[18] The model for D imposes restrictions on counterfactual choices, and our analysis exploits these restrictions. See Vytlacil (2002) for an analysis of the restrictions on counterfactuals imposed by this model, and see Heckman and Vytlacil (2000) for an analysis of the role of this assumption in treatment effect analysis.

[19] In the Roy Model $U_D = U_1 - U_0$ in the notation of Section 1. In the Generalized Roy Model, $U_D = U_1 - U_0 - U_C$.

[20] As noted in Section 2.1, a much weaker set of conditions is required to define the parameters than is required to identify them. As noted in Section 5, stronger conditions are required for policy forecasting.

[21] Heckman and Vytlacil (1999, 2000) assume (U_1, U_D) and (U_0, U_D) are independent of (Z, X) but this assumption is only made for notational convenience and is easily relaxed.

Assumptions (A-1) and (A-2) are "instrumental variable" assumptions that there is at least one variable that determines participation in the program and that is independent of potential outcomes given X. These are also the assumptions used in the natural and social experiment literatures. (A-2) also assumes that U_D is independent of Z given X and is used to generate counterfactuals. Assumption (A-3) is a technical assumption made primarily for expositional convenience. Assumption (A-4) guarantees that the conventional treatment parameters are well defined. Assumption (A-5) is the assumption in the population of both a treatment and a control group for each X. Observe that there are no exogeneity requirements for X. This is in contrast with the assumptions made in the conventional structural literature and the semiparametric selection literature (see, e.g. Powell, 1994). A counterfactual "no feedback" condition facilitates interpretability so that conditioning on X does not mask the effects of D . Letting X_d denote a value of X if D set to d, a sufficient condition that rules out feedback from D to X is:

(A-6) $X_1 = X_0$ *a.e.*

Condition (A-6) is not strictly required to formulate an evaluation model, but it enables an analyst who conditions on X to capture the "total" or "full effect" of D.[22] In this paper, we examine treatment effects conditional on X, and thus we maintain assumption (A-6) in this paper.

To satisfy (A-1) for a nonparametric $\mu_D(Z)$, we need some variable in Z that is not in X. Define $P(Z)$ as the probability of receiving treatment given Z: $P(Z) \equiv \Pr(D = 1 \mid Z) = F_{U_D|X}(\mu_D(Z))$, where $F_{U_D|X}(\cdot)$ denotes the distribution of U_D conditional on X.[23] We often denote $P(Z)$ by P, suppressing the Z argument. As a normalization, we may always impose $U_D \sim Unif[0,1]$ and $\mu_D(Z) = P(Z)$ because if the latent variable generating choices is $D^* = \nu(Z) - V$, where V is a general continuous random variable we can apply a probability transform to reparameterize the model so that $\mu_D(Z) = F_{V|X}(\nu(Z))$ and $U_D = F_{V|X}(V)$.[24]

Vytlacil (2002) establishes that assumptions (A-1)–(A-5) for selection model (3a), (3b) and (4) are equivalent to the assumptions used to generate the *LATE* model of Imbens and Angrist (1994). Thus the nonparametric selection model for treatment effects developed in this paper is equivalent to an influential instrumental variable model for treatment effects. Our latent variable model satisfies their assumptions and their assumptions generate our latent variable model.

[22] See Pearl (2000).

[23] Throughout this paper, we will refer to cumulative distribution function of a random vector A by $F_A(\cdot)$ and to the cumulative distribution function of a random vector A conditional on random vector B by $F_{A|B}(\cdot)$. We will write the cumulative distribution function of A conditional on $B = b$ by $F_{A|B}(\cdot \mid b)$.

[24] This representation is valid whether or not (A-2) is true. However, (A-2) imposes restrictions on counterfactual choices. For example, if a change in government policy changes the distribution of Z by an external manipulation, under (A-2) the model can be used to generate the choice probability from $P(z)$ evaluated at the new arguments i.e., the model is invariant with respect to the distribution Z.

Our model and assumptions (A-1)–(A-5) impose two testable restrictions on the distribution of (Y, D, Z, X). The model imposes an index sufficiency restriction: for any measurable set $\mathcal{A}$ and for $j = 0, 1, \Pr(Y_j \in \mathcal{A} \mid Z, D = j) = \Pr(Y_j \in \mathcal{A} \mid P(Z), D = j)$. The model also imposes a monotonicity restriction. For $j = 0, 1$, let $g_0(Y_0, X)$ and $g_1(Y_1, X) \geq 0$ with probability 1. Then under our assumptions $E[(1-D)g_0(Y, X) \mid X, P(Z) = p]$ is weakly decreasing in p and $E[Dg_1(Y, X) \mid X, P(Z) = p]$ is weakly increasing in p.[25] For example, if Y_1 and Y_0 are known to be nonnegative w.p.1, then choosing $g_j(Y, X) = Y$ results in the monotonicity restriction that $E[(1-D)Y \mid X, P(Z) = p]$ is weakly decreasing in p and $E[DY \mid X, P(Z) = p]$ is weakly increasing in p. As another example, without any assumptions on the support of the distributions of Y_0 and Y_1, we can choose $g_j(Y, X) = \mathbf{1}[Y \leq t]$, implying the testable restriction that $\Pr(D = 0, Y \leq t \mid X, P(Z) = p)$ is weakly decreasing in p and $\Pr(D = 1, Y \leq t \mid X, P(Z) = p)$ is weakly increasing in p.[26]

The model of treatment effects presented in this paper is not the most general possible model because it has testable implications and hence empirical content. It unites various literatures and produces a nonparametric version of the widely used selection model. It links the treatment literature to economics.

2.1 Definitions of Treatment Effects

The difficulty of observing the same individual in both treatment states leads to the use of various population level treatment effects widely used in the biostatistics literature and applied in economics.[27] The most commonly invoked treatment effect is the Average Treatment Effect (ATE): $\Delta^{ATE}(x) \equiv E(\Delta \mid X = x)$ where $\Delta = Y_1 - Y_0$. This is the effect of assigning treatment randomly to everyone of type X assuming full compliance, and ignoring general equilibrium effects. The impact of treatment on persons who actually take the treatment is Treatment on the Treated (TT): $\Delta^{TT}(x) \equiv E(\Delta | X = x, D = 1)$, which can also be defined conditional on $P(Z)$: $\Delta^{TT}(x, p) \equiv E(\Delta | X = x, P(Z) = p, D = 1)$.[28]

The mean effect of treatment on those for whom $X = x$ and $U_D = u_D$, the Marginal Treatment Effect (MTE), plays a fundamental role in our analysis.

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta | X = x, U_D = u_D). \tag{5}$$

[25] The monotonicity restriction is stated under the condition that $g_0(Y, X)$ and $g_1(Y, X)$ are nonnegative w.p.1. If the condition is strengthened to be that g_0 and g_1 are strictly positive w.p.1, then the result is strengthened to be that $E[(1-D)g_0(Y, X) \mid X, P(Z) = p]$ is strictly decreasing in p and $E[Dg_1(Y, X) \mid X, P(Z) = p]$ is strictly increasing in p.

[26] In Appendix A, we formally state and prove the monotonicity restriction, give more examples of how g_1 and g_0 might be chosen, and show that this restriction includes the restriction on IV tested by Imbens and Rubin (1997) as a special case by taking Z to be binary and by making an appropriate choice of g_0 and g_1 functions.

[27] Heckman, LaLonde and Smith (1999) discussed panel data cases where it is possible to observe both Y_0 and Y_1.

[28] The two are related in a simple way: $\Delta^{TT}(x) = \int_0^1 \Delta^{TT}(x, p) dF_{P(Z)|X,D}(p|x, 1)$ where $F_{P(Z)|X,D}(\cdot|x, 1)$ is the distribution of $P(Z)$ given $X = x$ and $D = 1$.

As a consequence of (A-2), if $u_D = \mu_D(Z)$, $\Delta^{MTE}(x, u_D)$ equals $E(\Delta|X = x,\, U_D = u_D, \mu_D(Z) = u_D)$, which is the mean gain measured in terms of $Y_1 - Y_0$ for persons with observed characteristics X at the margin of indifference at $U_D = u_D = \mu_D(Z)$. When Y_1 and Y_0 are value outcomes, MTE is a mean willingness level of utility $\mu_D(Z)$. MTE is a choice-theoretic building block that unites the treatment effect, selection and matching literatures.

The $LATE$ parameter of Imbens and Angrist (1994) is a version of MTE. Define D_z as a counterfactual choice variable, with $D_z = 1$ if D would have been chosen if Z had been set to z, and $D_z = 0$ otherwise. Let $\mathcal{Z}(x)$ denote the support of the distribution of $P(Z)$ conditional on $X = x$. For any $(z, z') \in \mathcal{Z}(x) \times \mathcal{Z}(x)$ s.t. $P(z) > P(z')$, $LATE$ is $E(\Delta|X = x, D_z = 1, D_{z'} = 0) = E(Y_1 - Y_0|X = x, D_z = 1, D_{z'} = 0)$, the mean gain to persons who would be induced to switch from $D = 0$ to $D = 1$ if Z were manipulated externally from z' to z. It follows from the latent index model that

$$E(Y_1 - Y_0|X = x, D_z = 1, D_{z'} = 0) = E(Y_1 - Y_0|X = x, u'_D \leq U_D < u_D) = \Delta^{LATE}(x, u_D, u'_D)$$

for $u_D = \Pr(D_z = 1) = P(z)$, $u'_D = \Pr(D_{z'} = 1) = P(z')$.[29] Imbens and Angrist define the $LATE$ parameter as an estimand. Their analysis conflates issues of definition of parameters with issues of identification. Our representation of $LATE$ allows us to separate these two issues and to define the $LATE$ parameter more generally, since one can imagine evaluating the right hand side of this equation at any u_D, u'_D points in the unit interval and not only at points in the support of the distribution of the propensity score $P(Z)$ conditional on $X = x$ where it is identified. Notice that $\lim_{u'_D \uparrow u_D} \Delta^{LATE}(x, u_D, u'_D) = \Delta^{MTE}(x, u_D)$. [30] As a consequence of the index structure, the following relationships among the various treatment effect parameters can be established.[31]

$$\Delta^{ATE}(x) = \int_0^1 \Delta^{MTE}(x, u)\, du,$$

$$\Delta^{TT}(x, p) = \frac{1}{p}\int_0^p \Delta^{MTE}(x, u)\, du, \text{ and}$$

$$\Delta^{LATE}(x, u_D, u'_D) = \left[\int_{u'_D}^{u_D} \Delta^{MTE}(x, u)\, du\right] \frac{1}{u_D - u'_D}.$$

$\Delta^{TT}(x)$ is a weighted average of Δ^{MTE}:

$$\Delta^{TT}(x) = \int_0^1 \Delta^{MTE}(x, u_D)\, h_{TT}(x, u_D)\, du_D,$$

[29] Assumption (A-2) implies that $\Pr(D_z = 1) = \Pr(D = 1|Z = z)$, $\Pr(D_{z'} = 1) = \Pr(D = 1|Z = z')$

[30] Given our assumptions (A-2), (A-3), and (A-4), one can apply Lebesgue's theorem for the derivative of an integral to show that $\Delta^{LATE}(x, u_D, u'_D)$ is continuous in u_D and u'_D outside of a set of Lebesgue measure zero.

[31] These relationships were first presented in Heckman and Vytlacil (1999).

where

$$h_{TT}(x, u_D) = \frac{1 - F_{P|X}(u_D \mid x)}{\int_0^1 \left(1 - F_{P|X}(t \mid x)\, dt\right)} = \frac{S_{P|X}(u_D \mid x)}{E\left(P(Z) \mid X = x\right)}, \tag{6}$$

and $S_{P|X}(u_D \mid x)$ is the survivor function for the distribution of $P(Z)$ conditional on $X = x$ ($\Pr(P(Z) > u_D \mid X = x)$) and $h_{TT}(x, u_D)$ is a weighted distribution.[32] $\Delta^{TT}(x)$ oversamples $\Delta^{MTE}(x, u_D)$ for those with low values of u_D that make them more likely to participate in the program being evaluated. These results are summarized in Table IA. The various weights are collected in Table IB. The other weights, treatment effects and estimands shown in this table are discussed later.

Observe that if $E(\Delta \mid X = x, U_D = u_D) = E(\Delta \mid X = x)$, so Δ is mean independent of U_D given $X = x$, then $\Delta^{MTE} = \Delta^{ATE} = \Delta^{TT} = \Delta^{LATE}$. Therefore in cases where there is no heterogeneity (Δ constant conditional on $X = x$) or agents do not act on it (mean independence), so that marginal treatment effects are average treatment effects, all of the evaluation parameters are the same. Otherwise, they are different. Only in the case where the marginal treatment effect is the average treatment effect will the "effect" of treatment be uniquely defined.

Figure 1A plots weights for the parametric normal Generalized Roy Model generated from parameters shown at the base of the figure. A high u_D is associated with higher cost, relative to return, and less likelihood of choosing $D = 1$. The decline of MTE in terms of higher values of u_D means that people with progressively higher u_D have lower gross returns. TT overweights low values of u_D (i.e., it oversamples U_D that make it likely to have $D = 1$). ATE samples U_D uniformly. Treatment on the Untreated ($E(Y_1 - Y_0 \mid X = x, D = 0)$) or TUT, oversamples the values of U_D unlikely to have $D = 1$.

Table II shows the treatment parameters produced from the different weighting schemes. Given the decline of the MTE in u_D, it is not surprising that $TT > ATE > TUT$. The difference between TT and ATE is a sorting gain: $E(Y_1 - Y_0 \mid X, D = 1) - E(Y_1 - Y_0 \mid X)$, the average gain experienced by people who sort into treatment compared to what the average person would experience. Purposive selection on the basis of gains should lead to positive sorting gains of the sort found in the table. The other numbers in the table are discussed later.

Heckman (2001a) presents evidence on the nonconstancy of the MTE drawn from a variety of studies of schooling, job training, migration and unionism. With the exception of studies of unionism, a common finding in the empirical literature is the nonconstancy of MTE given X.[33] The evidence from the literature suggests that different treatment parameters measure different effects, and persons participate in programs based on heterogeneity in responses to the program being studied. The phenomenon of nonconstancy of the MTE that we analyze in this paper is of substantial empirical interest.

[32] See Heckman and Vytlacil (2000) for its derivation.

[33] However, most of the empirical evidence is based on parametric models.

Assumptions (A-1)–(A-5) are far stronger than what is required to define the parameters and establish relationships among them. Defining $\Delta^{MTE}(x, u_D)$ as in (5), the treatment effect parameters are defined and the relationships among them remain valid even if Z is not independent of U_D, if there are no variables in Z which are not also contained in X, or if there is no additively separable version of the latent index.[34] Thus the various evaluation parameters can be defined, and the relationships among parameters presented in this section remain valid, even if Z is not a valid instrument and even if the choice equation is not additively separable in U_D and Z. Assumptions (A-1)–(A-5) will be used to interpret the instrumental variables estimand, to define policy relevant treatment effects, and to relate instrumental variables to the policy relevant treatment effects. They are also sufficient to identify $\Delta^{MTE}(x, u_D)$ at any u_D evaluation point that is a limit point of the support of the distribution of $P(Z)$ conditional on $X = x$.[35]

The literature on structural econometrics is clear about the basic parameters of interest although it is not always clear about the exact combinations of parameters needed to answer specific policy problems.[36] The literature on treatment effects offers a variety of evaluation parameters. Missing from that literature is an algorithm for defining a treatment effect that answers a precisely formulated policy question. The *MTE* provides a framework for developing such an algorithm which we now develop.

3 Policy Relevant Treatment Parameters

The conventional treatment parameters do not always answer the economically interesting questions. For example, their link to cost benefit analysis and interpretable economic frameworks is often obscure.[37] Ignoring general equilibrium effects, Δ^{TT} is one ingredient for determining whether or not a given program should be shut down or retained. It is informative on the question of whether the persons participating in

[34] In other words, the parameters can be defined if $D = \mathbf{1}[\Omega(Z, U_D) \geq 0]$ without additive separability of Ω in terms of U_D and Z. Recall that U_D is defined in a model for D that is valid for counterfactual choices. The assumptions of additive separability between U_D and Z in the latent index and the assumption that $U_D \perp\!\!\!\perp Z|X$ are not innocuous, as these assumptions impose restrictions on counterfactual choices even though they impose no restrictions on the observed distribution of (D, Z) (Vytlacil, 2002). Heckman and Vytlacil (2000) define $\Delta^{MTE}(x, u_D)$ as $E(Y_1 - Y_0|X = x, U_D = u_D)$ for the case where $Z \not\perp\!\!\!\perp U_D|X$. If $Z \not\perp\!\!\!\perp U_D|X$, then the interpretation of $\Delta^{MTE}(x, u_D)$ as the mean value of $Y_1 - Y_0$ for those indifferent at $U_D = u_D$ and $\mu_D(Z) = u_D$ is incorrect. Thus, the economic interpretation of the parameter is altered. For further discussion of these issues, see Heckman and Vytlacil (2000,2004).

[35] For example, if we additionally impose that the distribution of $P(Z)$ conditional on X has a density with respect to Lebesgue measure, then (A-1)–(A-5) will enable us to identify $\Delta^{MTE}(x, u_D)$ at all (x, u_D) such that x is in the support of the distribution of X and u_D is in the support of the distribution of $P(Z)$ conditional on $X = x$.

[36] In a fundamental paper, Marschak (1953) shows how different combinations of structural parameters are required to forecast the impacts of different policies. It is possible to answer many policy questions without identifying any of the structural parameters individually. The treatment effect literature partially embodies this vision, but typically does not define the economic question being answered, in contrast to Marschak's approach. See Heckman (2001a) and Heckman and Vytlacil (2004).

[37] Heckman (1997), Heckman and Smith (1998) and Heckman and Vytlacil (2004) develop the relationship between these parameters and the requirements of cost benefit analysis.

a program benefit from it in gross terms.[38] Δ^{MTE} estimates the gross gain from a marginal expansion of a program.

A more promising approach to defining parameters is to postulate a policy question or decision problem of interest and to derive the treatment parameter that answers it. Taking this approach does not in general produce the conventional treatment parameters or the estimands produced from instrumental variables, a point we illustrate below. Consider a class of policies that affect P, the probability of participation in a program, but do not affect Δ^{MTE}. We characterize this class of policies more precisely in Section 5. An example from the schooling literature would be policies that change tuition or distance to school but do not directly affect the gross returns to schooling.[39]

The policies analyzed in the treatment effect literature that change the Z not in X are more restrictive than the general policies that shift X and Z analyzed in the structural literature. Let a and a' denote two potential policies and let D_a and $D_{a'}$ denote the counterfactual choices that would be made under policies a and a'. Let the corresponding decision rules be $D_a = \mathbf{1}[P_a(Z_a) \geq U_D]$, $D_{a'} = \mathbf{1}[P_{a'}(Z_{a'}) \geq U_D]$, where $P_a(Z_a) = \Pr(D_a = 1|Z_a)$ and $P_{a'}(Z_{a'}) = \Pr(D_{a'} = 1|Z_{a'})$. For ease of exposition, we will suppress the arguments of these functions and write P_a and $P_{a'}$ for $P_a(Z_a)$ and $P_{a'}(Z_{a'})$. We assume that the policy does not change (Y_0, Y_1, X, U_D), and that Z_a and $Z_{a'}$ are independent of (Y_0, Y_1, U_D) conditional on X.

Using a Benthamite social welfare criterion as a prototype widely used in applied work, and comparing policies using mean outcomes, and considering the effect for individuals with a given level of $X = x$, we obtain, defining $E_a(Y \mid X = x)$ as $E(Y \mid X = x, \text{ under policy } a)$, the policy relevant treatment effect, *PRTE*, denoted $\Delta^{PRTE}(x)$:

$$E_a(Y \mid X = x) - E_{a'}(Y|X = x) = \int_0^1 \Delta^{MTE}(x, u_D)\{F_{P_{a'}|X}(u_D \mid x) - F_{P_a|X}(u_D \mid x)\}\, du_D, \tag{7}$$

where $F_{P_a|X}(\cdot \mid x)$ and $F_{P_{a'}|X}(\cdot \mid x)$ are the distributions of P_a and $P_{a'}$ conditional on $X = x$, respectively, defined for the different policy regimes. These weights apply to the entire population. This can also be defined for a general criterion $V(Y)$ where V is an evaluation function. The derivation of these weights is given in Appendix B.

Define $\Delta\bar{P}(x) = E(P_{a'}|X = x) - E(P_a|X = x)$, the change in the proportion of people induced into the program due to the intervention. Assuming $\Delta\bar{P}(x)$ is positive, we may define *per person affected weights* as $h_{PRT}(x, u_D) = \frac{F_{P_{a'}|X}(u_D|x) - F_{P_a|X}(u_D|x)}{\Delta\bar{P}(x)}$. As demonstrated in the next section, in general, conventional *IV* weights Δ^{MTE} differently than either the conventional treatment parameters (Δ^{ATE} or Δ^{TT}) or the policy

[38] It is necessary to account for costs to conduct a proper cost benefit analysis.

[39] Recall that we abstract from general equilibrium effects in this paper so effects on (Y_0, Y_1) from changes in the level of education are ignored.

relevant parameters, and so does not recover these parameters.

Instead of hoping that conventional treatment parameters or favorite estimators identify interesting economic questions, the approach developed in this paper is to estimate Δ^{MTE} and weight it by the appropriate weight determined by how the policy changes the distribution of P to construct $\Delta^{PRTE}(x)$. An alternative approach is to produce a policy weighted instrument to identify Δ^{PRTE} by standard instrumental variables. We develop both approaches in the next section. Before doing so, we first consider what conventional IV estimates and conditions for identifying Δ^{MTE}. We also consider matching methods and *OLS*.

4 Instrumental Variables, Local Instrumental Variables, *OLS* and Matching

In this section, we use Δ^{MTE} to organize the literature on econometric evaluation estimators. We focus primarily on instrumental variable estimators but also briefly consider matching. We develop the method of local instrumental variables. Well established intuitions about instrumental variable identification strategies break down when Δ^{MTE} is nonconstant in u_D given X. Two sets of instrumental variable conditions are presented in the current literature for this more general case: those associated with conventional instrumental variable assumptions which are implied by the assumption of "no selection on heterogenous gains" and those which permit selection on heterogeneous gains. Neither set implies the other, nor does either identify the policy relevant treatment effect in the general case. Each set of conditions identifies different treatment parameters.

In place of standard instrumental variables methods, we advocate a new approach to estimating policy impacts by estimating Δ^{MTE} using local instrumental variables (LIV) to identify all of the treatment parameters from a generator Δ^{MTE}. Δ^{MTE} can be weighted in different ways to answer different policy questions. For certain classes of policy interventions discussed in Section 5, Δ^{MTE} possesses an invariance property analogous to the invariant parameters of traditional structural econometrics. We also develop a new instrumental variable procedure that directly estimates $\Delta^{PRTE}(x)$.

4.1 Conventional Instrumental Variables

In the general case with a nonconstant (in u_D) Δ^{MTE}, linear IV does not estimate any of the treatment effects previously defined. Let $J(Z)$ denote a potential instrument. We sometimes denote $J(Z)$ by J, leaving implicit that J is a function of Z. Conditions $J(Z) \not\perp\!\!\!\perp (U_1, U_0)$, and $Cov(J(Z), D) \neq 0$ do not, by

themselves, identify conventional or policy-relevant treatment effects. We must supplement the standard conditions to identify interpretable parameters. To link our analysis to conventional analyses of IV, we assume additive separability of outcomes in terms of (U_1, U_0) so $Y_1 = \mu_1(X) + U_1$ and $Y_0 = \mu_0(X) + U_0$, but this is not strictly required.[40]

The two sets of instrumental variable conditions in the literature are due to Heckman and Robb (1985, 1986) and Heckman (1997) and Imbens and Angrist (1994). In the case of a nonconstant (in u_D) Δ^{MTE}, linear IV estimates different parameters depending on which assumptions are maintained. To establish this point, it is useful to briefly review the IV method in the case of a common treatment effect conditional on X, where $Y_1 - Y_0 = \Delta$, with Δ a deterministic function of X, and we assume additive separability in outcomes. Using (1a) and (1b) with $U_1 = U_0 = U$, and assuming $E(U|X) = 0$, we may write $Y = \mu_0(X) + D\Delta + U$ where $\Delta = \mu_1(X) - \mu_0(X)$. $Z \perp\!\!\!\perp U|X$ implies $E(UJ(Z)|X) = 0$. The standard instrumental variables intuition is that when $E(UJ|X) = 0$ and $Cov\,(J, D|X) \neq 0$, linear IV identifies Δ:

$$\frac{Cov\,(J, Y|X)}{Cov\,(J, D|X)} = \Delta = \mu_1(X) - \mu_0(X).$$

These intuitions break down in the heterogeneous response case where the outcomes are generated by different unobservables $(U_0 \neq U_1)$ so $Y = \mu_0(X) + D\Delta + U_0$,where $\Delta = \mu_1(X) - \mu_0(X) + U_1 - U_0$. This is a variable response model. There are two important cases of the variable response model. The first case arises when responses are heterogeneous, but conditional on X, people do not base their participation on these responses. In this case

(C-1) $D \perp\!\!\!\perp \Delta \mid X \Longrightarrow E(\Delta \mid X, U_D = u_D) = E(\Delta \mid X)$, Δ^{MTE} is constant in $U_D = u_D$ given X and $\Delta^{MTE} = \Delta^{ATE} = \Delta^{TT} = \Delta^{LATE}$.

The second case arises when

(C-2) $D \not\perp\!\!\!\perp \Delta|X$ and $E(\Delta \mid X, U_D = u_D) \neq E(\Delta \mid X)$.

In this case the Δ^{MTE} is nonconstant and the treatment parameters differ among each other.

Application of the standard IV equation to the general variable coefficient model produces

$$\frac{Cov\,(J, Y|X)}{Cov\,(J, D|X)} = \frac{Cov\,(J, D\Delta|X)}{Cov\,(J, D|X)}$$

[40] It is important to note that all derivations and results in this section hold without any additive separability assumption if $\mu_1(x)$ and $\mu_0(x)$ are replaced by $E(Y_1|X = x)$ and $E(Y_1|X = x)$, respectively, and U_1 and U_0 are replaced by $Y_1 - E(Y_1|X = x)$ and $Y_0 - E(Y_0|X = x)$, respectively.

where $J = J(Z)$. Under additive separability we obtain

$$\frac{Cov\,(J, Y|X)}{Cov\,(J, D|X)} = \mu_1(X) - \mu_0(X) + \frac{Cov\,(J, D(U_1 - U_0)|X)}{Cov\,(J, D|X)}.$$

Knowledge of (X, Z, D) and $(X, Z, (U_0, U_1))$ dependencies are not enough to determine the covariance in the second term. We need to know joint (X, Z, D, U_0, U_1) dependencies.

A sufficient condition for satisfying (C-1) is the strong information condition that decisions to participate in the program are not made on the basis of $U_1 - U_0$:

(I-1) $\Pr(D = 1 \mid Z, X, U_1 - U_0) = \Pr(D = 1 \mid Z, X)$.

Given our assumption that $(U_1 - U_0)$ is independent of Z given X, one can use Bayes' Theorem to show that (I-1) implies the weaker mean independence condition:

(I-2) $E(U_1 - U_0 \mid Z, X, D = 1) = E(U_1 - U_0|X, D = 1)$

which is generically necessary and sufficient for linear IV to identify Δ^{TT} and Δ^{ATE}.

Case (C-2) is inconsistent with (I-2). IV estimates Δ^{LATE} under the conditions of Imbens and Angrist (1994). Δ^{LATE}, selection models, and LIV, introduced below, analyze the more general case (C-2).[41] Different assumptions define different parameters, and the treatment effect literature is balkanized. Not only are there many different parameters but, as we establish in Section 4.3, different instruments define different parameters and traditional intuitions about instrumental variables break down.

4.2 Estimating The *MTE* Using Local Instrumental Variables

Heckman and Vytlacil (1999; 2000) resolve the confusion in the instrumental variables estimation literature using the Local Instrumental Variable (LIV) estimator to recover Δ^{MTE} pointwise. Conditional on X, LIV is the derivative of the conditional expectation of Y with respect to $P(Z) = p$:

$$\Delta^{LIV}(X, p) \equiv \frac{\partial E(Y \mid X, P(Z) = p)}{\partial p}. \tag{8}$$

$E(Y_1 - Y_0|X, P(Z))$ exists $(a.e.)$ by assumption (A-4), and $E(Y|X, P(Z))$ can be recovered over the support of $(X, P(Z))$. (A-2), (A-3) and (A-4) jointly allow one to use Lebesgue's theorem for the derivative of an integral to show that $E(Y_1 - Y_0|X, P(Z) = p)$ is differentiable in p, and thus we can recover $\frac{\partial}{\partial p}E(Y \mid X, P(Z) = p)$ for almost all p that are limit points of the support of distribution of $P(Z)$ conditional on X.

[41] Heckman and Vytlacil (2004) discuss these conditions in greater detail.

For example, if the distribution of $P(Z)$ conditional on X has a density with respect to Lebesgue measure, then all points in the support of the distribution of $P(Z)$ conditional on X are limit points of that support and we can identify $\Delta^{LIV}(X,p) = \dfrac{\partial E(Y \mid X, P(Z)=p)}{\partial p}$ for $(a.e.)$ p.

Under our assumptions, LIV identifies MTE for all limit points in the support of the distribution of $P(Z)$ conditional on X:

$$\frac{\partial E(Y \mid X, P(Z)=p)}{\partial p} = \Delta^{MTE}(X,p) = E(Y_1 - Y_0 \mid X, U_D = p).^{42}$$

This expression does not require additive separability of $\mu_1(X,U_1)$ or $\mu_0(X,U_0)$.

Under standard regularity conditions, a variety of nonparametric methods can be used to estimate the derivative of $E(Y \mid X, P(Z))$ and thus to estimate Δ^{MTE}. With Δ^{MTE} in hand, if the support of the distribution of $P(Z)$ conditional on X is the full unit interval, one can generate all the treatment parameters defined in Section 2 as well as the policy relevant treatment parameter presented in Section 3 as weighted versions of Δ^{MTE}. When the support of the distribution of $P(Z)$ conditional on X is not full, it is still possible to identify some parameters.[43] Heckman and Vytlacil (2000, 2001, 2004) construct sharp bounds on the the treatment parameters under the same assumptions imposed in this paper without imposing support conditions. The resulting bounds are simple and easy to apply compared with those presented in the previous literature.

In order to establish the relationship between LIV and ordinary IV based on $P(Z)$ and to motivate how LIV identifies Δ^{MTE}, notice that from the definition of Y, the conditional expectation of Y given Z is

$$E(Y \mid Z=z) = E(Y_0 \mid Z=z) + E(Y_1 - Y_0 \mid Z=z, D=1)\Pr(D=1 \mid Z=z)$$

where we keep the conditioning on X implicit. Our model and conditional independence assumption (A-2) imply an index sufficiency restriction, so that we may rewrite this expectation as

$$E(Y \mid Z=z) = E(Y_0) + E(\Delta \mid P(z) \geq U_D)P(z) = E(Y \mid P(Z) = P(z)).$$

Applying the IV or Wald estimator for two different values of Z, z and z', assuming $P(z) \neq P(z')$, we

[42] The ideas of the marginal treatment effect and the limit form of $LATE$ were first introduced in the context of a parametric normal Generalized Roy model by Björklund and Moffitt (1987), and were also used more generally in Heckman (1997). Angrist, Graddy and Imbens (2000) also define and develop a limit form of $LATE$.

[43] For example, Heckman and Vytlacil (2000, 2004) show that to identify ATE under our assumptions, it is necessary and sufficient that the support of the distribution of $P(Z)$ conditional on X includes 0 and 1. Thus, identification of ATE requires a very strong condition, but does not require that distribution of $P(Z)$ conditional on X be the full unit interval or that the distribution of $P(Z)$ conditional on X contain any limit points.

obtain:

$$\begin{aligned} & \frac{E(Y \mid P(Z)=P(z))\text{-}E(Y \mid P(Z)=P(z'))}{P(z)\text{-}P(z')} \\ = \quad & \Delta^{ATE}+\frac{E(U_1\text{-}U_0 \mid P(z) \geq U_D)P(z)\text{-}E(U_1\text{-}U_0 \mid P(z') \geq U_D)P(z')}{P(z)-P(z')} \end{aligned}$$

where the last expression is obtained under the assumption of additive separability in the outcomes so (1a) and (1b) apply.[44] When $U_1 \equiv U_0$ or $(U_1 - U_0) \perp\!\!\!\perp U_D$, corresponding to case (C-1), IV based on $P(Z)$ estimates Δ^{ATE} because the second term on the right hand side of this expression vanishes. Otherwise, IV estimates a difficult-to-interpret combination of MTE parameters.

Another representation of $E(Y \mid P(Z) = p)$ that reveals the index structure more explicitly writes under additive separability (keeping the conditions on X explicit) that

$$E(Y|P(Z)=p) = E\left(Y_0\right) + \Delta^{ATE}p + \int_0^p E(U_1 - U_0|U_D = u_D)du_D. \tag{9}$$

We can differentiate with respect to p and use LIV to identify the Δ^{MTE}:

$$\frac{\partial E(Y \mid P(Z)=p)}{\partial p} = \Delta^{ATE} + E(U_1 - U_0|U_D = p) = \Delta^{MTE}(p).$$

Notice that IV estimates Δ^{ATE} when $E(Y \mid P(Z) = p)$ is a linear function of p. Thus a test of the linearity of $E(Y \mid P(Z) = p)$ in p is a test of the validity of linear IV for Δ^{ATE}. More generally, a test of the linearity of $E(Y \mid P(Z) = p)$ in p is a test of whether or not the data are consistent with a correlated random coefficient model. The nonlinearity of $E(Y \mid P(Z) = p)$ in p affords a way to distinguish whether Case (C-1) or Case (C-2) describes the data. It is also a test of whether or not agents can at least partially anticipate future unobserved (by the econometrician) gains (the $Y_1 - Y_0$ given X) at the time they make their participation decisions. These results generalize to the nonseparable case and so apply more generally. We use separability only to simplify the exposition.[45]

Figure 2A plots two cases of $E(Y \mid P(Z) = p)$ based on the Generalized Roy Model used to generate the example in Figure 1A and 1B. When Δ^{MTE} does not depend on u_D the expectation is a straight line. Figure 2B plots the derivatives of the two curves in Figure 2A. When Δ^{MTE} depends on u_D, people sort into the program being studied positively on the basis of gains from the program, and one gets the

[44] The same equation holds without additive separability if one replaces U_1 and U_0 with $Y_1 - E(Y_1|X)$ and $Y_0 - E(Y_0|X)$.

[45] The same expression holds with the same derivation for the nonseparable case if we replace U_1 and U_0 with $Y_1 - E(Y_1|X)$ and $Y_0 - E(Y_0|X)$, respectively. Making the conditioning on X explicit, we obtain again that $E(Y \mid X = x, P(Z) = p) = E(Y_0 \mid X = x) + \Delta^{ATE}(x)p + \int_0^p E(U_1 - U_0|X = x, U_d = u_D)du_D$, with derivative with respect to p given by $\Delta^{MTE}(x, p)$.

curved line depicted in Figure 2A. The levels and derivatives of $E(Y \mid P(Z) = p)$ and standard errors can be estimated using a variety of semiparametric methods. The derivative estimator of Δ^{MTE} is the local instrumental variable (LIV) estimator of Heckman and Vytlacil (1999, 2000). Thus it is possible to test condition (C-1) using simple econometric methods.

4.3 What does linear *IV* estimate?

Given the popularity of linear IV, it is instructive to consider what linear IV estimates when Δ^{MTE} is nonconstant, and conditions (A-1)–(A-5) hold. We consider the general non-separable case. We consider instrumental variables conditional on $X = x$ using a general function of Z as an instrument, and then specialize our result using $P(Z)$ as the instrument. Let $J(Z)$ be any function of Z such that $Cov\left(J(Z), D \mid X = x\right) \neq 0$. Define

$$\beta_{IV}\left(x; J\right) \equiv \left[Cov\left(J\left(Z\right), Y \mid X = x\right)\right] \,/\, \left[Cov\left(J\left(Z\right), D \mid X = x\right)\right].$$

First consider the numerator of this expression,

$$\begin{aligned} Cov\left(J\left(Z\right), Y \mid X = x\right) &= E\left(\left[J\left(Z\right) - E\left(J\left(Z\right) \mid X = x\right)\right] Y \mid X = x\right) \\ &= E\left(\left(J\left(Z\right) - E\left(J\left(Z\right) \mid X = x\right)\right)\left(Y_0 + D\left(Y_1 - Y_0\right)\right) \mid X = x\right) \\ &= E\left(\left(J\left(Z\right) - E\left(J\left(Z\right) \mid X = x\right)\right) D\left(Y_1 - Y_0\right) \mid X = x\right) \end{aligned}$$

where the second equality comes from substituting (1c) for Y and the third equality follows from assumption (A-2). Define $\tilde{J}(Z) \equiv J(Z) - E(J(Z) \mid X = x)$. Then

$$\begin{aligned} &Cov\left(J\left(Z\right), Y \mid X = x\right) \\ &\quad = E\left(\tilde{J}(Z)\ \mathbf{1}[U_D \leq P(Z)]\ \left(Y_1 - Y_0\right)\ \middle|\ X = x\right) \\ &\quad = E\left(\tilde{J}(Z)\ \mathbf{1}[U_D \leq P(Z)]\ E\left(Y_1 - Y_0 \mid X = x, Z, U_D\right)\ \middle|\ X = x\right) \\ &\quad = E\left(\tilde{J}(Z)\ \mathbf{1}[U_D \leq P(Z)]\ E\left(Y_1 - Y_0 \mid X = x, U_D\right)\ \middle|\ X = x\right) \\ &\quad = E\left(E\left[\tilde{J}(Z)\ \mathbf{1}[U_D \leq P(Z)]\ \middle|\ X = x, U_D\right] E\left(Y_1 - Y_0 \mid X = x, U_D\right)\ \middle|\ X = x\right) \\ &\quad = \int \left\{E(\tilde{J}(Z) \mid X = x, P(Z) \geq u_D) \Pr(P(Z) \geq u_D) E\left(Y_1 - Y_0 \mid X = x, U_D = u_D\right)\right\} du_D \\ &\quad = \int \Delta^{MTE}(x, u_D) E(\tilde{J}(Z) \mid X = x, P(Z) \geq u_D) \Pr(P(Z) \geq u_D) du_D \end{aligned}$$

where the first equality follows from plugging in the model for D; the second equality follows from the law of iterated expectations with the inside expectation conditional on $(X = x, Z, U_D)$; the third equality

follows from assumption (A-2); the fourth equality follows from the law of iterated expectations with the inside expectation conditional on $(X = x, U_D = u_D)$; the fifth equality follows from Fubini's Theorem and the normalization that U_D is distributed unit uniform conditional on X; and the final equality follows from plugging in the definition of Δ^{MTE}. Now consider the denominator of the IV estimand. Observe that by iterated expectations

$$Cov\left(J\left(Z\right), D \mid X = x\right) = Cov\left(J\left(Z\right), P\left(Z\right) \mid X = x\right).$$

Thus

$$\beta_{IV}\left(x; J\right) = \int \Delta^{MTE}(x, u_D) h_{IV}\left(u_D \mid x; J\right) du_D \tag{10}$$

where

$$h_{IV}\left(u_D \mid x; J\right) = \frac{E(\tilde{J}(Z) \mid X = x, P(Z) \geq u_D) \Pr(P(Z) \geq u_D)}{Cov\left(J\left(Z\right), P\left(Z\right) \mid X = x\right)} \tag{11}$$

where by assumption $Cov\left(J\left(Z\right), P\left(Z\right) \mid X = x\right) \neq 0$. Note that the weights integrate to unity:

$$\int_0^1 h_{IV}\left(u_D \mid x; J\right) du_D = 1.$$

We now discuss the properties of the weights for the special case where $J(Z) = P(Z)$ (the propensity score itself is used as the instrument), and then analyze the properties of the weights for a general instrument $J(Z)$. From equation (11), we obtain

$$h_{IV}\left(u_D | x; P(Z)\right) = \frac{\left[E(P(Z) \mid X = x, P(Z) \geq u_D) - E(P(Z) \mid X = x)\right] \Pr(P(Z) \geq u_D)}{Var\left(P\left(Z\right) \mid X = x\right)}.$$

Let p_x^{Min} and p_x^{Max} denote the minimum and maximum points in the support of the distribution of $P(Z)$ conditional on $X = x$. For u_D evaluation points between p_x^{Min} and p_x^{Max}, $u_D \in (p_x^{\text{Min}}, p_x^{\text{Max}})$, we have that $E\left(P\left(Z\right) \mid P\left(Z\right) \geq u_D, X = x\right) > E\left(P\left(Z\right) \mid X = x\right)$ and $\Pr(P(Z) \geq u_D) > 0$, so that $h_{IV}\left(u_D \mid x; P(Z)\right) > 0$ for any $u_D \in (p_x^{\text{Min}}, p_x^{\text{Max}})$. For $u_D \leq p_x^{\text{Min}}$, $E\left(P\left(Z\right) \mid \left(P\left(Z\right) \geq u_D, X = x\right) = E\left(P\left(Z\right) \mid X = x\right)\right.$. For any $u_D > p_x^{\text{Max}}$, $\Pr(P(Z) \geq u_D) = 0$. Thus, $h_{IV}\left(u_D \mid x; P(Z)\right) = 0$ for any $u_D \leq p_x^{\text{Min}}$ and for any $u_D > p_x^{\text{Max}}$. If the distribution of $P(Z)$ conditional on $X = x$ does not place a mass point at p_x^{Max}, then $h_{IV}\left(u_D \mid x; P(Z)\right) = 0$ for any u_D outside of $(p_x^{\text{Min}}, p_x^{\text{Max}})$. Thus, the weights for using $P(Z)$ as the instrument are nonnegative for all evaluation points, and are strictly positive for $u_D \in (p_x^{\text{Min}}, p_x^{\text{Max}})$.

Our expression for the weights does not impose any support conditions on the distribution of $P(Z)$ conditional on X, and thus does not require either that $P(Z)$ be continuous or discrete. To demonstrate this, consider two extreme special cases: (a) when the distribution of $P(Z)$ conditional on X has a density

with respect to Lebesgue measure ($P(Z)$ is a continuous random variable), and (b) when the distribution of $P(Z)$ conditional on X has a density with respect to counting measure ($P(Z)$ is a discrete random variable). For ease of exposition, we continue to consider the case where $J(Z) = P(Z)$, so that the propensity score is used as the instrument.

With $J(Z) = P(Z)$, consider the case where the distribution of $P(Z)$ conditional on X has a density with respect to Lebesgue measure with nonnegative density on the interval $(p_x^{\text{Min}}, p_x^{\text{Max}})$. In this case, $\Delta^{LIV}(x, u_D)$ is well defined for all $u_D \in (p_x^{\text{Min}}, p_x^{\text{Max}})$ and is thus well defined for all u_D such that $h_{IV}(u_D \mid x; P(Z)) > 0$. Therefore, using the fact that $\Delta^{LIV}(x, u_D) = \Delta^{MTE}(x, u_D)$ at any evaluation points where LIV is well defined, we can rewrite our general result as

$$\beta_{IV}\left(x; P(Z)\right) = \int_{p_x^{\text{Min}}}^{p_x^{\text{Max}}} \Delta^{LIV}(x, u_D) h_{IV}(u_D \mid x; P(Z)) du_D.$$

In this special case, our result becomes a latent variable version of the formulae in Yitzhaki (1996, 1999) and Angrist, Graddy and Imbens (2000).

With $J(Z) = P(Z)$, next consider the case where the distribution of $P(Z)$ conditional on X has density with respect to counting measure. For simplicity, assume that the support of the distribution of $P(Z)$ conditional on X contains a finite number of values, $\{p_1, ..., p_K\}$ with $p_1 < p_2 < \cdots < p_K$. Then $E(P(Z) \mid X = x, P(Z) \geq u_D)$ is constant in u_D for u_D within any (p_j, p_{j+1}) interval, and $\Pr(P(Z) \geq u_D)$ is constant in u_D for u_D within any (p_j, p_{j+1}) interval, and thus $h_{IV}(u_D \mid x; P(Z))$ is constant in u_D over any (p_j, p_{j+1}) interval. Let q_j denote the value taken by $h_{IV}(u_D \mid x; P(Z))$ for $u_D \in (p_j, p_j + 1)$. We obtain

$$\begin{aligned}
\beta_{IV}(x; P(Z)) &= \int E(\Delta | X = x, U_D = u_D) h_{IV}(u_D \mid x; P(Z)) du_D \\
&= \sum_{j=1}^{K-1} \int_{p_j}^{p_{j+1}} E(\Delta | X = x, U_D = u_D) q_j du_D \\
&= \sum_{j=1}^{K-1} q_j (p_{j+1} - p_j) \int_{p_j}^{p_{j+1}} E(\Delta | X = x, U_D = u_D) \frac{1}{(p_{j+1} - p_j)} du_D \\
&= \sum_{j=1}^{K-1} \Delta^{LATE}(x, p_j, p_{j+1}) \tilde{q}_j
\end{aligned}$$

where $\tilde{q}_j = q_j(p_{j+1} - p_j)$. In this special case, our analysis is a latent variable version of the formula in Imbens and Angrist (1994).

We now consider the properties of the weights for general $J(Z)$. They depend critically on the relationship between $J(Z)$ and $P(Z)$. Define $T(p \mid x; J) = E(J \mid P(Z) = p, X = x) - E(J | X = x)$. In this

notation

$$h_{IV}(u_D \mid x; J) = \frac{\int_{u_D}^{1} T(t \mid x; J) dF_{P|X}(t|x)}{Cov\left(J, P \mid X = x\right)}.$$

From this expression, we see that the IV estimator using $J(Z)$ as an instrument satisfies the following properties: (a) $h_{IV}(u_D \mid x; J) = h_{IV}(u_D \mid x; T\left(P(Z) \mid x; J\right))$; (b) $h_{IV}(u_D \mid x; J)$ is non-negative for all u_D if $E(J \mid X = x, P(Z) \geq p)$ is weakly monotonic in p; and (c) the support of $h_{IV}(u_D \mid x; J)$ is contained in $(p_x^{\text{Min}}, p_x^{\text{Max}}]$.[46]

Property (a) states that any instrument J leads to the same weights on Δ^{MTE} as using $T(p \mid x; J)$ as an instrument. Two instruments J and J^* weight MTE equally at all u_D if and only if $E(J|X = x, P(Z) = p) - E(J|X = x) = E(J^*|X = x, P(Z) = p) - E(J^*|X = x)$ for all p in the support of $P(Z)$ conditional on $X = x$. Property (b) states that using J as an instrument yields nonnegative weights on Δ^{MTE} if $E(J \mid X = x, P(Z) \geq p)$ is weakly monotonic in p. This condition is satisfied when $J(Z) = P(Z)$. As another special case, if J is a monotonic function of $P(Z)$, then using J as the instrument will lead to nonnegative weights on Δ^{MTE}. There is no guarantee that the weights for a general $J\left(Z\right)$ will be nonnegative for all u_D, although the weights integrate to unity and thus must be positive over some range of evaluation points. We produce examples below where the instrument leads to negative weights for some evaluation points.

Restriction (c) states that using any other instrument leads to nonzero weights only on a subset of $(p_x^{\text{Min}}, p_x^{\text{Max}}]$. Thus, for example, $h_{IV}\left(0 \mid x; J\right) = 0 = h_{IV}\left(1 \mid x; J\right)$.[47] More generally $h_{IV}\left(t, J \mid X = x\right) = 0$ for $t \leq p_x^{\text{Min}}$ and for $t > p_x^{\text{Max}}$. Using the propensity score as an instrument leads to nonnegative weights on a larger range of evaluation points than using any other instrument. Figure 1B plots the IV weight for $J = P(Z)$ and the MTE for our Roy model example.

Observe that from (11) the interpretation placed on the IV estimand depends on the specification of $P(Z)$ even if $J(Z)$ (e.g. a coordinate of Z) is used as the instrument. This drives home the point about the difference between IV in the traditional model and IV in the model analyzed in this paper. In the traditional model, the choice of among valid instruments and the specification of the instruments used in $P(Z)$ does not affect the IV estimand (the probability limit of the IV estimator). In the more general model analyzed in this paper, these choices matter. Two economists, using the same $J(Z) = Z_1$, say, will obtain the same IV point estimate, but the interpretation placed on that estimate will depend on the specification of Z in the $P(Z)$ even if $P(Z)$ is not used as an instrument.

Table II gives the IV estimand for the Generalized Roy Model used to generate Figures 1A and 1B using

[46] If the distribution of $P(Z)$ conditional on X places a mass point at p_x^{Max}, then the weights will be nonzero on $(p_x^{\text{Min}}, p_x^{\text{Max}}]$ and zero outside of $(p_x^{\text{Min}}, p_x^{\text{Max}}]$.

[47] If $\Pr(P(Z) = 1) > 0$, then it is possible to have $h_{IV}\left(1 \mid x; J\right) \neq 0$.

$P(Z)$ as the instrument. The model for generating $D = \mathbf{1}[\alpha' Z > V]$ is given at the base of Figure 1B (Z is a scalar, α' is 1, V is normal, $U_D = \Phi\left(\frac{V}{\sigma_\varepsilon \sigma_V}\right)$). We compare the IV estimand with the policy relevant treatment effect for a policy defined at the base of Table II. If $Z > 0$, persons get a bonus Zt. Their participation decision rule if $Z > 0$ is $D = \mathbf{1}[Z(1+t) > V]$. For those with $Z < 0$, $t = 0$ and $D = \mathbf{1}[Z > V]$. Given the assumed distribution of Z, and the other parameters of the model, we obtain $h_{PRTE}(u_D)$ as plotted in Figures 3A-3C (the scales differ across the graphs). We use the *per capita PRTE* and consider three instruments.

The first is $P(Z)$, which ignores the policy (t) effect on choices. Its weight is plotted in 3A which also has the *OLS* weight (discussed later) imposed. The IV weights for $P(Z)$ and the weights for Δ^{PRTE} do not agree. Given the shape of $\Delta^{MTE}(u_D)$, it is not surprising that the estimand for IV based on $P(Z)$ is so much above the Δ^{PRTE} which weights a lower-valued segment of $\Delta^{MTE}(u_D)$ more heavily.

The second instrument exploits the variation induced by the policy in place. On intuitive grounds this instrument should work well. The instrument is

$$\tilde{P}(Z,t) = P(Z(1 + t(\mathbf{1}[Z > 0]))$$

which jumps in value when $Z > 0$. This is the choice probability in the regime with the policy in place. Figure 3B plots the weight for this IV along with the weight for $P(Z)$ as an IV (repeated from 3A). While this weight looks a bit more like the weight for Δ^{PRTE}, it is clearly different. Table IIIB, which reports the estimands for the instruments used in this simulation (including the one to be discussed next), reflects the movement of these IV weights toward the Δ^{PRTE} weights, but the distance between the IV estimand and the policy relevant treatment effect is still substantial. Figure 3C plots the weight for an ideal instrument which is a randomization of eligibility. We use an instrument B such that

$$B = \begin{cases} 1 & \text{if a person is eligible to participate in the program} \\ 0 & \text{otherwise.} \end{cases}$$

If $B = 1$, persons make their participation choices under the rules previously discussed. If $B = 0$, $t = 0$ and there is no bonus. We assume $P(B) = 0.5$ so persons are equally likely to receive or not receive eligibility for the bonus. The IV weight for this case can be derived from (11):

$$h_{IV}(u_D \mid B) = \frac{E\left(B - E(B) \mid \hat{P}(Z) \geq u_D\right) \Pr\left(\hat{P}(Z) \geq u_D\right)}{Cov\left(B, \hat{P}(Z)\right)}$$

where $\hat{P}(Z) = P\left(Z\left(1 + t\left(\mathbf{1}[Z > 0]\right)\right)\right)^B P(Z)^{(1-B)}$. This IV corresponds exactly to the policy. Indeed it

is equivalent to a social experiment that identifies

$$\frac{E(Y \mid B=1) - E(Y \mid B=0)}{\Pr(D=1 \mid B=1) - \Pr(D=0 \mid B=0)}.$$

Thus it is not surprising that this IV weight and h_{PRTE} are identical.

Monotonicity property (b) is strong. For a general $J(Z)$, there is no guarantee that it will be satisfied even if $J(Z)$ is independent of (Y_0, Y_1) given X and if $J(Z)$ is correlated with D given $X = x$ so that standard IV conditions are satisfied. Thus if Z is a K-dimensional vector and $J(Z) = Z_1$, even if conditional on $Z_2 = z_2, \ldots, Z_K = z_K$, $P(Z)$ is monotonic in Z_1, there is no guarantee that Z_1 as an instrument for D has positive weights.[48] Figure 4 demonstrates this possibility for the model given at its base. We work with V rather than normalized $F_V(V) = U_D$ in this example. This simulation is generated from a classical normal error term selection model with nonnormal instruments. The instruments are generated as mixtures of normals from two underlying populations. One can think of this as a two-component ecological model with different $J(Z), P(Z)$ covariance relationships in the two components. Alternatively, there are different $J(Z), \alpha' Z$ covariance relationships in the two subpopulations. In the first component the covariance is .98. In the second, the covariance varies as shown in Table IV. The IV is Z_1 but the choice probability depends on Z_1 and Z_2 ($\mu_D(Z) = \gamma' Z$). *Ceteris paribus*, increasing Z_1 increases the probability that $D = 1$. Symmetrically, increasing Z_2 holding Z_1 constant also increases the probability. Yet, since Z_1 and Z_2 covary, varying Z_1 implicitly varies Z_2, which may offset the *ceteris paribus* effect of Z_1 and produce non-monotonicity and negative weights. In this example there are different covariance relationships in different normal subcomponents of the data. As Z_1 increases $P(Z)$ sometimes increases and sometimes decreases leading to two-way flows into and out of treatment for different people. IV estimates the effect of Z_1 on outcomes *not controlling* for the other elements of Z. For the configuration of parameters shown there (and for numerous other configurations), the IV weight is negative over a substantial range of values.

The negativity of the weights over certain regions makes clear that Z_1 (and more generally $J(Z)$) fails the monotonicity condition (b) and does not estimate a gross treatment effect. Some agents withdraw from participation in the program when Z_1 is raised (not holding constant Z_2) while others enter, even though *ceteris paribus* a higher Z_1 raises participation (D). Thus the widely held view that IV estimates some treatment effect of a change in D induced by a change in Z_1, is in general false.[49] It estimates a net effect

[48] If we redefine IV for Z_1 to be conditional on $Z_2 = z_2, \ldots, Z_K = z_K$ and $P(Z)$ is monotonic in Z_1, holding the other arguments fixed, then the weights are positive. Conditioning on instruments not used to form the primary covariance relationship is a new concept that does not appear in the conventional IV literature. In conventional cases governed by condition (C-1), any valid instrument identifies the same parameter. In the general case analyzed in this paper, the choice of an instrument and the conditioning set of other instruments defines a different parameter. Of course if $P(Z)$ is not monotonic in Z_1 given $Z_2 = z_2, \ldots, Z_K = z_K$, then obviously, the monotonicity condition is also violated.

[49] See, e.g., Kling (2001) or DeLong, Goldin and Katz (2003) who, among a legion of applied economists, assume that IV

and not a treatment effect, because monotonicity may be violated.

Monotonicity condition (b) is testable under independence assumption (A-2). Monotonicity is equivalent to the nonnegativity of the weights for Δ^{MTE}. If the weights are negative, the change in $J(Z)$ induces two way flows into and out of treatment. Since it is possible to estimate the joint density of $(J(Z), P(Z))$ given X nonparametrically, under independence (A-2) it is possible to test for the positivity of the weights which under our assumptions is also a test for monotonicity condition (b). See Heckman and Vytlacil (2004) for further discussion.

4.4 *OLS* Weights

The *OLS* estimator can also be represented as a weighted average of Δ^{MTE}. Straightforward manipulation reveals that the weight is

$$h_{OLS}(u_D|x) = 1 + \frac{E(U_1 \mid X = x, U_D = u_D)h_1(u_D|x) - E(U_0 \mid X = x, U_D = u_D)h_0(u_D|x)}{\Delta^{MTE}(x, u_D)}$$

for (x, u_D) such that $\Delta^{MTE}(x, u_D) \neq 0$, and $h_{OLS}(u_D|x) = 0$ otherwise, where $h_1(u_D \mid x) = \left[\int_{u_D}^{1} f_{P|X}(t|x)dt\right] \cdot \frac{1}{E(P \mid X = x)}$, and $h_0(u_D \mid x) = \left[\int_0^{u_D} f_{P|X}(t \mid x)dt\right] \frac{1}{E((1-P) \mid X = x)}$ where $f_{P|X}(t|x)$ is the density of P given X. Unlike the weights for *IV*, Δ^{TT} and Δ^{ATE}, these weights do not necessarily integrate to 1 and they are not necessarily nonnegative. The *OLS* weights for the Generalized Roy Model are plotted in Figure 1B. The negative component of the *OLS* weight leads to a smaller *OLS* treatment estimate compared to the other treatment effects in Table II.

Table II shows the estimated *OLS* treatment effect for the Generalized Roy Example. For a binary regressor, D, *OLS* conditional on X identifies $\Delta^{OLS}(X) = E(Y_1 \mid X, D = 1) - E(Y_0 \mid X, D = 0) = E(Y_1 - Y_0 \mid X, D = 1) + \{E(Y_0 \mid X, D = 1) - E(Y_0 \mid X, D = 0)\}$ where the term in braces is the "selection bias" term—the difference in pretreatment outcomes between treated and untreated individuals. It is also the bias for Δ^{TT}. The large negative selection bias in this example is consistent with comparative advantage as emphasized by Roy (1951). People who are good in Sector 1 may be very poor in Sector 0. The differences among the policy relevant treatment effects, the conventional treatment effects and the *OLS* estimand is illustrated in Figure 3A and Tables II and IIIA.

estimates a gross treatment effect.

4.5 Policy Relevant Instrumental Variables

The analysis of Subsection 4.3 answers the question: "If we use a particular function of Z as an instrument, what weights on the Δ^{MTE} produce the estimand?" It is natural to ask the reverse question. Suppose there is a particular parameter of interest, defined by a given weighted average of Δ^{MTE} conditional on $X = x$: "Can we construct a function of Z to use as an ordinary instrument so that the resulting estimand corresponds to the desired weighted average of Δ^{MTE}?" This question is especially interesting if the estimand is a policy counterfactual: "Can we construct an instrument so that the resulting IV estimand corresponds to a desired policy counterfactual?" We examine this issue for the case where the distribution of $P(Z)$ conditional on X has a density with respect to Lebesgue measure.

Suppose that we seek to recover a parameter defined by $\int \Delta^{MTE}(x,u)w(u|x)du$ for some weighting function $w(u|x)$ using linear instrumental variables. We know from equation (11) the form of the weights corresponding to the IV estimator for any particular instrument $J(Z)$. We seek an instrument $J(Z)$ that has associated weights on MTE that are the same as those on the desired parameter:

$$w(u|x) = \frac{\int_u^1 T(t \mid x; J) dF_{P|X}(t|x)}{Cov\left(J, P \mid X = x\right)},$$

where $T(t \mid x; J) \equiv E(J \mid X = x, P(Z) = t) - E(J \mid X = x)$. Assuming that $F_{P|X}$ has a density with respect to Lebesgue measure, the right hand side of this expression is differentiable in $U = u$ $(a.e.)$. Assuming that $w(u|x)$ is also differentiable at all points of evaluation, it follows that, $w'(u|x) = -\frac{T(u|x;J)f_{P|X}(u|x)}{Cov\left(J,P|X=x\right)}$. The following proposition provides conditions under which an instrument exists with the desired properties.

Proposition 1: Under the following conditions

(a) $F_{P|X}(\cdot)$ has a density with respect to Lebesgue measure;

(b) $w(\cdot|x)$ satisfies the following properties: $w(u|x)$ differentiable in u for all $u \in [0,1]$, $\int_0^1 w(u|x)du = 1$, and $w(1|x) - w(0|x) = 0$;

(c) $f_{P|X}(t|x) = 0$ implies $w'(t|x) = 0$;

(d) $\int_0^1 t w'(t|x)dt = -1$

there exists an instrument $J(Z)$ such that $Cov\left(J, D \mid X = x\right) \neq 0$ and $w(u|x) = \frac{\int_u^1 T(t \mid x; J) dF_{P|X}(t|x)}{Cov\left(J, P \mid X = x\right)}$.

An instrument that satisfies these conditions is

$$J(Z) = \begin{cases} \frac{w'(P(Z)|x)}{f_{P|X}(P(Z)|x)} & \text{if } f_{P|X}(P(Z)|x) > 0 \\ 0 & \text{if } f_{P|X}(P(Z)|x) = 0^{50}. \end{cases}$$

Given assumptions (a) and (b), assumptions (c) and (d) are necessary and sufficient for the existence of such an instrument.

Proof See Appendix C. ■

When such an instrument exists, it will not be unique, since the IV estimand will be invariant to rescaling or location shifts for the instrument. Condition (c) is strong but natural. It requires that the support of the propensity score includes the support of $w'(\cdot|x)$. This condition will always be satisfied if $f_{P|X}(t|x) > 0$ for all $t \in [0,1]$. Given (a) and (b), if (c) fails, no instrument exists. A second, less natural condition is that $\int_0^1 w'(t|x)tdt = -1$. An instrument with the desired weights does not exist if this condition fails. While this condition may seem unnatural, it is always satisfied for the weights implied by the policy relevant treatment effect given in (7). We specialize the previous proposition for the special case of policy weights:

Proposition 2: Assume

(a) $F_{P_{a'}|X}(\cdot)$ and $F_{P_a|X}(\cdot)$ have densities with respect to Lebesgue measure;

(b) $E(P_a|X = x) \neq E(P_{a'}|X = x)$;

(c) for any t, $f_{P|X}(t|x) = 0$ implies $f_{P_a|X}(t|x) - f_{P_{a'}|X}(t|x) = 0$;

Define J to be a policy relevant instrument if it satisfies $Cov\,(J, D \mid X = x) \neq 0$ and

$$\frac{\int_u^1 T(t \mid x; J) dF_{P|X}(t|x)}{Cov\,(J, P \mid X = x)} = \frac{F_{P_{a'}|X}(t \mid x) - F_{P_a|X}(t \mid x)}{E(P_{a'} \mid X = x) - E(P_a \mid X = x)}.$$

Given assumptions (a) and (b), assumption (c) is necessary and sufficient for the existence of such an instrument. If the instrument exists, it is

$$J(Z) = \begin{cases} \frac{f_{P_{a'}|X}(P(Z)) - f_{P_a|X}(P(Z))}{f_{P|X}(P(Z))} & \text{if } f_{P|X}(P(Z)|x) > 0 \\ 0 & \text{if } f_{P|X}(P(Z)|x) = 0.^{51} \end{cases}$$

Proof: Follows by verifying the conditions of Proposition 1. See Appendix C. ■

[50] Note that $f_{P|X}(P(Z)|x) > 0$ w.p.1 so that $J(Z) = w'(P(Z)|x)/f_{P|X}(P(Z)|x)$ w.p.1.

[51] Note that $f_{P|X}(P(Z)|x) > 0$ w.p.1 so that $J(Z) = \frac{f_{P_{a'}|X}(P(Z)) - f_{P_a|X}(P(Z))}{f_{P|X}(P(Z))}$ w.p.1.

Again, if such a $J(Z)$ exists, then any linear function of $J(Z)$ will also produce the desired set of weights. Using the proposition, we immediately obtain the corollary that IV using the propensity score as the instrument recovers the policy relevant parameter if

$$P(Z) = \alpha(X) + \beta(X)\left[\frac{f_{P_{a'}|X}(P(Z)) - f_{P_a|X}(P(Z))}{f_{P|X}(P(Z))}\right]$$

where $\alpha(X) = E(P(Z)|X)$ and $\beta(X) = -Var(P(Z))$. In words, using the propensity score as an instrument consistently estimates the policy counterfactual only if the propensity score happens to be linear in $\{f_{P_{a'}|X}(P(Z)) - f_{P_a|X}(P(Z))\}/f_{P|X}(P(Z))$.

A related question asks whether a given instrument corresponds to the weighting required for some policy counterfactual. In other words, given an instrument, does there exist a policy counterfactual such that the given instrument is the policy relevant instrument for that counterfactual? We investigate this question for policy counterfactuals starting from a baseline current distribution of $P(Z)$ (the base policy is the policy currently in place so that $P_a(Z_a) = P(Z)$) to some new policy characterized by $P_{a'}(Z_{a'})$. We first answer the question for the special case where the propensity score is the instrument. Solving for $f_{P_{a'}|X}(P(Z))$ in the above equation, we have that the propensity score will be the policy relevant instrument for the policy with the weights on $P_{a'}$ characterized by

$$f_{P_{a'}|X}(u_D) = f_{P|X}(u_D)\left(1 - \frac{u_D - E(P(Z)|X)}{Var(P(Z) \mid X)}\right).$$

Clearly, $f_{P_{a'}|X}(\cdot)$ always integrates to one.[52] It will be nonnegative and thus a proper density if and only if $u_D - E(P(Z)|X) \leq Var(P(Z))$ for all u_D such that $f_{P|X}(u_D) > 0$. If we let p_x^{Max} denote the maximum of the support of $P(Z)$ conditional on X, we can thus rewrite this condition as $p_x^{\text{Max}} - E(P(Z)|X) \leq Var(P(Z))$. If this condition holds, then using the propensity score as the instrument identifies the $PRTE$ going from the current distribution of $P(Z)$ to a new distribution of $P(Z)$, for a policy distribution of a very particular form. Nothing guarantees the existence of this density so one cannot guarantee that an instrument produces any policy counterfactual. Not all instruments answer well posed policy questions.

We next consider the question of whether a general instrument is the policy relevant instrument for some policy. Following the same series of steps just used, if the instrument $J(Z)$ has a corresponding policy, then the policy must be characterized by

$$f_{P_{a'}|X}(u_D) = f_{P|X}(u_D)\left(1 - \frac{E(J(Z)|X, P(Z) = u_D) - E(J(Z)|X)}{Cov(J, P(Z))}\right)$$

[52] $E(P(Z) \mid X = x) = \int_0^1 u_D f_{P|X}(u_D) du_D$.

Once more, the implied $f_{P_{a'}}(\cdot)$ always integrates to one. It is nonnegative for all evaluation points if and only if $\frac{E(J(Z)|X,P(Z)=u_D)-E(J(Z)|X)}{Cov\,(J,P(Z))} \leq 1$ for all u_D such that $f_{P|X}(u_D) > 0$. If this condition fails, then the instrument is not the policy relevant instrument for any policy.

The preceding analysis conditions on X. Suppose that we wish to recover parameters, e.g., defined by $\int \left[\int \Delta^{MTE}(x,u)w(u|x)du\right] dF_X(x)$. If the conditions of Proposition 1 hold for $X = x$ (a.e.), then one solution would be to construct $J(Z)$ for each x, estimate the parameter conditional on X for each x, and then average over x values. However, from the construction of $J(Z)$, one can use instrumental variables unconditional on X with the constructed $J(Z)$ as the instrument to obtain the desired parameter in one step as stated by the following proposition.

Proposition 3: Assume that the conditions of Proposition 1 hold for a.e. X. Construct

$$J(Z) = \begin{cases} \frac{w'(P(Z)|X)}{f_{P|X}(P(Z))} & \text{if } f_{P|X}(P(Z)) > 0 \\ 0 & \text{if } f_{P|X}(P(Z)) = 0.^{53} \end{cases}$$

Then

$$\frac{Cov\,(J(Z),Y)}{Cov\,(J(Z),D)} = \int \left[\int \Delta^{MTE}(x,u)w(u|x)du\right] dF_X(x).$$

Proof See Appendix C. ■

We next briefly consider the assumptions about Δ^{MTE} imposed in one widely used version of the method of matching.

4.6 Matching

The method of matching is coming into wideere is no purposeful selection into the program based on unmeasured (by the econometrician) components of gain. This is condition (C-1) discussed in our analysis of instrumental variables. It implies under (A-1)–(A-5) that $E(Y \mid X, P(Z) = p)$ is linear in p. For a model satisfying (A-1)–(A-5) with full support of $P(Z)$, $\Delta^{MTE}(x,u_D)$ does not vary with u_D.[54,55] Otherwise fortuitous balancing is required with increasing and decreasing components just offsetting each other. Such

[53] Note that $f_{P|X}(P(Z)|X) > 0$ w.p.1 so that $J(Z) = w'(P(Z)|X)/f_{P|X}(P(Z)|X)$ w.p.1.

[54] Heckman, Ichimura, Smith and Todd (1998) and Heckman, Ichimura and Todd (1997) show that in place of (M-??) one can work with weaker mean independence assumptions: $E\,(Y_0 \mid X = x, D = 0) = E\,(Y_0 \mid X = x)$ and $E\,(Y_1 \mid X = x, D = 1) = E\,(Y_1 \mid X = x)$. Generically, these still imply that $E\,(Y_1 - Y_0 \mid X = x, D = 1) = E\,(Y_1 - Y_0 \mid X = x)$ so $\Delta^{MTE}(x,u_D)$ does not depend on u_D, under our assumptions (A-1)–(A-5). If the goal of the analysis is to estimate Δ^{TT}, one can get by with the weaker assumption $E\,(Y_0 \mid X = x, D = 0) = E\,(Y_0 \mid X = x)$ so that under our assumptions there is no selection of Y_0 on U_D given X.

[55] In other words, $\Delta^{MTE}(X,u_D) = \Delta^{MTE}(X,u'_D)$ for $(a.e.)\ u_D, u'_D$.

balancing is ruled out in the Roy model and in the Generalized Roy Model.[56,57]

5 Policy Invariant Parameters, Out of Sample Policy Forecasting, Forecasting the Effects of New Policies and Structural Models

Thus far we have been concerned with the problem of estimating the impact of a program in place in a particular environment free of bias. This is the problem of "internal validity". Extrapolating internally valid estimates to new environments, "external validity", or forecasting the effects of new policies, are also important problems which we now address.[58]

Let $a \in \mathcal{A}$ denote a policy characterized by vector Z_a. Let $e \in \mathcal{E}$ denote an environment characterized by vector X_e. A history, $\mathcal{H}$, is a collection of policy-environment (a, e) pairs that have been experienced and documented. We assume that the environment is autonomous so the choice of a does not affect X_e. Letting $X_{e,a}$ denote the value of X_e under policy a, autonomy requires that

(A-7) $X_{e,a} = X_e \quad \forall a, e$ (autonomy).

Autonomy is a more general notion than the concept introduced in (A-6), but they are the same when the policy is a treatment. General equilibrium feedback effects can cause a failure of autonomy. In this paper we assume autonomy, in accordance with the partial equilibrium tradition in the treatment effect literature.[59]

Evaluating a particular policy a' in environment e' is straightforward if $(a', e') \in \mathcal{H}$. One simply looks at the associated outcomes and treatment effects formed in that policy environment and applies the methods previously discussed to obtain internally valid estimates. The new challenge comes in forecasting the impacts of policies (a') in environments (e') not in $\mathcal{H}$.

[56] In particular, assume $Y_j = \mu_j(X) + U_j$ for $j = 0, 1$, assume $D = \mathbf{1}[Y_1 - Y_0 \geq C(Z) + U_C]$, and let $U_D = U_C - (U_1 - U_0)$. Then if $U_C \perp\!\!\!\perp U_1 - U_0$, and U_C has a log concave density, then $E(Y_1 - Y_0 | X, U_D = u_D)$ is decreasing in u_D, $\Delta^{TT}(x) > \Delta^{ATE}(x)$, and matching cannot hold. If $U_C \perp\!\!\!\perp U_1 - U_0$ but U_C does not have a log concave density, then it is still the case that $(U_1 - U_0, U_D)$ is negative quadrant dependent. One can show that $(U_1 - U_0, U_D)$ being negative quadrant dependent implies that $\Delta^{TT}(x) > \Delta^{ATE}(x)$, and thus again that the matching conditions cannot hold. See Heckman and Vytlacil (2004).

[57] It is sometimes said that the matching assumptions are "for free" (See Gill and Robins, 2001) because one can always replace unobserved $F(Y_1 \mid X = x, D = 0)$ with $F(Y_1 \mid X = x, D = 1)$ and unobserved $F(Y_0 \mid X = x, D = 1)$ with $F(Y_0 \mid X = x, D = 0)$. This ignores the counterfactual states generated under the matching assumptions that (C-1) is true in the population. The assumed absence of selection is not a "for free" assumption, and produces fundamentally different counterfactual states for the same model under matching and selection assumptions.

[58] The terms "internal" and "external" validity were first defined in Campbell and Stanley (1966).

[59] But see Heckman and Vytlacil (2004) for the general case and Heckman, Lochner and Taber (1998b) for an example of a nonautonomous treatment model.

We show how Δ^{MTE} plays the role of a policy-invariant functional that aids in creating counterfactual states never previously experienced. We focus on the problem of constructing the policy relevant treatment effect Δ^{PRTE} but our discussion applies more generally to the other treatment parameters.

We seek to form (7) for a general valuation function $V(Y)$ for a never-previously experienced configuration of a' and e'. Representation (7) demonstrates the value of our approach. Under (A-1)–(A-5), and the autonomy assumption, the general problem of constructing

$$E\left(V\left(Y\right) \mid X_e = x, \text{ under policy } a'\right) \equiv E_{a'}\left(V(Y) \mid X_e = x\right)$$

can be broken down into the task of identifying Δ^{MTE} and constructing h_{PRT} for a new policy (we are maintaining the assumption that the baseline policy has already been observed, $(a, e) \in \mathcal{H}$). In turn, h_{PRT} can be constructed from the distributions of

$$P_{a,e} = \Pr\left(D_a = 1 \mid X_e, Z_a\right) \qquad \text{and} \qquad P_{a',e} = \Pr\left(D_{a'} = 1 \mid X_e, Z_{a'}\right)$$

from the two policy regimes. Thus for environment e, the policy forecasting problem is to construct $F_{P_{a'}|X_e}$ since $F_{P_a|X_e}$ is known under the conditions of the preceding analysis. This task is more focused, and tractable, than the general problem of constructing $E_{a'}(V(Y))$. The separation of tasks entailed in this approach requires the invariance of Δ^{MTE} when policies change. It also requires that we construct $F_{P_{a'}|X_e}$ for policies never previously experienced. When we allow environments to change, we must also extend both Δ^{MTE} and $F_{P_{a'}|X}$ to new environments.

There are three distinct problems: (a) establishing invariance of Δ^{MTE} to policies; (b) constructing $F_{P_{a'}|X_e}(\cdot|x)$ for a fixed evaluation point x in the support of the distribution of X_e; and (c) extending Δ^{MTE} and $F_{P_a|X_e}(\cdot \mid x)$ to evaluation points x not in the support for X_e. We discuss these problems in turn. We first discuss conditions under which Δ^{MTE} is policy invariant. An invariant Δ^{MTE} can be used to evaluate a whole menu of policies characterized by different $F_{P_{a'}|X_e}$. In addition, under invariance, because of the index structure, we can focus on how a', which is characterized by $Z_{a'}$, produces the distribution $F_{P_{a'}|X}$ which weights an invariant Δ^{MTE} without having to conduct a new investigation of (Y, Z) relationships for each proposed policy.[60]

[60] Ichimura and Taber (2002) present a discussion of policy analysis in a more general framework without the *MTE* structure, using a framework anticipated by Hurwicz (1962).

5.1 Policy Invariance

We define a policy applied to a population, conditional on $X_e = x$, as a conditional distribution of Z. Two policies, a and a', are associated with two different distributions of Z: $F_{Z_a|X_e}(\cdot \mid x)$ and $F_{Z_{a'}|X_e}(\cdot \mid x)$. In order to use the same Δ^{MTE} to predict conventional and policy relevant treatment effects under alternative policies, it must be policy invariant.[61] Parameters are policy invariant (conditional on $X_e = x$) if they are unaffected by the choice of the distribution of Z:

Δ^{MTE} is policy invariant if

(A-8) $E(Y_1 - Y_0 \mid U_D = u_D, X_e = x)$ is invariant to the choice of the conditional distribution of Z given $X_e = x$.

This is implied by assumption (A-2). But (A-8) is the basic condition.

Δ^{LATE} is formulated by Imbens and Angrist (1994) as the estimand associated with linear IV for a given Z and so is not policy invariant, except in the case of binary Z, because it depends on the distribution of Z. Under our definition, Δ^{LATE} in terms of U_D presented in Subsection 2.1 is policy invariant. The existence of such a U_D is implied by the Imbens-Angrist monotonicity and independence conditions (Vytlacil, 2002). This point demonstrates the value of defining parameters in terms of underlying economic primitives, separating definition from identification and issues of estimation.

It is instructive to consider what (A-8) rules out. It excludes non-monotonicity in the response of treatment choices to Z, except in special cases. Monotonicity is the requirement that as Z is changed from $Z = z$ to $Z = z'$, all persons shift in the same direction toward $D = 1$ or $D = 0$. More precisely, in the notation of Imbens and Angrist (1994) as defined in Section 2.1, if the Z are changed for everyone from $Z = z$ to $Z = z'$, $D_z \geq D_{z'}$ or $D_z \leq D_{z'}$ for all U_D conditional on X. This monotonicity restriction along with D_z independent of Z for all z in the support of the distribution of Z are jointly guaranteed by representation (4) and assumption (A-2).[62]

A general random coefficient version of $\mu_D(Z)$ in (4) violates (A-2). Consider the following example. If $\mu_D(Z) = Z\gamma$, where γ is a common coefficient shared by everyone, the choice model satisfies the monotonicity property. If γ is a random coefficient (i.e. has a nondegenerate distribution) that can take both negative and positive values, monotonicity is clearly violated, but it can be violated even when all components are of the same sign if Z is a vector. For the model $D = \mathbf{1}[Z\gamma > V]$, if $E(|\gamma|) < \infty$ and $\gamma \perp\!\!\!\perp (Z, V) \mid X$, defining $\varepsilon = \gamma - E(\gamma)$, we obtain $U_D = V + Z\varepsilon$, where $E(U_D) = 0$ if $E(V) = 0$. In the general case for this model, the distribution of U_D depends on the distribution of Z. Under monotonicity

[61] Hendry (1995) discusses the role of policy invariant parameters in macro-forecasting and policy evaluation.

[62] More precisely, by the assumption that $U_D \perp\!\!\!\perp Z \mid X$.

and independence, Vytlacil's theorem (2002) implies that there exists a scalar U_D such that conditioning on it and X produces a policy-invariant Δ^{MTE}. If monotonicity fails, the U_D used in representation (A-8) with independence assumption (A-2) may not exist and hence the notion of policy invariance is not well defined. If independence in Z fails, policy invariance clearly fails.

For example, in the case where $(\varepsilon, U_0, U_1, V) \sim N(0, \Sigma)$, monotonicity is violated and

$$\begin{aligned} \Delta^{MTE}(x, u_D) &= \mu_1(x) - \mu_0(x) + E(U_1 - U_0 \mid U_D = u_D, X = x) \\ &= \mu_1(x) - \mu_0(x) + (\frac{\sigma_{1D} - \sigma_{0D}}{\sigma_D^2}) u_D \end{aligned} \tag{12}$$

where

$$\sigma_D^2 = \sigma_V^2 + \sum_{i=1}^{K} \sum_{j=1}^{K} \{[Cov\,(Z_i, Z_j) + E(Z_i)E(Z_j)]\, Cov\,(\varepsilon_i, \varepsilon_j)\}$$

and K is the number of components of Z. In this case, Δ^{MTE} depends on the distribution of Z, and is not policy invariant, provided $K \geq 2$ and more than one component of γ is nondegenerate. Even if we restrict $\gamma > 0$ (or $\gamma < 0$), we do not obtain monotonicity and we violate policy invariance in the case of vector Z.[63]

Thus, in the general case, the requirement of policy invariance (A-8) rules out non-monotonicity in the choice equation arising from a random coefficient structure because it leads to a violation of independence (A-2). It rules out heterogeneity in treatment choice responses with respect to Z ($\mu_D(Z)$ random given $Z = z$) although it clearly is compatible with heterogeneity in responses to treatment in general cases. This is a fundamental asymmetry imposed by the requirement of policy invariance.

The monotonicity condition and the additional condition of positive weights for MTE are both required to obtain gross treatment effects using IV. If these conditions are violated, changes in Z induce two way flows with some people changing into treatment and others leaving it. Thus we do not identify the "gross effect" of treatment. Recall from our discussion in Section 4.3 that even if we have monotonicity as defined in this section (a necessary condition for the existence of representation (4)) we may still obtain negative IV weights.[64]

These conditions are invoked when the treatment (indicated by D) is the policy being evaluated. But treatments are only a subset of all possible policies of interest and if the goal is to evaluate the effects of a policy on aggregate outcomes, as in Δ^{PRTE}, the monotonicity requirement is artificial. One is interested in the net impact of the policy and not the impact of treatment operating through a particular mechanism.

Consider the case where D indicates schooling, which is the treatment. $D = 1$ if the person goes to

[63] In the scalar case, with $\gamma > 0$, we can write the choice rule as $Z\gamma > V \Longrightarrow Z > \frac{V}{\gamma}$. In our notation, $U_D = \frac{V}{\gamma}$ and we obtain both monotonicity and policy invariance. In the case of vector γ that is nondegenerate, this trick is not possible unless only one component of Z is nondegenerate.

[64] Indeed, the term monotonicity has multiple meanings in this literature, and they should be carefully distinguished.

college; $D = 0$ otherwise. Suppose that the policy being studied is the introduction of a physical education (PE) requirement in colleges along with mandatory augmented athletics facilities. The policy has no effect on (Y_1, Y_0) (e.g. potential earnings) but it affects the choice of college, so it is a valid Z. Some people hate PE while others love it and are attracted by colleges with good gyms, so monotonicity is violated. If $Z_a = z$ is the policy with PE and $Z_{a'} = z'$ is the policy without PE, $E(Y|Z_a = z) - E(Y|Z_{a'} = z')$ is a perfectly valid policy parameter—the effect of the policy on aggregate outcomes—even if monotonicity is violated and Δ^{MTE} is not policy invariant. From the vantage point of Δ^{PRTE}, monotonicity is an unnecessary requirement.

The treatment effect literature focuses on a class of policies that move treatment choices in the same direction for everyone. We have shown that general instruments do not have universally positive weights on Δ^{MTE} and hence are not guaranteed to shift everyone in the same direction and hence estimate a gross treatment effect. However, the effect of treatment is not always the parameter of policy interest. Thus, in our example, schooling is the vehicle through which policy operates. One might be interested in the effect of schooling (the treatment effect) or the effect of the policy. These are separate questions unless the policy is the treatment.

It should also be noted that violation of monotonicity and non-invariance are not necessarily fatal to the use of Δ^{MTE} in the fashion we advocate, if the analyst models how Δ^{MTE} is affected by shifts in the distribution of Z. As (12) makes clear, it is possible to model how Δ^{MTE} is affected by these shifts. But such modeling is currently out of bounds in the treatment effect literature and takes us into the structural equations literature which explicitly models such effects.

5.2 Constructing Weights for New Policies in a Common Environment

The problem of constructing Δ^{PRTE} for policy a' (compared to a) in environment e when $(a', e) \notin \mathcal{H}$ entails constructing $E_{a'}(V(Y))$. (We maintain the assumption that the baseline policy is observed, so $(a, e) \in \mathcal{H}$.) Under policy invariance for Δ^{MTE}, this entails constructing $F_{P_{a'}|X_e}$ from the policy histories $\mathcal{H}_e$, defined as the elements of $\mathcal{H}$ for a particular environment e.

Associated with the policy histories $a \in \mathcal{H}_e$ is a collection of policy variables $\{Z_a : a \in \mathcal{H}_e\}$. Suppose that a new policy a' can be written as $Z_{a'} = T_{a',j}(Z_j)$ for some $j \in \mathcal{H}_e$ where $T_{a',j}$ is a known deterministic transformation and $Z_{a'}$ has the same coordinates as Z_j. Examples of policies that can be characterized in this way are tax and subsidy policies on wages, prices and incomes that affect unit costs (wages or prices) and transfers. Tuition might be shifted upward for everyone by the same amount, or tuition might be shifted according to a nonlinear function of current tuition, parents' income, and other observable characteristics in Z_j.

To construct $F_{P_{a'}|X_e}$ from data in the policy history entails two distinct steps. From the definitions, $\Pr(P_{a'} \leq t \mid X_e) = \Pr(Z_{a'} : \Pr(D_{a'} = 1 \mid Z_{a'}, X_e) \leq t \mid X_e)$. If (a) we know the distribution of $Z_{a'}$, and (b) we know the function $\Pr(D_{a'} = 1 \mid Z_{a'} = z, X_e = x)$ over the appropriate support, we then can recover the distribution of $P_{a'}$ conditional on X_e. Given that $Z_{a'} = T_{a',j}(Z_j)$ for a known function $T_{a',j}(\cdot)$, step (a) is straightforward since we recover the distribution of $Z_{a'}$ from the distribution of Z_j using the fact that $\Pr(Z_{a'} \leq t) = \Pr(Z_j : T_{a',j}(Z_j) \leq t)$. Alternatively, part of the specification of the policy a' might be distribution $\Pr(Z_{a'} \leq t)$. Then constructing $F_{P_{a'}|X_e}$ is straightforward if $P_{a'}$ is known. We now turn to the second step, recovering the function $\Pr(D_{a'} = 1 \mid Z_{a'} = z, X_e = x)$ over the appropriate support.

If $Z_{a'}$ and Z_j contain the same elements though with possibly different distributions, then a natural approach is to postulate that

$$P_j(z) = \Pr(D_j = 1 \mid Z_j = z, X_e = x) = \Pr(D_{a'} = 1 \mid Z_{a'} = z, X_e = x) = P_{a'}(z). \tag{13}$$

i.e. that over a common support for Z_j and $Z_{a'}$ the known probability and the desired probability agree. Condition (13) will hold, for example, if $D_j = D_{Z_j} = \mathbf{1}[\mu_D(Z_j) - U_D \geq 0]$, $D_{a'} = D_{Z_{a'}} = \mathbf{1}[\mu_D(Z_{a'}) - U_D \geq 0]$, $Z_j \perp\!\!\!\perp U_D$, and $Z_{a'} \perp\!\!\!\perp U_D$. Even if condition (13) is satisfied on a common support, the support of Z_j and $Z_{a'}$ may not be the same. If the support of the distribution of $Z_{a'}$ is not contained in the support of the distribution of Z_j, then some form of extrapolation is needed. Alternatively, if we strengthen our assumptions to have (13) hold for all $j \in \mathcal{H}$, then we can identify $P_{a'}(z)$ for all z in $\bigcup_{j \in \mathcal{H}_e} Supp\,(Z_j)$. However, there is no guarantee that the support of the distribution of $Z_{a'}$ will be contained in $\bigcup_{j \in \mathcal{H}_e} Supp\,(Z_j)$, in which case again some form of extrapolation is needed.

If extrapolation is required, then one approach is to assume a parametric functional form for $P_j(\cdot)$. Given a parametric function form, one can use the joint distribution of (D_j, Z_j) to identify the unknown parameters of $P_j(\cdot)$ and then extrapolate the parametric functional form to evaluate $P_j(\cdot)$ for all evaluation points in the support of $Z_{a'}$. Alternatively, if there is overlap between the support of $Z_{a'}$ and Z_j,[65] so there is some overlap in the historical and policy a' supports of Z, we may use nonparametric methods presented in Matzkin (1994) with functional restrictions (e.g. homogeneity) to construct the desired probabilities on new supports or to bound them. Under the appropriate conditions, we may use analytic continuation to extend $\Pr(D_j = 1 | Z_j = z, X_e = x)$ to a new support. (Rudin, 1974).

The above approach is based on the assumption stated in equation (13). That assumption is quite natural when $Z_{a'}$ and Z_j all contain the same elements, say both contain tuition and parent's income. However, in some cases $Z_{a'}$ might contain additional elements not contained in Z_j. As an example, $Z_{a'}$

[65] If we strengthen condition (13) to hold for all $j \in \mathcal{H}$, then the condition becomes that $Supp\,(Z_{a'}) \cap \bigcup_{j \in \mathcal{H}_e} Supp\,(Z_j)$ is not empty

might include new user fees while Z_j consists of taxes and subsidies but does not include user fees. In this case, the assumption stated in equation (13) is not expected to hold and is not even well defined if $Z_{a'}$ and Z_j contain a different number of elements.

A more basic approach analyzes a class of policies that operate on constraints, prices and endowments arrayed in vector C. Given the preferences and technology of the agent, a given $C = c$, however arrived at, generates the same choices for the agent. Thus a wage tax offset by a wage subsidy of the same amount produces a wage that has the same effect on choices as a no-policy wage. Policy j affects C (e.g. it affects prices paid, endowments and constraints). Define a map $\Phi_j : Z_j \longrightarrow C_j$ which maps a policy j, described by Z_j, into its consequences (C_j) for the baseline, fixed-dimensional vector C. A new policy a', characterized by $Z_{a'}$, produces $C_{a'}$ that is possibly different from all previous policies $j \in \mathcal{H}_e$.

To construct random variable $P_{a'} = \Pr(D_{a'} = 1 \mid Z_{a'}, X_e)$, we postulate that

$$\begin{aligned}
\Pr(D_j = 1 \mid Z_j = z_j, X_e = x) &= \Pr(D_j = 1 \mid C_j = \Phi_j(Z_j) = c, X_e = x) \\
&= \Pr(D_{a'} = 1 \mid C_{a'} = c, X_e = x) \\
&= \Pr(D_{a'} = 1 \mid \Phi_{a'}(Z_{a'}) = c, X_e = x).
\end{aligned}$$

Given these assumptions, our ability to recover $\Pr(D_{a'} = 1 \mid Z_{a'} = z, X_e = x)$ for all z in the support of $Z_{a'}$ depends on what Φ_j functions have been historically observed-how rich the histories are of C_j, $j \in \mathcal{H}_e$. For each $Z_{a'} = z_{a'}$, there is a corresponding $\Phi_{a'}(z_{a'}) = c$. If, in the policy histories, there is at least one $Z_j = z_j$, $j \in \mathcal{H}_e$ such that $\Phi_j(z_j) = c$ then

$$\begin{aligned}
\Pr(D_{a'} = 1 \mid Z_{a'} = z_{a'}, X_e = x) &= \Pr(D_{a'} = 1 \mid \Phi_{a'}(Z_{a'}) = c, X_e = x) \\
&= \Pr(D_j = 1 \mid C_j = c, X_e = x) \\
&= \Pr(D_j = 1 \mid \Phi_j(Z_j) = c, X_e = x)
\end{aligned}$$

and we can construct the probability of the new policy from data in the policy histories. The methods used to extrapolate $P_{a'}(\cdot)$ over new regions, previously discussed, apply here. If the distribution of $C_{a'}$ (or $Z_{a'}$) is known as part of the specification of the proposed policy, the distribution of $F_{P_{a'}|X_e}$ can be constructed using the constructed $P_{a'}$. Alternatively, if we can relate $C_{a'}$ to C_j by $C_{a'} = \Psi_{a',j}(C_j)$ or $Z_{a'}$ to Z_j by $Z_{a'} = T_{a',j}(Z_j)$ and the distributions of C_j and/or Z_j are known for some $j \in \mathcal{H}_e$, we can apply the method previously discussed to derive $F_{P_{a'}|X_e}$ and hence the policy weights for the new policy.

This approach assumes that a new policy acts on components of C like a policy in $\mathcal{H}_e$, so it is possible to forecast the effect of a policy with nominally new aspects. The essential idea is to recast the new

aspects of policy in terms of old aspects previously measured. Thus in a model of schooling, let $D = \mathbf{1}[Y_1 - Y_0 - B \geq 1]$ where $Y_1 - Y_0$ is the discounted gain in earnings from going to school and B is the tuition cost. Here the effect of cost is just the negative of the effect of return. Historically we might only observe variation in $Y_1 - Y_0$ (say tuition has never previously been charged). But B is on the same footing (has the same effect on choice, except for sign) as $Y_1 - Y_0$. This identified historical variation in $Y_1 - Y_0$ can be used to non-parametrically forecast the effect of introducing B provided that the support of $P_{a'}$ is in the historical support generated by the policy histories in $\mathcal{H}_e$. Otherwise, some functional structure (parametric or semi-parametric) must be imposed to solve the support problem for $P_{a'}$.

As another example, following Marschak (1953), consider the introduction of wage taxes in a world where there has never before been a tax. Let Z_j be the wage without taxes. We seek to forecast a post-tax net wage $Z_{a'} = (1 - \tau) Z_j + b$ where τ is the tax rate and b is a constant shifter. Thus $Z_{a'}$ is a known linear transformation of policy Z_j. We can construct $Z_{a'}$ from Z_j. We can forecast under (A-2) using $\Pr(D_j = 1 \mid Z_j = z) = \Pr(D_{a'} = 1 \mid Z_{a'} = z)$. This assumes that the response to after tax wages is the same as the response to wages at the after tax level. At issue is whether $P_{a'|X_e}$ lies in the historical support, or whether extrapolation is needed. Nonlinear versions of this example can be constructed.

As a final example, environmental economists use variation in one component of cost (e.g. travel cost) to estimate the effect of a new cost (*e.g.* a park registration fee). See Smith and Banzhaf (2003). Relating the costs and characteristics of new policies to the costs and characteristics of old policies is a standard, but sometimes controversial method for forecasting the effects of new policies.

In the context of our model, extrapolation and forecasting are confined to constructing $P_{a'}$ and its distribution. If policy a', characterized by vector $Z_{a'}$, consists of new components that cannot be related to Z_j, $j \in \mathcal{H}_e$, or a base set of characteristics whose variation can be identified, the problem is intractable. Then $P_{a'}$ and its distribution cannot be formed using econometric methods applied to historical data.

When it can be applied, our approach allows us to simplify the policy forecasting problem and concentrate our attention on forecasting choice probabilities and their distribution in solving the policy forecasting problem. We can use choice theory and choice data to construct these objects to forecast the impacts of new policies, by relating new policies to previously experienced policies.

5.3 Forecasting the Effects of Policies in New Environments

When the effects of policy a are forecast for a new environment e' from baseline environment e, and $X_e \neq X_{e'}$, in general both $\Delta^{MTE}(x, u_D)$ and $F_{P_a|X_e}$ will change. In general, neither object is environment invariant. The new $X_{e'}$ may have a different support than X_e or any other environment in $\mathcal{H}$. In addition, the new $(X_{e'}, U_D)$ stochastic relationship may be different from the historical (X_e, U_D) stochastic

relationship. Constructing $P_a|X_{e'}$ from $P_a|X_e$ and $F_{Z_a|X_{e'}}$ from $F_{Z_a|X_e}$ can be done using (a) functional form (including semiparametric functional restrictions) or (b) analytic continuation methods. Notice that the maps $T_{a,j}$ and Φ_a may depend on X_e and so the induced changes in these transformations must also be modeled.

Forecasting new stochastic relationships between $X_{e'}$ and U_D is a difficult task. It can be avoided if we invoke the traditional exogeneity assumptions of classical econometrics:

(A-9) $V \perp\!\!\!\perp (X_e, Z_a) \quad \forall e, a \quad \text{so} \quad U_D \perp\!\!\!\perp X_e, Z_a$

where $D_a = 1\left(\mu_D\left(Z_a\right) > V \mid X_e = x\right)$. Under (A-9), we only encounter the support problems for both Δ^{MTE} and the distribution of $\Pr\left(D_a = 1 \mid Z_a, X_e\right)$.

Conditions (A-8) and (A-9) are unnecessary if the only goal of the analysis is to establish internal validity, the standard objective of the treatment effect literature. Autonomy and exogeneity conditions become important issues if we seek external validity. An important lesson from this section is that as we try to make the treatment effect literature do the tasks of structural econometrics (i.e. make out of sample forecasts), the assumptions invoked in the two literatures come together.

5.4 A Comparison of Three Approaches

Table V compares the strengths and limitations of the three approaches texogeneity assumptions, unless nonparametric versions of invariance and exogeneity assumptions are made. However, Δ^{MTE} is comparable across populations with different distributions of P (conditional on X_e) and results from one population can be applied to another population under the conditions presented in this section. Analysts can use Δ^{MTE} to forecast a variety of policies. This invariance property is shared with conventional structural parameters. Our framework solves the problem of external validity which is ignored in the standard treatment effect approach. The price of these advantages of the structural approach is the greater range of econometric problems that must be solved. They are avoided in the conventional treatment approach at the cost of producing parameters that cannot be linked to well-posed economic models and hence do not provide building blocks for an empirically motivated general equilibrium analysis or for investigation of the impacts of new public policies. Δ^{MTE} estimates the preferences of the agents being studied and provides a basis for integration with well posed economic models. If the goal of a study is to examine one policy in place (the problem of internal validity) the stronger assumptions invoked in this section of the paper, and in structural econometrics, are unnecessary. Even if this is the only goal of the analysis, however, our approach allows the analyst to generate all treatment effects and IV estimands from a common parameter and provides a basis for unification of the treatment effect literature.

6 Ordered Choice Extensions

The preceding analysis was for binary choices and outcomes. Yet in a variety of contexts, multiple outcomes arise. For example, there are many grades of school and these may have different returns.

In this section, we develop an ordered choice model extension of the *MTE*. Ordered choice models arise in many settings. In schooling models, there are multiple grades. One has to complete grade $s-1$ to attain grade s. The ordered choice model has been shown to fit well data on schooling transitions and its nonparametric identifiability has been studied (Cameron and Heckman, 1998, and Carneiro, Hansen and Heckman, 2003). There is a sound economic justification for it.

Potential outcomes are written as

$$Y_s = \mu_s(X, U_s) \qquad s = 1, ..., \bar{S}.$$

We define latent variables

$$D_S^* = \mu_D(Z) - V$$

where

$$D_s = \mathbf{1}[C_{s-1}(W_{s-1}) < \mu_D(Z) - V \leq C_s(W_s)], \qquad s = 1, ..., \bar{S},$$

and

$$C_{s-1}(W_{s-1}) \leq C_s(W_s), \quad C_0(W_0) = -\infty \quad \text{and} \quad C_{\bar{S}}(W_{\bar{S}}) = \infty.$$

Observed outcomes are:

$$Y = \sum_{s=1}^{\bar{S}} Y_s D_s.$$

The Z shift the index generally, the W_s affect s-specific transitions. Thus, in a schooling example, the Z could be family background while a W_s could be college tuition or wage opportunities in unskilled labor.[66] Collect the W_s into $W = (W_1, ..., W_{\bar{S}})$, and the U_s into $U = (U_1, ..., U_{\bar{S}})$. Larger values of $C_s(W_s)$ make it more likely that $D_s = 1$. The inequality restrictions on the $C_s(W_s)$ functions play a critical role in defining the model and producing its statistical implications.

We assume

(A-10) $E(|Y_s|) < \infty$, $s = 1, ..., \bar{S}$

(A-11) The distribution of V is absolutely continuous with respect to Lebesgue measure

[66] Many of the instruments studied by Card (2001) are transition-specific but his model is not sufficiently rich to make the distinction between the Z and the W.

(A-12) $(U_s, V) \perp\!\!\!\perp (Z, W)|X$, $s = 1, ..., \bar{S}$

(A-13) $\mu_D(Z)$ is a nondegenerate random variable conditional on X and W.

(A-14) For $s = 1, ..., \bar{S} - 1$, the distribution of $C_s(W_s)$ conditional on X, Z and the other $C_j(W_j)$, $j = 1, ..., \bar{S}$ $j \neq s$, is nondegenerate and is absolutely continuous with respect to Lebesgue measure.

(A-15) $0 < \Pr(D_s = 1|X) < 1$ for $s = 1, ..., \bar{S}$.

Assumption (A-10), (A-11), (A-12), (A-13) and (A-15) play roles analogous to their counterparts in the two outcome model. (A-14) is a new condition that is key to identification of the Δ^{MTE} defined below for each transition. A necessary condition for (A-14) to hold is that there is at least one continuous element of W_s that is not an element of X, Z, or W_j for $j \neq s$. Intuitively, one needs an instrument (or source of variability) for each transition. The continuity of the regressor allows us to differentiate with respect to $C_s(W_s)$, and is not needed except when we seek to estimate *MTE*.

Angrist and Imbens (1995) consider an ordered model and present independence and monotonicity conditions that generalize their earlier work, but they do not consider estimation of transition-specific parameters as we do, or even transition-specific *LATE*. Vytlacil (2003) shows that their conditions imply (and are implied by) a more general version of the ordered choice model with stochastic thresholds, which appear in Heckman, LaLonde and Smith (1999) and Carneiro, Hansen and Heckman (2003). We develop the analysis of this section for the more general model in Appendix D.

The conditional probability of $D_s = 1$ is

$$\Pr(D_s = 1 \mid W, Z, X) \equiv P_s(Z, W) = \Pr(C_{s-1}(W_{s-1}) < \mu_D(Z) - V \leq C_s(W_s)).$$

Analogous to the binary case, we can define

$$U_D = F_V(V)$$

so $U_D \sim \text{Unif}[0, 1]$ under our assumption that the distribution of V is absolutely continuous with respect to Lebesgue measure. The probability integral transformation is somewhat less useful in the ordered choice case so we work with both U_D and V in this section. Monotonic transformations of V induce monotonic transformations of $\mu_D(Z) - C_s(W_s)$, so the relative scale of $\mu_D(Z)$ and $C_s(W_s)$ is invariant to monotonic transformations and one is not free to form arbitrary monotonic transformations of $\mu_D(Z)$ and $C_s(W_s)$ separately. The expression for choice s is

$$D_s = \mathbf{1}\left[F_V\left(\mu_D(Z) - C_{s-1}(W_{s-1})\right) > U_D \geq F_V\left(\mu_D(Z) - C_s(W_s)\right)\right].$$

Keeping the conditioning on X implicit, we define

$$P_s(Z,W) = F_V(\mu_D(Z) - C_{s-1}(W_{s-1})) - F_V(\mu_D(Z) - C_s(W_s)).$$

It is convenient to work with

$$\pi_s(Z,W_s) = F_V(\mu_D(Z) - C_s(W_s)) = \Pr\left(\left[\sum_{j=s+1}^{\bar{S}} D_j = 1\right] \middle| Z, W_s\right)$$

so $\pi_{\bar{S}}(Z,W_{\bar{S}}) = 0$, $\pi_0(Z,W_0) = 1$ and $P_s(Z,W) = \pi_{s-1}(Z,W_{s-1}) - \pi_s(Z,W_s)$. The transition-specific Δ^{MTE} for transition s to $s+1$ is defined as

$$\Delta^{MTE}_{s,s+1}(x,u_D) = E(Y_{s+1} - Y_s \mid X = x, U_D = u_D), \qquad s = 1, ..., \bar{S} - 1.$$

When we set $u_D = \pi_s(Z,W_s)$, as a consequence of (A-13) we obtain the mean return to persons indifferent between s and $s+1$ at mean level of utility $\pi_s(Z,W_s)$. While we have defined MTE in terms of U_D and not in terms of V, there is a one-to-one relationship between U_D and V so that

$$E(Y_{s+1} - Y_s \mid X = x, V = v) = E(Y_{s+1} - Y_s \mid X = x, U_D = F_V(v)) = \Delta^{MTE}_{s,s+1}(x, F_V(v)).$$

In this notation, keeping X implicit,

$$\begin{aligned} E(Y|Z,W) &= \sum_{s=1}^{\bar{S}} E(Y_s \mid D_s = 1, Z, W) \Pr(D_s = 1 \mid Z, W) \\ &= \sum_{s=1}^{\bar{S}} \int_{\pi_s(Z,W_s)}^{\pi_{s-1}(Z,W_{s-1})} E(Y_s \mid U_D = u_D) du_D \end{aligned} \tag{14}$$

where we have used conditional independence assumption (A-12). We thus obtain the index sufficiency restriction $E(Y|Z,W) = E(Y \mid \pi(Z,W))$, where $\pi(Z,W) = [\pi_1(Z,W_1), ..., \pi_{\bar{S}-1}(Z,W_{\bar{S}-1})]$. This restriction is the ordered choice analog of the index sufficiency restriction in the binary outcome model.

We can identify $\pi_s(z,w_s)$ for (z,w_s) in the support of the distribution of (Z,W_s) from the relationship $\pi_s(z,w_s) = \Pr(\sum_{j=s+1}^{\bar{S}} D_j = 1 \mid Z = z, W_s = w_s)$. Thus $E(Y \mid \pi(Z,W) = \pi)$ is identified for all π in the support of $\pi(Z,W)$. Assumptions (A-10), (A-11), and (A-12) jointly allow one to use Lebesgue's theorem for the derivative of an integral to show that $E(Y \mid \pi(Z,W) = \pi)$ is differentiable in π for (*a.e.*) π, and thus we can recover $\frac{\partial}{\partial \pi} E(Y \mid \pi(Z,W) = \pi)$ for almost all π that are limit points of the support of distribution of $\pi(Z,W)$. Under assumption (A-14), all points in the support of the distribution of $\pi(Z,W)$

will be limit points of that support, and we thus have that $\frac{\partial}{\partial \pi} E(Y \mid \pi(Z,W) = \pi)$ is well defined and is identified for $(a.e.)$ π, with

$$\frac{\partial E(Y \mid \pi(Z,W) = \pi)}{\partial \pi_s} = \Delta^{MTE}_{s,s+1}(U_D = \pi_s) = E(Y_{s+1} - Y_s \mid U_D = \pi_s). \tag{15}$$

Equation (15) is the basis for identification of the transition-specific *MTE*.

We may use (14) to obtain

$$\begin{aligned} E(Y \mid \pi(Z,W) = \pi) &= \sum_{s=1}^{\bar{S}} E(Y_s \mid \pi_s \leq U_D < \pi_{s-1})(\pi_{s-1} - \pi_s) \\ &= \sum_{s=1}^{\bar{S}-1} \left[E(Y_{s+1} \mid \pi_{s+1} \leq U_D < \pi_s) - E(Y_s \mid \pi_s \leq U_D < \pi_{s-1}) \right] \pi_s + E(Y_1 \mid \pi_1 \leq U_D < 1) \\ &= \sum_{s=1}^{\bar{S}-1} \{ m_{s+1}(\pi_{s+1}, \pi_s) - m_s(\pi_s, \pi_{s-1}) \} \pi_s + E(Y_1 \mid \pi_1 \leq U_D < 1) \end{aligned}$$

where $m_s(\pi_s, \pi_{s-1}) = E[Y_s \mid \pi_s \leq U_D < \pi_{s-1}]$. In general this expression is a nonlinear function of (π_s, π_{s-1}). This model has a testable restriction of index sufficiency in the general case: $E(Y|\pi(Z,W) = \pi)$ is a nonlinear function additive in functions of (π_s, π_{s-1}) so there are no interactions between π_s and $\pi_{s'}$ if $|s - s'| > 1$, i.e.,

$$\frac{\partial^2}{\partial \pi_s \pi_{s'}} E(Y \mid \pi(Z,W) = \pi) = 0 \qquad \text{if} \qquad |s - s'| > 1.$$

Observe that if $U_D \perp\!\!\!\perp U_s$ for $s = 1, ..., \bar{S}$,

$$E(Y \mid \pi(Z,W) = \pi) = \sum_{s=1}^{\bar{S}} E(Y_s)(\pi_{s-1} - \pi_s) = \sum_{s=1}^{\bar{S}-1} \left[E(Y_{s+1}) - E(Y_s) \right] \pi_s + E(Y_1).$$

Defining $E(Y_{s+1}) - E(Y_s) = \Delta^{ATE}_{s,s+1}$,

$$E(Y \mid \pi(Z,W) = \pi) = \sum_{s=1}^{\bar{S}-1} \Delta^{ATE}_{s,s+1} \pi_s + E(Y_1).$$

Thus, under full independence, we obtain linearity of the conditional mean in the π_s.

The policy relevant treatment effect is defined analogously to the way it was defined in the binary case. As before, let a, a' be two policies. They are defined by distributions of (Z,W), $F^a(Z,W)$. We compare two policies by forming

$$E_a(Y) = \int E(Y \mid Z = z, W = w) dF^a_{Z,W}(z,w)$$

for each policy a. We assume that (Z,W) are nondegenerate random vectors. Let $l_s(Z, W_s) = \mu_D(Z) -$

$C_s(W_s)$, and let $H_s^a(\cdot)$ denote the cumulative distribution function of $l_s(Z, W_s)$ under regime a,

$$H_s^a(t) = \int \mathbf{1}[\mu_D(z) - C_s(w_s) \leq t] dF_{Z,W}^a(z, w).$$

Given that $C_0(W_0) = -\infty$ and $C_{\bar{S}}(W_{\bar{S}}) = \infty$, $l_0(Z, W_0) = \infty$ and $l_{\bar{S}}(Z, W_{\bar{S}}) = -\infty$, and thus $H_0^a(t) = 0$ and $H_{\bar{S}}^a(t) = 1$ for any policy a and for all evaluation points. Since $l_{s-1}(Z, W_{s-1})$ is always larger than $l_s(Z, W_s)$, we have that

$$\mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] = \mathbf{1}[V < l_{s-1}(Z, W_{s-1})] - \mathbf{1}[V \leq l_s(Z, W_s)]. \tag{16}$$

Using this equality along with independence assumption (A-12), we obtain

$$E_a\left(\mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] \mid V\right) = H_s^a(V) - H_{s-1}^a(V). \tag{17}$$

Thus we obtain

$$\begin{aligned}
E_a(Y) &= E_a\left[E(Y \mid V, Z, W)\right] \\
&= E_a\left[\sum_{s=1}^{\bar{S}} \mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] E(Y_s | V, Z, W)\right] \\
&= \sum_{s=1}^{\bar{S}} E_a\left[\mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] E(Y_s | V)\right] \\
&= \sum_{s=1}^{\bar{S}} E_a\left[E(Y_s | V)\{H_s^a(V) - H_{s-1}^a(V)\}\right] \\
&= \sum_{s=1}^{\bar{S}} \int \left[E(Y_s | V = v)\{H_s^a(v) - H_{s-1}^a(v)\}\right] dF_V(v),
\end{aligned}$$

where the first equality is from the law of iterated expectations; the second equality comes from the definition of Y; the third equality follows from linearity of expectations and independence assumption (A-12); the fourth equality applies the law of iterated expectations and equation (17); and the final equality rewrites the expectation explicitly as an integral against the distribution of V. Recalling that $H_0^a(v) = 0$ and $H_{\bar{S}}^a(v) = 1$, we may rewrite this result as

$$E_a(Y) = \sum_{s=1}^{\bar{S}-1} \int E(Y_s - Y_{s+1} | V = v) H_s^a(v) dF_V(v) + \int E\left(Y_{\bar{S}} | V = v\right) dF_V(v),$$

where the last term is $E\left(Y_{\bar{S}}\right)$. Hence, comparing two policies under a and a', we obtain

$$\begin{aligned}\Delta_{a,a'}^{PRTE} &= E_{a'}(Y) - E_a(Y) \\ &= \sum_{s=1}^{\bar{S}-1} \int E(Y_{s+1} - Y_s | V = v) \left[H_s^a(v) - H_s^{a'}(v) \right] dF_V(v).\end{aligned}$$

Alternatively, we can express this in terms of Δ^{MTE}:

$$\Delta_{a,a'}^{PRTE} = \sum_{s=1}^{\bar{S}-1} \int \Delta_{s,s+1}^{MTE}(u) \left[\tilde{H}_s^a(u) - \tilde{H}_s^{a'}(u) \right] du$$

where $\tilde{H}_s^a(t)$ is the cumulative distribution function of $F_V(\mu_D(Z) - C_s(W_s))$ under policy a, $\tilde{H}_s^a(t) = \int \mathbf{1}[F_V(\mu_D(z) - C_s(w_s)) \leq t] dF_{Z,W_s}^a(z, w_s)$.

Paralleling our discussion in Section 4, we next characterize what instrument $J(Z, W)$ identifies and contrast it with what is required to determine $\Delta_{a,a'}^{PRTE}$. It is common practice to regress Y on D where D is completed years of schooling and call the coefficient on D a rate of return when Y is log earnings. This is the standard approach to earnings functions pioneered by Mincer (1974). We seek an expression for the instrumental variable estimator of the effect of D on Y:

$$\frac{Cov\left(J(Z, W), Y\right)}{Cov\left(J(Z, W), D\right)} \tag{18}$$

where $D = \sum_{s=1}^{\bar{S}} s D_s$. We keep conditioning on X implicit. We first derive $Cov\left(J(Z, W), Y\right)$. Its derivation is typical of the other terms needed to form (18). Defining $\tilde{J}(Z, W) = J(Z, W) - E(J(Z, W))$, we obtain, since $Cov\left(J(Z, W), Y\right) = E\left(\tilde{J}\left(Z, W\right), Y\right)$,

$$\begin{aligned}E(\tilde{J}(Z, W)Y) &= E\left[\tilde{J}(Z, W) \sum_{s=1}^{\bar{S}} \mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] E(Y_s | V, Z, W) \right] \\ &= \sum_{s=1}^{\bar{S}} E\left[\tilde{J}(Z, W) \mathbf{1}[l_s(Z, W_s) \leq V < l_{s-1}(Z, W_{s-1})] E(Y_s | V) \right]\end{aligned}$$

where the first equality comes from the definition of Y and the law of iterated expectations, and the second equality follows from linearity of expectations and independence assumption (A-12). Let $H_s(\cdot)$ equal $H_s^a(\cdot)$ for a equal to the policy that characterizes the observed data, i.e., $H_s(\cdot)$ is the cumulative distribution function of $l_s(Z, W_s)$,

$$H_s^a(t) = \Pr(l_s(Z, W_s) \leq t) = \Pr(\mu_D(Z) - C_s(W_s) \leq t).$$

Using the law of iterated expectations and equation (16), we obtain

$$\begin{aligned} E(\tilde{J}(Z,W)Y) &= \sum_{s=1}^{\bar{S}} E\left[E\left(\tilde{J}(Z,W)\left\{\mathbf{1}[V < l_{s-1}(Z,W_{s-1})] - \mathbf{1}[V \leq l_s(Z,W_s)]\right\} \,\middle|\, V\right) E(Y_s|V)\right] \\ &= \sum_{s=1}^{\bar{S}} \int \left[E(Y_s \mid V=v)\{K_{s-1}(v) - K_s(v)\}\right] dF_V(v) \\ &= \sum_{s=1}^{\bar{S}-1} \int \left[E(Y_{s+1} - Y_s \mid V=v) K_s(v)\right] dF_V(v) \end{aligned}$$

where $K_s(v) = E\left(\tilde{J}(Z,W) \mid l_s(Z,W_s) > v\right)(1 - H_s(v))$ and we use the fact that $K_{\bar{S}}(v) = K_0(v) = 0$. Now consider the denominator of the IV estimand,

$$\begin{aligned} E(D\tilde{J}(Z,W)) &= E\left[\tilde{J}(Z,W)\sum_{s=1}^{\bar{S}} s\mathbf{1}[l_s(Z,W_s) \leq V < l_{s-1}(Z,W_{s-1})]\right] \\ &= \sum_{s=1}^{\bar{S}} sE\left[\tilde{J}(Z,W)\mathbf{1}[l_s(Z,W_s) \leq V < l_{s-1}(Z,W_{s-1})]\right] \\ &= \sum_{s=1}^{\bar{S}} sE\left[E\left(\tilde{J}(Z,W)\left\{\mathbf{1}[V < l_{s-1}(Z,W_{s-1})] - \mathbf{1}[V \leq l_s(Z,W_s)]\right\} \,\middle|\, V\right)\right] \\ &= \sum_{s=1}^{\bar{S}} s\int \left[K_{s-1}(v) - K_s(v)\right] dF_V(v) = \sum_{s=1}^{\bar{S}-1} \int K_s(v)\, dF_V(v)\,. \end{aligned}$$

Collecting results, we obtain an expression for the IV estimand (18):

$$\frac{Cov(J,Y)}{Cov(J,D)} = \sum_{s=1}^{\bar{S}-1} \int E(Y_{s+1} - Y_s \mid V=v)\omega(s,v)\, dF_V(v)$$

where

$$\omega(s,v) = \frac{K_s(v)}{\sum_{s=1}^{\bar{S}} s\int \left[K_{s-1}(v) - K_s(v)\right] dF_V(v)} = \frac{K_s(v)}{\sum_{s=1}^{\bar{S}-1} \int K_s(v)\, dF_V(v)}$$

and clearly

$$\sum_{s=1}^{\bar{S}-1} \int \omega(s,v)\, dF_V(v) = 1, \qquad \omega(0,v) = 0, \qquad \text{and} \qquad \omega(\bar{S},v) = 0.$$

We can re-express the result in terms of $E(Y_{s+1} - Y_s \mid U_D = u_D) = \Delta^{MTE}_{s-1,s}(u_D)$:

$$\frac{Cov(J,Y)}{Cov(J,D)} = \sum_{s=1}^{\bar{S}-1} \int \Delta^{MTE}_{s,s+1}(u)\tilde{\omega}(s,u)\, du$$

where

$$\tilde{\omega}(s,u) = \frac{\tilde{K}_s(u)}{\sum_{s=1}^{\bar{S}} s \int_0^1 \left[\tilde{K}_{s-1}(u) - \tilde{K}_s(u)\right] du} = \frac{\tilde{K}_s(u)}{\sum_{s=1}^{\bar{S}-1} \int_0^1 \tilde{K}_s(u)\, du} \tag{19}$$

and

$$\tilde{K}_s(u) = E\left(\tilde{J}(Z,W) \mid \pi_s(Z,W_s)) > u\right) \Pr(\pi_s(Z,W_s) \geq u) \tag{20}$$

We again have that the weights integrate to unity and that $\tilde{\omega}(0,v) = 0$, $\tilde{\omega}(\bar{S},v) = 0$.

Compare equations (19, 20) for the ordered choice model to equation (11) for the binary choice model. The numerator of the weights implied on Δ^{MTE} for a particular transition in the ordered choice model are exactly the numerator of the weights implied for the binary choice model, only substituting $\pi_s(Z,W_s) = \Pr(D > s \mid Z,W_s)$ in place of $P(Z) = \Pr(D = 1 \mid Z)$. While the numerator for the weights for IV in the binary choice model is driven by the connection between the instrument and $P(Z)$, the numerator for the weights for IV in the ordered choice model for a particular transition is driven by the connection between the instrument and $\pi_s(Z,W_s)$. The denominator of the weights is the covariance between the instrument and D for both the binary and ordered cases. However, in the binary case the covariance between the instrument and D is completely determined by the covariance between the instrument and $P(Z)$, while in the ordered choice case the covariance depends on the relationship between the instrument and the full vector $[\pi_1(Z,W_1), ..., \pi_{\bar{S}-1}(Z,W_{\bar{S}-1})]$.

From equation (20), the IV estimator using $J(Z,W)$ as an instrument satisfies the following properties: (a) the weights on $\Delta^{MTE}_{s,s+1}$ implied by using $J(Z,W)$ as an instrument are the same as the weights on $\Delta^{MTE}_{s,s+1}$ implied by using $E(J(Z,W) \mid \pi_s(Z,W))$ as the instrument; (b) the numerator of the weights on $\Delta^{MTE}_{s,s+1}(u)$ are non-negative for all u if $E(J(Z,W) \mid \pi_s(Z,W_s) \geq \pi_s)$ is weakly monotonic in π_s; and (c) the support of the weights on $\Delta^{MTE}_{s,s+1}$ using $\pi_s(Z,W_s)$ as the instrument is $(\pi_s^{\text{Min}}, \pi_s^{\text{Max}})$ where π_s^{Min} and π_s^{Max} are the minimum and maximum values in the support of $\pi_s(Z,W_s)$, respectively, and the support of the weights on $\Delta^{MTE}_{s,s+1}$ using any other instrument is a subset of $(\pi_s^{\text{Min}}, \pi_s^{\text{Max}})$.

Restriction (a) states that for any instrument $J(Z,W)$, using $J(Z,W)$ as an instrument leads to the same weights on $\Delta^{MTE}_{s,s+1}(u)$ as using $E\left(J(Z,W) \mid \pi_s(Z,W_s)\right)$ as an instrument. Two instruments J and J^* weight MTE for the s to $s+1$ transition equally at all u_D if and only if $E(J(Z,W) \mid \pi_s(Z,W_s)) - E(J(Z,W)) = E(J^*(Z,W) \mid \pi_s(Z,W_s)) - E(J^*(Z,W))$. However, the two instruments may weight MTE for the s to $s+1$ transition equally and still provide very different weights on MTE for other transitions. Restriction (b) states that, if the covariance between J and D is positive, then using J as an instrument will lead to nonnegative weights on $\Delta^{MTE}_{s,s+1}(u)$ if $E(J(Z,W_s) \mid \pi_s(Z,W_s) \geq \pi_s)$ is weakly monotonic in π_s. For example, if $Cov\left(\pi_s(Z,W_s), D\right) > 0$, then setting $J(Z,W) = \pi_s(Z,W_s)$ will lead to nonnegative weights on $\Delta^{MTE}_{s,s+1}(u)$ (though it may lead to negative weights on other transitions). Restriction (c) states

that using $\pi_s(Z, W_s)$ as an instrument leads to nonzero weights on $\Delta^{MTE}_{s,s+1}(u)$ for $u \in (\pi_s^{\text{Min}}, \pi_s^{\text{Max}})$. Using any other instrument leads to nonzero weights on a subset of $(\pi_s^{\text{Min}}, \pi_s^{\text{Max}})$. Thus, using $\pi_s(Z, W_s)$ as an instrument leads to the nonnegative weights on a larger range of evaluation points for the s to $s+1$ transition than using any other instrument.

The weights are not necessarily positive. Recall that, if the covariance between J and D is positive, then using J as an instrument will lead to nonnegative weights on $\Delta^{MTE}_{s,s+1}(u)$ if $E(J(Z, W) \mid \pi_s(Z, W_s) \geq \pi_s)$ is weakly monotonic in π_s. For example, if the covariance between $J(Z, W)$ and D is positive, then using $J(Z, W)$ will lead to nonnegative weights on $\Delta^{MTE}_{s,s+1}(u)$ if $J(Z, W)$ is a monotonic function of $\pi_s(Z, W_S)$. Clearly, an instrument may produce positive weights on the Δ^{MTE} for one transition but still produce negative weights for the Δ^{MTE} on another transition. In order to gain a greater understanding about the structure of these weights, it is useful to consider a variety of special cases. Suppose first that the the distributions of W_s, $s = 1, \ldots, \bar{S}$, are degenerate so that the C_s are constants with $C_1 < \cdots < C_{\bar{S}-1}$. In this case, $\pi_s(Z, W_s) = F_V(\mu_D(Z) - C_s)$ for any $s = 1, ..., \bar{S}$, and we can specialize the above result to say that using J as an instrument will lead to nonnegative weights on all transitions if $J(Z, W_s)$ is a monotonic function of $\mu_D(Z)$. For example, using $\mu_D(Z)$ itself as the instrument leads to weights on $\Delta^{MTE}_{s,s+1}(u)$ of the form specified above with

$$\tilde{K}_s(u) = \Big[E(\mu_D(Z) \mid \mu_D(Z) > F_V^{-1}(u) + C_s) - E(\mu_D(Z))\Big] \Pr(\mu_D(Z) > F_V^{-1}(u) + C_s).$$

Clearly, these weights will be nonnegative for any evaluation point and will be strictly positive for any evaluation point u such that $1 > \Pr(\mu_D(Z) > F_V^{-1}(u) + C_s) > 0$.

Next consider the case where $C_s(W_s) = W_s$, $s = 1, \ldots, \bar{S} - 1$, and where $\mu_D(Z) = 0$. Suppose that each W_s is a scalar, and consider $J(Z, W) = W_j$, a pure transition-specific instrument. In this case, the weight on $\Delta^{MTE}_{s,s+1}(u)$ is of the form given above with

$$\tilde{K}_s(u) = \Big[E(W_s \mid W_s > F_V^{-1}(u)) - E(W_S)\Big] \Pr(W_s > F_V^{-1}(u)),$$

which will be nonnegative for all evaluation points and strictly positive for any evaluation point such that $1 > \Pr(W_s > F_V^{-1}(u)) > 0$. What will be the implied weights on $\Delta^{MTE}_{s',s'+1}(u)$ for $s' \neq s$? First, consider the case where W_s is independent of $W_{s'}$ for $s \neq s'$. This independence can hold if the supports of W_s and $W_{s'}$ do not overlap for any $s' \neq s$. In this case, the weight on $\Delta^{MTE}_{s',s'+1}(u)$ for $s' \neq s$ is of the form given above with

$$\tilde{K}_{s'}(u) = \Big[E(W_s \mid W_{s'} > F_V^{-1}(u)) - E(W_s)\Big] \Pr(W_{s'} > F_V^{-1}(u)) = 0.$$

Thus, in this case, the instrument will only be weighting the *MTE* for the s to $s+1$ transition. Note that this result relies critically on the assumption that W_s is independent of $W_{s'}$ for $s' \neq s$.

Consider another version of this case where $C_s(W_s) = W_s$, $s = 1, \ldots, \bar{S}-1$, with W_s a scalar, but now allow $\mu_D(Z)$ to have a nondegenerate distribution and allow there to be dependence across the W_s. In particular, consider the case where $F_{W_1,\ldots,W_{\bar{S}-1}}(t_1, ..., t_{\bar{S}-1})$ has a density with respect to Lebesgue measure given by

$$\frac{\prod_{i=1}^{\bar{S}-1} f_i(w_i)}{\int \cdots \int \left[\mathbf{1}[w_1 < w_2 < \cdots < w_{\bar{S}-1}] \prod_{i=1}^{\bar{S}-1} f_i(w_i)\right] dw_1 \cdots dw_{\bar{S}}}.$$

In this case, since w_j is the instrument, we have

$$\omega(s,v) = \frac{\int \cdots \int_{-\infty<w_1<\cdots<w_{\bar{S}-1}<\infty} (w_j - E(w_j)) F_{\mu_D(Z)}(w_s+v) f_1(w_1) \cdots f_{\bar{S}-1}(w_{\bar{S}-1})\, dw_1 \cdots dw_{\bar{S}-1}\, dF_V(v)}{\sum_{s=1}^{\bar{S}-1} \int \cdots \int_{-\infty<w_1<\cdots<w_{\bar{S}-1}<\infty} (w_j - E(w_j)) F_{\mu_D(Z)}(w_s+v) f_1(w_1) \cdots f_{\bar{S}-1}(w_{\bar{S}-1})\, dw_1 \cdots dw_{\bar{S}-1}\, dF_V(v)}.$$

In the special case where $\mu_D(Z) \sim U(-K, K)$, with $Z \perp\!\!\!\perp W_s$ for $s = 1, ..., \bar{S}-1$, the numerator is

$$\begin{aligned} &\int \cdots \int_{-\infty<w_1<\cdots<w_{\bar{S}-1}<\infty} (w_j - E(w_j)) \frac{(w_s+v)}{2K} f_1(w_1) \cdots f_{\bar{S}-1}(w_{\bar{S}-1})\, dw_1 \cdots dw_{\bar{S}-1}\, dF_V(v) \\ = \;& \frac{1}{2K} Cov\,(W_j, W_s | W_1 < \cdots < W_{\bar{S}-1}). \end{aligned}$$

Observe that when $f_s(W) = f_j(W)$ for all j, s, by Bickel's Theorem (1967), we know that this expression is positive. (This is trivial when $j = s$.) The ordering $W_1 < \cdots < W_{\bar{S}-1}$ implies that as $W_j \uparrow$, W_l for $l < j$ is stochastically increasing (the lower boundary is shifted to the right). As $W_l \uparrow$ for $l > j$, W_j is also stochastically increasing. Hence because of the order on the W implied by the ordered discrete choice model, a positive weighting is produced. This result can be overturned when $dF_W(w)$ has a general structure. The positive dependence induced by the order on the components of W can be reversed by the negative dependence in the structure of $dF_W(w)$. We present an example below.

This analysis departs substantially from that of Imbens and Angrist (1994) and Angrist and Imbens (1995) by introducing both transition-specific instruments (the W) and general instruments (Z) across all transitions. In general, the method of linear instrumental variables applied to D does not estimate anything that is economically interpretable. It is not guaranteed to estimate a positive number since the weights can be negative. In contrast, we can use (15), under conditions (A-10)–(A-15), to apply *LIV* to identify Δ^{MTE} transition by transition which can be used to build up Δ^{PRTE}. We develop the more general case of this model with stochastic thresholds in Appendix D.

To illustrate the contrast between what *IV* estimates and what is required for Benthamite policy

analysis, we consider a three-outcome ($\bar{S} = 3$) model with both common instruments (Z) and transition-specific instruments (W). In this example, $\mu_D(Z) = Z$, $C(W_s) = W_s$, $f(w_1, w_2) = N(0, \Sigma_W)$ and $(U_1, U_2, U_3, V) \perp\!\!\!\perp (W, Z)$, $(U_1, U_2, U_3, V) \sim N(0, \Sigma)$. We assume $dF_{W,\mu_D(Z)}(\mu_D(z), w) = dF_{\mu_D}(\mu_D)\, dF_W(w)$. One can think of D as completed schooling, Y_1 as the potential earnings of a dropout, Y_2 as the potential earnings of a high school graduate and Y_3 as the potential earnings of a college graduate. There are two transitions, $1 \to 2$ and $2 \to 3$. The policy consists of changing W_2 to $W_2 - t$. One can think of this as a college tuition reduction policy. We ask how well IV identifies Δ^{PRTE} in this simulated data and we study the weights associated with IV and Δ^{PRTE}. Specific values for the policy IV and counterfactual models are given below. Figure 5A plots the policy invariant Δ^{MTE} for the two transitions. A higher value of $V = v$ is associated with greater cost and a lower probability of being in states $D = 2$ or $D = 3$. We first consider a simulation where W_1 is the instrument and then consider the case where Z is the instrument. The case with W_2 as an instrument is similar and for the sake of brevity is deleted.

Figure 5A plots the Δ^{MTE} for the $1 \to 2$ and $2 \to 3$ transitions. Specific parameter values are presented at the base of the figure. Both of the Δ^{MTE} parameters have the typical shape of declining returns for people less likely to make the transition, i.e., those who have a higher $V = v$. Even though the levels are higher for outcomes 2 and 3, the marginal returns are higher for the transition $1 \to 2$. Figure 5B plots the policy weights for the two transitions for a policy that lowers W_2 ("reduces tuition"). It also plots the IV weights for the two Δ^{MTE} functions for the case where W_1 is the instrument. The correlation pattern for (w_1, w_2) is positive with specific values given below the figure. The policy studied in 5B shifts 42.8% of the $D_1 = 1$ people into the category $D_3 = 1$ and 92.4% of D_2 people into D_3. In this simulation, the IV weights are positive. The IV weights and Δ^{PRTE} weights are distinctly different and the IV estimate is 0.201 vs. Δ^{PRTE} of 0.166.

When we change the correlation structure between W_1 and W_2, so they are negatively correlated (Figure 5C), the IV weight for $\Delta^{MTE}_{1,2}$ becomes *negative* while that for $\Delta^{MTE}_{2,3}$ remains positive. The contrast in these figures between negative and positive IV weights depends on the correlation structure between W_1 and W_2. The stochastic order ($W_2 > W_1$) is a force toward positive weights, which can be undone when the dependence induced by the density ($f(w_1, w_2)$) is sufficiently negative. The discord between the IV and Δ^{PRTE}weights is substantial and is reflected in the estimates ($\Delta^{PRTE} = 0.159$ vs. $IV = 0.296$). As Figure 5D illustrates, the weights on $\Delta^{PRTE}$are not guaranteed to be positive either. Thus neither the IV weights nor the weights on $\Delta^{PRTE}$are guaranteed to be positive or negative and the relation between the two sets of weights can be quite weak.

Figures 6A-6D present a parallel set of simulations when $Z$is used as an instrument. Changes in Zshift persons across all transitions whereas $W_1$is a transition-specific shifter. Figure 6A reproduces the policy

invariant Δ^{MTE}parameters from Figure 5A. Figure 6B shows that the IV weights for $\Delta^{MTE}_{2,3}$assume both positive and negative values. The IV weights for $\Delta^{MTE}_{1,2}$are positive but not monotonic. In Figure 6C, where there is negative dependence between $W_1$and W_2, both sets of IV weights assume both positive and negative values. In the case where $f(w_1, w_2) = f_1(w_1)f_2(w_2)$, the weights on $\Delta^{MTE}_{1,2}$for $\Delta^{PRTE}$are negative.

These simulations, and others not shown that are based on W_2as an instrument, show a rich variety of shapes and signs for the weights. They illustrate a main point of this paper—that standard IV methods are not guaranteed to weight marginal treatment effects positively or to produce estimates close to policy relevant treatment effects or even to produce any gross treatment effect. Estimators based on LIV and its extension to the ordered model (15) identify $\Delta^{MTE}$for each transition and answer policy relevant questions. For further discussion of this model, see Heckman and Vytlacil (2004).

7 Summary and Proposed Extensions

This paper develops an approach to policy evaluation based on the marginal treatment effect (Δ^{MTE}), which provides a choice-theoretic foundation for organizing the treatment effect literature. All of the conventional treatment effect parameters can be expressed as different weighted averages of Δ^{MTE}. These conventional treatment effect parameters do not, in general, answer economically interesting questions. We define the policy relevant treatment effect as the solution to a Benthamite policy criterion for policies operating on decisions to participate, but not on potential outcomes. The policy relevant treatment effect can be represented as a weighted average of $\Delta^{MTE}$where the weights differ, in general from the weights used to generate conventional treatment effects. Thus the conventional treatment effects are not guaranteed to answer policy relevant questions.

Instrumental variable estimators and OLS estimators converge to expressions that can be represented as weighted averages of Δ^{MTE} parameters, with the weights in general different from those used to define the various treatment effects and the weights not necessarily positive so they do not identify a gross treatment effect. We show how to test whether the weights are positive. Conventional IV and matching assumptions impose a strong condition on the Δ^{MTE}—that selection into programs is not made in terms of any unobservable gain from program participation.

We present methods for estimating Δ^{MTE}based on local instrumental variables and we develop a new instrumental variable for recovering policy relevant treatment effects using standard instrumental variable methods. We develop the conditions required to forecast the effects of old policies on new environments and the effects of new policies. These issues are typically ignored in the treatment effect literature but are central to the structural policy evaluation literature. We extend our analysis to a multiple treatment

setting, focusing on an ordered choice model which is a natural generalization of our binary choice model. Carneiro (2002) and Carneiro, Heckman, and Vytlacil (2003), apply these methods to study the marginal and average returns to college attendance for high school graduates using a widely used data set. They find evidence of comparative advantage in the labor market. Their analysis implies that, for the data they analyze, conventional IV methods do not estimate policy relevant treatment effects or conventional treatment effects, and that the method of matching does not recover any of these treatment parameters in this prototypical example.

Appendices

A Monotonicity Restriction

The assumptions imposed in this paper lead to a testable monotonicity restriction. We analyze this restriction after formally proving that the restriction holds under our assumptions.

Proposition 4: *Monotonicity Condition:* If $D = \mathbf{1}[P(Z) \geq U_D]$, with $U_D \sim$Unif$[0,1]$, and conditions (A-1) to (A-5) apply, for $j = 0, 1$, let g_0, g_1 be any functions such that $g_0(Y_0, X), g_1(Y_1, X) \geq 0$w.p.1, then $E((1-D)g_0(Y,X)|X, P(Z) = p)$is weakly decreasing in $p$and $E(Dg_1(Y,X)|X, P(Z) = p)$is weakly increasing in p.

Proof

Consider $E(Dg_1(Y,X)|X = x, P(Z) = p)$for some x. Let p_1, p_0 denote any two points in the support of the distribution of $P(Z)$ conditional on $X = x$such that $p_1 > p_0$. Then

$$\begin{aligned}
E(Dg_1(Y,X)|X &= x, P(Z)=p_1) - E(Dg_1(Y,X)|X = x, P(Z) = p_0) \\
&= E(\mathbf{1}[U_D \leq P(Z)]g_1(Y_1,X)|X = x, P(Z) = p_1) \\
&\quad -E(\mathbf{1}[U_D \leq P(Z)]g_1(Y_1,X)|X = x, P(Z) = p_0) \\
&= E(\mathbf{1}[U_D \leq p_1]g_1(Y_1,X)|X = x) - E(\mathbf{1}[U_D \leq p_0]g_1(Y_1,X)|X = x) \\
&= E(\{\mathbf{1}[U_D \leq p_0] + \mathbf{1}[p_0 < U_D \leq p_1]\}g_1(Y_1,X)|X = x) - E(\mathbf{1}[U_D \leq p_0]g_1(Y_1,X)|X = x) \\
&= E(\mathbf{1}[p_0 < U_D \leq p_1]g_1(Y_1,X)|X = x) \geq 0
\end{aligned}$$

where the first equality follows from the definition of D and uses $Dg_1(Y,X) = Dg_1(Y_1,X)$. The second equality uses independence condition (A-2); the third equality uses the fact that $p_0 < p_1$ and thus that $\mathbf{1}[U_D \leq p_1] = \mathbf{1}[U_D \leq p_0] + \mathbf{1}[p_0 < U_D \leq p_1]$. The fourth equality follows from linearity of expectations. The final inequality follows from $g_1(Y_1,X) \geq 0$ w.p.1. The proof that $E((1-D)g_0(Y,X)|X, P(Z) = p)$ is decreasing in p is symmetric. ■

The proposition is stated for nonnegative $g_0(Y_0,X), g_1(Y_1,X)$ functions. If the condition is strengthened to apply to strictly positive $g_0(Y_0,X), g_1(Y_1,X)$ functions, than a trivial modification of the last line of the proof results in strengthened conclusion that $E((1-D)g_0(Y,X)|X, P(Z) = p)$ is strictly decreasing in p and $E(Dg_1(Y,X)|X, P(Z) = p)$ is strictly increasing in p. Consider the following examples of g_0 and g_1: If Y_1, Y_0 are known to be non-negative (for example, Y_1, Y_0 are indicator variables, or Y_1, Y_0 are wages), then choosing $g_j(Y,X) = Y$, $E((1-D)Y|X, P(Z) = p)$ is weakly decreasing in p and $E(DY|X, P(Z) = p)$ is weakly increasing in p. If Y_1, Y_0 are known to be bounded from below by a function of X, $Y_1 \geq l_1(X), Y_0 \geq l_0(X)$ w.p.1, then choosing $g_j(Y,X) = Y - l_j(X)$ results in $E((1-D)(Y - l_0(X))|X, P(Z) = p)$ being weakly decreasing in p and $E(D(Y - l_1(X))|X, P(Z) = p)$ being weakly increasing in p. Without any

assumptions on Y_1, Y_0, and so relaxing (A-4), let t denote a real number and take $g_j(Y,X) = \mathbf{1}[Y \leq t]$ for $j = 0, 1$. Then the result implies that $\Pr(D = 0, Y \leq t|X, P(Z) = p)$ is weakly decreasing in p and $\Pr(D = 1, Y \leq t|X, P(Z) = p)$ is weakly increasing in p. More generally, let $\mathcal{A}$ denote any measurable subset of the real line, and take $g_j(Y,X) = \mathbf{1}[Y \in \mathcal{A}]$ for $j = 0, 1$. Then the result can be rewritten as $\Pr(D = 0, Y \in \mathcal{A}|X, P(Z) = p)$ is weakly decreasing in p and $\Pr(D = 1, Y \in \mathcal{A}|X, P(Z) = p)$ is weakly increasing in p.

This restriction includes the Imbens-Rubin (1997) restrictions on IV as a special case. Imbens and Rubin assume a binary Z, and obtain the density of Y_1 and Y_0 from the observed data. They derive the testable restriction that these densities be nonnegative. Our analysis is more general.

For ease of exposition, suppress conditioning on X. Take the case where $Z = 0, 1$, and with $P(1) > P(0)$. Consider the Y_1 outcome; the analysis for Y_0 is completely symmetric. For binary Z with $P(1) > P(0)$, our restriction can be rewritten as $E(Dg_1(Y)|Z = 1) \geq E(Dg_1(Y)|Z = 0)$. Take $g_1(Y) = \mathbf{1}[Y \in \mathcal{A}]$ for any pre-specified set $\mathcal{A}$ (for example, the intervals they examine in their histogram). Then in this special case, our monotonicity restriction is that $\Pr(D = 1, Y \in \mathcal{A}|Z = 1) - \Pr(D = 0, Y \in \mathcal{A})|Z = 0) > 0$. This restriction is the same as that of the Imbens and Rubin restriction of a nonnegative density. The only difference is that we replace their densities with the probability that Y lies in any given set. Thus, their restriction is a very special case of the general monotonicity restriction developed in this paper.

B Proof of Equation (7)

Define $\mathbf{1}_{\mathcal{A}}(t)$ to be the indicator function for the event $t \in \mathcal{A}$. Then

$$\begin{aligned}
E_a\left(Y \mid X = x\right) &= \int_0^1 E(Y|X, P_a(Z_a) = p)\, dF_{P_a|X}(p) \\
&= \int_0^1 \left[\int_0^1 [\mathbf{1}_{[0,p]}(u)E(Y_1 \mid X, U_D = u) + \mathbf{1}_{(p,1]}(u)E(Y_0 \mid X, U_D = u)\, du\right] dF_{P_a|X}(p) \\
&= \int_0^1 \left[\int_0^1 [\mathbf{1}_{[u,1]}(p)E(Y_1 \mid X, U_D = u) + \mathbf{1}_{(0,u]}(p)E(Y_0 \mid X, U_D = u)\, dF_{P_a|X}(p)\right] du \\
&= \int_0^1 \left[(1 - F_{P_a|X}(u))E(Y_1 \mid X, U_D = u) + F_{P_a|X}(u)E(Y_0 \mid X, U_D = u)\right] du.
\end{aligned}$$

This derivation involves changing the order of integration. Note that

$$E|\mathbf{1}_{[0,p]}(u)E(Y_1 \mid U_D = u) + \mathbf{1}_{(p,1]}(u)E(Y_0 \mid U_D = u)| \leq E(|Y_1| + |Y_0|) < \infty$$

by assumption (A-4), and thus the change in the order of integration is valid by Fubini's theorem. Thus comparing policy a to policy a',

$$E_a\left(Y \mid X = x\right) - E_{a'}\left(Y \mid X = x\right) = \int_0^1 E(\Delta \mid X, U_D = u_D)(F_{P_{a'}|X}(u_D) - F_{P_a|X}(u_D))\, du_D$$

which gives the required weights. (Recall $D = Y_1 - Y_0$.) Under the assumption that the new policy does not change the distribution of X (this requirement is "autonomy" defined in Section 5), the policy counterfactual not conditioning on X is given by

$$E_a\left(Y\right) - E_{a'}\left(Y\right) = E_X\left[\int_0^1 E(\Delta \mid X, U_D = u_D)(F_{P_{a'}|X}(u_D) - F_{P_a|X}(u_D))\, du_D\right].$$

C Proofs of Propositions

Proof of Proposition 1. We first show that, given (a) and (b), assumptions (c) and (d) are sufficient for the instrument $J(Z)$ defined by the proposition to have the desired properties. As a preliminary step, note that with this definition of $J(Z)$,

$$E(J|X=x) = \int_0^1 \mathbf{1}[f_{P|X}(p|x) > 0] w'(p|x) dp = \int_0^1 w'(p|x) dp = 0$$

where the first equality comes from plugging in the proposed J; the second equality follows from assumption (b); and the final equality follows from assumption (b). We now check that the proposed J is correlated with D under assumptions (a) to (d).

$$Cov\left(J(Z), D \mid X\right) = Cov\left(J(Z), P(Z) \mid X\right) = \int_0^1 \mathbf{1}[f_{P|X}(p|x) > 0] w'(p|x) p dp = \int_0^1 w'(p|x) p dp = -1$$

where the first equality follows from the law of iterated expectations; the second equality comes from plugging in the proposed J and using $E(J|X=x) = 0$; the third equality uses assumption (c) and the final equality follows from assumption (d). We now check that the proposed instrument J implies the desired weights on Δ^{MTE}. With the proposed J, we have that, for u such that $f_{P|X}(u|x) > 0$,

$$-\frac{T(u \mid x; J) f_{P|X}(u|x)}{Cov\left(J, P \mid X = x\right)} = -\frac{w'(u|x)}{-1} = w'(u|x)$$

where the first equality comes from plugging in the proposed J and using $E(J|X=x) = 0$ and $Cov\left(J, P \mid X = x\right) = -1$. Thus, for u such that $f_{P|X}(u|x) > 0$, we have that $-T(u \mid x; J) f_{P|X}(u|x) / Cov\left(J, P \mid X = x\right) = w'(u|x)$ as desired. For u such that $f_{P|X}(u|x) = 0$, assumption (c) implies that $w'(u|x) = 0$, and thus trivially $-T(u \mid x; J) f_{P|X}(u|x) / Cov\left(J, P \mid X = x\right) = w'(u|x)$ for u such that $f_{P|X}(u|x) = 0$.

We now show that, given assumptions (a) and (b), assumptions (c) and (d) are necessary. Suppose that (c) does not hold, so that there exists a set of t values such that $f_{P|X}(t|x) = 0$ but $w'(t|x) > 0$. Then, for such values of t, $-\frac{T(t|x;J) f_{P|X}(t|x)}{Cov\left(J,P|X=x\right)} = 0$ for any potential instrument J while $w'(t|x) > 0$, and thus trivially there cannot exist an instrument J such that $-\frac{T(t|x;J) f_{P|X}(t|x)}{Cov\left(J,P|X=x\right)} = w'(t|x)$ for all t. Thus (c) is a necessary condition. Now suppose that (c) holds but (d) does not hold, so that $\int_0^1 w'(t|x) t dt \neq -1$. We will now use a proof by contradiction to show that there cannot exist a J with the desired properties. Assume that there exists a J such that $-T(t \mid x; J) f_{P|X}(t|x) / Cov\left(J, P \mid X = x\right) = w'(t|x)$. For any t such that $f_{P|X}(t|x) > 0$, we can solve for $T(t \mid x; J)$ to obtain $T(t \mid x; J) = \alpha(x) w'(t|x) / f_{P|X}(t|x)$ where $\alpha(x) = -Cov\left(J, P \mid X = x\right)$. We thus have

$$\begin{aligned}
\alpha(x) &= -Cov\left(J, P \mid X = x\right) = -Cov\left(E(J|X=x, P), P \mid X = x\right) \\
&= -Cov\left(T(P \mid x; J), P \mid X = x\right) = -Cov\left(\alpha(x) w'(P|x) / f_{P|X}(t|x), P \mid X = x\right) \\
&= -\alpha(x) \int_0^1 w'(t|x) t dt.
\end{aligned}$$

Thus, $\alpha(x) = -\alpha(x) \int w'(t|x) t dt$. Since (d) does not hold by assumption, this equality implies that $\alpha(x) = 0$. But $\alpha(x) = -Cov\left(J, P|X = x\right)$, so that J cannot be a proper instrument. Thus, given conditions (a) and (b), conditions (c) and (d) are necessary for the existence of an instrument with the desired properties. ■

Proof of Proposition 2.

Define $w(\cdot|x) \equiv \frac{F_{P_{a'}|X}(\cdot) - F_{P_a|X}(\cdot)}{E(P_a|X) - E(P_{a'}|X)}$. We now show that assumptions (a) and (b) of Proposition 2 imply as-

sumptions (a), (b), and (d) of Proposition 1 when $w(\cdot|x)$ is defined in this manner. Note that assumption (a) of Proposition 2 immediately implies that assumption (a) of Proposition 1 holds. Assumption (a) of Proposition 2 implies that the $w(\cdot|x)$ is differentiable for all evaluation points with $w'(\cdot|x) = \frac{f_{P_{a'}|X}(\cdot) - f_{P_a|X}(\cdot)}{E(P_a|X) - E(P_{a'}|X)}$. Using that $F_{P_a|X}(\cdot)$ and $F_{P_{a'}|X}(\cdot)$ are distribution functions, one can directly verify that $\int_0^1 w(u|x)du = 1$ and $w(1|x) - w(0|x) = 0$. Now consider $\int_0^1 w'(t|x)tdt$. We have that

$$\begin{aligned}
\int_0^1 w'(t|x)tdt &= \int_0^1 \frac{f_{P_{a'}}(t) - f_{P_a}(t)}{E(P_a \mid X) - E(P_{a'} \mid X)} tdt \\
&= \frac{1}{E(P_a \mid X) - E(P_{a'} \mid X)} \left[\int_0^1 t f_{P_{a'}|X}(t|x)dt - \int_0^1 t f_{P_a|X}(t|x)dt \right] \\
&= -1
\end{aligned}$$

Thus, defining $w(\cdot|x)$ in this manner, we have that assumptions (a) and (b) of Proposition 2 imply assumptions (a), (b) and (d) of Proposition 1. Given $w'(\cdot|x) = \frac{f_{P_{a'}|X}(\cdot) - f_{P_a|X}(\cdot)}{E(P_a|X) - E(P|X)}$, we have that assumption (c) of Proposition 2 is equivalent to assumption (c) of Proposition 1 for this choice of $w(\cdot|x)$. The result now follows directly from Proposition 1. ■

Proof of Proposition 3. Assume that the conditions of Proposition 1 hold for *a.e.* x. From the proof of Proposition 1, under the stated conditions, $E(J(Z) \mid X = x) = 0$, $Cov\,(J(Z), D \mid X = x) = -1$, and $\frac{Cov\,(J(Z),Y|X)}{Cov\,(J(Z),D|X)} = \int \Delta^{MTE}(X,u)w(u|X = x)du$. It follows that $Cov\,(J(Z), D) = Cov\,(J(Z), D \mid X = x) = -1$, that $Cov\,(J(Z), Y) = E(J(Z), Y) = E\big[E(J(Z)Y \mid X = x)\big]$, and thus that

$$\begin{aligned}
\frac{Cov\,(J(Z), Y)}{Cov\,(J(Z), D)} &= \frac{E(J(Z)Y)}{-1} = E\big[-E(J(Z)Y \mid X)\big] \\
&= E\left[\frac{Cov\,(J(Z), Y \mid X)}{Cov\,(J(Z), D \mid X)}\right] = \int \left[\int_0^1 \Delta^{MTE}(x,u)w(u|x)du\right] dF_X(x). \blacksquare
\end{aligned}$$

D Generalized Ordered Choice Model

The ordered choice model presented in the text with parameterized, but nonstochastic, thresholds is analyzed in Cameron and Heckman (1998) who establish its nonparametric identifiability under the conditions they specify. See also Heckman and Vytlacil (2003). Treating the W_s (or components of it) as unobservables, we obtain the generalized ordered choice model analyzed in Heckman and Vytlacil (2003), Vytlacil (2003) and Carneiro, Hansen and Heckman (2003). In this Appendix, we present the main properties of this model.

The thresholds are now written as $Q_s + C_s(W_s)$ in place of $C_s(W_s)$, where Q_s is a random variable. In addition to the order on the $C_s(W_s)$ in the text, we impose the order $Q_s \geq Q_{s-1}$, $s = 2, \ldots, \bar{S} - 1$. We impose the requirement that $Q_{\bar{S}} = \infty$ and $Q_0 = -\infty$. The latent index D_s^* is as defined in the text, but now

$$\begin{aligned}
D_s &= \mathbf{1}\left[C_{s-1}(W_{s-1}) + Q_{s-1} < \mu_D(Z) - V \leq C_s(W_s) + Q_s\right] \\
&= \mathbf{1}[l_{s-1}(Z, W_{s-1}) - Q_{s-1} > V \geq l_s(Z, W_s) - Q_s],
\end{aligned}$$

where $l_s = \mu_D(Z) - C_s(W_s)$. Using the fact that $l_s(Z, W_s) - Q_s < l_{s-1}(Z, W_{s-1}) - Q_{s-1}$, we obtain

$$\mathbf{1}[l_{s-1}(Z, W_{s-1}) - Q_{s-1} > V \geq l_s(Z, W_s) - Q_s] = \mathbf{1}[V + Q_{s-1} < l_{s-1}(Z, W_{s-1})] - \mathbf{1}[V + Q_s \leq l_s(Z, W_s)]$$

The nonparametric identifiability of this choice model is established in Carneiro, Hansen and Heckman (2003) and Heckman and Vytlacil (2003). We retain assumptions (A-10)–(A-11) and (A-13)–(A-15), but

alter (A-12) to

(A-12)′ $(Q_s, U_s, V) \perp\!\!\!\perp (Z, W) \mid X, \qquad s = 1, \ldots, \bar{S}.$

Vytlacil (2003) shows that this model with no transition specific instruments (with W_s degenerate for each s) implies and is implied by the independence and monotonicity conditions of Angrist and Imbens (1995) for an ordered model. Define $Q = (Q_1, \ldots, Q_{\bar{S}})$. Redefine $\pi_s(Z, W_s) = F_{V+Q_s}(\mu_D(Z) - C_s(W_s))$ and define $\pi(Z, W) = [\pi_1(Z, W_1), ..., \pi_{\bar{S}-1}(Z, W_{\bar{S}-1})]$. Redefine $U_{D,s} = F_{V+Q_s}(V + Q_s)$. We have that

$$\begin{aligned}
E(Y \mid Z, W) &= E\left(\sum_{s=1}^{\bar{S}} \mathbf{1}[l_{s-1}(Z, W_{s-1}) - Q_{s-1} > V \geq l_s(Z, W_s) - Q_s] Y_s \;\middle|\; Z, W\right) \\
&= \sum_{s=1}^{\bar{S}} \Big(E\left(\mathbf{1}[V + Q_{s-1} < l_{s-1}(Z, W_{s-1})] Y_s \mid Z, W\right) - E\left(\mathbf{1}[V + Q_s \leq l_s(Z, W_s)] Y_s \mid Z, W\right) \Big) \\
&= \sum_{s=1}^{\bar{S}} \Bigg(\int_{-\infty}^{l_{s-1}(Z, W_{s-1})} E\left(Y_s \mid V + Q_{s-1} = t\right) dF_{V+Q_{s-1}}(t) \\
&\quad - \int_{-\infty}^{l_s(Z, W_s)} E\left(Y_s \mid V + Q_s = t\right) dF_{V+Q_s}(t) \Bigg) \\
&= \sum_{s=1}^{\bar{S}} \left(\int_0^{\pi_{s-1}(Z, W_{s-1})} E\left(Y_s \mid U_{D,s-1} = t\right) dt - \int_0^{\pi_s(Z, W_s)} E\left(Y_s \mid U_{D,s} = t\right) dt \right) \\
&= \sum_{s=1}^{\bar{S}-1} \int_0^{\pi_s(Z, W_s)} E\left(Y_{s+1} - Y_s \mid U_{D,s} = t\right) dt.
\end{aligned}$$

We thus have the index sufficiency restriction that

$$E(Y \mid Z, W) = E(Y \mid \pi(Z, W)).$$

and in the general case

$$\frac{\partial}{\partial \pi_s} E(Y \mid \pi(Z, W) = \pi) = E(Y_{s+1} - Y_s \mid U_{D,s} = \pi_s).$$

Also, notice that we have the restriction that

$$\frac{\partial^2}{\partial \pi_s \partial \pi_{s'}} E(Y \mid \pi(Z, W) = \pi) = 0$$

if $|s - s'| > 1$. Under full independence between U_s and $V + Q_s$, $s = 1, \ldots, \bar{S}$, we can test full independence for the more general choice model by testing for linearity of $E(Y \mid \pi(Z, W) = \pi)$ in π.

Define

$$\Delta^{MTE}_{s+1,s}(x, u) = E(Y_{s+1} - Y_s \mid X = x, U_{D,s} = u),$$

so that our result above can be rewritten as

$$\frac{\partial}{\partial \pi_s} E(Y \mid \pi(Z, W) = \pi) = \Delta^{MTE}_{s+1,s}(x, \pi_s).$$

Since $\pi(Z, W)$ can be nonparametrically identified immediately from $\pi_s(Z, W_s) = \Pr\left(\sum_{j=s+1}^{\bar{S}} D_j = 1 \mid Z, W_s\right)$,

we have that the above offset equality immediately implies identification of *MTE* for all evaluation points within the appropriate support.

The policy relevant treatment effect is defined analogously. Recall that H_s^a is defined as the cumulative distribution function of $\mu_D(Z) - C_s(W_s)$. We have that

$$\begin{aligned}
E_a(Y_a) &= E_a\left(E(Y \mid V, Q, Z, W)\right) \\
&= E_a\left(\sum_{s=1}^{\bar{S}} \mathbf{1}[l_{s-1}(Z, W_{s-1}) - Q_{s-1} > V \geq l_s(Z, W_s) - Q_s] E(Y_s \mid V, Q, Z, W)\right) \\
&= E_a\left(\sum_{s=1}^{\bar{S}} \mathbf{1}[l_{s-1}(Z, W_{s-1}) - Q_{s-1} > V \geq l_s(Z, W_s) - Q_s] E(Y_s \mid V, Q)\right) \\
&= \sum_{s=1}^{\bar{S}} E_a\left(E(Y_s \mid V, Q)\{H_s^a(V + Q_s) - H_{s-1}^a(V + Q_{s-1})\}\right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{s=1}^{\bar{S}} \int \left(E(Y_s \mid V = v, Q = q)\{H_s^a(v + q_s) - H_{s-1}^a(v + q_{s-1})\}\right) dF_{V,Q}(v, q) \\
&= \sum_{s=1}^{\bar{S}} \Bigg(\int E(Y_s \mid V + Q_s = t) H_s^a(t) dF_{V+Q_s}(t) \\
&\quad - \int E(Y_s \mid V + Q_{s-1} = t) H_{s-1}^a(t) dF_{V+Q_{s-1}}(t) \Bigg)
\end{aligned}$$

so that

$$\begin{aligned}
\Delta_{a,a'}^{PRTE} &= E_{a'}(Y) - E_a(Y) \\
&= \sum_{s=1}^{\bar{S}-1} \int \left(E(Y_{s+1} - Y_s \mid V + Q_s = t)\{H_s^a(t) - H_s^{a'}(t)\}\right) dF_{V+Q_s}(t).
\end{aligned}$$

Alternatively, we can express this result in terms of *MTE*,

$$E_a(Y_a) = \sum_{s=1}^{\bar{S}} \left(\int E(Y_s \mid U_{D,s} = t) \tilde{H}_s^a(t) dt - \int E(Y_s \mid U_{D,s-1} = t) \tilde{H}_{s-1}^a(t) dt \right)$$

so that

$$\begin{aligned}
\Delta_{a,a'}^{PRTE} &= E_{a'}(Y) - E_a(Y) \\
&= \sum_{s=1}^{\bar{S}-1} \int \left(E(Y_{s+1} - Y_s \mid U_{D,s} = t)\{\tilde{H}_s^a(t) - \tilde{H}_s^{a'}(t)\}\right) dt
\end{aligned}$$

where $\tilde{H}_s^a$ is the cumulative distribution function of the random variable $F_{U_{D,s}}(\mu_D(Z) - C_s(W_s))$. For further discussion, see Heckman and Vytlacil (2003).

Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; Telephone: (773) 702-0634, Fax: (773) 702-8490, E-mail: jjh@uchicago.edu

and

Department of Economics, Stanford University, 579 Serra Mall, Stanford CA 94305, U.S.A.; Telephone: (650) 725-7836, Fax: (650) 725-5702, vytlacil@stanford.edu

References

[1] Ahn, H. and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

[2] Amemiya, T. (1985): *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

[3] Andrews, D. W. K and M. M. A. Schafgans (1998): "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65, 497-517.

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

[5] Angrist, J. and G. 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.

[6] Angrist, J., and A. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of labor economics*. Volume 3A. ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier Science, 1277-1366.

[7] Bickel, P. J. (1967): "Some Contributions to the Theory of Order Statistics," *Proceedings of the Fifth Berkeley Symposium*, 575-591.

[8] Björklund, A. and R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, 69, 42-49.

[9] Cameron, S. and J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 262-333.

[10] Campbell, D. T., and Stanley, J. C. (1966): *Experimental and quasi-experimental designs for research.* Skokie, Illinois: Rand McNally.

[11] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127-60.

[12] Carneiro, P. (2002): "Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation," Unpublished Ph.D. Thesis University of Chicago.

[13] Carneiro, P., K. Hansen and J. Heckman (2003): "Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice," *International Economic Review*, 44, 361-422.

[14] Carneiro, P., J. Heckman and E. Vytlacil (2003): "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," Unpublished working paper, University of Chicago, Department of Economics.

[15] DeLong, J. B., C. Goldin and L. Katz (2003). "Sustaining U.S. Economic Growth," in *Agenda for the Nation*, ed. by H. J Aaron, J. M Lindsay, and P. S. Nivola. Washington, D.C.: The Brookings Institution, 17-60.

[16] Florens, J.-P., J. Heckman, C. Meghir and E. Vytlacil (2001): "Instrumental Variables, Local Instrumental Variables, and Control Functions," Unpublished working paper, University of Chicago.

[17] Gill, R. D. and J. M. Robins (2001): "Causal inference for complex longitudinal data: the continuous case," *The Annals of Statistics,* 29, 1-27.

[18] Hansen, L. and T. Sargent (1981): "Linear Rational Expectations Models of Dynamically Interrelated Variables," in *Rational Expectations and Econometric Practice*, ed. by R. Lucas and T. Sargent. Minneapolis: University of Minnesota Press, 127–56.

[19] Heckman, J. (1974): "Shadow Prices, Market Wages and Labor Supply," *Econometrica* 42, 679-94.

[20] _____ (1976): "Simultaneous Equation Models with both Continuous and Discrete Endogenous Variables With and Without Structural Shift in the Equations," in *Studies in Nonlinear Estimation*, ed. by S. Goldfeld and R. Quandt. Cambridge, MA: Ballinger.

[21] _____ (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313-318.

[22] _____ (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441-462.

[23] _____ (2001a): "Accounting for Heterogeneity, Diversity and Social Policy Evaluation," *Economic Journal.* 111, F654-99.

[24] _____ (2001b): "Micro Data, Heterogeneity and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109, 673-748.

[25] Heckman, J., H. Ichimura and P. Todd (1997): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261-294.

[26] Heckman, J. H. Ichimura, J. Smith and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.

[27] Heckman, J., R. LaLonde and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. Vol. 3A. Amsterdam: Elsevier Science, 1865-2097.

[28] Heckman, J., L. Lochner and C. Taber (1998a): "General Equilibrium Treatment Effects: A Study of Tuition Policy," *American Economic Review*, 88, 381-386.

[29] _____ (1998b): "Tax Policy and Human Capital Formation," *American Economic Review*, 88, 293-297.

[30] _____ (1999): "Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy," *Fiscal Studies*, 20, 25-40.

[31] Heckman, J. and R. Robb (1985): "Alternative Methods for Estimating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data,* ed. by J. Heckman and B. Singer. New York: Cambridge University Press, 156-245.

[32] _____ (1986; 2000): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer. New York: Springer-Verlag, 63-107. (2000): Mahwah, N.J. : Lawrence Erlbaum Associates.

[33] Heckman, J. and J. Smith (1998): "Evaluating The Welfare State," in *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium, Econometric Society Monograph Series*, 16, ed. by S. Strom, Cambridge, UK: Cambridge University Press, Chapter 8, 241-318.

[34] Heckman, J. and E. Vytlacil (1999): "Local Instrumental Variable and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730-4734.

[35] _____ (2000): "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 C. Hsiao, K. Morimune, and J. Powell. Cambridge: Cambridge University Press, 1-46.

[36] _____ (2001): "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, 1-23.

[37] _____ (2003): "Extending the Marginal Treatment Effect to an Ordered Choice Model," Unpublished manuscript, University of Chicago, Department of Economics.

[38] _____ (2004): "Econometric Evaluation of Social Programs," forthcoming in *Handbook of Econometrics*, Volume 6, ed. by J. Heckman and E. Leamer. Amsterdam: Elsevier Science.

[39] Hendry, D. (1995): *Dynamic Econometrics*. Oxford, UK: Oxford University Press.

[40] Horowitz, J. (1998): *Semiparametric Methods in Econometrics*. Volume 131. Berlin: Springer-Verlag.

[41] Hurwicz, L. (1962) "On the Structural Form of Interdependent Systems," in *Logic, Methodology and Philosophy of Science*, ed. by E. Nagel, P. Suppes, and A. Tarski. Stanford: Stanford University Press, 232-239.

[42] Ichimura, H. and C. Taber (2002): "Direct Estimation of Policy Impacts," Unpublished working paper, University College London, Department of Economics.

[43] Imbens, G. and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

[44] Imbens, G. and D. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555-74.

[45] Kling, J. (2001): " Interpreting Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business and Economic Statistics*, 19, 358-64.

[46] Marschak, J. (1953): "Economic Measurements for Policy and Predictions", in *Studies in Econometric Method*, ed. by W. C. Hood and T. C. Koopmans. Cowles Commission for Research in Economics Monograph; no. 14. New York: Wiley.

[47] Matzkin, R. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Volume 4, ed. by R F Engle; and D. McFadden. New York: North-Holland, 2523-58.

[48] Mincer, J. (1974): *Schoolings, Experience, and Earnings*. Cambridge, MA: NBER, distributed by Columbia University Press.

[49] Pearl, J. (2000): *Causality.* Cambridge, UK: Cambridge University Press.

[50] Powell, J. L. (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics*, Volume 4, ed. by R F Engle; and D. McFadden. New York: North-Holland, 2443-2521.

[51] Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

[52] Roy, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135-146.

[53] Rudin, W. (1974): *Real and Complex Analysis*. New York: McGraw Hill.

[54] Smith, V. K. and H. S. Banzhaf (2003): "A Diagrammatic Exposition of Weak Complementarity and the Willig Condition," Unpublished manuscript, North Carolina State University.

[55] Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica,* 70(1): 331-41

[56] _____. (2003): "Ordered Discrete Choice Selection Models and *LATE* Assumptions: Equivalence, Nonequivalence and Representation Results," Unpublished manuscript, Stanford University, Department of Economics.

[57] Yitzhaki, S. (1996): "On Using Linear Regression in Welfare Economics," *Journal of Business and Economic Statistics*, 14, 478-486.

[58] _____ (1999): "The Gini Instrumental Variable, or 'The Double *IV* Estimator,'" unpublished manuscript, Hebrew University, Department of Economics.

Table IA
Treatment Effects and Estimands as Weighted Averages of the Marginal Treatment Effect

$$ATE(x) = \int_0^1 \Delta^{MTE}(x, u_D) du_D$$

$$TT(x) = \int_0^1 \Delta^{MTE}(x, u_D) h_{TT}(x, u_D) du_D$$

$$TUT\,(x) = \int_0^1 \Delta^{MTE}\,(x, u_D)\, h_{TUT}\,(x, u_D)\, du_D$$

$$\Delta^{PRTE}(x) = \int_0^1 \Delta^{MTE}(x, u_D) h_{PRT}(x, u_D) du_D$$

$$IV(x) = \int_0^1 \Delta^{MTE}(x, u_D) h_{IV}(x, u_D) du_D$$

$$OLS(x) = \int_0^1 \Delta^{MTE}(x, u_D) h_{OLS}(x, u_D) du_D$$

Table IB
Weights

$$h_{ATE}(x, u_D) = 1$$

$$h_{TT}(x, u_D) = \left[\int_{u_D}^1 f(p \mid X = x) dp\right] \frac{1}{E(P \mid X = x)}$$

$$h_{TUT}\,(x, u_D) = \left[\int_0^{u_D} f\,(p|X = x)\, dp\right] \frac{1}{E\,((1 - P)\,|X = x)}$$

$$h_{PRT}(x, u_D) = \left[\frac{F_{P^*,X}(u_D) - F_{P,X}(u_D)}{\Delta \overline{P}}\right] \text{ where } \Delta \bar{P} = P^* - P$$

$$h_{IV}(x, u_D) = \left[\int_{u_D}^1 (p - E(P \mid X = x)) f(p \mid X = x) dp\right] \frac{1}{Var(P \mid X = x)} \text{ for } P(Z) \text{ as an instrument}$$

$$h_{OLS}(x, u_D) = \frac{E(U_1 \mid X = x, U_D = u_D) h_1(x, u_D) - E(U_0 \mid X = x, U_D = u_D) h_0(x, u_D)}{\Delta^{MTE}(x, u_D)}$$

$$h_1(x, u_D) = \left[\int_{u_D}^1 f(p \mid X = x) dp\right] \left[\frac{1}{E(P \mid X = x)}\right]$$

$$h_0(x, u_D) = \left[\int_0^{u_D} f(p \mid X = x) dp\right] \frac{1}{E((1 - P) \mid X = x)}$$

Table II

Treatment Parameters and Estimands
in the Generalized Roy Example

Treatment on the Treated	0.2353
Treatment on the Untreated	0.1574
Average Treatment Effect	0.2000
Sorting Gain[(1)]	0.0353
Policy Relevant Treatment Effect ($PRTE$)	0.1549
Selection Bias[(2)]	-0.0628
Linear Instrumental Variables[(3)]	0.2013
Ordinary Least Squares	0.1725

[(1)] $TT - ATE = E(Y_1 - Y_0 \mid X = x, D = 1) - E(Y_1 - Y_0 \mid X = x)$

[(2)] $OLS - TT = E(Y_0 \mid X = x, D = 1) - E(Y_0 \mid X = x, D = 0)$

[(3)] Using Propensity Score as the Instrument

Note: The model used to create Table II is the same as those used to create Figures 1A and 1B. The $PRTE$ is computed using a policy t characterized as follows:

$$\begin{aligned} \text{If } Z &> 0 \text{ then } D = 1 \text{ if } Z(1+t) - V > 0 \\ \text{If } Z &\leq 0 \text{ then } D = 1 \text{ if } Z - V > 0 \end{aligned}$$

For this example t is set equal to 1.2.

TableIIIA

Treatment Parameters and Estimands in the Generalized Roy Example
When $P\left(Z\left(1+t\left(\mathbf{1}\left[Z>0\right]\right)\right)\right)$ is the Instrument

Ordinary Least Squares	0.1725
Treatment on the Treated	0.2353
Treatment on the Untreated	0.1574
Average Treatment Effect	0.2000
Linear Instrumental Variables[(1)]	0.1859
Policy Relevant Treatment Effect ($PRTE$)	0.1549

(1) Propensity Score $P\left(Z\left(1+t\left(\mathbf{1}\left[Z>0\right]\right)\right)\right)$ as the Instrument

Note: Parameters used to create figures Table IIIA are the same as those used in Figures 1A and 1B. The $PRTE$ and the Linear Instrumental variables estimator are computed using the policy described previously. (See Table II notes)

Table IIIB

Linear Instrumental Variables vs Policy Relevant Treatment Effect

Linear Instrumental Variables[(1)]	0.2013
Linear Instrumental Variables[(2)]	0.1859
Linear Instrumental Policy[(3)]	0.1549
Policy Relevant Treatment Effect ($PRTE$)	0.1549

[(1)]Propensity Score $P(Z)$ as the Instrument

[(2)]Propensity Score $P(Z(1+t(\mathbf{1}(Z>0))))$ as the Instrument

[(3)]Uses a dummy B as an Instrument. The dummy B is such that $B=1$ if it belongs to a randomly assigned eligible population, 0 otherwise.

Table IV

The IV Estimator and $Cov(Z_1, \alpha' Z)$ Associated with each Value of Σ_2

Group 2 Covariance

Weights	Σ_2	IV	$Cov(Z_1, \alpha' Z) = \alpha' \Sigma_2^1$
h_1	$\begin{bmatrix} 0.6 & -0.3 \\ -0.3 & 0.6 \end{bmatrix}$	0.133	-0.30
h_2	$\begin{bmatrix} 0.6 & -0.1 \\ -0.1 & 0.6 \end{bmatrix}$	0.177	-0.02
h_3	$\begin{bmatrix} 0.6 & 0.1 \\ 0.1 & 0.6 \end{bmatrix}$	0.194	0.26

Weights for Mixture of Normals IV:

$$h_{IV}(v) = \frac{\frac{P_1 \alpha' \Sigma_1^1}{(\alpha' \Sigma_1 \alpha)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{v - \alpha' \mu_1}{(\alpha' \Sigma_1 \alpha)^{1/2}}\right)^2\right] + \frac{P_2 \alpha' \Sigma_2^1}{(\alpha' \Sigma_2 \alpha)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{v - \alpha' \mu_2}{(\alpha' \Sigma_2 \alpha)^{1/2}}\right)^2\right]}{\frac{P_1 \alpha' \Sigma_1^1}{\left(\alpha' \Sigma_1 \alpha + \sigma_V^2\right)^{1/2}} \exp\left[-\left(\frac{-\alpha' \mu_1}{\left(\alpha' \Sigma_1 \alpha + \sigma_V^2\right)^{1/2}}\right)^2\right] + \frac{P_2 \alpha' \Sigma_2^1}{\left(\alpha' \Sigma_2 \alpha + \sigma_V^2\right)^{1/2}} \exp\left[-\left(\frac{-\alpha' \mu_2}{\left(\alpha' \Sigma_2 \alpha + \sigma_V^2\right)^{1/2}}\right)^2\right]}$$

where Σ_1^1 and Σ_2^1 are the first rows of Σ_1 and Σ_2, respectively. Clearly, $h_{IV}(-\infty) = 0$, $h_{IV}(\infty) = 0$. The weights clearly integrate to one over the support of $V = (-\infty, \infty)$. Observe that if $P_2 = 0$, the weights must be positive. Thus the structure of the data, and in particular, the structure of the covariances of the instruments is a key determinant of the positivity of the weights for any instrument. It has nothing to do with the ceteris paribus effect of Z_1 on $P(Z)$ in the general case. Now observe that a necessary condition for $h_{IV} < 0$ is that $\text{sign}(\alpha' \Sigma_1^1) = -\text{sign}(\alpha' \Sigma_2^1)$, i.e., that the covariance between Z_1 and $\alpha' Z$ be of opposite signs in the two populations. Without loss of generality assume that $\alpha' \Sigma_1^1 > 0$. If it equals zero, we fail the rank condition.

Table V

	Structural Econometric Approach	Treatment Effect Approach	Approach Based on MTE
Interpretability	Well defined economic parameters and welfare comparisons	Link to economics and welfare comparisons obscure	Interpretable in terms of willingness to pay; weighted averages of the MTE answer well-posed economic questions
Range of Questions Addressed	Answers many counterfactual questions	Focuses on one treatment effect or narrow range of effects	With support conditions, generates all treatment parameters
Extrapolation to New Environments	Provides ingredients for extrapolation	Evaluates one program in one environment	Can be partially extrapolated; extrapolates to policy new environments with different distributions of the probability of participation due solely to differences in distributions of Z;
Comparability Across Studies	Policy invariant parameters comparable across studies	Not generally comparable	Partially comparable; comparable across environments with different distributions of the probability of participation due solely to differences in distributions of Z.
Key Econometric Problems	Exogeneity, policy invariance and selection bias	Selection bias	Selection bias
Range of Policies that Can Be Evaluated	Programs with either partial or universal coverage, depending on variation in data (prices/endowments)	Programs with partial coverage (treatment and control groups)	Programs with partial coverage (treatment and control groups)
Extension to General Equilibrium Evaluation	Need to link to time series data; parameters compatible with general equilibrium theory	Difficult because link to economics is not precisely specified	Can be linked to nonparametric general equilibrium models

Figure 1A

Weights for the Marginal Treatment Effect for Different Parameters

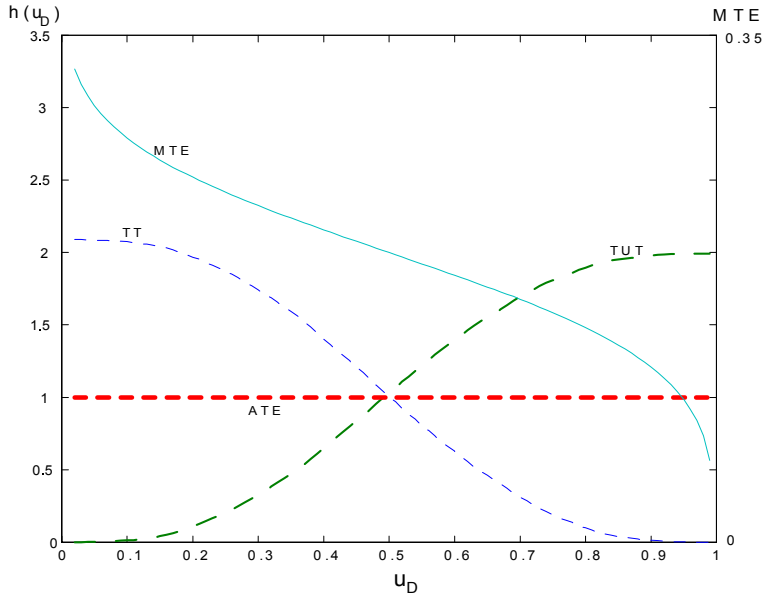
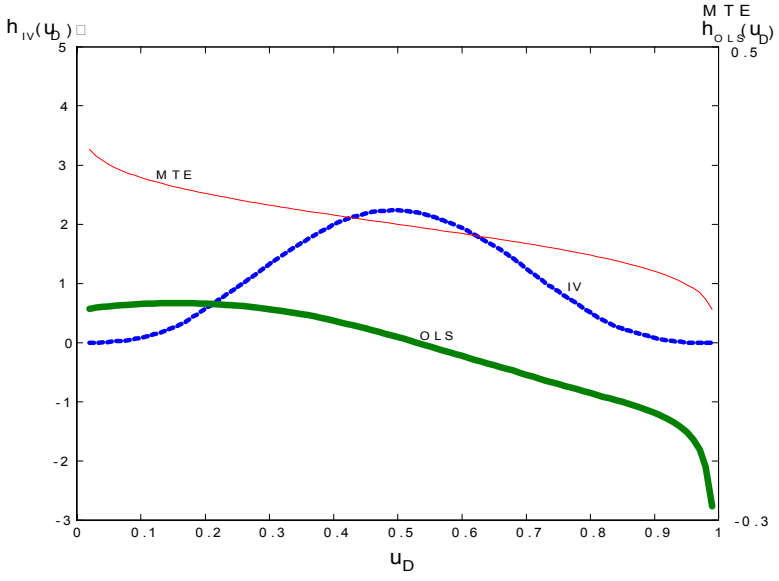


Figure 1B

Marginal Treatment Effect vs Linear Instrumental Variables and Ordinary Least Squares Weights



$$\ln Y_1 = \alpha + \beta + U_1 \qquad U_1 = \sigma_1 \varepsilon \qquad \alpha = 0.67 \qquad \sigma_1 = 0.012$$
$$\ln Y_0 = \alpha + U_0 \qquad U_0 = \sigma_0 \varepsilon \qquad \beta = 0.2 \qquad \sigma_0 = -0.050$$
$$D = 1 \text{ if } \; Z - V > 0 \qquad V = \sigma_V \varepsilon \qquad \varepsilon \sim N(0,1) \qquad \sigma_V = -1.000$$
$$U_D = \Phi\left(\frac{V}{\sigma_V \sigma_\varepsilon}\right) \qquad Z \sim N(-0.0026, 0.2700)$$

Figure 2A

Plot of the $E(Y|P(Z)=p)$

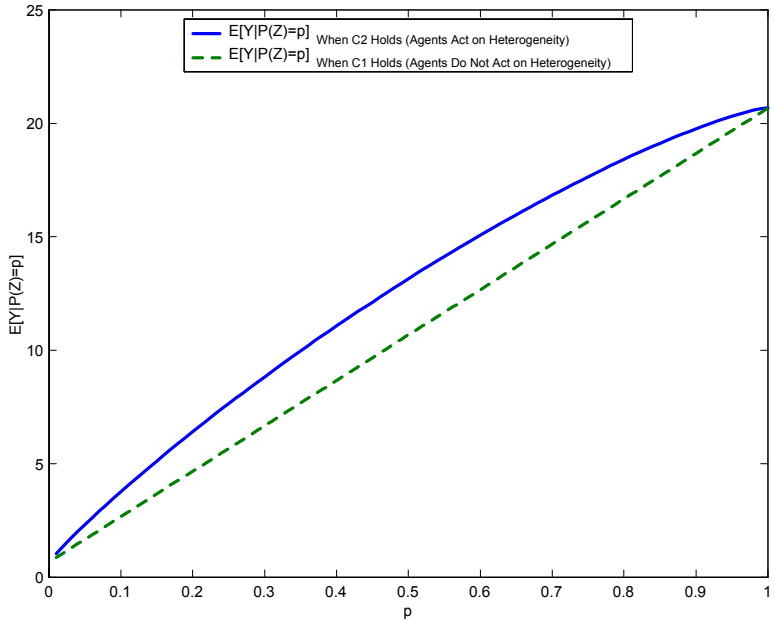
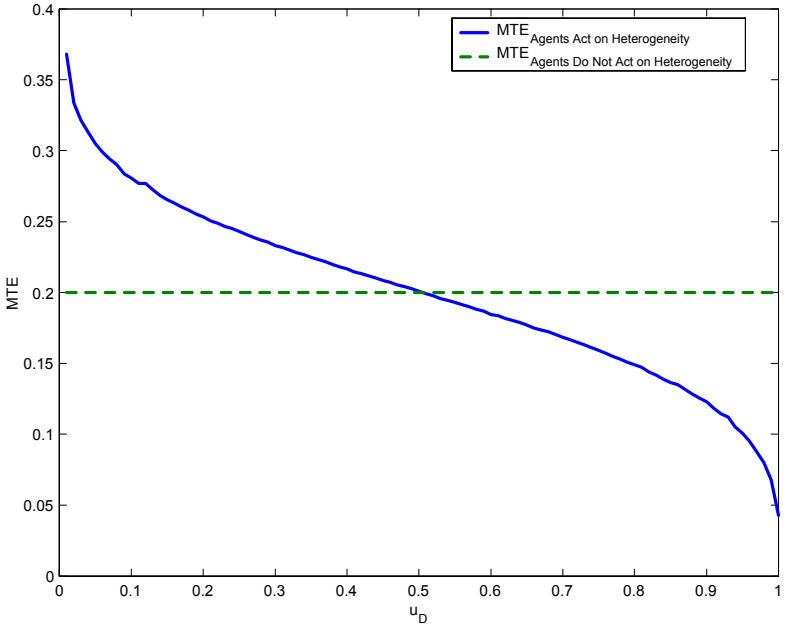


Figure 2B

Plot of the Identified Marginal Treatment Effect from Figure 2A (the Derivative).



Note: Parameters for the general heterogeneous case are the same as those used in Figures 1A and 1B. For the homogeneous case we impose $U_1 = U_0$ ($\sigma_1 = \sigma_0 = 0.012$).

Figure 3A

Marginal Treatment Effect vs Linear Instrumental Variables, Ordinary Least Squares, and Policy Relevant Treatment Effect Weights: When $P(Z)$ is the Instrument

The Policy is Given at the Base of Table 2

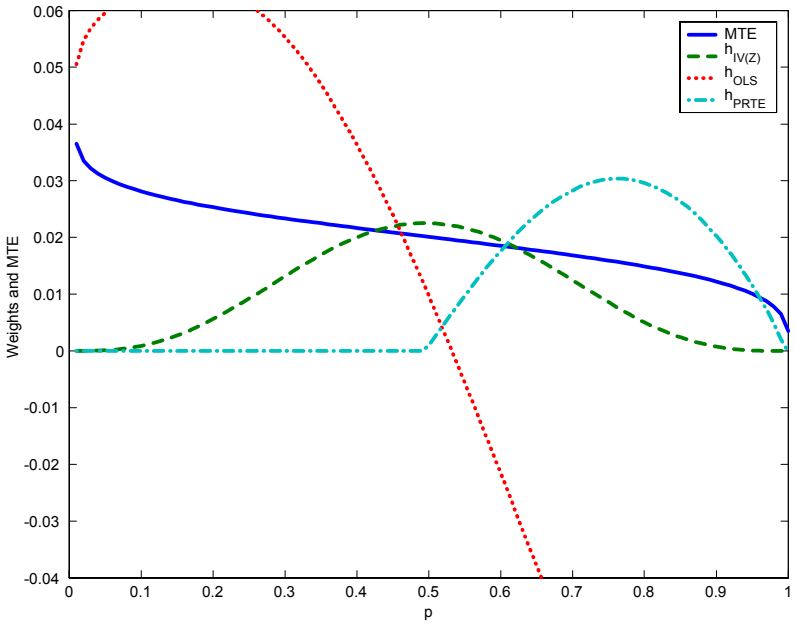


Figure 3B

Marginal Treatment Effect vs Linear IV with Z as an Instrument, Linear IV with $P(Z(1+t(\mathbf{1}[Z>0]))) = \tilde{P}(Z,t)$ as an Instrument, and Policy Relevant Treatment Effect Weights

For The Policy Defined at the Base of Table 2

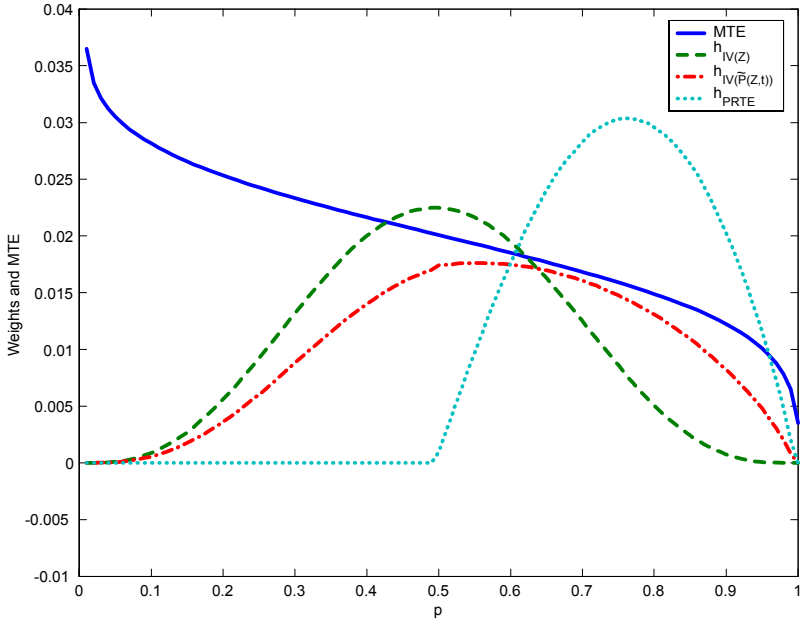


Figure 3C

Marginal Treatment Effect vs IV Policy and Policy Relevant Treatment Effect Weights

For The Policy Defined at the Base of Table 2

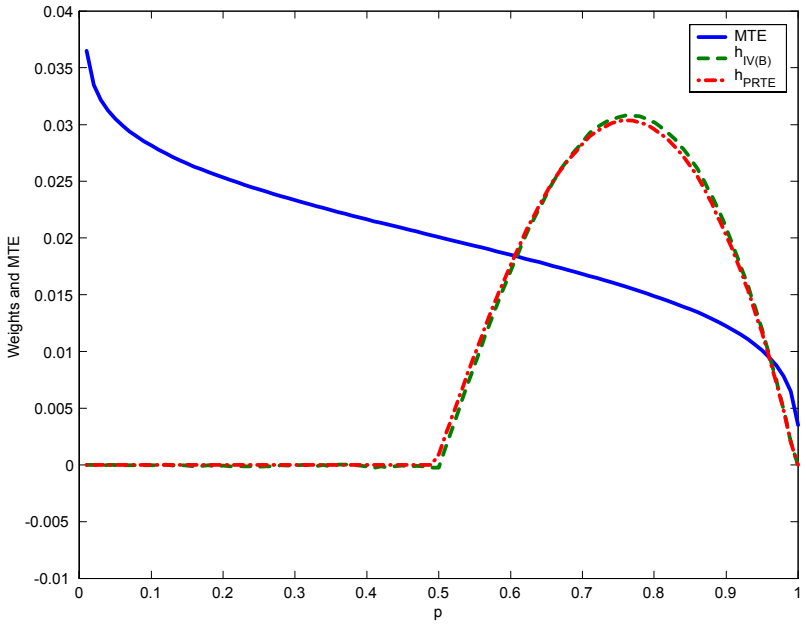
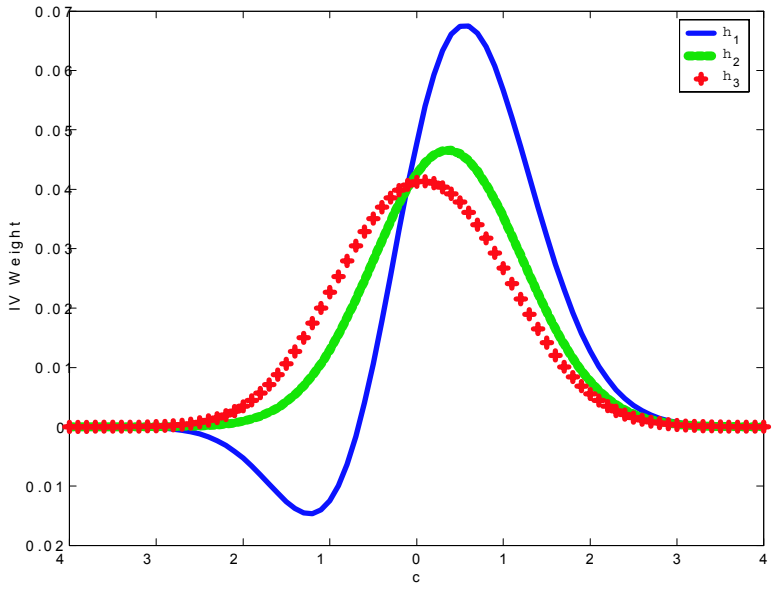


Figure 4

IV Weights when $Z \sim p_1 N(\mu_1, \Sigma_1) + p_2 N(\mu_2, \Sigma_2)$ for Different Values of Σ_2



$$\begin{array}{lll} Y_1 = \gamma + \beta + U_1 & U_1 = \sigma_1 \epsilon & \epsilon \sim N(0,1) \\ Y_0 = \gamma + U_0 & U_0 = \sigma_0 \epsilon & \sigma_1 = 0.012, \sigma_0 = -0.05, \sigma_V = -1 \\ I = \alpha' Z - V & V = \sigma_V \epsilon & \gamma = 0.67, \beta = 0.2 \end{array}$$

$$D = \begin{cases} 1 \text{ if } I > 0 \\ 0 \text{ if } I \leq 0 \end{cases}$$

$$Z \sim p_1 N(\mu_1, \Sigma_1) + p_2 N(\mu_2, \Sigma_2)$$

$$\mu_1 = \begin{bmatrix} 0 & -1 \end{bmatrix}, \mu_2 = \begin{bmatrix} 0 & 1 \end{bmatrix} \qquad \Sigma_1 = \begin{bmatrix} 1.4 & 0.5 \\ 0.5 & 1.4 \end{bmatrix}$$

$$p_1 = 0.45, p_2 = 0.55 \qquad \alpha = \begin{bmatrix} 0.2 & 1.4 \end{bmatrix}$$

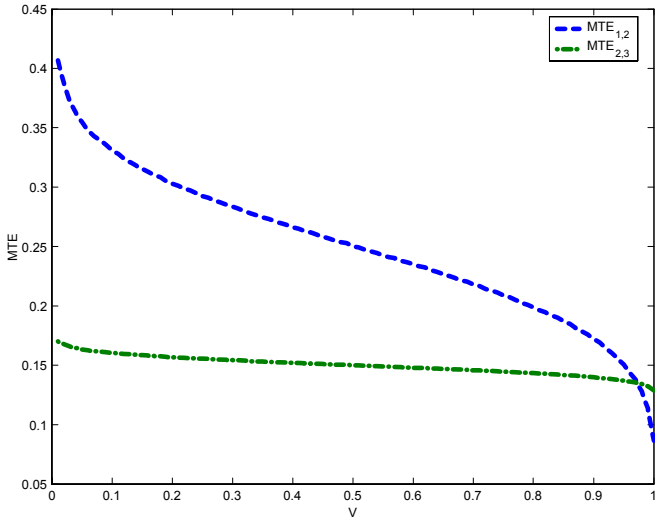
$$Cov(Z_1, \alpha' Z) = \alpha' \Sigma_1^1 = 0.98 \text{ (Group 1)}$$

$$\begin{aligned} \Delta^{MTE}(v) &= \beta + \left[\frac{Cov(U_1 - U_0, V)}{Var(V)} \right] v \\ h_{IV}(v) &= \frac{E(Z_1 | \alpha' Z > v) \Pr(\alpha' Z > v)}{Cov(Z_1, D)} \\ \beta_{IV} &= \int_{-\infty}^{\infty} \Delta^{MTE}(v) h_{IV}(v) dv \end{aligned}$$

The Policy: $W_2 - t$ where t=1.2 and W_1 is the instrument

Figure 5A

Marginal Treatment Effects by Transition



$Y_3 = \alpha + \beta_3 + U_3$; $D_3 = 1$ if $W_2 < I < \infty$; $U_3 = \sigma_3\epsilon$; $\sigma_3 = 0.02, \sigma_2 = 0.012, \sigma_1 = -0.05, \sigma_V = -1$

$Y_2 = \alpha + \beta_2 + U_2$; $D_2 = 1$ if $W_1 < I \leq W_2$; $U_2 = \sigma_2\epsilon$; $\alpha = 0.67, \beta_2 = 0.25, \beta_3 = 0.4$

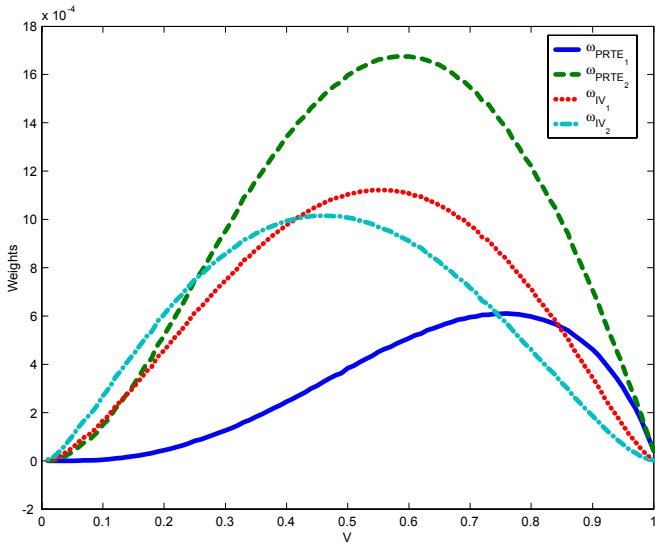
$Y_1 = \alpha + U_1$; $D_1 = 1$ if $-\infty < I \leq W_1$; $U_1 = \sigma_1\epsilon$; $Z \sim N(-0.0026, 0.27)$ and $Z \perp\!\!\!\perp V$

$I = Z - V$ $V = \sigma_V\epsilon$; $\epsilon \sim N(0,1)$

Sample Size = 1500

Figure 5B

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$$

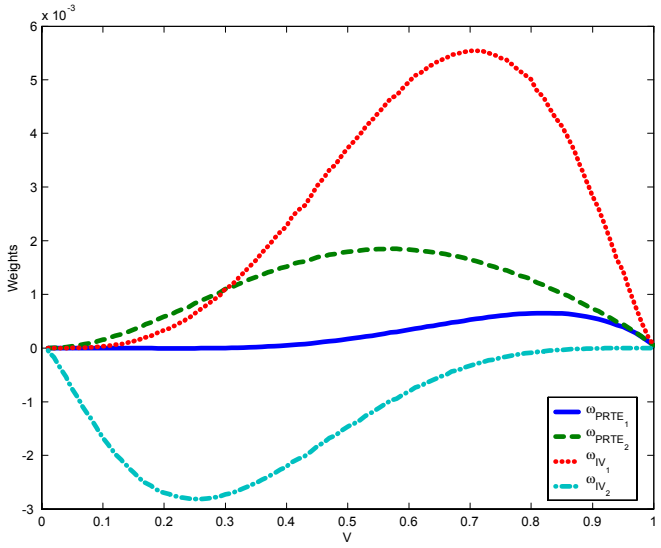
$\Delta^{PRTE} = 0.166$ $IV = 0.201$

Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 42.8\%$

Proportion Induced to Change from $D_2 = 1$ to $D_3 = 1$ $= 92.4\%$

Figure 5C

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}\right)$$

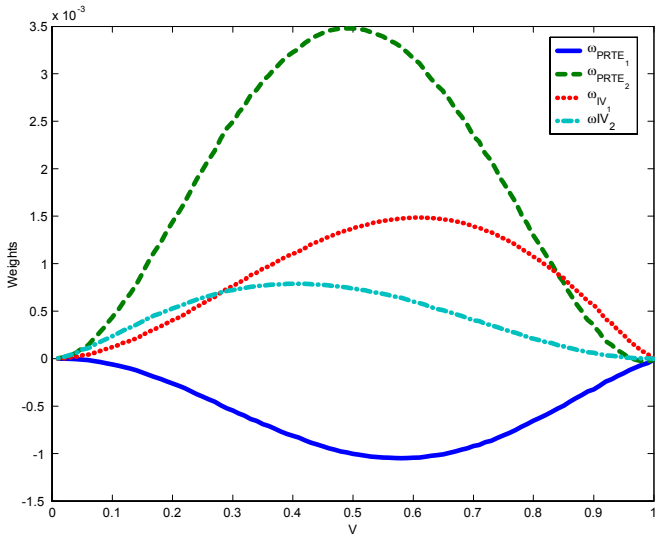
$$\Delta^{PRTE} = 0.159 \qquad IV = 0.296$$

Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 32.1\%$

Proportion Induced to Change from $D_2 = 1$ to $D_3 = 1$ $= 64.7\%$

Figure 5D

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$\Delta^{PRTE} = 0.110 \qquad IV = 0.210$$

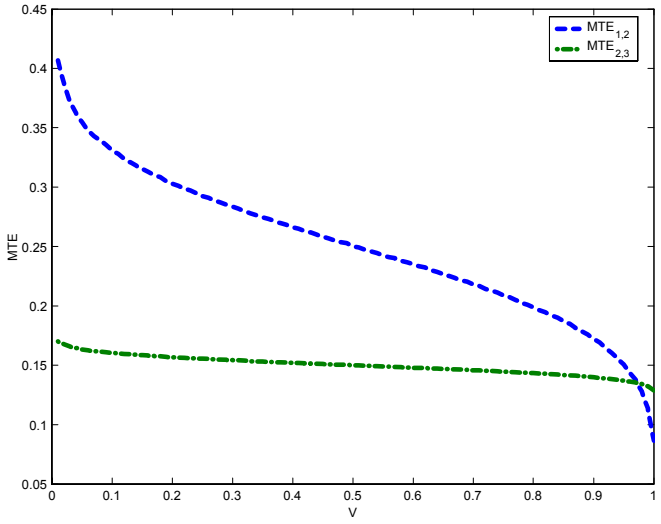
Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 27.5\%$

Proportion Induced to Change from $D_2 = 1$ to $D_3 = 1$ $= 76.8\%$

The Policy: $W_2 - t$ where $t = 1.2$ and Z is the instrument

Figure 6A

Marginal Treatment Effects by Transition



$Y_3 = \alpha + \beta_3 + U_3$; $D_3 = 1$ if $W_2 < I < \infty$; $U_3 = \sigma_3 \epsilon$; $\sigma_3 = 0.02, \sigma_2 = 0.012, \sigma_1 = -0.05, \sigma_V = -1$

$Y_2 = \alpha + \beta_2 + U_2$; $D_2 = 1$ if $W_1 < I \leq W_2$; $U_2 = \sigma_2 \epsilon$; $\alpha = 0.67, \beta_2 = 0.25, \beta_3 = 0.4$

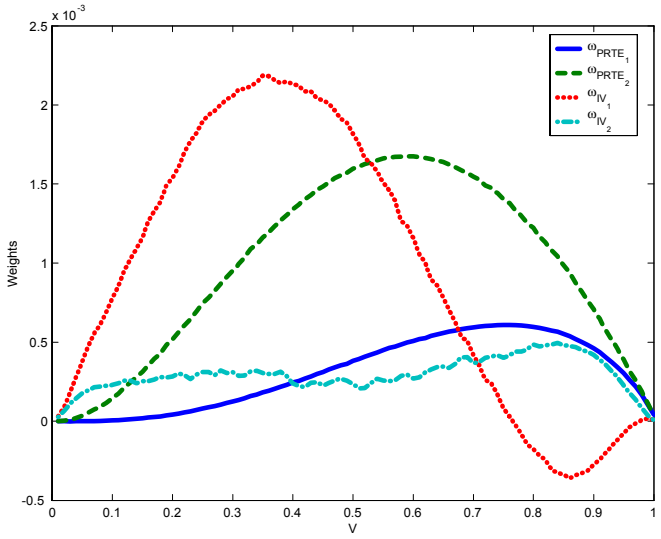
$Y_1 = \alpha + U_1$; $D_1 = 1$ if $-\infty < I \leq W_1$; $U_1 = \sigma_1 \epsilon$; $Z \sim N(-0.0026, 0.27)$ and $Z \perp\!\!\!\perp V$

$I = Z - V$ $V = \sigma_V \epsilon$; $\epsilon \sim N(0, 1)$

Sample Size = 1500

Figure 6B

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$$

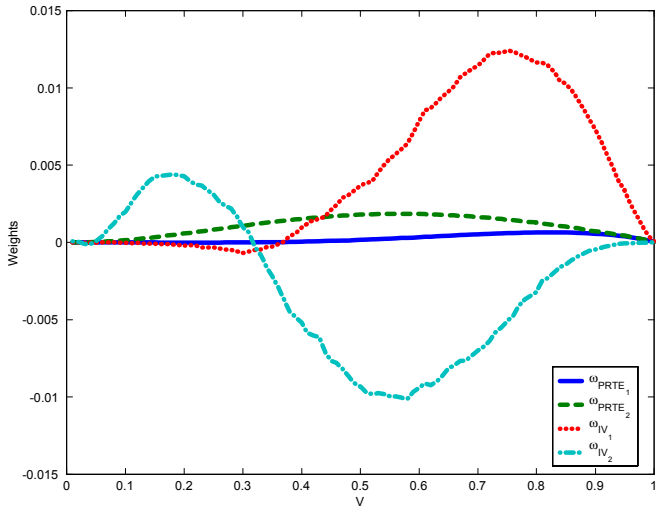
$\Delta^{PRTE} = 0.166$ $IV = 0.247$

Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 42.8\%$

Proportion Induced to Change from $D_2 = 1$ to $D_3 = 1$ $= 9.2\%$

Figure 6C

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}\right)$$

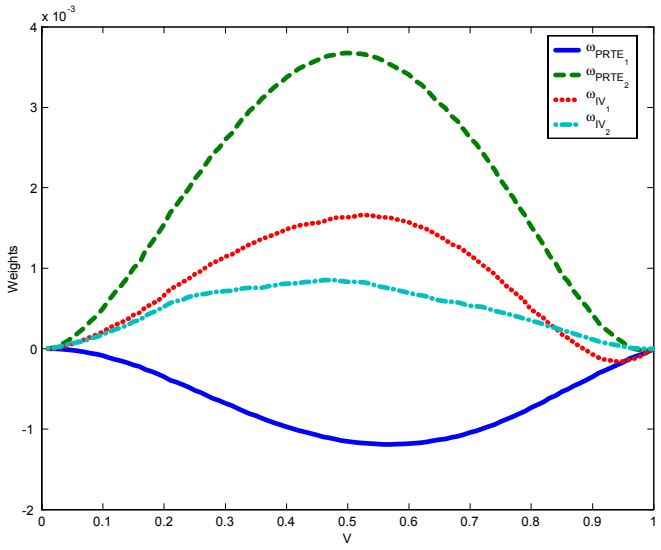
$$\Delta^{PRTE} = 0.159 \qquad IV = 0.346$$

Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 32.1\%$

Proportion Induced to Change form $D_2 = 1$ to $D_3 = 1$ $= 64.7\%$

Figure 6D

Policy Relevant Treatment Effect vs Instrumental Variables Weights by Transition



$$(W_1, W_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$\Delta^{PRTE} = 0.104 \qquad IV = 0.215$$

Proportion Induced to Change from $D_1 = 1$ to $D_3 = 1$ $= 27.3\%$

Proportion Induced to Change from $D_2 = 1$ to $D_3 = 1$ $= 69.3\%$