

# Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy

JAMES J. HECKMAN\*

*This paper compares the structural approach to economic policy analysis with the program evaluation approach. It offers a third way to do policy analysis that combines the best features of both approaches. I illustrate the value of this alternative approach by making the implicit economics of LATE explicit, thereby extending the interpretability and range of policy questions that LATE can answer. (JEL C21, E61)*

## 1. Introduction

Few topics in economics evoke more passion than discussions about the correct way to do empirical policy analysis.<sup>1</sup> These discussions are sometimes framed as

conflicts between “structural” and “reduced form” approaches.<sup>2</sup> In current usage, “structural” is taken to mean parametric, explicitly formulated, empirical economic models.

“Reduced form” has multiple and sometimes conflicting meanings. One strand of the

\* University of Chicago, University College Dublin, Cowles Foundation, Yale University, and the American Bar Foundation. I thank Ed Vytlačil for numerous discussions on the topics discussed in this paper over the years. I also thank him for his comments on this paper. I thank Philipp Eisenhauer, Miriam Gensowski, Tim Kautz, Rodrigo Pinto, Steve Stern, and Ben Williams for helpful comments. I draw on comments and suggestions made on precursors to this paper by Steve Durlauf, Lars Hansen, Richard Robb, and Jose Scheinkman, although none have read this version. The precursor papers are Heckman (1992, 1997, 2000, 2008). I thank Philipp Eisenhauer for his help with constructing the figures in this paper. I thank students in Econ 373 at the University of Chicago, Winter 2010, for serving as guinea pigs for testing the arguments of this paper. Heckman thanks the National Institutes of Health (R01-HD054702), the JB and MK Pritzker Family Foundation, and the American Bar Foundation. Heckman also thanks the Cowles Foundation at Yale University, which supported a visit that facilitated completion of this research. Supplementary material is presented at [http://jenni.uchicago.edu/building\\_bridges/](http://jenni.uchicago.edu/building_bridges/).

<sup>1</sup> See the essays in the symposium “Con out of Economics,” *Journal of Economic Perspectives*, Vol. 24, No. 2, Spring 2010 (Joshua D. Angrist and Jorn-Steffen Pischke 2010; Liran Einav and Jonathan D. Levin 2010; Michael P. Keane 2010; Edward E. Leamer 2010; Aviv Nevo and Michael D. Whinston 2010; Christopher A. Sims 2010; James H. Stock 2010).

<sup>2</sup> James J. Heckman and Edward J. Vytlačil (2007a) discuss the concept of structure and reduced form as defined by the pioneering Cowles Commission econometricians who developed the first rigorous framework for inference and policy analysis. This concept received its clearest statement in a classic paper by Leonid Hurwicz (1962). A structural relationship in its original usage is a relationship invariant to a class of policy interventions and can be used to make valid policy forecasts for policies in that class. The explicit parametrizations used in the modern version of the “structural” literature are intended to represent policy invariant parameters. Reduced forms are one representation of a structure that represent endogenous variables in terms of exogenous variables. Current meanings of “structure” and “reduced form” have changed greatly from their original meanings, but that is not the point of this essay.

reduced form approach uses explicit economic models to motivate and interpret empirical analyses and approximates the economic models using simple econometric techniques. Arnold C. Harberger (1964), Robert Shimer and Ivan Werning (2008), Raj Chetty (2009), and Einav, Amy Finkelstein, and Mark R. Cullen (2009) are good examples of this approach. Chetty (2009) surveys a large literature in this tradition.

Another strand is the “program evaluation” approach surveyed in Guido W. Imbens and Jeffrey M. Wooldridge (2009). This approach focuses on “effects” defined by experiments or surrogates for experiments as the objects of interest, and not the parameters of explicit economic models. It often leaves implicit the economics and policy relevance of the output from its procedures. This paper compares structural and program evaluation approaches—the contrast between the *economic parameters* featured in the structural approach and the “effects” featured in the program evaluation approach.

Explicit structural models facilitate policy analysis. However, after sixty years of experience with fitting explicit structural models on a variety of data sources, empirical economists have come to appreciate the practical difficulties that arise in identifying, and precisely estimating, the parameters of fully specified structural models that can answer a wide variety of policy questions. There have been many demonstrations of the sensitivity of estimates of structural models to assumptions about functional forms and distributions of unobservables. Ta-Chung Liu (1960), David F. Hendry (1980), Sims (1980), and Leamer (1983) gave early warnings about the fragility of standard econometric estimates of explicit economic models. Mark R. Killingsworth and Heckman (1986) and John H. Pencavel (1986) summarize structural estimates of the effects of taxes and wages on labor supply and report estimates from the literature that are sometimes absurd. H. Gregg Lewis (1986) reports the sensitivity

of structural estimates of the causal effect of unionism on wages to the use of alternative methodologies and reports many estimates that are incredibly large. An influential paper by Robert J. LaLonde (1986) is widely interpreted as having demonstrated that standard structural estimation methods applied to non-experimental data cannot duplicate the estimates obtained from a job training experiment.

These and other studies reported in the literature more than twenty years ago fueled the flight of many empirical economists from structural models, even though Heckman and V. Joseph Hotz (1989) cautioned that many applications of the structural approach by those comparing structural estimates with experimental estimates did not perform specification tests to see if the estimated structural models were concordant with the preprogram data. They show that when such tests are performed, the surviving structural models closely match the estimates produced from the experiment analyzed by LaLonde, findings duplicated for other experiments (see Petra E. Todd and Kenneth I. Wolpin 2006, Orazio Attanasio, Costas Meghir, and Ana Santiago 2009, and the discussion in Keane, Todd, and Wolpin forthcoming).

The perceived failures of the structural methods of the 1970s and 1980s produced two different methodological responses.<sup>3</sup> One response—the “program evaluation” approach—was a retreat to statistics, away from the use of explicit economic models either in formulating economic policy questions or in suggesting frameworks for estimating models to answer such questions. The second

<sup>3</sup> Throughout this essay, I consider methodologies that conduct primary empirical analyses. I do not discuss calibration. Practitioners of the calibration approach use well-posed economic models but typically use estimates of key parameters taken from the literature with all of the attendant problems that the parameters utilized are not necessarily appropriate for the model being calibrated. For discussions of calibration, see Lars Peter Hansen and Heckman (1996), Finn E. Kydland and Edward C. Prescott (1996), and Sims (1996).

response was development of a more robust version of the structural approach.

The program evaluation approach replaces the traditional paradigm of economic policy evaluation with the paradigm of the randomized controlled trial. In place of economic models of counterfactuals, practitioners of this approach embrace a statistical model of experiments due to Jerzy Neyman (1923) and David R. Cox (1958) that was popularized by Donald B. Rubin (1974, 1978, 1986), and Paul W. Holland (1986). In this approach, the parameters of interest are *defined* as summaries of the outputs of experimental interventions. This is more than just a metaphorical usage. Rubin and Holland argue that causal effects are defined only if an experiment can be performed.<sup>4</sup> This conflation of the separate tasks of defining causality and identifying causal parameters from data is a signature feature of the program evaluation approach. It is the consequence of the absence of clearly formulated economic models. The probability limits of estimators, and not the parameters of well-defined economic models, are often used to define causal effects or policy effects.

The retreat to statistics in the program evaluation literature left a lot of economics behind. A big loss was the abandonment of economic choice theory. Important distinctions about ex ante and ex post outcomes and subjective and objective evaluations that are central to structural econometrics were forgotten.

The influence of the program evaluation approach is widespread. It is now commonplace for many empirical economists to use the language of “treatment” and “control” to describe the comparisons made in empirical policy studies.

The structural response to the perceived failures of the 1970s and 1980s structural models has focused on nonparametric identification and estimation of well-posed economic

models within which to conduct policy analyses. This line of work preserves the goals of the Cowles Commission pioneers of the structural approach—to estimate models that can make forecasts for a range of widely different policies and criteria. It is more explicit than the program evaluation approach in articulating economic models. It embraces developments in dynamic economics, game theory, auction theory, and the theory of mechanism design. The richness of the theoretical models (in contrast to the intuitive “effects” promoted in the program evaluation literature) make the fruit of this approach more intellectually interesting. It also produces estimates that cumulate across studies.<sup>5</sup>

On the negative side of the ledger on structural estimation, the often complex computational methods that are required to implement this approach make it less transparent. Replication and sensitivity analyses are often more difficult in this approach than in the program evaluation approach. Economists advocating the program evaluation approach dismiss the structural approach as overly complex and not “credible,” focusing on the statistical and computational properties of estimators as the measure of the credibility of procedures (see, e.g., Angrist and Pischke 2008, 2010).<sup>6</sup>

An important paper by Jacob Marschak (1953) suggests a middle ground between these two camps, and is a motivation for the present paper. Writing in the early 1950s

<sup>5</sup> Rosa L. Matzkin (2007) provides a valuable overview of the literature. See also Daniel Akerberg et al. (2007) and Susan Athey and Philip A. Haile (2007) for developments in IO and auction theory respectively. See Jaap H. Abbring and Heckman (2007) for discussion of identification of models of dynamic discrete choice and Victor Aguirregabiria and Pedro Mira (2010) for a survey of computational methods.

<sup>6</sup> For brevity, in this paper my emphasis is on micro-econometric approaches. There are parallel developments and dichotomies in the macro time series and policy evaluation literatures. See Heckman (2000) for a discussion of that literature.

<sup>4</sup> For a recent statement of this position, see Richard Berk, Azusa Li, and Laura J. Hickman (2005).

during the first wave of structural econometrics, Marschak noted that for many problems of policy analysis, it is not necessary to identify or estimate fully specified structural models—the goal of structural analysis as conceived by the Cowles pioneers and successor generations of structural economists. *Marschak's Maxim* suggests that economists should solve well-posed economic problems with minimal assumptions. All that is required to conduct many policy analyses or to answer many well-posed economic questions are policy invariant *combinations* of the structural parameters that are often much easier to identify than the individual parameters themselves and that do not require knowledge of individual structural parameters. This approach advocates transparency and empirical robustness as does the program evaluation approach, but it also focuses attention on answering clearly stated economic and policy questions.

This approach is often less computationally intensive and focuses on a more limited range of policy questions than the very large range of policy questions contemplated by the Cowles pioneers. The computationally less demanding models, more transparent sources of identifiability and the relative ease of performing replication and sensitivity analyses give credibility to this approach. At the same time, this approach improves on the program evaluation approach by producing estimates that have clear economic and policy relevance.

The plan of this paper is as follows. In the next section, I review the range of questions that arise in evaluating economic policies and how the program evaluation approach and the structural approach address them. I use the Roy model (1951) and its extensions as examples of widely used and prototypical structural models. In the following section, I apply Marschak's Maxim to the analysis of the Roy model. This produces an empirical approach that simplifies policy analysis for a certain class of policies. It links the

Roy model to the Local Average Treatment (LATE) framework of Imbens and Angrist (1994). This approach facilitates the interpretation of LATE and extends the range of questions LATE answers. The final section of the paper summarizes the argument.

This paper is not a position piece for or against any particular statistical methodology. It is about the interpretability of estimates and their policy relevance. It is an exploration of how to get the most out of the data using economics to define the questions of interest and statistics to help answer them.

## 2. *The Structural versus the Program Evaluation Approach to Evaluating Economic Policies*

Policy analysis is all about identifying counterfactual states. Counterfactual policy states are possible outcomes in different hypothetical states of the world. An example is the set of outcomes of the same persons in different tax regimes. Causal comparisons entail contrasts between outcomes in alternative possible states holding factors other than the features of the policy being analyzed the same across the contrasts. Thus the person subject to a particular policy is the same as the person who is not, except for treatment status and, possibly, the outcome associated with treatment status.

The concept of causality at the individual level is based on the notion of controlled variation—variation in treatment holding other factors constant. This is Alfred Marshall's (1890) *ceteris paribus* clause that has been the operational definition of causality in economics for over a century.<sup>7</sup> It is distinct from other notions of causality sometimes used in

<sup>7</sup> See Heckman (2000) for a discussion of the intellectual history of causality in economics.

economics that are based on prediction (e.g., Clive W. J. Granger, 1969, and Sims, 1972).<sup>8</sup>

There are two distinct tasks in causal inference and policy analysis: (a) Defining counterfactuals and (b) Identifying causal models from data. Table 1 delineates the two distinct problems.

The first task entails the application of economic models. Models are descriptions of hypothetical worlds obtained by varying—*hypothetically*—the factors determining outcomes. Models may be motivated by empirical evidence, and they may crystalize evidence. They are, however, abstract representations of the evidence with an internal logic of their own.

The second task is inference from data. It requires solving the identification problem including solving practical problems of inference from empirical samples.<sup>9</sup> Economists sometimes differ over what constitutes admissible data for examining any policy question, what prior information should be used and how the prior information should be used. There are no sharp rules to settle these differences.

Part of the controversy surrounding the construction of policy counterfactuals for evaluating policies is a consequence of analysts being unclear about the two distinct tasks represented in table 1 and sometimes confusing them. Particular methods of estimation (e.g., randomization, matching or instrumental variable estimation) have become associated with the definition of “causal parameters” because issues of definition and identification are often conflated.

<sup>8</sup> Holland (1986) makes useful distinctions among commonly used definitions of causality. Nancy Cartwright (2004) discusses a variety of definitions of causality from a philosopher's perspective.

<sup>9</sup> Many econometricians, but not all, distinguish the task of identification from the task of inference. In this distinction, identification is about recovering parameters from population data distributions, where sampling variation is not an issue, and inference is about properties of sampling distributions.

The structural econometric approach to policy evaluation separates these two tasks and emphasizes the role of models in defining hypotheticals and causal effects. Some statisticians reject the use of models in defining causality and seek an assumption-free approach to causal inference and policy analysis (see, e.g., John W. Tukey, 1986).

Any estimator makes assumptions (often implicit) about the behavior of the agents being analyzed. For example, the ability of a randomized controlled trial to identify parameters of interest depends on assumptions about the behavior of the agents being subject to randomization.<sup>10</sup> The structural approach is explicit about these assumptions. The program evaluation approach is often not. Some economists confuse the absence of explicit statements of assumptions with the absence of assumptions.<sup>11</sup>

The “causal models” advocated in the program evaluation literature are motivated by the experiment as an ideal. They do not clearly specify the theoretical mechanisms determining the set of possible counterfactual outcomes, how hypothetical counterfactuals are realized or how hypothetical interventions are implemented except to compare “randomized” with “nonrandomized” interventions. They focus on outcomes, leaving the model for selecting outcomes and the preferences of agents over expected outcomes unspecified.

The emphasis on randomization or its surrogates, like matching or instrumental variables, rules out a variety of alternative channels of identification of policy effects from data. The emphasis on randomization has practical consequences leading to the conflation of Task 1 with Task 2 in table 1.

<sup>10</sup> For example, risk-averse agents may not participate in randomized controlled trials. For discussion of this and other examples, see Heckman (1992) and Heckman and Jeffrey A. Smith (1995).

<sup>11</sup> Mark R. Rosenzweig and Wolpin (2000) present a catalogue of examples of this practice.



TABLE 1  
TWO DISTINCT TASKS THAT ARISE IN THE ANALYSIS OF CAUSAL MODELS

Task	Description	Requirements
1	Defining the Set of Hypotheticals or Counterfactuals	A Well-Specified Economic Theory
2	Identifying Causal Parameters from Data	Mathematical Analysis of Point or Set Identification Joined with Estimation and Testing Theory

Since a randomized protocol is used to define the parameters of interest, this practice sometimes leads to the confusion that randomization is the only way—or at least the best way—to identify causal parameters from data.

The models in the program evaluation literature do not specify the sources of randomness generating variability among agents, i.e., they do not specify why otherwise observationally identical people make different choices. They do not distinguish what is in the agent's information set from what is in the observing economist's information set, although the distinction is fundamental in justifying the properties of any estimator for solving selection and evaluation problems. They do not allow for interpersonal interactions inside and outside of markets in determining outcomes that are at the heart of game theory, general equilibrium theory, and models of social interactions and contagion (see, e.g., Aguirregabiria and Mira 2010; Donald J. Brown and Matzkin 1996; William A. Brock and Steven N. Durlauf 2001; Durlauf and H. Peyton Young 2001; Charles F. Manski 1993; Elie Tamer 2003).

The goal of the structural econometrics literature, like the goal of all science, is to understand the causal mechanisms producing effects so that one can use empirical versions of models to forecast the effects of interventions never previously experienced,

to calculate a variety of policy counterfactuals and to use theory to guide choices of estimators to interpret evidence and to cumulate evidence across studies. These activities require models for understanding “causes of effects” in contrast to the program evaluation literature that focuses only on the “effects of causes” (Holland 1986).

Before turning to a specific comparison of the two approaches, it is useful to review the variety of questions that arise in policy analysis. What are the economically interesting questions? The success or failure of any methodology hinges on how well it answers substantive policy questions. Thus, it is helpful to have a list in front of us to examine which questions are addressed or ignored by different approaches.

### 2.1 *Policy Evaluation Problems and Criteria of Interest*

Three broad classes of policy evaluation problems arise in economics. Policy evaluation problem one is:

**P1** *Evaluating the Impacts of Implemented Interventions on Outcomes Including Their Impacts on the Well-Being of the Treated and Society at Large.*

It is useful to distinguish objective or public outcomes that can in principle be measured

by all external observers from “subjective” outcomes that are the evaluations of the agents experiencing treatment (e.g., patients) or the agents prescribing treatment (e.g., physicians). Objective outcomes are intrinsically ex post (“after the fact”) in nature. The literature on program evaluation focuses on ex post objective outcomes.

The structural literature studies both subjective and objective outcomes. Subjective outcomes can be ex ante (“anticipated”) or ex post. The outcome of a medical trial produces both a cure rate and the pain and suffering of the patient. Ex ante anticipated pain and suffering from a procedure may be different from ex post realized pain and suffering. A similar distinction arises in an analysis of the returns to schooling. Monetary returns are only part of total benefits which include important elements of psychic cost.<sup>12</sup> Ex ante evaluations of outcomes by agents may differ from their ex post evaluations. Thus in advance of going to school, students may have expectations about rewards and costs that differ from their realizations. Expected rewards govern responses to market incentives. The impacts in **P1** include individual level or population level counterfactuals and their valuations. “Well-being” in **P1** includes the valuations of the outcomes of interventions by the agents being analyzed or other parties (e.g., the parents of the agent or “society” at large). They may be ex ante or ex post, and both are of interest in evaluating policy. Regret and anticipation are important aspects of public approval of policies.

**P1** is the problem of *internal validity*. It is the problem of identifying the impacts of an intervention conducted in a given environment. Solving just this problem can be a challenging task and good answers are valuable. However, most economic policy

evaluation is conducted with an eye toward the future and toward informing decisions about new policies and applications of old policies to new environments.

There is a second problem frequently encountered in policy analysis.

**P2** *Forecasting the Impacts (Constructing Counterfactual States) of Interventions Implemented in One Environment in Other Environments, including Impacts on Well-Being.*

This is the problem of *external validity*: taking a treatment parameter or a set of parameters identified in one environment to another environment.<sup>13</sup>

The most ambitious problem is forecasting the effects of a new policy, never previously experienced:

**P3** *Forecasting the Impacts of Interventions (Constructing Counterfactual States Associated with Interventions) Never Historically Experienced, including Their Impacts on Well-Being.*

**P3** is a problem that economic policy analysts have to solve daily. Structural econometrics addresses this question. The program evaluation approach does not.

## 2.2 A Prototypical Economic Model for Policy Evaluation

Abstract discussions of policy evaluation problems become very tedious very fast. To be specific and to keep the discussion focused, consider the following version of the Roy model, which is a useful framework for policy evaluation.<sup>14</sup> The Roy model and

<sup>12</sup> See, e.g., Gary S. Becker (1964) and Flavio Cunha, Heckman, and Salvador Navarro (2005).

<sup>13</sup> See, e.g., Donald T. Campbell and Julian C. Stanley (1963), Heckman, LaLonde, and Smith (1999), or William R. Shadish, Thomas D. Cook, and Campbell (2002).

<sup>14</sup> Heckman (2008) and Heckman and Vytlačil (2007a, 2007b) present general discussions of the policy evaluation problem.

its extensions undergird a huge literature in microeconometrics.<sup>15</sup> It was applications of the Roy model that fueled the flight from structural econometrics in the 1980s.

The Roy model is a model of hypothetical outcomes. Economic theory defines the space of possible counterfactual outcomes. It also specifies agent decision rules. It executes task 1 of table 1.

A. D. Roy (1951) considered an economy where agents face two potential outcomes  $(Y_0, Y_1)$  with distribution  $F_{Y_0, Y_1}(y_0, y_1)$  where “0” refers to the no treatment state and “1” refers to the treated state and  $(y_0, y_1)$  are particular values of random variables  $(Y_0, Y_1)$ . More generally, the set of potential outcomes is  $\{Y_s\}_{s \in \mathcal{S}}$  where  $\mathcal{S}$  is the set of indices of potential outcomes. In the Roy model,  $\mathcal{S} = \{0, 1\}$ . I focus on two outcome models to simplify the exposition. In the application of Reuben Gronau (1974) and Heckman (1974),  $Y_0$  is the value of nonmarket time, and  $Y_1$  is the value of market time. In Robert J. Willis and Sherwin Rosen (1979),  $Y_0$  is the present value of high school earnings, and  $Y_1$  is the present value of college earnings. In the application of Insan Tunali (2000),  $Y_0$  and  $Y_1$  are incomes in two regions. The central question recognized in this literature is that analysts observe either  $Y_0$  or  $Y_1$ , but not both, for any person. In the program evaluation literature, this is called the **evaluation problem**.

In addition to this problem, there is the **selection problem**. The values of  $Y_0$  or  $Y_1$  that are observed are not necessarily a random sample of the potential  $Y_0$  or  $Y_1$  distributions. In the original Roy model, an agent selects into sector 1 if  $Y_1 > Y_0$ . Let  $D$  be an indicator or dummy variable. Then

$$(1) \quad D = \mathbf{1}(Y_1 > Y_0),$$

where “ $\mathbf{1}$ ” is a function that is 1 if the condition in the argument holds and is 0 otherwise.

A variety of more general decision rules have been considered in the structural literature. A simple generalization of the Roy model adds cost  $C$ . This can be thought of as a cost of moving from “0” to “1,” e.g., tuition in the Willis–Rosen model, costs of work in the Gronau–Heckman model, or costs of migration in the Tunali model. The decision rule becomes

$$(2) \quad D = \mathbf{1}(Y_1 - Y_0 - C > 0).$$

This framework defines a set of counterfactual outcomes and costs  $(Y_0, Y_1, C)$  with distribution  $F_{Y_0, Y_1, C}(y_0, y_1, c)$  and a mechanism for selecting which element of  $(Y_0, Y_1)$  is observed for any person. The outcome observed for any person,  $Y$ , can be written as

$$(3) \quad Y = DY_1 + (1 - D)Y_0.$$

This representation is the Richard E. Quandt (1958, 1972) switching regression framework.

Agents may make their choices under imperfect information. Let  $\mathcal{I}$  denote the agent’s information. In advance of participation, the agent may be uncertain about all components of  $(Y_0, Y_1, C)$ . The expected benefit is  $I_D = E(Y_1 - Y_0 - C | \mathcal{I})$ . Then

$$(4) \quad D = \mathbf{1}(I_D > 0).$$

Moreover, the decision maker selecting “treatment” may be different than the person who experiences the outcomes  $(Y_0, Y_1)$ . Thus parents may make schooling decisions for their children; doctors may make treatment decisions for patients. More generally,

<sup>15</sup> See Heckman (2001) and Heckman and Vytlačil (2007a, 2007b) for surveys. See also Francis Vella (1998) and Keane, Todd, and Wolpin (forthcoming).



decisions to participate may entail joint approval of all parties.<sup>16</sup>

The ex post objective outcomes are  $(Y_0, Y_1)$ . The ex ante outcomes are  $E(Y_0|\mathcal{I})$  and  $E(Y_1|\mathcal{I})$ . The ex ante subjective evaluation is  $I_D$ . The ex post subjective evaluation is  $Y_1 - Y_0 - C$ . Agents may regret their choices because realizations may differ from anticipations.

The ex ante versus ex post distinction is essential for understanding behavior. In environments of uncertainty, agent choices are made in terms of ex ante calculations. Yet the treatment effect literature largely reports ex post returns. For example, the recent literature on the returns to schooling reports ex post returns (David H. Autor, Lawrence F. Katz, and Melissa S. Kearney 2005; Katz and Autor 1999; Katz and Kevin M. Murphy 1992). Yet it is the analysis of ex ante returns that is needed to understand why, over time, responses to increases in ex post returns to schooling have been so sluggish.<sup>17</sup>

Pedro Carneiro, Karsten T. Hansen, and Heckman (2001, 2003), Cunha, Heckman, and Navarro (2005, 2006) and Cunha and Heckman (2007) develop econometric methods for distinguishing ex ante from ex post evaluations of social programs. Abbring and Heckman (2007) provide an extensive survey of this literature.<sup>18</sup>

In the language of the program evaluation literature,  $Y_1 - Y_0$  is the individual level treatment effect. It is also the Marshallian ceteris paribus causal effect. Because of the evaluation problem, it is generally impossible to identify individual level treatment

effects (Task 2). Even if it were possible,  $Y_1 - Y_0$  does not reveal the ex ante subjective evaluation  $I_D$  or the ex post assessment  $Y_1 - Y_0 - C$ .

Economic policies can operate through changing  $(Y_0, Y_1)$  or through changing  $C$ . Thus, in the college going example, policies may reduce tuition costs or reduce commuting costs (David Card 2001). They may tax future earnings. The structural approach considers policies affecting both returns and costs.<sup>19</sup>

### 2.3 Population Parameters of Interest

Because it is generally impossible to identify individual-level treatment effects, analysts typically seek to identify parameters defined at the population level. Conventional parameters include the Average Treatment Effect ( $ATE = E(Y_1 - Y_0)$ ), the effect of Treatment on The Treated ( $TT = E(Y_1 - Y_0|D = 1)$ ), or the effect of Treatment on the Untreated ( $TUT = E(Y_1 - Y_0|D = 0)$ ).

However, in addressing economic policy evaluation questions, a variety of other population level parameters are often more interesting. In positive political economy, the fraction of the population that perceives a benefit from treatment is of interest. This is called the **voting criterion** and is

$$\Pr(I_D > 0) = \Pr(E(Y_1 - Y_0 - C|\mathcal{I}) > 0).$$

In gauging support for a policy in place, the percentage of the population that ex post perceives a benefit may also be of interest:  $\Pr(Y_1 - Y_0 - C > 0)$ .

More generally, for evaluation of the distribution of welfare, knowledge of the ex ante and ex post joint distributions of outcomes are

<sup>16</sup> See Dale J. Poirier (1980) and Henry S. Farber (1983).

<sup>17</sup> The econometrician may possess a different information set,  $\mathcal{I}_e$ . Choice probabilities computed against one information set are not generally the same as those computed against another information set. Operating with hindsight, the econometrician may be privy to some information not available to agents when they make their choices.

<sup>18</sup> Manski (2004) surveys a rich literature on the elicitation of expectations.

<sup>19</sup> See, e.g., Heckman, Lance Lochner, and Christopher Taber (1998a, 1998b, 1998c), Esther Dufo (2004), Jeremy Lise, Shannon Seitz, and Jeffrey Smith (2005, 2006), James Albrecht, Gerard Van den Berg, and Susan Vroman (2009), or Donghoon Lee and Wolpin (2006).

of interest.<sup>20</sup> Because of the evaluation problem, it is very difficult to identify the joint distribution because we generally do not observe  $Y_0$  and  $Y_1$  together. This problem plagues all methodologies including social experiments.<sup>21</sup>

Determining marginal returns to a policy is a central goal of economic analysis. In the generalized Roy model, the margin is specified by people who are indifferent between “1” and “0,” i.e., those for whom  $I_D = 0$ . The mean effect of treatment for those at the margin of indifference is

$$E(Y_1 - Y_0 | I_D = 0).$$

I discuss approaches for identifying this parameter in section 3.

#### 2.4 Treatment Effects versus Policy Effects

Different policies can affect treatment choices and outcomes differently. Each of the population-level treatment effects discussed in the previous subsection can be defined for different policy regimes. Economists are often more interested in the effects of policies on outcomes than in conventional treatment effects.

To illustrate this point, consider the Policy Relevant Treatment Effect (Heckman and Vytlačil, 2001c) which extends the Average Treatment Effect by accounting for voluntary participation in programs. It is designed to address problems **P2** and **P3**. Let “ $b$ ” represent a baseline policy (“before”) and “ $a$ ” represent a policy being evaluated (“after”). Let  $Y^a$  be the outcome under policy  $a$ , while  $Y^b$  is the outcome under the baseline.  $(Y_0^a, Y_1^a, C^a)$  and  $(Y_0^b, Y_1^b, C^b)$  are outcomes under the two policy regimes.

Policy invariance facilitates the job of answering problems **P2** and **P3**. If some

parameters are invariant to policy changes, they can be safely transported to different policy environments. Structural econometricians search for policy invariant “deep parameters” that can be used to forecast policy changes.<sup>22</sup>

Under one commonly invoked form of policy invariance, policies keep the potential outcomes unchanged for each person:  $Y_0^a = Y_0^b$ ,  $Y_1^a = Y_1^b$ , but affect costs ( $C^a \neq C^b$ ).<sup>23</sup> A tuition policy in the absence of general equilibrium effects is an example. Invariance of this type rules out social effects including peer effects and general equilibrium effects. Let  $D^a$  and  $D^b$  be the choice taken under each policy regime. Invoking invariance of potential outcomes, the observed outcomes under each policy regime are  $Y^a = Y_1 D^a + Y_0(1 - D^a)$  and  $Y^b = Y_1 D^b + Y_0(1 - D^b)$ . The **Policy Relevant Treatment Effect** (PRTE) is

$$\text{PRTE} = E(Y^a - Y^b).$$

This is the Benthamite comparison of aggregate outcomes under policies “ $a$ ” and “ $b$ ”.<sup>24</sup> PRTE extends ATE by recognizing that policies affect incentives to participate ( $C$ ) but do not force people to participate. Only if  $C$  is very large under  $b$  and very small under  $a$ , so there is universal nonparticipation under  $b$  and universal participation under  $a$ , would ATE and PRTE be the same parameter.

#### 2.5 The Structural Approach versus the Program Evaluation Approach

The recent literature on program evaluation in economics draws on a model of counterfactuals and causality attributed to

<sup>20</sup> See Heckman, Smith, and Nancy Clements (1997) and Abbring and Heckman (2007) for discussions of these parameters.

<sup>21</sup> See Abbring and Heckman (2007) for discussions of alternative approaches to identify or bounding these joint distributions.

<sup>22</sup> Ragnar Frisch (1933, translated in 2009) considered these notions under the concept of “autonomy.” See Marschak (1953) and Hurwicz (1962), who develop refinements of this concept.

<sup>23</sup> Heckman and Vytlačil (2007a, 2007b) discuss a variety of invariance assumptions.

<sup>24</sup> If potential outcomes are not policy invariant, one would work with  $(Y_0^b, Y_1^b)$  and  $(Y_0^a, Y_1^a)$ .

Rubin by Holland (1986).<sup>25</sup> It defines causality using experimental manipulations and thereby creates the impression in the minds of many followers of this approach that random assignment is the most convincing way to identify causal models.

Neyman and Rubin postulate counterfactuals  $\{Y_s\}_{s \in \mathcal{S}}$ .<sup>26</sup> They do not develop choice mechanisms that determine which outcome is selected or the subjective evaluations of treatments. There is no discussion of the mechanisms producing the outcomes studied or the relationship between outcomes and choice mechanisms.

Rubin (1986) invokes a portion of the traditional econometric invariance assumptions developed by Hurwicz (1962).<sup>27</sup> Since he does not develop models for choice or subjective evaluations, he does not consider the more general invariance conditions for both objective and subjective evaluations that are features of the structural literature.<sup>28</sup> The range of issues covered by the two approaches is given in table 2.<sup>29</sup>

The Neyman–Rubin model does not consider many issues discussed in structural econometrics. It is at best an incomplete introduction to some of the important issues in evaluating social policies.<sup>30</sup> The analysis in Rubin’s 1974 and 1978 papers is a dichotomy between randomization and

nonrandomization, and not an explicit treatment of particular selection mechanisms in the nonrandomized case as is developed in the structural literature.

The statisticians who have had the greatest impact on the program evaluation literature in economics conflate the two tasks stated in table 1. The discussion of Holland (1986) illustrates this point and the central role of the randomized controlled trial to the Holland–Rubin analysis. After explicating the “Rubin model,” Holland makes a very revealing claim: there can be no causal effect of gender on earnings because analysts cannot randomly assign gender. This statement confuses the act of defining a causal effect (a purely mental act performed within a model) with empirical difficulties in estimating it.<sup>31</sup> The local average treatment effect “LATE” parameter of Imbens and Angrist (1994), discussed in section 3, follows in this tradition and uses instrumental variables as surrogates for randomization. LATE is defined by an instrument and conflates tasks 1 and 2 of table 1. In section 3, I present a framework that defines the LATE parameter within the generalized Roy model discussed in subsection 2.2 that separates issues of definition of parameters from issues of identification.

## 2.6 Identifying Policy Parameters

The structural approach to policy evaluation addresses policy evaluation questions

<sup>25</sup> See, e.g., Imbens and Wooldridge (2009).

<sup>26</sup>  $\mathcal{S} = \{0, 1\}$  in the Roy example.

<sup>27</sup> He calls it “SUTVA” for Stable Unit Treatment Value Assumption.

<sup>28</sup> See, e.g., Heckman (2008) or Heckman and Vytlačil (2007a, 2007b) for discussions of these conditions.

<sup>29</sup> Not every paper in the empirical structural literature addresses all of the issues in table 2 in deriving its estimates, but most papers in this tradition are explicit in noting which questions are not addressed.

<sup>30</sup> It is a mark of their detachment from economics that advocates of the program evaluation approach in economics claim that Marshall’s *ceteris paribus* concept (Marshall 1890) for defining causality was developed in a 1974 paper by Rubin, and that they also attribute versions of the Cowles Commission policy invariance assumptions to Rubin.

<sup>31</sup> As another example of the same point, Rubin denies that it is possible to define a causal effect of sex on intelligence because a randomization cannot *in principle* be performed. “Without treatment definitions that specify actions to be performed on experimental units, I cannot unambiguously discuss causal effects of treatments” (Rubin 1978, p. 39). In this and many other passages in the statistics literature, a causal effect is defined by a randomization. Issues of definition and identification are confused. This confusion continues to flourish in the literature in applied statistics. For example, Berk, Li, and Hickman (2005) echo Rubin and Holland by insisting that if an experiment cannot “in principle” be performed, a causal effect cannot be defined.

TABLE 2  
COMPARISON OF THE ASPECTS OF EVALUATING SOCIAL POLICIES THAT ARE COVERED BY THE  
NEYMAN–RUBIN APPROACH AND THE STRUCTURAL APPROACH

	Neyman–Rubin Framework	Structural Framework
Counterfactuals for objective outcomes ( $Y_0, Y_1$ )	Yes	Yes
Agent valuations of subjective outcomes ( $I_D$ )	No (choice-mechanism implicit)	Yes
Models for the causes of potential outcomes	No	Yes
Ex ante versus ex post counterfactuals	No	Yes
Treatment assignment rules that recognize the voluntary nature of participation	No	Yes
Social interactions, general equilibrium effects and contagion	No (assumed away)	Yes (modeled)
Internal validity (problem <b>P1</b> )	Yes	Yes
External validity (problem <b>P2</b> )	No	Yes
Forecasting effects of new policies (problem <b>P3</b> )	No	Yes
Distributional treatment effects	No <sup>a</sup>	Yes (for the general case)
Analyze relationship between outcomes and choice equations	No (implicit)	Yes (explicit)

<sup>a</sup>An exception is the special case of common ranks of individuals across counterfactual states: “rank invariance.” See the discussion in Abbring and Heckman (2007).

**P1–P3** by estimating models for  $Y_0$ ,  $Y_1$ , and  $C$  in different economic environments. Commonly used specifications write

$$\begin{aligned}
 (5) \quad Y_1 &= \mu_1(X) + U_1, \\
 Y_0 &= \mu_0(X) + U_0, \\
 C &= \mu_C(Z) + U_C,^{32}
 \end{aligned}$$

where  $(X, Z)$  are observed by the analyst, and  $U_0, U_1, U_C$  are unobserved. Economic theory specifies the ingredients in  $Z$  and  $X$ . In general, there is no “objective” way to choose these conditioning variables. Any argument for inclusion or exclusion of variables has to be made by an appeal to theory—implicit or explicit.

To simplify notation, I define  $Z$  to include all of  $X$ . Variables in  $Z$  not in  $X$  are instruments. Write  $I_D = E(Y_1 - Y_0 - C | \mathcal{I}) = \mu_D(Z) - V$  where  $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z) | \mathcal{I})$  and  $V = -E(U_1 - U_0 - U_C | \mathcal{I})$ . In this notation, choice equation (3) can be expressed as

$$(6) \quad D = 1(\mu_D(Z) > V).$$

In the early literature that implemented this approach,  $\mu_0(X)$ ,  $\mu_1(X)$ , and  $\mu_C(Z)$  were assumed to be linear in the parameters, and the unobservables were assumed to be normal and distributed independently of  $X$  and  $Z$ .

The caricature of the structural approach in the recent program evaluation literature is that linearity and normality are essential to this approach. In truth, the essential aspect of the structural approach is joint modeling

<sup>32</sup>  $\mu_1(X) = E(Y_1 | X)$ ;  $\mu_0(X) = E(Y_0 | X)$ ;  $\mu_C(Z) = E(C | Z)$ .

of outcome and choice equations. Structural econometricians have developed nonparametric identification analyses for the Roy and generalized Roy models. See Heckman and Bo E. Honoré (1990), Heckman (1990), Hyungtaik Ahn and James L. Powell (1993), Donald W. K. Andrews and Marcia M. A. Schafgans (1998), and Mitali Das, Whitney K. Newey, and Vella (2003). The field has moved well beyond the parametric functional forms used in the early papers that were the targets of the 1980's criticism. Traditional distributional and parametric assumptions are relaxed in the recent structural econometric literature. (See Akerberg et al. 2007; Athey and Haile 2007; Matzkin 1992, 1993, 1994, 2007, 2010a, 2010b; Powell 1994; Vella 1998; Abbring and Heckman 2007; and Keane, Todd, and Wolpin forthcoming; for reviews.)

As an illustration of the benefits of the structural approach for solving policy problem **P3**, consider the analysis of college choice by Stephen V. Cameron and Taber (2004). Suppose that one seeks to know the effects of increases in the expected gross returns to college  $E(Y_1 - Y_0 | \mathcal{I})$  on college choices. From equation (6), if one knows the effects of variations in tuition ( $C$ ) on college choices, one can use the choice outcomes associated with variations in  $C$  to accurately predict the response to changes of equal magnitude (in opposite sign) in expected mean gross returns, even if returns to schooling have never varied in the past.

## 2.7 *Marschak's Maxim and the Relationship Between the Structural Economics Literature and the Program Evaluation Literature: A Synthesis*

Structural models make explicit the preferences and constraints that govern individual decisions, the rules of interaction among agents in market and social settings, and the sources of variability across agents. These features facilitate finding answers to policy

questions **P1–P3**. They are absent from the program evaluation literature.

At the same time, that literature makes fewer *statistical* assumptions in terms of independence, functional form, exclusion, and distributional assumptions than the standard structural estimation literature in econometrics. This is an attractive feature of the program evaluation approach.<sup>33</sup> The greater simplicity of its estimation schemes fosters transparency, replicability, and sensitivity analyses.<sup>34</sup> While the structural economics literature has advanced greatly in recent years in terms of producing a robust version of its product, it is an unpleasant fact that fully-specified structural models are often harder to compute. It is more difficult to replicate the estimates from them and to test the sensitivity of estimates to assumptions.

The two approaches can be reconciled by noting that for many policy questions, it is not necessary to identify fully specified models to answer a range of policy questions. It is often sufficient to identify policy-invariant *combinations* of structural parameters. These combinations are often much easier to identify (i.e., require fewer and weaker assumptions), and do not require knowledge of the particular individual structural parameters that form the combination.

Marschak (1953) recognized that the answers to many policy evaluation questions do not require knowledge of all of the component parts of full structural models. I call this principle **Marschak's Maxim** in honor of his insight. Consider estimating the marginal effect of policy expansions. The traditional structural approach identifies the component parts of  $E(Y_1 - Y_0 | I_D = 0)$  constructed from

<sup>33</sup> A recent exception to this robust approach is the analysis of Angrist and Pischke (2008) who claim that policy evaluation models should be based on linear-in-parameter estimating equations.

<sup>34</sup> Manski (1995, 2003) has developed an elaborate methodology for sensitivity analysis in the program evaluation literature for certain classes of data.



estimates of the parameters of equations (5) and (6) and assembles them to estimate the marginal effect of the policy expansion (see Anders Björklund and Robert Moffitt 1987).

In the next section, I exposit an approach that is consistent with Marschak's Maxim that directly identifies the combination of parameters that define  $E(Y_1 - Y_0 | I_D = 0)$  to solve policy problems, rather than identifying the component parts of the structural model and building it up from the components. Marschak's Maxim is an application of Occam's Razor to policy evaluation.<sup>35</sup>

### 3. *Using Economics to Interpret What LATE Estimates and to Make it Useful for Evaluating a Broad Range of Policies*

This section presents an example of an approach to policy evaluation that implements Marschak's Maxim in the context of LATE. It makes the implicit economics in LATE explicit and thereby expands the range of policy questions that LATE can address.

In the economic theory of policy evaluation, a comparison between marginal benefits and marginal costs determines the optimal size of social programs. For example, to evaluate the optimality of a policy that promotes expansion in college attendance, analysts need to estimate the return to college for the marginal student, and compare it to the marginal cost of the policy. This task requires that analysts identify marginal returns.

In the spirit of the program evaluation literature, in the following discussion, I ignore general equilibrium effects, and I do not emphasize the ex ante and ex post distinction. Both topics are addressed in many

papers in the structural approach. To simplify the notation, I keep the conditioning variables  $X$  implicit unless it clarifies matters by making them explicit. I follow standard conventions and denote random variables by capital letters and their realizations by the corresponding lower case letters. Thus  $Z = z$  means that random variable  $Z$  takes the value  $z$ .  $Z$  is a vector with  $K$  components,  $Z = (Z_1, \dots, Z_K)$ .  $z^j$  means a particular realization of  $Z$ , i.e.,  $z^j = (z_1^j, \dots, z_K^j)$ . I assume for simplicity that all means are finite.

The equation for ex post outcome  $Y$  as a function of participation status is

$$(7) \quad Y = \alpha + \beta D + \varepsilon,$$

where  $D$  is a dummy variable indicating participation in a program,  $\beta$  is the individual return to participation or treatment effect and  $\varepsilon$  is an error term that is unobserved by the analyst. Equation (7) is one representation of the Quandt switching regression model (3). In terms of counterfactual notation,  $\alpha = \mu_0$ ,  $\varepsilon = U_0$  and  $Y_0 = \mu_0 + \varepsilon$ , and  $\beta = (Y_1 - Y_0) = \mu_1 - \mu_0 + U_1 - U_0$ .

Estimating marginal returns to a policy that changes  $D$  is a relatively simple task if the effect of the policy is the same for everyone (conditional on  $X$ ). This is the case when  $U_1 - U_0 = 0$ , and  $\varepsilon = U_0 = U_1$ . In this case, the mean marginal and average returns are the same for all people with the same  $X$ .

The recent literature on policy evaluation allows for the possibility that  $\beta$  varies among people even after conditioning on  $X$ . Denoting the mean of  $\beta$  by  $\bar{\beta}$ , the outcome equation can be written as

$$(8) \quad Y = \alpha + \bar{\beta} D + \{\varepsilon + (\beta - \bar{\beta})D\},$$

where  $\bar{\beta} = \mu_1 - \mu_0$ .<sup>36</sup> If  $\beta$  is uncorrelated with  $D$ , the only new econometric problem

<sup>35</sup> Heckman and Richard Robb (1985, p. 190) apply a version of Marschak's Maxim to methods in program evaluation. Thus matching can identify the average treatment effect or treatment on the untreated without identifying the component parts of equations (5) and (6).

<sup>36</sup> I assume that the mean is finite:  $E|\beta| < \infty$ .

that arises in the analysis of (8) that is not present in the traditional analysis of (7) is that the error term is heteroscedastic. As in the case where  $\beta$  is a common parameter shared by everyone with the same  $X$ , the main econometric problem for inference about  $\bar{\beta}$  is that  $D$  is correlated with  $\varepsilon$ .<sup>37</sup>

$\beta$  is statistically independent of  $D$  if, given  $X$ , agents cannot anticipate their ex post idiosyncratic gains from participation so  $\beta - \bar{\beta}$  is independent of  $D$  because it is not in the agent's information set  $\mathcal{I}$ . Another reason why  $\beta$  might be independent of  $D$  is that agents know  $\beta - \bar{\beta}$ , but do not act on it in choosing  $D$ . In both cases, mean marginal returns are the same as mean average returns. Under standard conditions, application of instrumental variables identifies  $\bar{\beta}$ .<sup>38</sup> One does not have to specify the model by which  $D$  is selected. All valid instruments identify  $\bar{\beta}$ .

The recent literature analyzes the less conventional case where agents know and make choices about  $D$  with at least partial knowledge of  $\beta = Y_1 - Y_0$ , and the agent knows more about  $Y_1 - Y_0$  than what is in the observing economists' conditioning set  $(X, Z)$ . Instrumental variables (IV) do not in general estimate  $\bar{\beta}$  and instrumental variables estimators using different instruments have different probability limits.<sup>39</sup> Structural selection models can estimate the distribution of  $\beta$  (and hence  $\bar{\beta}$ ) and answer a range of the public policy evaluation questions discussed in section 2 but under assumptions that are held to be "incredible" in the

program evaluation literature.<sup>40</sup> Angrist and Pischke (2008, 2010) offer the Local Average Treatment Effect (LATE) as a "credible" alternative to structural methods.

Under the conditions reviewed in the next subsection, Imbens and Angrist (1994) show that instrumental variable estimators identify LATE, which measures the mean gross return to treatment for individuals induced into treatment by a change in an instrument. The LATE parameter is widely interpreted as estimating the mean return at the margin defined by manipulation of the instrument.

In general, LATE is not the same as  $\bar{\beta}$ , but it might be all that is needed to evaluate any particular policy. The key question is "what question does LATE answer?" Unfortunately, the people induced to go into state 1 ( $D = 1$ ) by a change in any particular instrument need not be the same as the people induced to go to state 1 by policy changes other than those corresponding exactly to the variation in the instrument. A desired policy effect may not directly correspond to the variation captured by the IV. The people induced to change state by the instrument are not identified in LATE. Widely held intuitions about what IV identifies break down in this case since different instruments identify different parameters. Moreover, if there is a vector of instruments that generates choices and the components of the vector are intercorrelated, IV estimates using the components of  $Z$  as instruments, one at a time, do not, in general, identify the policy effect corresponding to varying that instrument, keeping all other instruments fixed, the *ceteris paribus* effect of the change in the instrument. Recent research that builds on and improves LATE shows how to use the generalized Roy model implicit in LATE to estimate the mean marginal returns to alternative ways of producing marginal expansions of programs when variation in the

<sup>37</sup> If this problem is solved, it is possible to estimate the distribution of  $\beta$  (see, e.g., Heckman and Smith 1998).

<sup>38</sup> Matching and selection methods also identify  $\bar{\beta}$  under their assumed conditions. Regression discontinuity methods are a local version of instrumental variables (Jinyong Hahn, Todd, and Wilbert Van der Klaauw 2001; Heckman and Vytlačil 2007b).

<sup>39</sup> See Heckman and Robb (1985), p. 196; Heckman (1997); Heckman, Daniel Schmieder, and Sergio Urzua (forthcoming).

<sup>40</sup> See Abbring and Heckman (2007) for a survey of methods for estimating and bounding the distribution of  $\beta$ .

available instruments does not correspond exactly to the variation induced by proposed policies.<sup>41</sup> This research also enables analysts to determine the people who are affected by changes in instruments. I first review LATE and then consider recent extensions of it.

### 3.1 LATE

LATE is defined by the variation of an instrument. The instrument in LATE plays the role of a randomized assignment. Indeed, randomized assignment is an instrument.<sup>42</sup> Using the notation of section 2,  $Y_0$  and  $Y_1$  are potential ex post outcomes. Instrument  $Z$  assumes values in  $\mathcal{Z}$ ,  $z \in \mathcal{Z}$ .  $D(z)$  is an indicator of hypothetical choice representing what choice the individual would have made had the individual's  $Z$  been exogenously set to  $z$ .  $D(z) = 1$  if the person chooses (is assigned to) 1.  $D(z) = 0$ , otherwise. One can think of the values of  $z$  as fixed by an experiment or by some other mechanism independent of  $(Y_0, Y_1)$ . All policies are assumed to operate through their effects on  $Z$ . It is assumed that  $Z$  can be varied conditional on  $X$ .

Imbens and Angrist (1994) make three assumptions to define LATE. Their first assumption is an instrumental variables assumption formulated in terms of a model of counterfactuals:

$$(IA-1) \quad (Y_0, Y_1, \{D(z)\}_{z \in \mathcal{Z}}) \perp\!\!\!\perp Z | X$$

where “ $\perp\!\!\!\perp$ ” denotes independence, and  $A \perp\!\!\!\perp B | X$  means  $A$  is independent of  $B$  conditional on  $X$ .  $(Y_1, Y_0, \{D(z)\}_{z \in \mathcal{Z}})$  are random variables defined over the population. Assumption (IA-1) states that the values of potential outcomes and potential choices are independent of  $Z$  (conditioning on  $X$ ).

Imbens and Angrist also assume a rank condition:

(IA-2)  $\Pr(D = 1 | Z = z)$  is a nontrivial function of  $z$  conditional on  $X$ .

This says that the distribution of  $P(Z) = \Pr(D = 1 | Z)$  is nondegenerate conditional on  $X$ .

To make IV identify a treatment effect, they invoke a monotonicity condition on the  $D(z)$  at the individual level.

(IA-3) For any two values of  $Z$ , say  $Z = z^1$  and  $Z = z^2$ , either  $D(z^1) \geq D(z^2)$  for all persons, or  $D(z^1) \leq D(z^2)$  for all persons.

This condition is a statement *across* people.  $z^1$  and  $z^2$  are two different values of vector  $Z$ . Fixing the instrument at two values  $z^1$  and  $z^2$  moves choices across people in the same direction (either in favor of 1 or against it). This condition does not require that for any other two values of  $Z$ , say  $z^3$  and  $z^4$ , the direction of the inequalities on  $D(z^3)$  and  $D(z^4)$  have to be ordered in the same direction as they are for  $D(z^1)$  and  $D(z^2)$ . It only requires that the direction of the inequalities are the *same across people*. Thus for any person,  $D(z)$  need not be monotonic in  $z$ .<sup>43</sup>

Under these conditions, Imbens and Angrist establish that for two distinct values of  $Z$ ,  $z^1$  and  $z^2$ , IV applied to (7) identifies

$$\begin{aligned} & \text{LATE}(z^2, z^1) \\ &= E(Y_1 - Y_0 | D(z^2) = 1, D(z^1) = 0), \end{aligned}$$

if the change from  $z^1$  to  $z^2$  induces people into the program ( $D(z^2) \geq D(z^1)$ ).<sup>44</sup> This is

<sup>41</sup> See Heckman and Vytlačil (2005, 2007a, 2007b), Heckman, Urzua, and Vytlačil (2006), and Carneiro, Heckman, and Vytlačil (2009, 2010).

<sup>42</sup> Heckman (1996).

<sup>43</sup> For this reason, Heckman, Urzua, and Vytlačil (2006) call this condition “uniformity.”

<sup>44</sup> This expression is easily modified to cover the opposite case where the change in  $Z$  reduces program participation. (IA-3) rules out both cases arising at the same time.

the mean return to participation in the program for people induced to switch treatment status by the change from  $z^1$  to  $z^2$ .<sup>45</sup>

LATE does not identify which people are induced to change their treatment status by the change in the instrument. It also leaves unanswered many of the policy questions discussed in section 2. For example, if a proposed program changes the same components of vector  $Z$  as used to identify LATE but at different values of  $Z$  (say  $z^4, z^3$ ),  $\text{LATE}(z^2, z^1)$  does not identify  $\text{LATE}(z^4, z^3)$ . If the policy operates on different components of  $Z$  than are used to identify LATE, one cannot safely use LATE to identify marginal returns to the policy. LATE answers a version of policy problem **P1** for objective outcomes, but ignores **P2** and **P3**. It does not, in general, identify treatment on the treated, ATE or the other parameters discussed in section 2.

### 3.2 Making Explicit the Implicit Economics of LATE

In a fundamental paper, Vytlacil (2002) shows that the LATE model is *equivalent* to a nonparametric version of the generalized Roy model. The Imbens–Angrist conditions imply the generalized Roy model, and the generalized Roy model implies the LATE model. Vytlacil's analysis is the basis for defining LATE abstractly within a well-posed economic model and separating the task of definition (Task 1 of table 1) from the task of identification (Task 2 of table 1). Vytlacil's analysis clarifies the implicit economic assumptions of LATE, what features of the generalized Roy model LATE estimates, and what policy questions LATE addresses. It also extends the range of policy questions that LATE can answer.

By Vytlacil's theorem, the Imbens–Angrist conditions imply (and are implied by) a con-

tinuous latent variable discrete choice model, which represents the individual's decision to enroll in the program being studied. Recall that  $I_D$  (in equation (4)) is the net benefit to the individual of enrolling in the program. A person takes treatment  $D = 1$  (e.g., goes to college) if  $I_D > 0$ ; otherwise  $D = 0$ . Vytlacil shows that the treatment choice equation underlying LATE can be expressed in terms of observed ( $Z$ ) and unobserved ( $V$ ) variables that can be represented by equation (6):  $I_D = \mu_D(Z) - V$  and  $D = 1$  if  $I_D > 0$ ;  $D = 0$  otherwise, where  $V$  is a continuous random variable with distribution function  $F_V$ .<sup>46</sup>  $\mu_D(Z)$  is defined in the discussion preceding equation (6).  $V$  may depend on  $U_0$  and  $U_1$  in a general way.<sup>47</sup>

LATE assumes that  $(U_0, U_1, V)$  are independent of  $Z$  given  $X$ . This relaxes the independence assumption (between  $X$  and the unobservables) that was frequently maintained in the early structural literature. The counterfactual choice indicator is generated by choice equation (6):  $D(z) = \mathbf{1}(\mu_D(z) > V)$ . This representation makes explicit the implicit random variable ( $V$ ) used to define  $D(z)$  in the analysis of Imbens and Angrist, and the independence between  $Z$  and  $V$  that is part of condition (IA–1).

The additive separability between  $\mu_D(Z)$  and  $V$  in the latent index model (6) plays an essential role in LATE. Model (6) is far from the most general possible representation of choices. If choice responses to variations in  $Z$  are heterogeneous in a general way, the same change in  $Z$  could lead some persons toward and other persons away from participation in the program, and the separability between  $\mu_D(Z)$  and  $V$  in (6) would break down.

<sup>45</sup> If “monotonicity” were not invoked, changes in  $Z$  from  $z^1$  to  $z^2$  could induce two way flows.

<sup>46</sup> Recall that I keep the  $X$  implicit, but it is implicitly conditioned on throughout this paper.

<sup>47</sup> In the original Roy model (1951),  $V = -(U_1 - U_0)$ .

Another way to say this is that monotonicity condition (IA–3) would be violated.<sup>48</sup>

To understand the economic model implicit in LATE, let  $P(z)$  denote the probability of taking treatment (e.g., attending college,  $D = 1$ ) conditional on  $Z = z$ :  $P(z) \equiv \Pr(D = 1 | Z = z)$ . From equation (6),  $P(z) = \Pr(\mu_D(z) > V) = F_V(\mu_D(z))$ .  $P(z)$  is a monotonic transformation of the mean utility function  $\mu_D(z)$  in discrete choice theory.  $P(z)$  is sometimes called the propensity score.

Define random variable  $U_D = F_V(V)$ , which is uniformly distributed over the interval  $[0, 1]$  and thus the  $p$ th quantile of  $U_D$  is  $p$ , i.e., the proportion of  $U_D$  that is  $p$  or lower. Different values of  $U_D$  correspond to different quantiles of  $V$ . We can rewrite (6) using  $F_V(\mu_D(Z)) = P(Z)$  so that

$$(9) \quad D = \mathbf{1}(P(Z) > U_D).$$

From the estimated propensity score, one can identify the ex ante net benefit  $I_D$  up to scale. Thus, one can determine for each value of  $Z = z$ , what proportion of people perceive that they will benefit from the program and the intensity of their benefit. Using the nonparametric identification analyses of Stephen R. Cosslett (1983), Manski (1988), Roger W. Klein and Richard H. Spady (1993), and Matzkin (1992, 1993, 1994, 2007), one can nonparametrically identify the distribution of  $V$  and the mean valuation  $\mu_D(Z)$  (up to scale).<sup>49</sup> Thus, from agent choices, one can supplement the information in LATE and ascertain ex ante subjective evaluations.

As a consequence of Vytlačil's theorem, the LATE assumptions imply the selection model representation (i.e., the generalized Roy model) and using the selection

model representation, one can establish that  $E(Y|Z = z) = E(Y|P(Z) = P(z))$ . Under the LATE assumptions,  $Z$  enters the model *only* through its effect on  $P(Z)$ . This property is called index sufficiency where  $P(Z)$  is the index. It is a central property of the LATE model.

As a consequence of Vytlačil's theorem, one can define LATE  $(z^2, z^1)$  using the latent variable  $U_D$  and the values taken by  $P(Z)$  when  $Z = z^1$  and  $Z = z^2$ . To do so, I use the property that the  $Z$  enter the model only through  $P(Z)$ .

$$(10) \quad \text{LATE}(z^2, z^1) \\ = E(Y_1 - Y_0 | P(z^1) \leq U_D \leq P(z^2)).$$

This is the mean gross return to persons whose  $U_D \in [P(z^1), P(z^2)]$ .<sup>50</sup>

The LATE parameter can be defined within the generalized Roy model, without reference to an instrument. Thus the LATE produced by economic theory can be expressed as

$$(11) \quad \text{LATE}(\bar{u}_D, \underline{u}_D) \\ = E(Y_1 - Y_0 | \underline{u}_D \leq U_D \leq \bar{u}_D),$$

the mean gross return to persons whose  $U_D \in [\underline{u}_D, \bar{u}_D]$ . This is a theoretical construct (task 1). Proceeding in this fashion, we separate task 1 of table 1 from task 2. A choice of two values of  $Z$  ( $z^1$  and  $z^2$ ) picks specific values of  $[\underline{u}_D, \bar{u}_D]$  that identify the model-generated LATE from data (say  $\Pr(D = 1 | Z = z^1) = p_1 = \underline{u}_D$  and  $\Pr(D = 1 | Z = z^2) = p_2 = \bar{u}_D$ ). This is task 2.

<sup>48</sup> See Heckman, Urzua, and Vytlačil (2006) for a discussion of this case. An example is  $\mu_D(Z) = \gamma Z$ , where  $\gamma$  varies among people so that the same change in  $Z$  can produce differences among people in choice responses to variations in  $Z$  in addition to the variation produced by  $V$ .

<sup>49</sup> The scale is the standard deviation of  $V$ ,  $\sigma_V$ .

<sup>50</sup> Because  $U_D$  is a continuous random variable, the distinction between strict and weak inequalities is irrelevant in defining the expressions in section 3 of this paper.



### 3.2.1 *The Surplus from Treatment and the Marginal Treatment Effect*

Using Vytlačil's theorem, it is possible to understand more deeply what economic questions LATE answers. Toward that end, it is useful to introduce the Marginal Treatment Effect (MTE) and show how it can be used to unify the literature on treatment effects and to make explicit the economic content of LATE.

For  $P(Z) = p$ , the mean gross gain of moving from "0" to "1" for people with  $U_D$  less than or equal to  $p$  is

$$\begin{aligned} (12) \quad E(Y_1 - Y_0 | P(Z) \geq U_D, P(Z) = p) \\ &= E(Y_1 - Y_0 | p \geq U_D) \\ &= E(Y_1 - Y_0 | \mu_D(z) \geq V).^{51} \end{aligned}$$

The first equality follows from the LATE assumption that  $(Y_0, Y_1)$  are independent of the instruments  $Z$  (IA-1) and hence any functions of  $Z$ . The second equality follows from the definition of the propensity score. The mean gross gain in the population (or gross surplus  $S(p)$ ) that arises from participation in the program for people whose  $U_D$  is at or below  $p$  is the product of the gain to people whose  $U_D$  is at or below  $p$  and the proportion of people whose  $U_D$  is at or below  $p$ :  $E(Y_1 - Y_0 | p \geq U_D)p = S(p)$ .

Using Vytlačil's theorem, we can move from the theory (task 1 of table 1) to the data (task 2 of table 1) to identify the gross surplus  $S(p)$ . The mean of  $Y$  given  $P(Z) = p$  depends on the gross surplus:

$$\begin{aligned} (13) \quad E(Y | P(Z) = p) \\ &= E(Y_0 + \mathbf{1}(p \geq U_D)(Y_1 - Y_0)) \\ &= E(Y_0) + \underbrace{E(Y_1 - Y_0 | p \geq U_D)p}_{S(p)}. \end{aligned}$$

<sup>51</sup> The mean net gain is  $E(Y_1 - Y_0 - C | p \geq U_D)$ .

We can identify the left-hand side of (13) for all values of  $p$  in the support of  $P(Z)$ .<sup>52</sup> This is task 2 in table 1. It is not necessary to impose functional forms to obtain this expression, and one can avoid one of the criticisms directed against 1980's structural econometrics. The surplus can be defined for all values of  $p \in [0, 1]$  whether or not the model is identified.

If  $p$  is increased by a small change in  $z$ , some people near the margin of indifference who chose not to participate in the program would now choose to participate. Small variations in  $p$  identify the mean marginal gross return to a policy expansion that changes  $P(Z)$ . Formally, the marginal increment in outcomes is

$$\begin{aligned} (14) \quad &\frac{\partial E(Y | P(Z) = p)}{\partial p} \\ &= E(Y_1 - Y_0 | U_D = p) \\ &= \frac{\partial S(p)}{\partial p}.^{53} \end{aligned}$$

This is the mean marginal gross return to treatment for persons indifferent between participation in the program or not at mean scale utility level  $p = U_D$ , and it is also the marginal change in the gross surplus. The sample analogue of (14) is the local instrumental variable (LIV) estimator of Heckman and Vytlačil (1999, 2005).<sup>54</sup> Adopting a nonparametric approach to estimating

<sup>52</sup> The support of a random variable is the region where it has positive density.

<sup>53</sup> For any two random variables  $M, N$  with density  $f(m, n)$  where  $m$  and  $n$  are realizations of  $M$  and  $N$ , where  $N$  is a uniform random variable in the interval  $[0, 1]$ ,  $E(M | r \geq N) \Pr(r \geq N) = \int_{-\infty}^{\infty} m \int_0^r f_{M,N}(m, n) dm dn$ , where  $\Pr(r > N) = \int_{-\infty}^{\infty} \int_0^r f_{M,N}(m, n) dm dn = \int_0^r f_N(n) dn = r$ , where  $f_N(n) = 1$ . Thus

$$\frac{\partial [E(M | r > N) \Pr(r > N)]}{\partial r} = E(M | N = r).$$

In the expression in the text,  $M = Y_1 - Y_0$  and  $N = U_D$ .

<sup>54</sup> Karim Chalak, Susanne Schennach, and Halbert White (2010) develop the sampling properties of this estimator under general conditions.

$E(Y|P(Z) = p)$  avoids extrapolation outside of the sample support of  $P(Z)$  and produces a data sensitive structural analysis.

A generalization of this parameter defined for other points of evaluation of  $u_D$  is the Marginal Treatment Effect (MTE):

$$\text{MTE}(u_D) \equiv E(Y_1 - Y_0 | U_D = u_D).$$

This parameter is very useful in understanding how to go from IV estimates to policy effects and in interpreting the economics of LATE.<sup>55</sup> Recall that  $U_D$  is a uniform random variable in the interval  $[0,1]$ , so that MTE for different  $u_D$  values shows how the mean gross returns to the program vary with different quantiles of the unobserved component of the utility of participation,  $U_D$ .

Expression (13) can be simplified to

$$(15) \quad E(Y|P(Z) = p) \\ = E(Y_0) + \underbrace{\int_0^p \text{MTE}(u_D) du_D}_{S(p)},$$

where  $S(p) = \int_0^p \text{MTE}(u_D) du_D$  and  $\partial S(p)/\partial p = \text{MTE}(p)$ . Figure 1 plots  $E(Y|P(Z) = p)$  (Figure 1(a)) and its derivative (Figure 1(b)) using values derived from a model discussed in Heckman and Vytlačil (2005). In this analysis,  $E(Y|P(Z) = p)$  increases at a diminishing rate in  $p$ , so  $\text{MTE}(u_D)$  is decreasing in  $u_D$ , i.e., there are diminishing returns to the marginal entrants attracted into the program by increasing  $P(z)$ .

Notice from (14) that persons with larger values of  $P(z)$  identify the return for those with larger values of  $U_D$ , i.e., values of  $U_D$  that make persons *less* likely to participate

in the program. This is so because marginal increases in  $P(z)$  at high levels of  $P(z)$  induce those individuals with high  $U_D$  values into treatment. This is a consequence of the economic choice model (9). Those with low values of  $U_D$  already participate in the program for low values of  $P(z) = p$ . A marginal increase in  $P(z)$  starting from a high value has no effect on the participation decision of those with low values of  $U_D$ . From LIV, it is possible to identify returns at all quantiles of  $U_D$  within the support of the distribution of  $P(Z)$  to determine which persons (identified by the quantile of the unobserved component of the desire to take the treatment,  $U_D$ ) are induced to go into the treatment ( $D = 1$ ) by a marginal change in  $P(z)$ , i.e., analysts can define the margins of choice traced out by variations in different instruments as they shift  $P(z)$ . This clarifies what empirical versions of LATE identify by showing that all instruments operate through  $P(z)$ , and variations around different levels of  $P(z)$  identify different stretches of the MTE. I now develop this point.

### 3.2.2. The Fundamental Role of the Choice Probability in Understanding What Instrumental Variables Estimate When $\beta$ Depends on $D$

For any two values of  $p$ , say  $p_1$  and  $p_2$ , generated by two different values of  $Z$ , where  $p_2 > p_1$ ,

$$\begin{aligned} S(p_2) - S(p_1) \\ &= E(Y_1 - Y_0 | p_1 \leq U_D \leq p_2) \\ &\quad \times \Pr(p_1 \leq U_D \leq p_2) \\ &= E(Y_1 - Y_0 | p_1 \leq U_D \leq p_2)(p_2 - p_1), \end{aligned}$$

where the last expression follows from the fact that  $\Pr(p_1 \leq U_D \leq p_2) = p_2 - p_1$ . Thus,

<sup>55</sup> MTE was introduced into the literature on policy evaluation by Björklund and Moffitt (1987) and extended in Heckman and Vytlačil (1999, 2001b, 2005, 2007b).

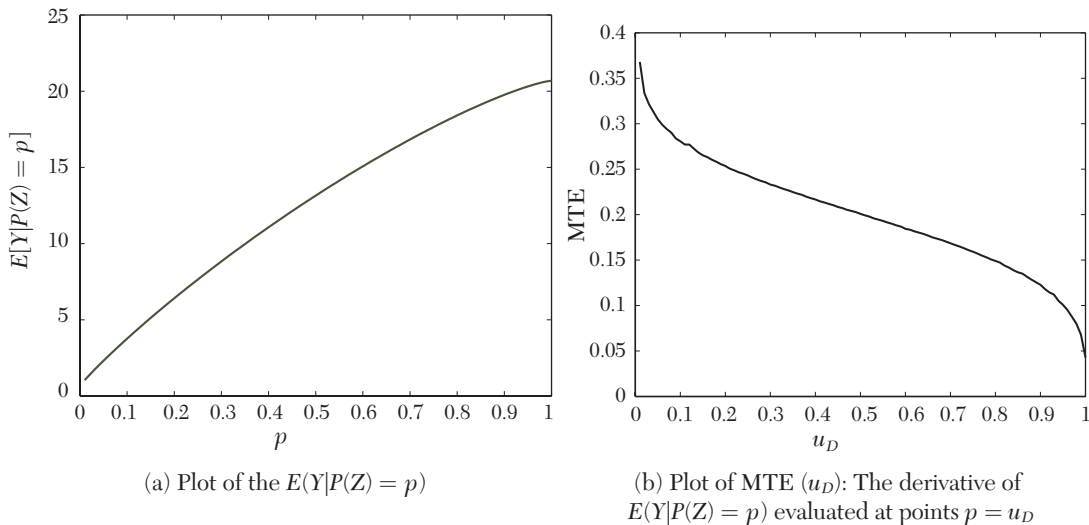


Figure 1. Plots of  $E(Y|P(Z) = p)$  and the MTE Derived from  $E(Y|P(Z) = p)$

Source: Heckman and Vytlacil (2005).

$$S(p_2) - S(p_1) = \int_{p_1}^{p_2} \text{MTE}(u_D) du_D.$$

This expression can be obtained directly from equation (15). From the definition of LATE,

$$(16) \quad \text{LATE}(p_2, p_1) = \frac{\int_{p_1}^{p_2} \text{MTE}(u_D) du_D}{p_2 - p_1} = \frac{S(p_2) - S(p_1)}{p_2 - p_1}.$$

Thus LATE is the chord of the gross surplus function over the interval  $[p_1, p_2]$ . The model-generated LATE approximates  $\text{MTE}(u_D)$  over an interval. By the mean value theorem,  $\text{LATE}(p_2, p_1) = \text{MTE}(u_D(p_2, p_1))$  where  $u_D(p_2, p_1)$  is a point of evaluation and  $u_D(p_2, p_1) \in [p_1, p_2]$ . The model-generated LATE can be identified if there are values of  $Z$ , say  $\tilde{z}$  and  $\tilde{\tilde{z}}$ , such that  $\Pr(D = 1 | Z = \tilde{z}) = p_1$  and  $\Pr(D = 1 | Z = \tilde{\tilde{z}}) = p_2$ . Under standard regularity conditions

$$\lim_{p_2 \rightarrow p_1} \text{LATE}(p_2, p_1) = \text{MTE}(p_1).$$

If we partition the support of  $u_D$  into  $M$  discrete and exhaustive intervals

$$[u_{D,0}, u_{D,1}), [u_{D,1}, u_{D,2}), \dots, [u_{D,M-1}, u_{D,M}],$$

where  $u_{D,0} = 0$  and  $u_{D,M} = 1$ , we can define

$$E(Y | U_D \leq u_{D,k}) = E(Y_0)$$

$$+ \sum_{j=1}^k \text{LATE}(u_{D,j}, u_{D,j-1}) \eta_j,$$

where  $\eta_j = u_{D,j} - u_{D,j-1}$ . Thus

$$(17) \quad E(Y) = E(Y_0) + \sum_{j=1}^M \text{LATE}(u_{D,j}, u_{D,j-1}) \eta_j,$$

which is the counterpart to expression (15) when  $p = 1$ . It shows how mean income can

be represented as a sum of incremental gross surpluses above  $E(Y_0)$ .

These expressions are derived from an underlying theoretical model. Whether or not the components can be identified from the data depends on the support of  $\Pr(D = 1 | Z)$ . If  $\Pr(D = 1 | Z = z)$  assumes values at only a discrete set of support points, say  $p_1 < p_2 < \dots < p_L$ , we can only identify LATE in intervals with boundaries defined by  $u_{D,\ell} = p_\ell, \ell = 1, \dots, L$ .

$\text{MTE}(u_D)$  and the model-generated LATE (10) are structural parameters in the sense that changes in  $Z$  (conditional on  $X$ ) do not affect  $\text{MTE}(u_D)$  or theoretical LATE. They are invariant with respect to all policy changes that operate through  $Z$ . Conditional on  $X$ , one can transport MTE and the derived theoretical LATEs across different policy environments and different data sets. These policy invariant parameters implement Marschak's Maxim since they are defined for combinations of the parameters of the generalized Roy model. Instead of separately estimating the components of the selection model presented in section 2.6, one can identify an interpretable marginal gross benefit function by using the derivative of  $E(Y | P(Z) = p)$ .

This deeper understanding of LATE facilitates its use in answering out-of-sample policy question **P3** for policies that operate through changing  $Z$ . Thus if one computes a LATE for any two pairs of values  $Z = z^1$ , and  $Z = z^2$ , with associated probabilities  $\Pr(D = 1 | Z = z^1) = P(z^1) = p_1$  and  $\Pr(D = 1 | Z = z^2) = P(z^2) = p_2$ , one can use it to evaluate any other pair of policies  $\tilde{z}$  and  $\tilde{z}$  such that

$$\begin{aligned} \Pr(D = 1 | Z = z^1) &= \Pr(D = 1 | Z = \tilde{z}) \\ &= p_1 \end{aligned}$$

and

$$\begin{aligned} \Pr(D = 1 | Z = z^2) &= \Pr(D = 1 | Z = \tilde{z}) \\ &= p_2. \end{aligned}$$

Thus, one can use an empirical LATE determined for one set of instrument configurations to identify outcomes for other sets of instrument configurations that produce the same  $p_1$  and  $p_2$ , i.e., one can compare any policy described by  $\tilde{z} \in \{z | P(z) = p_1\}$  with any policy  $\tilde{z} \in \{z | P(z) = p_2\}$  and not just the policies associated with  $z^1$  and  $z^2$  that identify the sample LATE. This is a powerful result and enables analysts to solve policy evaluation question **P3** to evaluate new policies never previously implemented if they can be cast in terms of variations in  $P(Z)$  over the empirical support on  $Z$ .<sup>56</sup>

Variation in different components of  $Z$  produce variation in  $P(Z)$ . Analysts can aggregate the variation in different components of  $Z$  into the induced variation in  $P(Z)$  to trace out  $\text{MTE}(u_D)$  over more of the support of  $u_D$  than would be possible using variation in any particular component of  $Z$ . The structural approach enables analysts to determine what stretches of the MTE different instruments identify and to determine the margin of  $U_D$  identified by the variation in an instrument.

Figure 2 reproduces the MTE displayed in figure 1(b) on a different scale. Consider values of instruments that are associated with  $P(z) = p_2$  and  $P(z) = p_1$ . They identify the MTE at a value of  $u_D$  in the interval  $u_D \in [p_1, p_2]$ , as depicted in the graph. This

<sup>56</sup> We only require  $Y$  observations for each value of  $P(Z) = p$  for each  $p$  in the target population, not values of  $Y$  for all  $Z$ . Assuming data on  $(Y, D, Z)$  triples, a completely nonparametric approach to identifying  $P(Z)$  would require that all LATEs required to answer **P3** would already be identified in the sample used to address **P1** and identify  $P(Z)$ , i.e., there is no distinction between **P1** and **P3**. However, one can imagine cases where the analyst has access to a richer set of data on  $(D, Z)$  where data on  $Y$  are not available. Using the index sufficiency property, analysts can determine  $E(Y | Z = z^*) = E(Y | P(Z) = P(z^*))$ , even if no  $Y$  is observed for a  $Z = z^*$ , so long as there is some value of  $Z = z^{**}$  in the sample such that  $P(z^*) = P(z^{**})$ . Moreover, if one adopts a parametric functional form for  $P(Z)$ , one can answer a much wider range of **P3** questions. The same is true if a nonparametric  $P(Z)$  is available from another sample. Structural invariance would justify combination of information across samples.

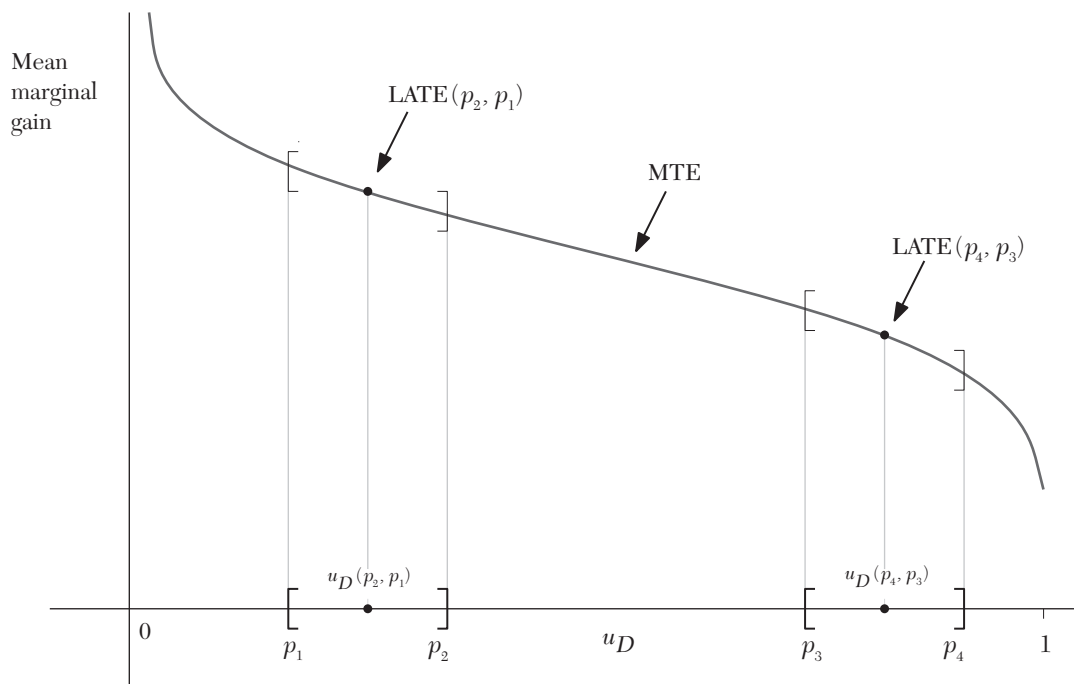


Figure 2. MTE as a Function of  $u_D$ : What Sections of the MTE Different Values of the Instruments and Different Instruments Approximate

is  $\text{LATE}(p_2, p_1)$ . If there is continuous variation in  $Z$ , it could be used to trace out the entire interval of MTE for  $P(Z) \in [p_1, p_2]$  using LIV. Independent of any instrument, we can define the LATE and MTE using the underlying economic model. In this fashion, we separate the task of definition of parameters from the task of identifying them.

Instruments associated with higher values of  $P(Z)$ ,  $[p_3, p_4]$ , identify the LATE in a different stretch of the MTE associated with higher values of  $u_D$ . Thus different instruments can identify different parameters. Continuous instruments identify entire stretches of the MTE while discrete instruments define the MTE at discrete points of the support (i.e., the LATE associated with the interval defined by the values assumed by  $P(Z)$ ). As a consequence

of Vytlačil's Theorem, one can identify the intervals of  $u_D$  implicit in using LATEs formed using different instruments.

If the MTE does not depend on  $u_D$ ,  $E(Y|P(Z) = p) = E(Y_0) + (\mu_1 - \mu_0)p$ , and all instruments identify the same parameter:  $\bar{\beta} = \mu_1 - \mu_0$ . In this case, MTE is a flat line parallel to the  $u_D$  axis. This is the case traditionally assumed in the analysis of instrumental variables.

A test of whether  $\text{MTE}(u_D)$  depends on  $u_D$ , or a test of nonlinearity of  $E(Y|P(Z) = p)$  in  $p$ , is a test of whether different instruments estimate the same parameter.<sup>57</sup> The LATE model and its extensions overturn the logic

<sup>57</sup> See Heckman, Schmierer, and Urzua (forthcoming) for a formal development of these tests and some Monte Carlo evidence on their performance.



of the J. Durbin (1954)–De-Min Wu (1973)–Jerry A. Hausman (1978) test for overidentification. Variability among the estimates from IV estimators based on different instruments may have nothing to do with the validity of any particular instrument, but may just depend on what stretch of the MTE they approximate.

### 3.3 All Treatment Effects Are Weighted Averages of the MTE

Using the economics implicit in LATE unifies and interprets the literature on treatment effects. All of the conventional treatment effects featured in the program evaluation literature can be written as weighted averages of the MTE or the structural LATEs where the weights can be estimated from the data over the sample support of  $P(Z)$ . Thus for treatment effect  $e$ ,

$$(18) \quad \text{Treatment Effect}(e) = \int_0^1 \text{MTE}(u_D) h^e(u_D) du_D,$$

where  $h^e(u_D)$  is a weighting function. For ATE,  $h^e(u_D) = 1$  and  $\text{ATE} = \int_0^1 \text{MTE}(u_D) du_D$ . Using the linearity of the integral, one can always break (18) into a sum of components over the intervals  $[u_{D,0}, u_{D,1}], [u_{D,1}, u_{D,2}], \dots, [u_{D,M-1}, u_{D,M}]$  to obtain

$$\begin{aligned} \text{Treatment Effect}(e) &= \sum_{j=1}^M \int_{u_{D,j-1}}^{u_{D,j}} \text{MTE}(u_D) h^e(u_D) du_D, \end{aligned}$$

where  $u_{D,j} > u_{D,j-1} > u_{D,j-2} > \dots$ . By the mean value theorem,<sup>58</sup> we may express each of the integrals in the sum as

<sup>58</sup> The mean value theorem for integrals states that, under standard regularity conditions (continuity and integrability),

$$\int_a^b G(t)\phi(t) dt = G(t^*) \int_a^b \phi(t) dt, \text{ for } t^* \in [a, b].$$

$$\begin{aligned} &\int_{u_{D,j-1}}^{u_{D,j}} \text{MTE}(u_D) h^e(u_D) du_D \\ &= h^e(u_{j,j-1}^*) \int_{u_{D,j-1}}^{u_{D,j}} \text{MTE}(u_D) du_D \end{aligned}$$

for some  $u_{j,j-1}^* \in [u_{D,j-1}, u_{D,j}]$ . From (16) we obtain

$$\begin{aligned} &h^e(u_{j,j-1}^*) \int_{u_{D,j-1}}^{u_{D,j}} \text{MTE}(u_D) du_D \\ &= h^e(u_{j,j-1}^*)(u_{D,j} - u_{D,j-1}) \text{LATE}(u_{D,j}, u_{D,j-1}). \end{aligned}$$

Thus we may write expression (18) as

$$\begin{aligned} \text{Treatment Effect}(e) &= \sum_{j=1}^M \text{LATE}(u_{D,j}, u_{D,j-1}) \tilde{h}^e(u_{j,j-1}^*), \end{aligned}$$

where  $\tilde{h}^e(u_{j,j-1}^*) = h^e(u_{j,j-1}^*)(u_{D,j} - u_{D,j-1})$ . In the special case of ATE,  $h^e(u_{j,j-1}^*) = 1$  and

$$\text{ATE} = \sum_{j=1}^M \text{LATE}(u_{D,j}, u_{D,j-1}) \eta_j,$$

where  $\eta_j = u_{D,j} - u_{D,j-1}$ .

The Policy Relevant Treatment Effect (PRTE), defined in section 2, is

$$E(Y | \text{Alternative Policy } a)$$

$$- E(Y | \text{Baseline Policy } b)$$

$$= E(Y^a) - E(Y^b)$$

$$= \int_0^1 \text{MTE}(u_D) h^{\text{PRTE}}(u_D) du_D,$$

where  $h^{\text{PRTE}}(u_D) = F_P^b(u_D) - F_P^a(u_D)$ , and  $F_P^b$  is the distribution of  $P(Z)$  under policy  $b$ , and  $F_P^a$  is the distribution of  $P(Z)$  under policy

$a$ .<sup>59</sup> Using the mean value theorem, one can generate a counterpart expression in terms of LATEs.

The PRTE weights  $MTE(u_D)$  by the change in the distribution of the probabilities of participation at different values of  $u_D$ .<sup>60</sup> Thus for a typical MTE as graphed in figure 1(a), if a policy shifts the distribution of participants toward low  $u_D$  values, it generates a positive PRTE, since  $MTE(u_D)$  is higher for low  $u_D$  values than it is for high  $u_D$  values.

Notice that the same  $MTE(u_D)$  can be used to evaluate the impacts of a variety of different policies.  $MTE(u_D)$  is a structural function since it is invariant across policies that affect the distribution of the  $P(Z)$  but not the distribution of the potential ex post outcomes. We can evaluate the effect of new policies never previously experienced if we can characterize the distributions of  $P(Z)$  for those policies.

Table 3 displays the weights of the MTE that produce the traditional treatment parameters. All of the weights can be estimated from the distribution  $P(Z)$ .<sup>61</sup> There are corresponding expressions for the case of discrete support for  $P(Z)$  that can be obtained using the mean value theorem. The weights integrate to 1. When  $\beta$  is independent of  $D$ ,  $MTE(u_D)$  does not depend on  $u_D$  ( $= \mu_1 - \mu_0 = \bar{\beta}$ ), so all treatment parameters equal  $\bar{\beta}$ .

Figure 3 plots an MTE taken from the analysis of Heckman and Vytlačil (2005) and the weights for MTE associated with ATE,

<sup>59</sup> Assuming policy invariance as defined in subsection 2.3,

$$\begin{aligned} E(Y) &= \int_0^1 E(Y|P(Z) = t) f_P(t) dt \\ &= \int_0^1 \left[ \int_0^1 \mathbf{1}_{[0,t]}(u_D) E(Y_1|U_D = u_D) du_D \right. \\ &\quad \left. + \int_0^1 \mathbf{1}_{(t,1)}(u_D) E(Y_0|U_D = u_D) du_D \right] f_P(t) dt, \end{aligned}$$

where

$$\mathbf{1}_{[0,t]}(u_D) = \begin{cases} 1 & \text{if } 0 \leq u_D \leq t \\ 0 & \text{otherwise.} \end{cases}$$

Changing the order of the integration,

TT, and TUT for a case where  $\beta$  is not independent of  $D$ . ATE weights  $u_D$  evenly. TT oversamples low values of  $u_D$  (associated with persons more likely to participate in the program). TUT oversamples high  $u_D$ . In this example, because  $MTE(u_D)$  is decreasing in  $u_D$ ,  $TT > ATE > TUT$ .

### 3.4 What Does Conventional IV Estimate?

In most empirical studies, more than the two values of  $Z$  are used to construct IV estimates. For this case, Imbens and Angrist (1994) use weights developed by Shlomo Yitzhaki (1989) to express IV as a weighted average of component LATEs defined for different values of the instruments. The weights used by Imbens–Angrist are positive for each component LATE for the special instrument they consider ( $P(Z)$  or some monotonic function of  $P(Z)$ ). Since  $MTE(u_D)$  may change sign over the interval  $u_D \in [0, 1]$ , the IV may be negative even if some portion of the  $MTE(u_D)$  is positive. For general instruments that are not monotonic functions of  $P(Z)$ , the output of IV is even more ambiguous. The IV weights can be negative over regions of  $u_D \in [0, 1]$ . Thus an IV based on general instruments may have a sign opposite to the true causal effect as defined by the MTE. Even if each component of LATE is positive, in the general case

---


$$\begin{aligned} E(Y) &= \int_0^1 \left[ \int_0^1 [\mathbf{1}_{[u_D,1]}(t) E(Y_1|U_D = u_D) \right. \\ &\quad \left. + \mathbf{1}_{[0,u_D]}(t) E(Y_0|U_D = u_D)] f_P(t) dt \right] du_D \\ &= \int_0^1 \left[ (1 - F_P(u_D)) E(Y_1|U_D = u_D) \right. \\ &\quad \left. + F_P(u_D) E(Y_0|U_D = u_D) \right] du_D. \end{aligned}$$

Comparing policy  $a$  to policy  $b$ ,  $E(Y^a|X) - E(Y^b|X) = \int_0^1 E((Y_1 - Y_0)|X, U_D = u_D) (F_P^a(u_D) - F_P^b(u_D)) du_D$ . An alternative proof is given in Appendix A.

<sup>60</sup> The PRTE can be interpreted as an economically more explicit version of Stock's (1989) nonparametric policy analysis parameter for a class of policy interventions with explicit agent preferences where the policies evaluated operate solely on agent choice sets.

<sup>61</sup> I discuss the relationship between MTE and IV in the next subsection. For the general case, knowledge of the joint distribution of  $Z, P(Z)$  is required.

TABLE 3  
MTE WEIGHTS FOR DIFFERENT TREATMENT PARAMETER AND IVs.  
( $F_p$  is the distribution of  $P$ ,  $f_p$  is its density.)

Treatment parameter weights	$h^{ATE}(u_D) = 1$
	$h^{TT}(u_D) = \left[ \int_{u_D}^1 f_p(p) dp \right] \frac{1}{E(P)}$
	$h^{TUT}(u_D) = \left[ \int_0^{u_D} f_p(p) dp \right] \frac{1}{E(1-P)}$
	$h^{PRTE}(u_D) = F_p^b(u_D) - F_p^a(u_D)$
IV weights	$h^{IV}(u_D) = \left[ \int_{u_D}^1 (p - E(P)) f_p(p) dp \right] \frac{1}{\text{Var}(P)}$ for $P(Z)$ as an instrument
	$h^{IV}(u_D) = \frac{\int_{u_D}^1 \int_{-\infty}^{\infty} (j - E(J)) f_{J,P}(j, p) dj dp}{\text{Cov}(J, P)}$ for a general instrument $J(Z)^*$ , a function of $Z$

\*  $f_{j,p}(j, p)$  is the joint density of  $J$  and  $P$ . For derivations of these weights, see Heckman and Vytlačil (1999, 2005, 2007b).

IV can be negative. Negative components of  $\text{MTE}(u_D)$  weighted by negative weights can generate a positive IV.

This analysis is constructive because the weights can be identified from the data. Analysts can ascertain whether or not the weights are negative and over what regions of  $u_D$ . In this subsection, I analyze the case where  $P(Z)$  is the instrument and the weights are positive. I analyze the general case in section 3.6 below.

To understand what IV identifies, consider a linear regression approximation of  $E(Y|P(Z) = p)$ :

$$E^*(Y|P(Z) = p) = a + bp,^{62}$$

<sup>62</sup>  $E^*(M|N)$  denotes linear projection, i.e., the linear regression of  $M$  on  $N$ .

where

$$\begin{aligned} b &= \frac{\text{Cov}(Y, P(Z))}{\text{Var}(P(Z))} \\ &= \frac{\text{Cov}(E(Y|P(Z)), P(Z))}{\text{Var}(P(Z))}. \end{aligned}$$

$b$  is the same as the IV estimate of “the effect” of  $D$  on  $Y$  using  $P(Z)$  as an instrument since  $\text{Cov}(P(Z), D) = \text{Var}(P(Z))$ .<sup>63</sup>

Using condition (IA–1) (in particular that  $P(Z)$  is independent of  $Y_0$ ), and expression (15), we obtain

<sup>63</sup>  $D(Z) = P(Z) + \tau$  where  $E(\tau|P(Z)) = 0$ , thus  $\text{Cov}(Y, P(Z))/\text{Cov}(D, P(Z)) = \text{Cov}(Y, P(Z))/\text{Var}(P(Z))$ .

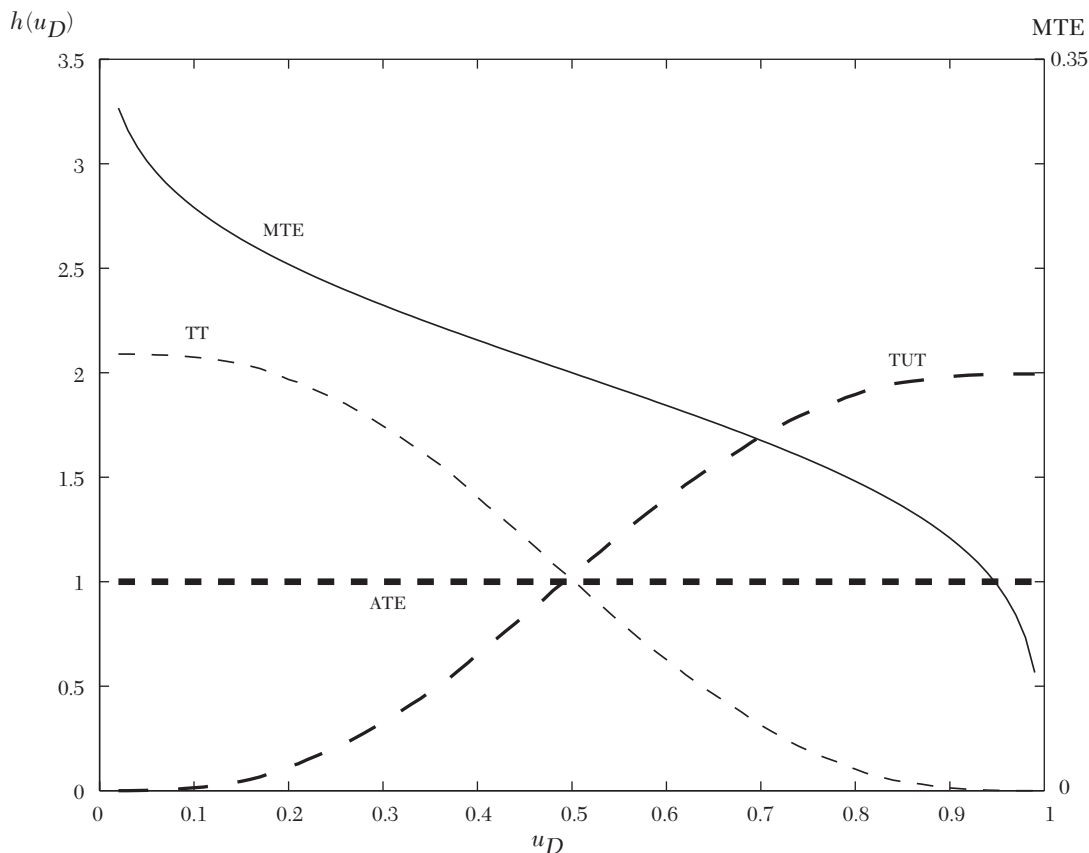


Figure 3. MTE and the Weights for the Marginal Treatment Effect for Different Parameters for the Model Graphed in Figure 1

Source: Heckman and Vytlacil (2005).

$$\begin{aligned}
 (19) \quad b &= \frac{\text{Cov}(Y, P(Z))}{\text{Var}(P(Z))} \\
 &= \frac{\text{Cov}(S(P(Z)), P(Z))}{\text{Var}(P(Z))} \\
 &= \frac{\text{Cov}\left(\int_0^{P(Z)} \text{MTE}(u_D) du_D, P(Z)\right)}{\text{Var}(P(Z))}.
 \end{aligned}$$

Note that when  $\text{MTE}(u_D)$  is constant in  $u_D$  ( $\text{MTE}(u_D) = \mu_1 - \mu_0 = \bar{\beta}$ ) so that  $\bar{\beta}$  is

independent of  $D$ , the numerator of the preceding expression simplifies to

$$\begin{aligned}
 &\text{Cov}\left(\int_0^{P(Z)} \text{MTE}(u_D) du_D, P(Z)\right) \\
 &= \text{Cov}(\bar{\beta}P(Z), P(Z)) = \bar{\beta}\text{Var}(P(Z))
 \end{aligned}$$

so  $b = \mu_1 - \mu_0 = \bar{\beta}$ . This is the traditional result for IV. In this case, the marginal

surplus is the same as the average surplus for all values of  $p$ . Expression (19) arises because  $D$  depends on  $\beta (= Y_1 - Y_0)$ , something assumed away in traditional applications of IV. As a consequence, in general, the marginal surplus is not the average surplus.

An explicit expression for the numerator of (19) is

$$\begin{aligned} & \text{Cov}(Y, P(Z)) \\ &= \int_0^1 \left[ \int_0^p \text{MTE}(u_D) du_D \right] (p - E(P)) f_P(p) dp. \end{aligned}$$

Reversing the order of the integration of the terms on the right-hand side and respecting the requirement that  $0 < u_D < p < 1$ , we obtain

$$\begin{aligned} b &= \frac{\text{Cov}(Y, P(Z))}{\text{Var}(P(Z))} \\ &= \frac{\int_0^1 \text{MTE}(u_D) \left[ \int_{u_D}^1 (p - E(P)) f_P(p) dp \right] du_D}{\text{Var}(P(Z))} \\ &= \int_0^1 \text{MTE}(u_D) h_{P(Z)}^{\text{IV}}(u_D) du_D, \end{aligned}$$

where

$$h_{P(Z)}^{\text{IV}}(u_D) = \frac{\int_{u_D}^1 (p - E(P)) f_P(p) dp}{\text{Var}(P(Z))}.^{64, 65}$$

<sup>64</sup> This result is due to Yitzhaki (1989) and is elaborated in Heckman, Urzua, and Vytlacil (2006) and Heckman and Vytlacil (2007a). The Yitzhaki paper is posted at the Web site for Heckman, Urzua, and Vytlacil (2006).

<sup>65</sup> Under the conditions of Fubini's theorem, it is valid to reverse the order of the integration. See the discussion in the preliminary remarks of Appendix A.

An alternative expression for the weight is as the mean of left truncated  $P(Z)$ :

$$\begin{aligned} & h_{P(Z)}^{\text{IV}}(u_D) \\ &= \frac{E(P(Z) | P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Var}(P(Z))}, \end{aligned}$$

which shows that the weight on the MTE ( $u_D$ ) is non-negative for all  $u_D$ .

The weights can be estimated from the sample distribution of  $P(Z)$ . The weights for  $P(Z)$  as an instrument have a distinctive profile. It is readily verified that they are non-negative, reach a peak at the mean of the distribution of  $P(Z)$ , and are zero at the extremes  $u_D = 0$  and  $u_D = 1$ . The weights integrate to 1.<sup>66</sup> Figure 4 plots the IV weights and the MTE from a study by Heckman and Vytlacil (2005). Comparing the IV weights with the weights for different treatment effects enables analysts to determine how closely IV approximates any particular mean treatment effect.

For discrete valued instruments mapped into  $P(\tilde{z}^1) = p_1 < P(\tilde{z}^2) = p_2 < \dots < P(\tilde{z}^L) = p_L$ ,

$$\text{IV} = \sum_{\ell=1}^{L-1} \text{LATE}(p_{\ell+1}, p_{\ell}) \lambda_{\ell},$$

where  $\lambda_{\ell} = (1/\text{Var}(P(Z))) \sum_{t>\ell}^L (p_t - E(P)) \times f_P(p_t)$  and  $f_P(p_t)$  is the probability that  $P(Z) = p_t$ . For a proof, see Heckman, Urzua, and Vytlacil (2006) or Appendix B.

### 3.5 The Problem of Limited Support

Before turning to the analysis of general instruments, I consider the problem of limited support for  $P(Z)$  for the special instrument  $P(Z)$  used by Imbens and Angrist.

<sup>66</sup>  $\int_0^1 \left[ \int_{u_D}^1 (p - E(P)) f_P(p) dp \right] du_D = \text{Var}(P(Z))$ .



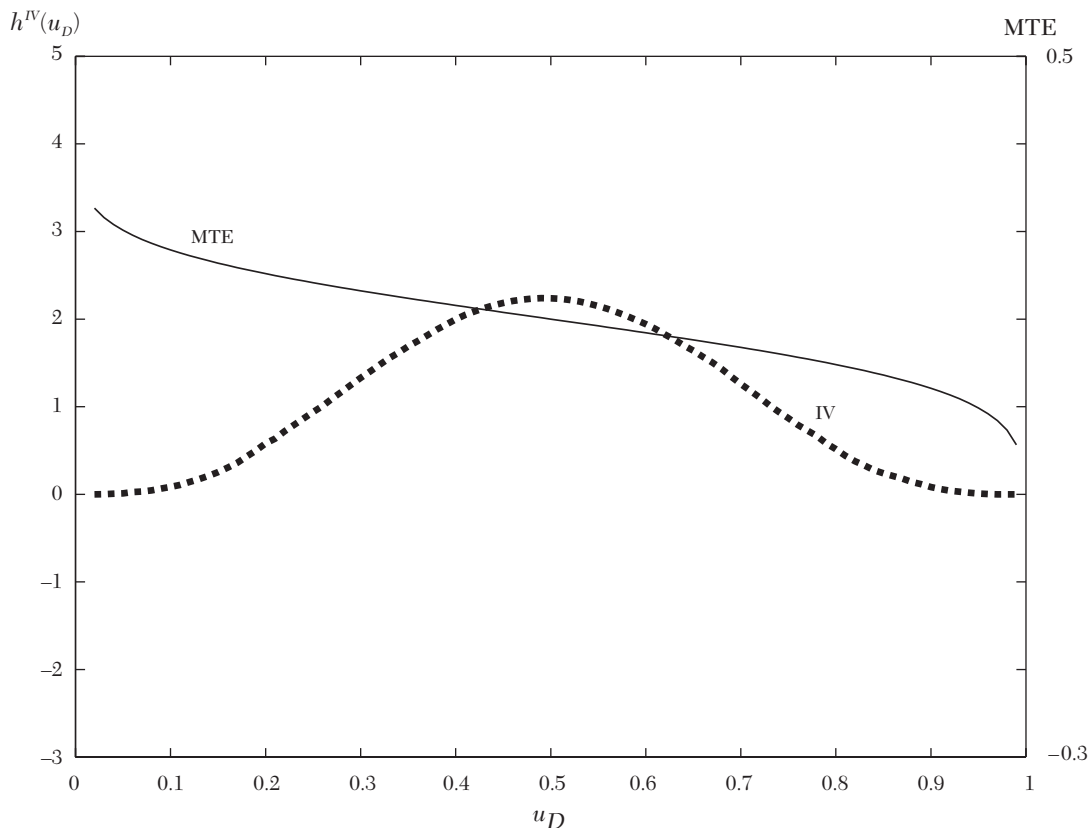


Figure 4. The MTE and IV Weights as a Function of  $u_D$

Source: Heckman and Vytlačil (2005).

Analysis of this simple case establishes principles that apply to more general instruments. While the various treatment parameters can be defined from the generalized Roy model, they may not necessarily be identified from the data. Both the nonparametric structural approach and the nonparametric program evaluation approach avoid the problem of extrapolating estimates outside the support of the data.<sup>67</sup> The early literature

in structural econometrics imposed strong functional forms (typically linearity) to obtain estimates.<sup>68</sup>  $P(Z)$  may not be identified over the full unit interval. Thus the lowest sample value of  $P(Z)$  may exceed zero and the largest value of  $P(Z)$  may be less than 1.

In addition,  $P(Z)$  may only assume discrete values. This limits the identifiability of MTE. In this case, only LATE over intervals of  $u_D \in [0, 1]$  can be identified from the values

<sup>67</sup> Angrist and Pischke (2008, 2010) are exceptions. They advocate use of linear equations in estimating treatment effects.

<sup>68</sup> As did the program evaluation literature. See Burt S. Barnow, Glen G. Cain, and Arthur S. Goldberger (1980).

of  $P(Z) = P(z)$  associated with the discrete instruments.<sup>69</sup>

One approach to this problem developed by Manski (1990, 1995, 2003) is to produce bounds on the treatment effects. Heckman and Vytlačil (1999, 2000, 2001a, 2001b, 2007b) develop specific bounds for the generalized Roy model that underlies the LATE model. The bounds developed in the literature are for conventional treatment effects and not for policy effects.

Carneiro, Heckman, and Vytlačil (2010) consider an alternative approach based on marginal policy changes. Many proposed policy changes are incremental in nature, and a marginal version of the PRTE is all that is required to answer questions of economic interest. When some instruments are continuous, it is possible under the conditions in their paper to identify a marginal version of PRTE (MPRTE). MPRTE is in the form of representation (18) where the weights can be identified from the data and the support requirements are more limited than the conditions required to identify PRTE for large changes in policies. Their paper presents a derivation of the weights for classes of policy expansions.<sup>70</sup> Application of these data sensitive nonparametric approaches enables analysts to avoid one source of instability of the estimates of policy effects that plagued 1980s econometrics.

### 3.6 More General Instruments

Typically, economists use a variety of instruments one at a time and not just  $P(Z)$ , or some function of  $P(Z)$ , as an instrument and compare the resulting estimates (see, e.g. Card, 1999, 2001). When there is

selection on the basis of gross gains ( $\beta \not\ll D$ ) so that the marginal gross surplus is not the same as the average gross surplus, different instruments identify different parameters. IV is a weighted average of MTEs where the weights integrate to 1 and can be estimated from sample data. However, in the case of general instruments, the weights can be negative over stretches of  $u_D$ .

Consider using the first component of  $Z$ ,  $Z_1$ , as an instrument for  $D$  in equation (7). Suppose that  $Z$  contains two or more elements ( $Z = (Z_1, \dots, Z_K)$ ,  $K \geq 2$ ). The economics implicit in LATE informs us that  $Z$  determines the distribution of  $Y$  through  $P(Z)$ . Any correlation between  $Y$  and  $Z_1$  arises from the statistical dependence between  $Z_1$  and  $P(Z)$  operating to determine  $Y$ .

The IV estimator based on  $Z_1$  is

$$\begin{aligned} IV_{Z_1} &= \frac{\text{Cov}(Y, Z_1)}{\text{Cov}(D, Z_1)} \\ &= \frac{\text{Cov}(E(Y|Z_1), Z_1)}{\text{Cov}(D, Z_1)}. \end{aligned}$$

Note, however, that choices (and hence  $Y$ ) are generated by the *full vector* of  $Z$  operating through  $P(Z)$ . The analyst may only use  $Z_1$  as an instrument but the underlying economic model informs us that the full vector of  $Z$  determines observed  $Y$ . Conditioning only on  $Z_1$  leaves uncontrolled the influence of the other elements of  $Z$  on  $Y$ . This is a new phenomenon in IV that would not be present if  $D$  did not depend on  $\beta (= Y_1 - Y_0)$ . An IV based on  $Z_1$  identifies an effect of  $Z_1$  on  $Y$  as it operates directly through  $Z_1$  ( $Z_1$  changing  $P(Z_1, \dots, Z_K)$ ) holding other elements in  $Z$  constant and indirectly through the effect of  $Z_1$  as it covaries with  $(Z_2, \dots, Z_K)$ , and how those variables affect  $Y$  through their effect on  $P(Z)$ .

A linear regression analogy helps to fix ideas. Suppose that outcome  $Q$  can be expressed

<sup>69</sup> Bounds for MTE and LATE in the case of limited support are presented in Heckman and Vytlačil (1999, 2001a, 2001b, 2007b). Manski (1990, 1995, 2003) presents bounds for a wide array of models.

<sup>70</sup> Hidehiko Ichimura and Taber (2002) develop an alternative local approach that does not exploit the structure of the generalized Roy model that underlies LATE.

as a linear function of  $W = (W_1, \dots, W_L)$ , an  $L$ -dimensional regressor:

$$Q = \sum_{\ell=1}^L \phi_{\ell} W_{\ell} + \varepsilon,$$

where  $E(\varepsilon | W) = 0$ . If we regress  $Q$  only on  $W_1$ , we obtain in the limit the standard omitted variable result that the estimated “effect” of  $W_1$  on  $Q$  is

$$(20) \quad \frac{\text{Cov}(Q, W_1)}{\text{Var}(W_1)} = \phi_1 + \sum_{\ell=2}^L \phi_{\ell} \frac{\text{Cov}(W_{\ell}, W_1)}{\text{Var}(W_1)}$$

where  $\phi_1$  is the ceteris paribus direct effect of  $W_1$  on  $Q$  and the summation captures the rest of the effect (the effect on  $Q$  of  $W_1$  operating through covariation between  $W_1$  and the other values  $W_{\ell}$ ,  $\ell \neq 1$ ). An analogous problem arises in using one instrument at a time to identify “the effect” of  $Z_1$ .

Thus if the analyst does not condition on the other elements of  $Z$  in using  $Z_1$  as an instrument, the margin identified by variations of  $Z_1$  does *not* in general correspond to variations arising solely from variations in  $Z_1$ , holding the other instruments constant. The margin of choice implicitly defined by the variation in  $Z_1$  is difficult to interpret and depends on the parameters of the generalized Roy model generating outcomes as well as on the sample dependence between instrument  $Z_1$  and  $P(Z)$ . Thus an IV based on  $Z_1$  mixes causal effects with sample dependence effects among the correlated regressors.<sup>71</sup> In a study of college going, if  $Z_1$  and  $Z_2$  are tuition and distance to college,

respectively, the instrument  $Z_1$  identifies the direct effect of variation in tuition on college attendance and the effect of distance to college on college attendance as distance covaries with tuition in the sample used by the analyst. This is not the ceteris paribus effect of a variation in tuition. It does not correspond to the answer needed to predict the effects of a policy that operates solely through an effect on tuition. In models in which  $D$  depends on  $\beta$ , the traditional instrumental variable argument that analysts do not need a model for  $D$  and can ignore other possible determinants of  $D$  other than the instrument being used, breaks down. To interpret which margin is identified by different instruments requires that the analyst specify and account for all of the  $Z$  that form  $P(Z)$ . Since different economists may disagree on the contents of  $Z$ , different economists using  $Z_1$  on the same data will obtain the same point estimate but may disagree about the interpretation of the margin identified by variation in  $Z_1$ .

To establish these points, note that as a consequence of Vytlačil's theorem,  $Z$  enters the distribution of  $Y$  only through  $P(Z)$ . Thus the conditional distribution of  $Y$  given  $Z_1 = z_1$  operates through the effect of  $Z_1$  as it affects  $P(Z)$ . That is a key insight from Vytlačil's theorem. Thus

$$E(Y | Z_1 = z_1)$$

$$= \int_0^1 E(Y | P(Z) = p) g_{P(Z)|Z_1}(p, z_1) dp$$

where  $g_{P(Z)|Z_1}(p, z_1)$  is the conditional density of  $P(Z)$  given  $Z_1 = z_1$ .<sup>72</sup> Putting all of these ingredients together, and using (15), we obtain

<sup>71</sup> Relationships that combine sample and structural relationships were called “mongrel” relationships by the early structural econometricians (see Lawrence R. Klein 1953).

<sup>72</sup>  $g_{P(Z)|Z_1}(p, z_1) = g_{P(Z), Z_1}(p, z_1) / g_{Z_1}(z_1)$ .

$$\begin{aligned}
E(Y|Z_1 = z_1) &= E(Y_0) + \int_0^1 S(p) g_{P(Z)|Z_1}(p, z_1) dp \\
&= E(Y_0) + \int_0^1 \underbrace{\left[ \int_0^p \text{MTE}(u_D) du_D \right]}_{S(p)} \\
&\quad \times g_{P(Z)|Z_1}(p, z_1) dp.
\end{aligned}$$

Using this expression to compute  $\text{Cov}(Y, Z_1)/\text{Cov}(D, Z_1)$ , we obtain

$$\begin{aligned}
IV_{Z_1} &= \frac{\int_{-\infty}^{\infty} (z_1 - E(Z_1)) \int_0^1 S(p) g_{P(Z)|Z_1}(p, z_1) dp dz_1}{\text{Cov}(Z_1, D)} \\
&= \frac{\int_{-\infty}^{\infty} (z_1 - E(Z_1)) \int_0^1 \left[ \int_0^p \text{MTE}(u_D) du_D \right] g_{P(Z)|Z_1}(p, z_1) dp dz_1}{\text{Cov}(Z_1, D)}
\end{aligned}$$

This expression integrates the argument in the numerator with respect to  $u_D$ ,  $p$ , and  $z_1$  in that order. Reversing the order of integration to integrate with respect to  $p$ ,  $z_1$ , and  $u_D$  in that order, we obtain

$$IV_{Z_1} = \int_0^1 \text{MTE}(u_D) h_{Z_1}^{IV}(u_D) du_D,$$

where

$$\begin{aligned}
h_{Z_1}^{IV}(u_D) &= \frac{\int_{-\infty}^{\infty} (z_1 - E(Z_1)) \int_{u_D}^1 g_{P(Z)|Z_1}(p, z_1) dp dz_1}{\text{Cov}(Z_1, D)}.
\end{aligned}$$

The weight integrates to 1 but can be negative over stretches of  $u_D$ .<sup>73</sup> At the extremes ( $u_D = 0, 1$ ), the weights are zero.

An illuminating way to represent this weight is

$$\begin{aligned}
h_{Z_1}^{IV}(u_D) &= \frac{E(Z_1 - E(Z_1)) \mathbf{1}(P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Cov}(Z_1, D)}.
\end{aligned}$$

As  $u_D$  is increased, the censored (by the condition  $P(Z) > u_D$ ) mean of  $(Z_1 - E(Z_1))$  may switch sign, and hence the weights may be negative over certain ranges. Thus the IV estimator may have a sign opposite to the true causal effect (defined by the MTE).

Figure 6 illustrates this possibility for the distribution of the data  $Z = (Z_1, Z_2)$  shown in figure 5, where  $Z$  is continuously distributed. The support of the data only permits identification of  $P(Z)$  over the interval  $[0.1, 0.9]$ . Thus none of the conventional treatment parameters is identified. From LIV, we can identify the MTE over the interval  $[0.1, 0.9]$ . We can also identify the weights over this interval. For values of  $u_D > 0.65$ , the weights are negative in this example. Thus it is possible that the IV based on  $Z_1$  can be negative even if the MTE is everywhere positive. Table 4, taken from Heckman, Urzua, and Vytlačil (2006), shows how three different distributions of

$$\begin{aligned}
&^{73} \text{Cov}(Z_1, Y) \\
&= E[(Z_1 - E(Z_1)) \mathbf{1}(U_D < P(Z))(Y_1 - Y_0)] \\
&= E[(Z_1 - E(Z_1)) \mathbf{1}(U_D < P(Z))E(Y_1 - Y_0 | Z, U_D)] \\
&= E[(Z_1 - E(Z_1)) \mathbf{1}(U_D < P(Z))E(Y_1 - Y_0 | U_D)] \\
&= E_{U_D}[E_Z[(Z_1 - E(Z_1)) \mathbf{1}(U_D < P(Z)) | U_D]E(Y_1 - Y_0 | U_D)] \\
&= \int_0^1 \{E(Z_1 - E(Z_1)) \mathbf{1}(P(Z) \geq u_D) \Pr(P(Z) \geq u_D) \\
&\quad \times E(Y_1 - Y_0 | U_D = u_D)\} du_D \\
&= \int_0^1 \text{MTE}(u_D) E(Z_1 - E(Z_1)) \mathbf{1}(P(Z) \geq u_D) \Pr(P(Z) \geq u_D) du_D.
\end{aligned}$$

The expression for  $\text{Cov}(Z_1, D)$  is  $\text{Cov}(Z_1, D) = E[(Z_1 - E(Z_1)) \mathbf{1}(U_D < P(Z))]$ . The expression for the weight for MTE can thus be written as  $h_{Z_1}^{IV}(u_D) = \text{Cov}(Z_1, \Pr(P(Z) > u_D)) / \text{Cov}(Z_1, \Pr(P(Z) > u_D))$  so it is easy to see that the weights integrate to 1.

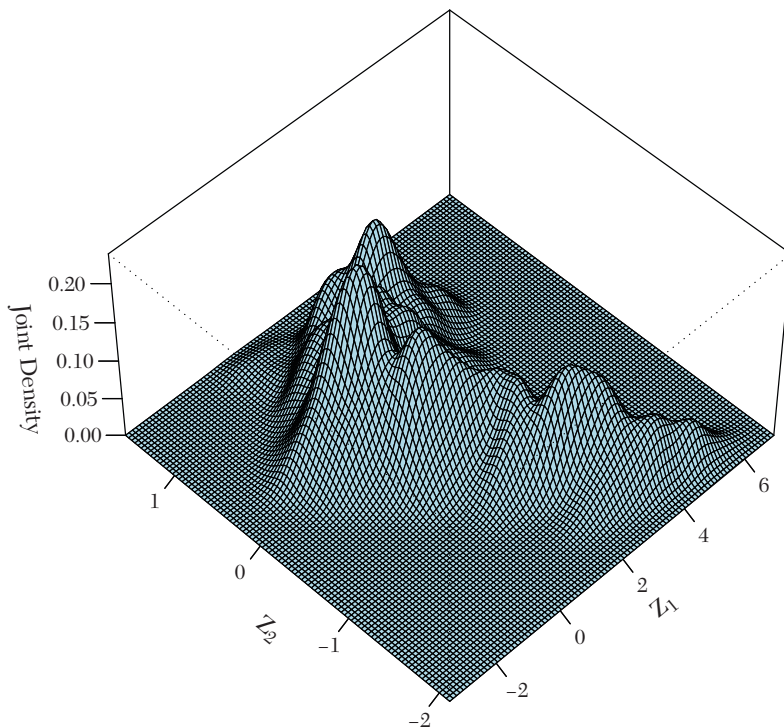


Figure 5. Joint Density of Instruments  $Z = (Z_1, Z_2)$

$Z$  for the same underlying policy-invariant model with the same ATE can produce very different IV estimates.

This analysis elucidates the benefits and limitations of the method of randomized controlled experiments. Experiments that manipulate  $Z_1$  independently of other components of  $Z$  isolate the effects of  $Z_1$  on outcomes in comparison with the effects obtained by sample variation in  $Z_1$  correlated with other components of  $Z$ . Neither set of variations may identify the returns to any given policy unless the experimentally induced variation corresponds exactly to the variation induced by the policy. Economists can use experimental variation to identify the MTE. The features of a proposed

policy are described by its effects on the PRTE weights as it affects the distribution of  $P(Z)$ . Proceeding in this way, one can use experiments to address a range of questions beyond the effects directly identified by the experiment.

Using the implicit economic theory underlying LATE, economists can do better than just report an IV estimate. We can be data sensitive but not at the mercy of the data. We can determine the MTE (or LATEs) over the identified regions of  $u_D$  in the empirical support of  $P(Z)$ . We can also determine the weights over the empirical support of  $P(Z)$  to determine whether they are negative or positive. We can bound estimates of the unidentified parameters. (See Heckman

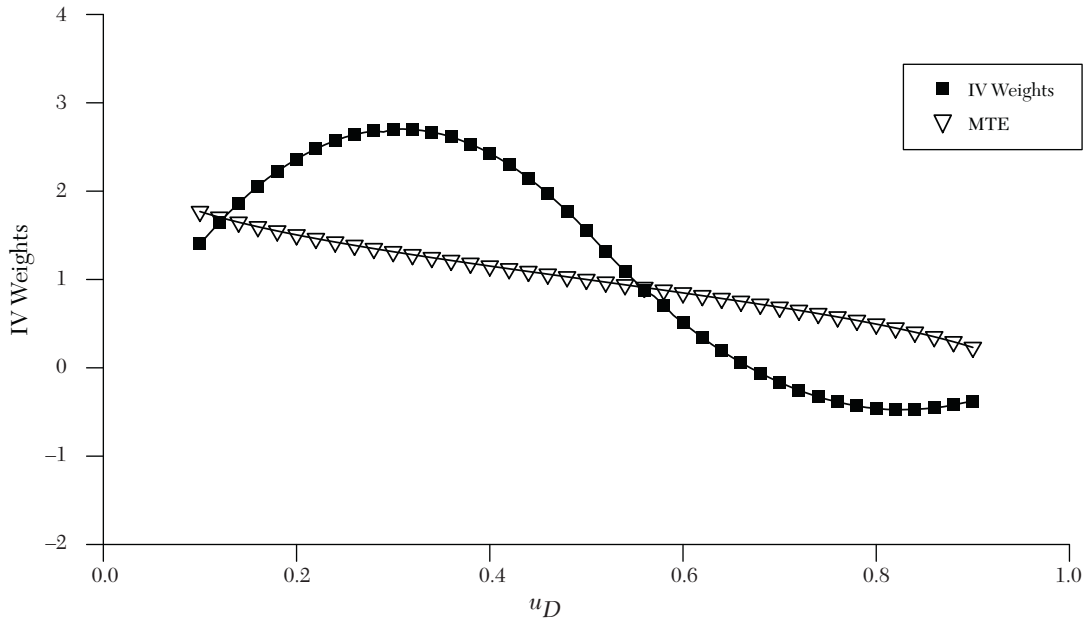


Figure 6. MTE and IV Weights for a General Instrument  $Z_1$ , a Component of  $Z = (Z_1, Z_2)$

and Vytlacil 1999, 2001a, 2001b, 2007b.) We can construct the effects of policy changes for new policies that stay within the support of  $P(Z)$  (see Carneiro, Heckman, and Vytlacil 2010).

### 3.7 Policy Effects, Treatment Effects, and IV

A main lesson of this paper is that policy effects are not generally the same as treatment effects and, in general, neither are produced from IV estimators. Since randomized assignments of components of  $Z$  are instruments, this analysis also applies to the output of randomized experiments. The economic approach to policy evaluation formulates policy questions using well-defined economic models. It then uses whatever statistical tools it takes to answer those questions. Policy questions and not statistical methods drive

analyses. Well-posed economic models are scarce in the program evaluation approach. Thus in contrast to the structural approach, it features statistical methods over economic content. “Credibility” in the program evaluation literature is assessed by statistical properties of estimators and not economic content or policy relevance.

We can do better than hoping that an instrument or an estimator answers policy problems. By recovering economic primitives, we can distinguish the objects various estimators identify from the objects needed to address policy problems. Constructing the PRTE is an example of this approach.<sup>74</sup>

<sup>74</sup> An alternative approach developed in Heckman and Vytlacil (2005) constructs combinations of instruments using sample data on  $Z$  that address specific policy questions.



TABLE 4  
IV ESTIMATOR FOR THREE DIFFERENT DISTRIBUTIONS OF Z BUT THE SAME GENERALIZED ROY MODEL.

Data Distribution	IV	ATE
1	0.434	0.2
2	0.078	0.2
3	-2.261	0.2

Source: Heckman, Urzua, and Vytlačil (2006, table 3).

Figure 7, taken from an analysis of the returns to attending college by Carneiro, Heckman, and Vytlačil (2009), plots the estimated weights for MTE from a marginal change in policy that proportionally expands the probability of attending college for everyone. The figure also plots the estimated MTE and the IV weight using  $P(Z)$  as an instrument. The IV weights and the policy weights are very different. The policy weights oversample high values of  $u_D$  compared to the IV weights. Since the MTE is declining in  $u_D$ , this translates into an IV estimate of 0.095 compared to a marginal policy effect of 0.015.<sup>75</sup> The IV estimate would suggest a substantial mean marginal gross return. The true marginal policy effect is much lower. Since the MTE can be estimated (or approximated) from the data and the policy weights constructed from the data, one can produce more accurate policy forecasts using the economics of the model.

### 3.8 Multiple Choices

Imbens and Angrist analyze a two choice model. Heckman, Urzua, and Vytlačil (2006, forthcoming) and Heckman and Vytlačil (2007b) extend their analysis to an ordered

choice model and to general unordered choice models.<sup>76</sup>

In the special case where the analyst seeks to estimate the mean return to those induced into a choice state by a change in an instrument compared to their next best option, the LATE framework remains useful (see Heckman, Urzua, and Vytlačil 2006, forthcoming; Heckman and Vytlačil 2007b). If, however, one is interested in identifying the mean returns to any pair of outcomes, unaided IV will not do the job. Structural methods are required.

In general unordered choice models, agents attracted into a state by a change in an instrument come from many origin states, so there are many margins of choice. Structural models can identify the gains arising from choices at these separate margins. This is a difficult task for IV without invoking structural assumptions. Structural models can also identify the fraction of persons induced into a state coming from each origin state. IV alone cannot. See Heckman and Urzua (2010).

<sup>75</sup> These estimated effects are statistically significantly different from each other (see Carneiro, Heckman, and Vytlačil 2009).

<sup>76</sup> Angrist and Imbens (1995) propose an ordered choice version of their 1994 paper. As shown by Heckman, Urzua, and Vytlačil (2006) and Heckman and Vytlačil (2007b), their proposed extension has some unsatisfactory features which can be removed by using an extension of the generalized Roy model to an ordered choice model using the choice framework of Cunha, Heckman, and Navarro (2007) and Carneiro, Hansen, and Heckman (2003).

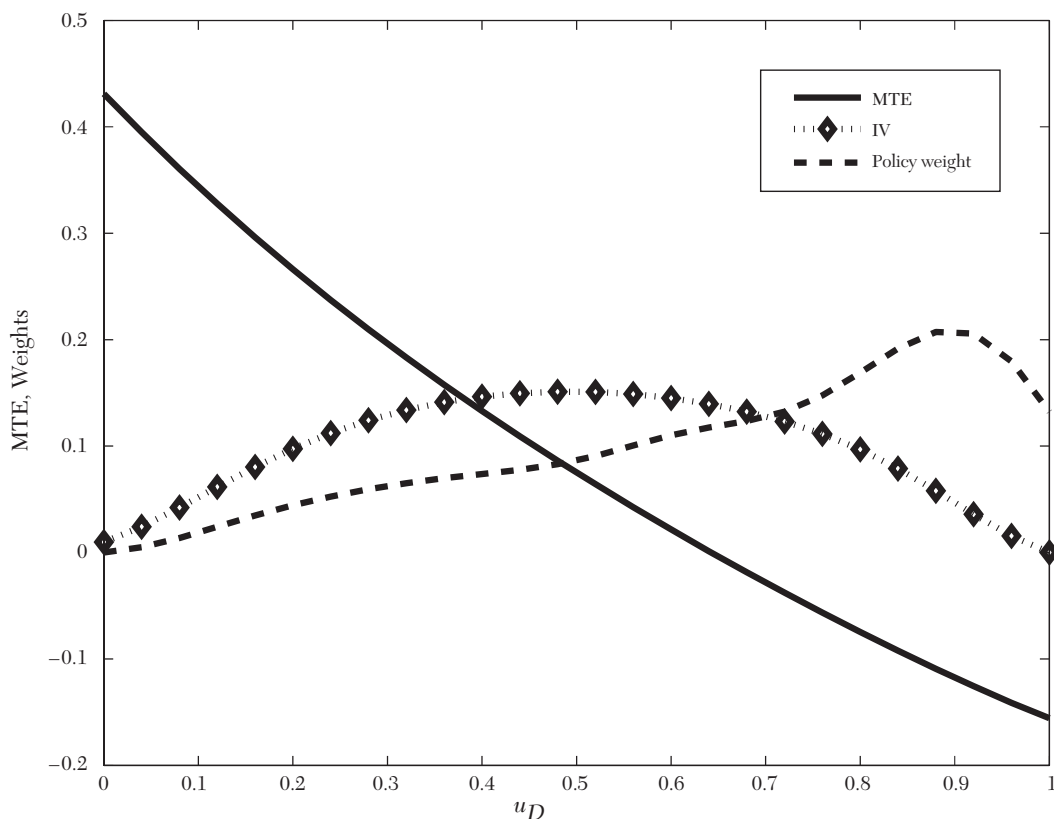


Figure 7. MTE and Weights for IV and MPRTE in the Carneiro–Heckman–Vytlačil (2009) Analysis of the Wage Returns to College

Source: Carneiro, Heckman, and Vytlačil (2009).

Notes: The scale of the  $y$ -axis is the scale of the MTE, not the scale of the weights, which are scaled to fit the picture. The IV is  $P(Z)$ .

#### 4. Conclusions

This paper compares the structural approach to empirical policy analysis with the program evaluation approach. It offers a third way to do policy analysis that combines the best features of both approaches. This paper does not endorse or attack any particular statistical methodology. Economists are fortunate to have a rich menu of estimation methods from which to choose.

This paper advocates placing the economic and policy questions being addressed front and center. Economic theory helps to sharpen statements of policy questions. Modern advances in statistics can make the theory useful in addressing these questions. A better approach is to use the economics to frame the questions and the statistics to help address them.

Both the program evaluation approach and the structural approach have desirable

features. Program evaluation approaches are generally computationally simpler than structural approaches, and it is often easier to conduct sensitivity and replication analyses with them. Identification of program effects is often more transparent than identification of structural parameters. At the same time, the economic questions answered and the policy relevance of the treatment effects featured in the program evaluation approach are often very unclear. Structural approaches produce more interpretable parameters that are better suited to conduct counterfactual policy analyses.

The third way advocated in this essay is to use Marschak's Maxim to identify the policy relevant *combinations* of structural parameters that answer well-posed policy and economic questions. This approach often simplifies the burden of computation, facilitates replication and sensitivity analyses, and makes identification more transparent. At the same time, application of this approach forces analysts to clearly state the goals of the policy analysis—something many economists (structural or program evaluation) have difficulty doing. That discipline is an added bonus of this approach.

I have illustrated this approach by using the economics implicit in LATE to interpret the margins of choice identified by instrument variation and to extend the range of questions LATE can answer. This analysis is a prototype of the value of a closer integration of theory and robust statistical methods to evaluate public policy.

## Appendices

### A. Derivation of the Weights for PRTE

**Preliminary Remarks:** Recall that, if for two random variables  $J$  and  $K$  for  $0 \leq j \leq 1$  and  $0 \leq k \leq 1$ , with density  $f_{j,k}(j, k)$

$$\int \int |f_{j,k}(j, k)| dj dk < \infty,$$

the value of the integral for the region  $0 < j < k < 1$

$$\begin{aligned} & \int_0^1 \int_0^k f_{j,k}(j, k) dj dk \\ &= \int_0^1 \int_j^1 f_{j,k}(j, k) dk dj. \end{aligned}$$

**Derivation:** We can write

$$\begin{aligned} & E(Y | \text{Baseline } b) \\ &= \int_0^1 E(Y | P(Z) = p) f_P^b(p) dp \\ &= E[\mathbf{1}(P(Z) \geq U_D) Y_1 + \mathbf{1}(P(Z) < U_D) Y_0] \\ &= \int_0^1 \int_0^p \int_{-\infty}^{\infty} y_1 f_{Y_1, U_D}(y_1, u_D) dy_1 \\ & \quad du_D f_P^b(p) dp \\ & \quad + \int_0^1 \int_p^1 \int_{-\infty}^{\infty} y_0 f_{Y_0, U_D}(y_0, u_D) dy_0 \\ & \quad du_D f_P^b(p) dp \end{aligned}$$

because  $P(Z) \perp U_D, Y_1, Y_0 | X$ .

Thus  $E(Y | \text{Baseline } b)$

$$\begin{aligned} &= \int_0^1 \int_0^p E(Y_1 | U_D = u_D) du_D f_P^b(p) dp \\ & \quad + \int_0^1 \int_p^1 E(Y_0 | U_D = u_D) du_D f_P^b(p) dp. \end{aligned}$$

Interchanging the limits of each integral

$E(Y | \text{Baseline})$

$$\begin{aligned}
 &= \int_0^1 E(Y_1 | U_D = u_D) \left[ \int_{u_D}^1 F_P^b(p) dp \right] du_D \\
 &+ \int_0^1 E(Y_0 | U_D = u_D) \left[ \int_0^{u_D} F_P^b(p) dp \right] du_D \\
 &= \int_0^1 E(Y_1 | U_D = u_D) [1 - F_P^b(u_D)] du_D \\
 &+ \int_0^1 E(Y_0 | U_D = u_D) F_P^b(u_D) du_D.
 \end{aligned}$$

By a parallel argument for  $E(Y | \text{Alternative Policy } a)$

$$\begin{aligned}
 &= \int_0^1 E(Y_1 | U_D = u_D) [1 - F_P^a(u_D)] du_D \\
 &+ \int_0^1 E(Y_0 | U_D = u_D) F_P^a(u_D) du_D.
 \end{aligned}$$

Subtracting the first expression from the second expression we obtain the expression in the text

$$\begin{aligned}
 \text{PRTE} &= \int_0^1 E(Y_1 - Y_0 | U_D = u_D) (F_P^b(u_D) \\
 &\quad - F_P^a(u_D)) du_D.
 \end{aligned}$$

### B. IV for Discrete Instruments

Suppose that the support of the distribution of  $P(Z)$  contains a finite number of values  $p_1 < p_2 < \dots < p_K$ . The support of the instrument  $Z_1$  is also discrete, taking  $I$  distinct values.  $E(Z_1 | P(Z) > u_D)$  is constant in  $u_D$  for  $u_D$  within any  $(p_\ell, p_{\ell+1})$  interval, and

$\Pr(P(Z) > u_D)$  is constant in  $u_D$  for  $u_D$  within any  $(p_\ell, p_{\ell+1})$  interval. Let  $\lambda_\ell$  denote the weight on the LATE for the interval  $(p_\ell, p_{\ell+1})$ .

Under monotonicity condition (IA-3),

$$\begin{aligned}
 \text{IV}_{Z_1} &= \int E(Y_1 - Y_0 | U_D = u_D) h_{Z_1}^{IV}(u_D) du_D \\
 &= \sum_{\ell=1}^{K-1} \lambda_\ell \int_{p_\ell}^{p_{\ell+1}} E(Y_1 - Y_0 | U_D = u_D) \\
 &\quad \times \frac{1}{(p_{\ell+1} - p_\ell)} du_D \\
 &= \sum_{\ell=1}^{K-1} \Delta^{\text{LATE}}(p_\ell, p_{\ell+1}) \lambda_\ell.
 \end{aligned}$$

Let  $z_1^i$  be the  $i$ th smallest value of the support of  $Z_1$ :

$$\begin{aligned}
 (21) \quad \lambda_\ell &= (p_{\ell+1} - p_\ell) \\
 &\quad \times \frac{\sum_{i=1}^I (z_1^i - E(Z_1)) \sum_{t \geq \ell}^K (f(z_1^i, p_t))}{\text{Cov}(Z_1, D)}.
 \end{aligned}$$

### REFERENCES

- Abbring, Jaap H., and James J. Heckman. 2007. "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation." In *Handbook of Econometrics, Volume 6B*, ed. James J. Heckman and Edward E. Leamer, 5145–5303. Amsterdam and Oxford: Elsevier, North-Holland.
- Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. "Econometric Tools for Analyzing Market Outcomes." In *Handbook of Econometrics, Volume 6A*, ed. James J. Heckman and Edward E. Leamer, 4171–4276. Amsterdam and Oxford: Elsevier, North-Holland.
- Aguirregabiria, Victor, and Pedro Mira. 2010. "Dynamic Discrete Choice Structural Models: A Survey." *Journal of Econometrics*, 156(1): 38–67.
- Ahn, Hyungha, and James L. Powell. 1993. "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism." *Journal of Econometrics*, 58(1–2): 3–29.
- Albrecht, James, Gerard J. van den Berg, and Susan Vroman. 2009. "The Aggregate Labor Market Effects

- of the Swedish Knowledge Lift Program." *Review of Economic Dynamics*, 12(1): 129–146.
- Andrews, Donald W. K., and Marcia M. A. Schafgans. 1998. "Semiparametric Estimation of the Intercept of a Sample Selection Model." *Review of Economic Studies*, 65(3): 497–517.
- Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430): 431–42.
- Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton and Oxford: Princeton University Press.
- Angrist, Joshua D., and Jorn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3–30.
- Athey, Susan, and Philip A. Haile. 2007. "Nonparametric Approaches to Auctions." In *Handbook of Econometrics, Volume 6A*, ed. James J. Heckman and Edward E. Leamer, 3847–3965. Amsterdam and Oxford: Elsevier, North-Holland.
- Attanasio, Orazio, Costas Meghir, and Ana Santiago. 2009. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progreso." IFS/EDEPO Working Paper.
- Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2005. "Trends in U.S. Wage Inequality: Reassessing the Revisionists." National Bureau of Economic Research Working Paper 11627.
- Barnow, Burt S., Glen G. Cain, and Arthur S. Goldberger. 1980. "Issues in the Analysis of Selectivity Bias." In *Evaluation Studies, Volume 5*, ed. E. Stromsdorfer and G. Farkas, 42–59. Beverly Hills, Calif: Sage Publications.
- Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: National Bureau of Economic Research.
- Berk, Richard, Azusa Li, and Laura J. Hickman. 2005. "Statistical Difficulties in Determining the Role of Race in Capital Cases: A Re-analysis of Data from the State of Maryland." *Journal of Quantitative Criminology*, 21(4): 365–90.
- Björklund, Anders, and Robert Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection." *Review of Economics and Statistics*, 69(1): 42–49.
- Brock, William A., and Steven N. Durlauf. 2001. "Interactions-Based Models." In *Handbook of Econometrics, Volume 5*, ed. James J. Heckman and Edward E. Leamer, 3297–3380. Amsterdam; London and New York: Elsevier Science, North-Holland.
- Brown, Donald J., and Rosa L. Matzkin. 1996. "Testable Restrictions on the Equilibrium Manifold." *Econometrica*, 64(6): 1249–62.
- Cameron, Stephen V., and Christopher Taber. 2004. "Estimation of Educational Borrowing Constraints Using Returns to Schooling." *Journal of Political Economy*, 112(1): 132–82.
- Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics, Volume 3A*, ed. Orley Ashenfelter and David Card, 1801–63. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.
- Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*, 69(5): 1127–60.
- Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2001. "Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies." *Swedish Economic Policy Review*, 8(2): 273–301.
- Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review*, 44(2): 361–422.
- Carneiro, Pedro, James J. Heckman, and Edward J. Vytlačil. 2009. "Estimating Marginal Returns to Education." Unpublished.
- Carneiro, Pedro, James J. Heckman, and Edward J. Vytlačil. 2010. "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin." *Econometrica*, 78(1): 377–94.
- Cartwright, Nancy. 2004. "Causation: One Word, Many Things." *Philosophy of Science*, 71(5): 805–19.
- Chalak, Karim, Susanne Schennach, and Halbert White. 2010. "Local Indirect Least Squares and Average Marginal Effects in Nonseparable Structural Systems." Boston College Department of Economics Working Paper 680.
- Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods." *Annual Review of Economics*, 1: 451–88.
- Cosslett, Stephen R. 1983. "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model." *Econometrica*, 51(3): 765–82.
- Cox, David R. 1958. *Planning of Experiments*. New York: Wiley.
- Cunha, Flavio, and James J. Heckman. 2007. "The Evolution of Inequality, Heterogeneity and Uncertainty in Labor Earnings in the U.S. Economy." National Bureau of Economic Research Working Paper 13526.
- Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers*, 57(2): 191–261.
- Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2006. "Counterfactual Analysis of Inequality and Social Mobility." In *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, ed. Stephen L. Morgan, David B. Grusky, and Gary S. Fields, 290–346. Stanford: Stanford University Press.

- Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2007. "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds." *International Economic Review*, 48(4): 1273–1309.
- Das, Mitali, Whitney K. Newey, and Francis Vella. 2003. "Nonparametric Estimation of Sample Selection Models." *Review of Economic Studies*, 70(1): 33–58.
- Duflo, Esther. 2004. "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia." *Journal of Development Economics*, 74(1): 163–97.
- Durbin, J. 1954. "Errors in Variables." *Review of the International Statistical Institute*, 22: 23–32.
- Durlauf, Steven N., and H. Peyton Young, eds. 2001. *Social Dynamics*. Washington, D.C.: Brookings Institution Press; Cambridge and London: MIT Press.
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen. Forthcoming. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics*, 125(3).
- Einav, Liran, and Jonathan D. Levin. 2010. "Empirical Industrial Organization: A Progress Report." *Journal of Economic Perspectives*, 24(2): 145–62.
- Farber, Henry S. 1983. "The Determination of the Union Status of Workers." *Econometrica*, 51(5): 1417–37.
- Frisch, Ragnar. 2009. *Problems and Methods of Econometrics: The Poincaré Lectures of Ragnar Frisch, 1933*. New York: Routledge.
- Granger, Clive W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, 37(3): 424–38.
- Gronau, Reuben. 1974. "Wage Comparisons—A Selectivity Bias." *Journal of Political Economy*, 82(6): 1119–43.
- Hahn, Jinyong, Petra E. Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–09.
- Hansen, Lars Peter, and James J. Heckman. 1996. "The Empirical Foundations of Calibration." *Journal of Economic Perspectives*, 10(1): 87–104.
- Harberger, Arnold C. 1964. "The Measurement of Waste." *American Economic Review*, 54(3): 58–76.
- Hausman, Jerry A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6): 1251–71.
- Heckman, James J. 1974. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica*, 42(4): 679–94.
- Heckman, James J. 1990. "Varieties of Selection Bias." *American Economic Review*, 80(2): 313–18.
- Heckman, James J. 1992. "Haavelmo and the Birth of Modern Econometrics: A Review of *The History of Econometric Ideas* by Mary Morgan." *Journal of Economic Literature*, 30(2): 876–86.
- Heckman, James J. 1996. "Randomization as an Instrumental Variable." *Review of Economics and Statistics*, 78(2): 336–41.
- Heckman, James J. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*, 32(3): 441–62.
- Heckman, James J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics*, 115(1): 45–97.
- Heckman, James J. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy*, 109(4): 673–748.
- Heckman, James J. 2008. "Econometric Causality." *International Statistical Review*, 76(1): 1–27.
- Heckman, James J., and Bo E. Honoré. 1990. "The Empirical Content of the Roy Model." *Econometrica*, 58(5): 1121–49.
- Heckman, James J., and V. Joseph Hotz. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association*, 84(408): 862–74.
- Heckman, James J., Robert J. LaLonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics, Volume 3A*, ed. Orley Ashenfelter and David Card, 1865–2097. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.
- Heckman, James J., Lance Lochner, and Christopher Taber. 1998a. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics*, 1(1): 1–58.
- Heckman, James J., Lance Lochner, and Christopher Taber. 1998b. "General-Equilibrium Treatment Effects: A Study of Tuition Policy." *American Economic Review*, 88(2): 381–86.
- Heckman, James J., Lance Lochner, and Christopher Taber. 1998c. "Tax Policy and Human-Capital Formation." *American Economic Review*, 88(2): 293–97.
- Heckman, James J., and Richard Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, ed. James J. Heckman and Burton Singer, 156–245. Cambridge; New York and Sydney: Cambridge University Press.
- Heckman, James J., Daniel Schmieder, and Sergio Urzua. Forthcoming. "Testing the Correlated Random Coefficient Model." *Journal of Econometrics*.
- Heckman, James J., and Jeffrey A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives*, 9(2): 85–110.
- Heckman, James J., and Jeffrey A. Smith. 1998. "Evaluating the Welfare State." In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. Steinar Strøm, 241–318. Cambridge; New York and Melbourne: Cambridge University Press.
- Heckman, James J., Jeffrey A. Smith, and Nancy Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*, 64(4): 487–535.



- Heckman, James J., and Sergio Urzua. 2010. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." *Journal of Econometrics*, 156(1): 27–37.
- Heckman, James J., Sergio Urzua, and Edward J. Vytlačil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics*, 88(3): 389–432.
- Heckman, James J., Sergio Urzua, and Edward J. Vytlačil. Forthcoming. "Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case." *Les Annales d'Economie et de Statistique*.
- Heckman, James J., and Edward J. Vytlačil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences*, 96(8): 4730–34.
- Heckman, James J., and Edward J. Vytlačil. 2000. "The Relationship between Treatment Parameters within a Latent Variable Framework." *Economics Letters*, 66(1): 33–39.
- Heckman, James J., and Edward J. Vytlačil. 2001a. "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect." In *Econometric Evaluation of Labour Market Policies*, ed. Michael Lechner and Friedhelm Pfeiffer, 1–15. Heidelberg and New York: Physica; Mannheim: Centre for European Economic Research.
- Heckman, James J., and Edward J. Vytlačil. 2001b. "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. Cheng Hsiao, Kimio Morimune, and James L. Powell, 1–46. Cambridge; New York and Melbourne: Cambridge University Press.
- Heckman, James J., and Edward J. Vytlačil. 2001c. "Policy-Relevant Treatment Effects." *American Economic Review*, 91(2): 107–11.
- Heckman, James J., and Edward J. Vytlačil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738.
- Heckman, James J., and Edward J. Vytlačil. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics, Volume 6B*, ed. James J. Heckman and Edward E. Leamer, 4779–4874. Amsterdam and Oxford: Elsevier, North-Holland.
- Heckman, James J., and Edward J. Vytlačil. 2007b. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments." In *Handbook of Econometrics, Volume 6B*, ed. James J. Heckman and Edward E. Leamer, 4875–5144. Amsterdam and Oxford: Elsevier, North-Holland.
- Hendry, David F. 1980. "Econometrics—Alchemy or Science?" *Economica*, 47(188): 387–406.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*, 81(396): 945–60.
- Hurwicz, Leonid. 1962. "On the Structural Form of Interdependent Systems." In *Logic, Methodology and Philosophy of Science*, ed. Ernest Nagel, P. Suppes, and A. Tarski, 232–39. Stanford: Stanford University Press.
- Ichimura, Hidehiko, and Christopher Taber. 2002. "Semiparametric Reduced-Form Estimation of Tuition Subsidies." *American Economic Review*, 92(2): 286–92.
- Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–75.
- Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1): 5–86.
- Katz, Lawrence F., and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics, Volume 3A*, ed. Orley Ashenfelter and David Card, 1463–1555. Amsterdam; New York and Oxford: Elsevier Science, North-Holland.
- Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 107(1): 35–78.
- Keane, Michael P. 2010. "A Structural Perspective on the Experimentalist School." *Journal of Economic Perspectives*, 24(2): 47–58.
- Keane, Michael P., Petra E. Todd, and Kenneth I. Wolpin. Forthcoming. "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." In *Handbook of Labor Economics, Volume 4*.
- Killingsworth, Mark R., and James J. Heckman. 1986. "Female Labor Supply: A Survey." In *Handbook of Labor Economics, Volume 1*, ed. Orley Ashenfelter and Richard Layard, 103–204. Amsterdam; Oxford and Tokyo: North-Holland.
- Klein, Lawrence R. 1953. *A Textbook of Econometrics*. Evanston: Row, Peterson and Co.
- Klein, Roger W., and Richard H. Spady. 1993. "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica*, 61(2): 387–421.
- Kydland, Finn E., and Edward C. Prescott. 1996. "The Computational Experiment: An Econometric Tool." *Journal of Economic Perspectives*, 10(1): 69–85.
- LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 76(4): 604–20.
- Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review*, 73(1): 31–43.
- Leamer, Edward E. 2010. "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives*, 24(2): 31–46.
- Lee, Donghoon, and Kenneth I. Wolpin. 2006. "Intersectoral Labor Mobility and the Growth of the

- Service Sector." *Econometrica*, 74(1): 1–46.
- Lewis, H. Gregg. 1986. *Union Wage Effects: A Survey*. Chicago: University of Chicago Press.
- Lise, Jeremy, Shannon Seitz, and Jeffrey Smith. 2005. "Equilibrium Policy Experiments and the Evaluation of Social Programs." Queen's University Department of Economics Working Paper 1076.
- Lise, Jeremy, Shannon Seitz, and Jeffrey Smith. 2006. "Evaluating Search and Matching Models Using Experimental Data." Queen's University Department of Economics Working Paper 1074.
- Liu, Ta-Chung. 1960. "Underidentification, Structural Estimation, and Forecasting." *Econometrica*, 28(4): 855–65.
- Manski, Charles F. 1988. "Identification of Binary Response Models." *Journal of the American Statistical Association*, 83(403): 729–38.
- Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review*, 80(2): 319–23.
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–42.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge and London: Harvard University Press.
- Manski, Charles F. 2003. *Partial Identification of Probabilities Distributions*. New York and Heidelberg: Springer.
- Manski, Charles F. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–76.
- Marschak, Jacob. 1953. "Economic Measurements for Policy and Prediction." In *Studies in Econometric Method*, ed. William C. Hood and Tjalling C. Koopmans, 1–26. New York: Wiley.
- Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan and Company.
- Matzkin, Rosa L. 1992. "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models." *Econometrica*, 60(2): 239–70.
- Matzkin, Rosa L. 1993. "Nonparametric Identification and Estimation of Polychotomous Choice Models." *Journal of Econometrics*, 58(1–2): 137–68.
- Matzkin, Rosa L. 1994. "Restrictions of Economic Theory in Nonparametric Methods." In *Handbook of Econometrics, Volume 4*, ed. Robert F. Engle and Daniel L. McFadden, 2523–58. Amsterdam; London and New York: Elsevier, North-Holland.
- Matzkin, Rosa L. 2007. "Nonparametric Identification." In *Handbook of Econometrics, Volume 6B*, ed. James J. Heckman and Edward E. Leamer, 5307–68. Amsterdam and Oxford: Elsevier, North-Holland.
- Matzkin, Rosa L. 2010a. "Estimation of Nonparametric Models with Simultaneity." Unpublished.
- Matzkin, Rosa L. 2010b. "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity." Unpublished.
- Nevo, Aviv, and Michael D. Whinston. 2010. "Taking the Dogma Out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2): 69–82.
- Neyman, Jerzy. 1923. "Statistical Problems in Agricultural Experiments." *Journal of the Royal Statistical Society B (Supplement 2)*: 107–80.
- Pencavel, John H. 1986. "Labor Supply of Men: A Survey." In *Handbook of Labor Economics, Volume 1*, ed. Orley Ashenfelter and Richard Layard, 3–102. Amsterdam; Oxford and Tokyo: North-Holland.
- Poirier, Dale J. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics*, 12(2): 209–17.
- Powell, James L. 1994. "Estimation of Semiparametric Models." In *Handbook of Econometrics, Volume 4*, ed. Robert F. Engle and Daniel L. McFadden, 2443–2521. Amsterdam; London and New York: Elsevier, North-Holland.
- Quandt, Richard E. 1958. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." *Journal of the American Statistical Association*, 53(284): 873–80.
- Quandt, Richard E. 1972. "A New Approach to Estimating Switching Regressions." *Journal of the American Statistical Association*, 67(338): 306–10.
- Rosenzweig, Mark R., and Kenneth I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature*, 38(4): 827–74.
- Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–46.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5): 688–701.
- Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics*, 6(1): 34–58.
- Rubin, Donald B. 1986. "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers." *Journal of the American Statistical Association*, 81(396): 961–62.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Shimer, Robert, and Ivan Werning. 2008. "Liquidity and Insurance for the Unemployed." *American Economic Review*, 98(5): 1922–42.
- Sims, Christopher A. 1972. "Money, Income, and Causality." *American Economic Review*, 62(4): 540–52.
- Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica*, 48(1): 1–48.
- Sims, Christopher A. 1996. "Macroeconomics and Methodology." *Journal of Economic Perspectives*, 10(1): 105–20.
- Sims, Christopher A. 2010. "But Economics Is Not an Experimental Science." *Journal of Economic Perspectives*, 24(2): 59–68.
- Stock, James H. 1989. "Nonparametric Policy Analysis." *Journal of the American Statistical Association*, 84(406): 567–75.

- Stock, James H. 2010. "The Other Transformation in Econometric Practice: Robust Tools for Inference." *Journal of Economic Perspectives*, 24(2): 83–94.
- Tamer, Elie. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies*, 70(1): 147–65.
- Todd, Petra E., and Kenneth I. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility." *American Economic Review*, 96(5): 1384–1417.
- Tukey, John W. 1986. "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes: Comment." In *Drawing Inferences from Self-Selected Samples*, ed. Howard Wainer, 108–110. New York: Springer-Verlag.
- Tunali, Insan. 2000. "Rationality of Migration." *International Economic Review*, 41(4): 893–920.
- Vella, Francis. 1998. "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources*, 33(1): 127–69.
- Vytlacil, Edward J. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica*, 70(1): 331–41.
- Willis, Robert J., and Sherwin Rosen. 1979. "Education and Self-Selection." *Journal of Political Economy*, 87(5): S7–36.
- Wu, De-Min. 1973. "Alternative Tests of Independence between Stochastic Regressors and Disturbances." *Econometrica*, 41(4): 733–50.
- Yitzhaki, Shlomo. 1989. "On Using Linear Regressions in Welfare Economics." Hebrew University Department of Economics Working Paper 217.

**This article has been cited by:**

1. Pedro Carneiro, , James J. Heckman, , Edward J. Vytlačil. 2011. Estimating Marginal Returns to Education. *American Economic Review* **101**:6, 2754-2781. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
2. François Claveau. 2011. Evidential variety as a source of credibility for causal inference: beyond sharp designs and structural models. *Journal of Economic Methodology* **18**:3, 233-253. [[CrossRef](#)]
3. Peter Arcidiacono, Paul B. Ellickson. 2011. Practical Methods for Estimation of Dynamic Discrete Choice Models. *Annual Review of Economics* **3**:1, 363-394. [[CrossRef](#)]
4. Jens Ludwig, , Jeffrey R. Kling, , Sendhil Mullainathan. 2011. Mechanism Experiments and Policy Evaluations. *Journal of Economic Perspectives* **25**:3, 17-38. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
5. Andrew Dillon. 2011. Do Differences in the Scale of Irrigation Projects Generate Different Impacts on Poverty and Production?. *Journal of Agricultural Economics* **62**:2, 474-492. [[CrossRef](#)]
6. Judea Pearl. 2011. Statistics and Causality: Separated to Reunite-Commentary on Bryan Dowd's "Separated at Birth". *Health Services Research* **46**:2, 421-429. [[CrossRef](#)]
7. C. B. Barrett, M. R. Carter. 2010. The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Applied Economic Perspectives and Policy* **32**:4, 515-548. [[CrossRef](#)]