

Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy

James J. Heckman
University of Chicago,
University College Dublin,
Cowles Foundation,
Yale University,
and the American Bar Foundation

Italian Statistical Society Scientific Meeting
Padua, Italy
June 16, 2010

Introduction

- A lot of division in the profession about the “right” way to do empirical policy analysis

Introduction

- A lot of division in the profession about the “right” way to do empirical policy analysis
- Structural vs. reduced form approaches

Introduction

- A lot of division in the profession about the “right” way to do empirical policy analysis
- Structural vs. reduced form approaches
- Structural used in the sense of Hurwicz (1962)

Introduction

- A lot of division in the profession about the “right” way to do empirical policy analysis
- Structural vs. reduced form approaches
- Structural used in the sense of Hurwicz (1962)
- As implemented, becomes parametric models with parameters invariant to classes of interventions

- An example of a structural relationship (Haavelmo, 1944, *Econometrica*)

$$Y = X\beta + Z\gamma + U \quad (*)$$

X background variables

Z policy variables

$*$ is an “all causes” model

External manipulations:

Variations in (X, Z) that hold U fixed

If the coefficients (β, γ) are invariant to shifts in (X, Z) , then $(*)$ is structural.

- Notice that OLS is

$$E^*(Y \mid X, Z) = X\beta + Z\gamma + E^*(U \mid X, Z)$$

where E^* is a linear projection.

- Notice that OLS is

$$E^*(Y \mid X, Z) = X\beta + Z\gamma + E^*(U \mid X, Z)$$

where E^* is a linear projection.

- OLS is not estimating a structural relationship.

- Notice that OLS is

$$E^*(Y \mid X, Z) = X\beta + Z\gamma + E^*(U \mid X, Z)$$

where E^* is a linear projection.

- OLS is not estimating a structural relationship.
- If $E(U \mid X, Z) = 0$, OLS gives a structural estimator for (β, Z) .

- If

$$E^*(U \mid X, Z) = E^*(U \mid X)$$

and the coefficient in the original model is invariant to manipulators in Z then OLS is structural for γ .

- If

$$E^*(U \mid X, Z) = E^*(U \mid X)$$

and the coefficient in the original model is invariant to manipulators in Z then OLS is structural for γ .

- But not necessarily for β .

- “Reduced form” has multiple meanings.

- “Reduced form” has multiple meanings.
- One strand of the reduced form approach uses explicit economic models to motivate and interpret empirical analyses and approximates the economic models using simple econometric techniques. (See, e.g., Chetty, 2009.)

- Another strand is the “program evaluation” approach surveyed in Imbens and Wooldridge (2009).

- Another strand is the “program evaluation” approach surveyed in Imbens and Wooldridge (2009).
- This is the approach to “causal effects” advocated in the statistics literature on causality.

- Another strand is the “program evaluation” approach surveyed in Imbens and Wooldridge (2009).
- This is the approach to “causal effects” advocated in the statistics literature on causality.
- 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.

- It often leaves implicit the economics and policy relevance of the output from its procedures.
- It leaves the choice of treatment equation implicit whereas the structural approach makes it explicit.

- It often leaves implicit the economics and policy relevance of the output from its procedures.
- It leaves the choice of treatment equation implicit whereas the structural approach makes it explicit.
- 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.

- The perceived failures of the structural methods of the 1970s and 1980s produced two different methodological responses.

- The perceived failures of the structural methods of the 1970s and 1980s produced two different methodological responses.
- 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 perceived failures of the structural methods of the 1970s and 1980s produced two different methodological responses.
- 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 response was development of a more robust version of the structural approach.

Building Bridges

- Writing in the early 1950s 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.

Building Bridges

- Writing in the early 1950s 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.

- 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.

- 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.

The Structural Versus the Program Evaluation Approach to Evaluating Economic Policies

- Causality at the individual level.

- Causality at the individual level.
- Based on the notion of controlled variation — variation in treatment holding other factors constant.

- Causality at the individual level.
- Based on the notion of controlled variation — variation in treatment holding other factors constant.
- This is Alfred Marshall's (1890) *ceteris paribus* clause which has been the operational definition of causality in economics for over a century.

- Causality at the individual level.
- Based on the notion of controlled variation — variation in treatment holding other factors constant.
- This is Alfred Marshall's (1890) *ceteris paribus* clause which has been the operational definition of causality in economics for over a century.
- It is distinct from other notions of causality sometimes used in economics that are based on prediction (e.g., Granger, 1969, and Sims, 1972).

- Two distinct tasks in causal inference and policy analysis:
(a) Defining counterfactuals and (b) Identifying causal models from data.

- 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.

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

Policy Evaluation Problems and Criteria of Interest

P1

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

- Objective evaluations

P1

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

- Objective evaluations
- Subjective evaluations

P1

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

- Objective evaluations
- Subjective evaluations
- Ex ante and ex post

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.

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.

- **P3** is a problem that economic policy analysts have to solve daily.
- Structural econometrics addresses this question.

- **P3** is a problem that economic policy analysts have to solve daily.
- Structural econometrics addresses this question.
- The program evaluation approach does not except through “demonstration programs” (i.e. that explicitly implement the policies).

A Prototypical Economic Model for Policy Evaluation

- Roy (1951): Agents face two potential outcomes (Y_0, Y_1) with distribution $F_{Y_0, Y_1}(y_0, y_1)$ where “0” refers to a 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).

- Roy (1951): Agents face two potential outcomes (Y_0, Y_1) with distribution $F_{Y_0, Y_1}(y_0, y_1)$ where “0” refers to a 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, set of potential outcomes is $\{Y_s\}_{s \in \mathcal{S}}$ where $\mathcal{S}$ is the set of indices of potential outcomes.

- Roy (1951): Agents face two potential outcomes (Y_0, Y_1) with distribution $F_{Y_0, Y_1}(y_0, y_1)$ where “0” refers to a 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, set of potential outcomes is $\{Y_s\}_{s \in \mathcal{S}}$ where $\mathcal{S}$ is the set of indices of potential outcomes.
- Roy model $\mathcal{S} = \{0, 1\}$.

- Analysts observe either Y_0 or Y_1 , but not both, for any person.

- 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**.

- The **selection problem**.

- The **selection problem**.
- 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.

- The **selection problem**.
- 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 $Y_1 > Y_0$.

- The **selection problem**.
- 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 $Y_1 > Y_0$.

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

- Generalized Roy model
(C is the cost of going from “0” to “1”)

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

- Generalized Roy model
(C is the cost of going from “0” to “1”)

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

- The outcome observed for any person, Y , can be written as

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

- $\mathcal{I}$ denotes agent information set.

- $\mathcal{I}$ denotes agent information set.
- In advance of participation, the agent may be uncertain about all components of (Y_0, Y_1, C) .

- $\mathcal{I}$ denotes agent information set.
- In advance of participation, the agent may be uncertain about all components of (Y_0, Y_1, C) .
- Expected benefit: $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I})$.

- $\mathcal{I}$ denotes agent information set.
- In advance of participation, the agent may be uncertain about all components of (Y_0, Y_1, C) .
- Expected benefit: $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I})$.
- Then

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

- The decision maker selecting “treatment” may be different than the person who experiences the outcomes (Y_0, Y_1).

- The *ex post* objective outcomes are (Y_0, Y_1) .

- The *ex post* objective outcomes are (Y_0, Y_1) .
- The *ex ante* outcomes are $E(Y_0 \mid \mathcal{I})$ and $E(Y_1 \mid \mathcal{I})$.

- The *ex post* objective outcomes are (Y_0, Y_1) .
- The *ex ante* outcomes are $E(Y_0 \mid \mathcal{I})$ and $E(Y_1 \mid \mathcal{I})$.
- The *ex ante* subjective evaluation is I_D .

- The *ex post* objective outcomes are (Y_0, Y_1) .
- The *ex ante* outcomes are $E(Y_0 \mid \mathcal{I})$ and $E(Y_1 \mid \mathcal{I})$.
- The *ex ante* subjective evaluation is I_D .
- The *ex post* subjective evaluation is $Y_1 - Y_0 - C$.

- The *ex post* objective outcomes are (Y_0, Y_1) .
- The *ex ante* outcomes are $E(Y_0 \mid \mathcal{I})$ and $E(Y_1 \mid \mathcal{I})$.
- 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.

- $Y_1 - Y_0$ is the individual level treatment effect.

- $Y_1 - Y_0$ is the individual level treatment effect.
- Also, the Marshallian ceteris paribus causal effect.

- $Y_1 - Y_0$ is the individual level treatment effect.
- Also, the Marshallian ceteris paribus causal effect.
- Because of the evaluation problem, it is generally impossible to identify individual level treatment effects (Task 2).

- $Y_1 - Y_0$ is the individual level treatment effect.
- 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 .

- Economic policies can operate through changing (Y_0, Y_1) or through changing C .
- The structural approach considers policies affecting both returns and costs.

Population Parameters of Interest

- Conventional parameters include the Average Treatment Effect ($ATE = E(Y_1 - Y_0)$), the effect of Treatment on The Treated ($TT = E(Y_1 - Y_0 \mid D = 1)$), or the effect of Treatment on the Untreated ($TUT = E(Y_1 - Y_0 \mid D = 0)$).

- In positive political economy, the fraction of the population that perceives a benefit from treatment is of interest and is called the **voting criterion** and is

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

- In positive political economy, the fraction of the population that perceives a benefit from treatment is of interest and is called the **voting criterion** and is

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

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

- Determining marginal returns to a policy is a central goal of economic analysis.

- 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 $I_D = 0$.

- 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

$$E(Y_1 - Y_0 \mid I_D = 0).$$

Treatment Effects Versus Policy Effects

- Policy Relevant Treatment Effect (Heckman and Vytlačil, 2001c) extends the Average Treatment Effect by accounting for voluntary participation in programs.

- Policy Relevant Treatment Effect (Heckman and Vytlačil, 2001c) extends the Average Treatment Effect by accounting for voluntary participation in programs.
- Designed to address problems **P2** and **P3**.

- Policy Relevant Treatment Effect (Heckman and Vytlacil, 2001c) extends the Average Treatment Effect by accounting for voluntary participation in programs.
- Designed to address problems **P2** and **P3**.
- “*b*”: baseline policy (“before”) and “*a*” represent a policy being evaluated (“after”).

- Policy Relevant Treatment Effect (Heckman and Vytlačil, 2001c) extends the Average Treatment Effect by accounting for voluntary participation in programs.
- Designed to address problems **P2** and **P3**.
- “ b ”: baseline policy (“before”) and “ a ” represent a policy being evaluated (“after”).
- Y^a : outcome under policy a ; Y^b is the outcome under the baseline.

- Policy Relevant Treatment Effect (Heckman and Vytlacil, 2001c) extends the Average Treatment Effect by accounting for voluntary participation in programs.
- Designed to address problems **P2** and **P3**.
- “ b ”: baseline policy (“before”) and “ a ” represent a policy being evaluated (“after”).
- Y^a : outcome under policy a ; 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**.

- 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.

- 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.

- 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 ($C^a \neq C^b$).

- 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 ($C^a \neq C^b$).
- Such invariance 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.

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

- The **Policy Relevant Treatment Effect** (PRTE) is

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

- The **Policy Relevant Treatment Effect** (PRTE) is

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

- Benthamite comparison of aggregate outcomes under policies “a” and “b”. PRTE extends ATE by recognizing that policies affect incentives to participate (C) but do not force people to participate.

- The **Policy Relevant Treatment Effect** (PRTE) is

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

- Benthamite comparison of aggregate outcomes under policies “ a ” and “ b ”. 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.

The Structural Approach Versus the Program Evaluation Approach

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
<i>Ex ante</i> versus <i>ex post</i> counterfactuals	No	Yes
Treatment assignment rules that recognize voluntary nature of participation	No	Yes
Social interactions, general equilibrium effects and contagion	No (assumed away)	Yes (modeled)
Internal validity (problem P1)	Yes	Yes
External validity (problem P2)	No	Yes
Forecasting effects of new policies (problem P3)	No	Yes
Distributional treatment effects	No ^a	Yes (for the general case)
Analyze relationship between outcomes and choice equations	No (implicit)	Yes (explicit)

^aAn exception is the special case of common ranks of individuals across counterfactual states: “rank invariance.” See the discussion in Abbring and Heckman (2007).

- 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.

- 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.

- The local average treatment effect “LATE” parameter of Imbens and Angrist (1994) follows in this tradition and uses instrumental variables as surrogates for randomization.

- The local average treatment effect “LATE” parameter of Imbens and Angrist (1994) 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.

- The local average treatment effect “LATE” parameter of Imbens and Angrist (1994) 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.
- Consider a framework that defines the LATE parameter within the generalized Roy model that separates issues of definition of parameters from issues of identification.

Identifying Policy Parameters

- The structural approach to policy evaluation addresses policy evaluation questions **P1-P3** by estimating models for Y_0 , Y_1 , and C in different economic environments.

Identifying Policy Parameters

- The structural approach to policy evaluation addresses policy evaluation questions **P1-P3** by estimating models for Y_0 , Y_1 , and C in different economic environments.
- Commonly used specifications write

$$Y_1 = \mu_1(X) + U_1, \quad Y_0 = \mu_0(X) + U_0, \quad C = \mu_C(Z) + U_C, \quad (5)$$

where (X, Z) are observed by the analyst, and U_0, U_1, U_C are unobserved.

- Define Z to include all of X .

- Define Z to include all of X .
- Variables in Z not in X are instruments.

- Define Z to include all of X .
- Variables in Z not in X are instruments.
- $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I}) = \mu_D(Z) - V$
 $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z) \mid \mathcal{I})$
 $V = -E(U_1 - U_0 - U_C \mid \mathcal{I})$.

- Define Z to include all of X .
- Variables in Z not in X are instruments.
- $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I}) = \mu_D(Z) - V$
 $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z) \mid \mathcal{I})$
 $V = -E(U_1 - U_0 - U_C \mid \mathcal{I})$.
- Choice equation (3):

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

- Define Z to include all of X .
- Variables in Z not in X are instruments.
- $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I}) = \mu_D(Z) - V$
 $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z) \mid \mathcal{I})$
 $V = -E(U_1 - U_0 - U_C \mid \mathcal{I})$.
- Choice equation (3):

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

- 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.

- 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 of outcome and choice equations.

- 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 of outcome and choice equations.
- Structural econometricians have developed nonparametric identification analyses for the Roy and generalized Roy models.

A useful fact:

$$\begin{aligned} P(z) &= \Pr(D = 1 \mid Z = z) \\ &= \Pr(\mu_D(z) \geq V) \\ &= \Pr\left(\frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V}\right) \end{aligned}$$

$$\begin{aligned} P(z) &= F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{\mu_D(z)}{\sigma_V}\right) \\ U_D &= F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{V}{\sigma_V}\right); \quad \text{Uniform}(0, 1) \end{aligned}$$

$$\begin{aligned} P(z) &= \Pr \left(F_{\frac{V}{\sigma_V}} \left(\frac{\mu_D(z)}{\sigma_V} \right) \right) \geq F_{\left(\frac{V}{\sigma_V}\right)} \left(\frac{V}{\sigma_V} \right) \\ &= \Pr(P(z) \geq U_D) \end{aligned}$$

$P(z)$ is the $p(z)^{\text{th}}$ quantile of U_D .

Recall

$$\begin{aligned} Y &= DY_1 + (1 - D)Y_0 \\ &= Y_0 + D(Y_1 - Y_0) \end{aligned}$$

Keep X implicit (condition on $X = x$)

$$\begin{aligned} E(Y \mid Z = z) &= E(Y_0) + \underbrace{E(Y_1 - Y_0 \mid D = 1, Z = z)P(z)}_{\text{from law of iterated expectations}} \\ &= E(Y_0) + E(Y_1 - Y_0 \mid P(z) \geq U_D)P(z) \end{aligned}$$

$\therefore$ It depends on Z only through $P(Z)$.

$$E(Y \mid Z = z') = E(Y_0) + E(Y_1 - Y_0 \mid P(z') \geq U_D)P(z')$$

- What is $E(Y_1 - Y_0 \mid P(z) \geq U_D)$?
- Let the joint density of $(Y_1 - Y_0, U_D)$ be

$$f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D).$$

- It does not depend on Z . It will, in general, depend on X .
- $E(Y_1 - Y_0 \mid P(z) \geq U_D)$

$$= \frac{\int_{-\infty}^{\infty} \int_0^{P(z)} (y_1 - y_0) f_{y_1 - y_0, u_D}(y_1 - y_0, u_D) du_D d(y_1 - y_0)}{\Pr(P(z) \geq U_D)}$$

- Now recall that

$$U_D = F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{V}{\sigma_V}\right).$$

- U_D is a quantile of the V/σ_V distribution.

- By construction, U_D is Uniform(0, 1) (this is the definition of a quantile).
- $\therefore f_{U_D}(u_D) = 1$.
- Also, $\Pr(P(z) \geq U_D) = P(z)$.
- Notice, by law of conditional probability,

$$f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D) = f_{Y_1 - Y_0, U_D}(y_1 - y_0 \mid U_D = u_D) f_{U_D}(u_D) = 1.$$

$$E(Y_1 - Y_0 \mid P(z) \geq U_D)$$

$$= \frac{\int_0^{P(z)} \int_{-\infty}^{\infty} (y_1 - y_0) f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D) d(y_1 - y_0) du_D}{P(z)}$$

$$E(Y_1 - Y_0 \mid P(z) \geq U_D)$$

$$= \frac{\int_0^{P(z)} \int_{-\infty}^{\infty} (y_1 - y_0) f_{Y_1 - Y_0, U_D}(y_1 - y_0 \mid U_D = u_D) d(y_1 - y_0) du_D}{P(z)}$$

$$= \frac{\int_0^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) du_D}{P(z)}$$

- $$\therefore E(Y \mid Z = z) = E(Y_0) + \int_0^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

$$\frac{\partial E(Y \mid Z = z)}{\partial P(z)} = \underbrace{E(Y_1 - Y_0 \mid U_D = P(z))}_{\text{EOTM or marginal gains}}$$

$$E(Y \mid Z = z') = E(Y_0) + \int_0^{P(z')} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

- Suppose $P(z) > P(z')$

$$\therefore E(Y \mid Z = z) - E(Y \mid Z = z') =$$

$$= \int_{P(z')}^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

$$= E(Y_1 - Y_0 \mid P(z) \geq U_D \geq P(z')) \Pr(P(z) \geq U_D \geq P(z'))$$

- Notice

$$\begin{aligned}\Pr(P(z) \geq U_D \geq P(z')) &= \int_{P(z')}^{P(z)} du_D \\ &= P(z) - P(z')\end{aligned}$$

$$\begin{aligned}E(Y \mid Z = z) - E(Y \mid Z = z') \\ = \underbrace{E(Y_1 - Y_0 \mid P(z) \geq U_D \geq P(z'))}_{\text{LATE}}(P(z) - P(z'))\end{aligned}$$

- $$\frac{E(Y \mid Z = z) - E(Y \mid Z = z')}{P(z) - P(z')} = \text{LATE}(z, z')$$

$$= \frac{\int_{P(z')}^{P(z)} \text{MTE}(u_D) du_D}{P(z) - P(z')}$$

Policy Relevant Treatment Effect

(Keep X implicit)

$$\begin{aligned}
E(Y_p) &= \int_0^1 E(Y_p \mid P_p(Z_p) = t) dF_{P_p}(t) \\
&= \int_0^1 \left[\int_0^1 [\mathbf{1}_{[0,t]}(u_D) E(Y_{1,p} \mid U_D = u_D) \right. \\
&\quad \left. + \mathbf{1}_{(t,1]}(u_D) E(Y_{0,p} \mid U_D = u_D)] du \right] dF_{P_p}(t) \\
&= \int_0^1 \left[\int_0^1 [\mathbf{1}_{[u_D,1]}(t) E(Y_{1,p} \mid U_D = u_D) \right. \\
&\quad \left. + \mathbf{1}_{(0,u_D]}(t) E(Y_{0,p} \mid U_D = u_D)] dF_{P_p}(t) \right] du_D \\
&= \int_0^1 [(1 - F_{P_p}(u_D)) E(Y_{1,p} \mid U_D = u_D) \\
&\quad + F_{P_p|X}(u_D) E(Y_{0,p} \mid U_D = u_D)] du_D.
\end{aligned}$$

- Comparing policy p to policy p' ,

$$E(Y_p) - E(Y_{p'}) = \int_0^1 \underbrace{E(Y_1 - Y_0 \mid U_D = u_D)}_{\text{MTE}(u_D)} (F_{P_{p'}}(u_D) - F_{P_p}(u_D)) du_D,$$

which gives the required weights.

- Policies shift the distribution of $P(Z)$.
- They keep the distribution of Y_1 and Y_0 unchanged.

- This derivation involves changing the order of integration.
- Note that from finiteness of the mean,

$$\begin{aligned} E \left| \mathbf{1}_{[0,t]}(u_D) E(Y_{1,p} \mid U_D = u_D) + \mathbf{1}_{(t,1]}(u_D) E(Y_{0,p} \mid U_D = u_D) \right| \\ \leq E(|Y_1| + |Y_0|) < \infty, \end{aligned}$$

$\therefore$ the change in the order of integration is valid by Fubini's theorem.

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

- For many policy questions, it is not necessary to identify fully specified models to answer a range of policy questions.

- 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.

- 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.

- 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.
- **Marschak's Maxim.**

- 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.
- **Marschak's Maxim.**
- Consider estimating the marginal effect of policy expansions.

- 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.
- **Marschak's Maxim.**
- Consider estimating the marginal effect of policy expansions.
- The traditional structural approach identifies the component parts of $E(Y_1 - Y_0 \mid I_D = 0)$ constructed from estimates of the parameters of equations (5) and (6) and assembles them to estimate the marginal effect of the policy expansion (see Björklund and Moffitt, 1987).

Using Economics to Interpret What LATE Estimates and to Make It Useful For Evaluating a Broad Range of Policies

- Make the implicit economics in LATE explicit.

Using Economics to Interpret What LATE Estimates and to Make It Useful For Evaluating a Broad Range of Policies

- Make the implicit economics in LATE explicit.
- This 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.

- In the economic theory of policy evaluation, a comparison between marginal benefits and marginal costs determines the optimal size of social programs.
- Consider how the IV estimator as embodied in LATE addresses this question, and how making explicit its implicit economics improves the range of policy questions LATE can answer.

- In the spirit of the program evaluation literature, ignore general equilibrium effects and do not emphasize the *ex ante* and *ex post* distinction.

- In the spirit of the program evaluation literature, ignore general equilibrium effects and do not emphasize the *ex ante* and *ex post* distinction.
- Both topics are addressed in many papers in the structural approach.

- In the spirit of the program evaluation literature, ignore general equilibrium effects and 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.

- $Z = z$: random variable Z takes the value z .

- $Z = z$: random variable Z takes the value z .
- Z is a vector with K components, $Z = (Z_1, \dots, Z_K)$.

- $Z = z$: 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., $z^j = (z_1^j, \dots, z_K^j)$.

- $Z = z$: 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., $z^j = (z_1^j, \dots, z_K^j)$.
- Assume all means are finite.

- Equation for *ex post* outcome Y as a function of participation status is

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

- Equation for *ex post* outcome Y as a function of participation status is

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

- In terms of counterfactual notation, $\alpha = \mu_0$, $\varepsilon = U_0$ and $Y_0 = \mu_0 + \varepsilon$, and $\beta = (Y_1 - Y_0) = \mu_1 - \mu_0 + U_1 - U_0$.

- Denote the mean of β by $\bar{\beta}$

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

where $\bar{\beta} = \mu_1 - \mu_0$.

- Instrumental variables (IV) do not in general estimate $\bar{\beta}$ and instrumental variables estimators using different instruments have different probability limits.

- Instrumental variables (IV) do not in general estimate $\bar{\beta}$ and instrumental variables estimators using different instruments have different probability limits.
- Structural selection models can estimate the distribution of β (and hence $\bar{\beta}$) and answer a range of the public policy evaluation questions but under assumptions that are held to be “incredible” in the program evaluation literature.

- Instrumental variables (IV) do not in general estimate $\bar{\beta}$ and instrumental variables estimators using different instruments have different probability limits.
- Structural selection models can estimate the distribution of β (and hence $\bar{\beta}$) and answer a range of the public policy evaluation questions but under assumptions that are held to be “incredible” in the program evaluation literature.
- Angrist and Pischke (2008, 2010) offer the Local Average Treatment Effect (LATE) as a “credible” alternative to structural methods.

- The LATE parameter is widely interpreted as estimating the mean return at the margin defined by manipulation of the instrument.

- The LATE parameter is widely interpreted as estimating the mean return at the margin defined by manipulation of the instrument.
- One must be very careful in making this interpretation.

- The LATE parameter is widely interpreted as estimating the mean return at the margin defined by manipulation of the instrument.
- One must be very careful in making this interpretation.
- Not generally true.

- In general, LATE is not the same as $\bar{\beta}$, but it might be all that is needed to evaluate any particular policy.

- 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?”

- 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?”
- 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.

- 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?”
- 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 in the IV.

- 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?”
- 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 in 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.

- 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.

- 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.
- Using the implicit economics of the model, one can do better than hope that LATE identifies a parameter of interest.

- 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.
- Using the implicit economics of the model, one can do better than hope that LATE identifies a parameter of interest.
- One can also identify the set of persons shifted by the instrument.

- 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.
- Using the implicit economics of the model, one can do better than hope that LATE identifies a parameter of interest.
- One can also identify the set of persons shifted by the instrument.
- One can also extend LATE to address problems **P2** and **P3**.

LATE

- LATE is defined by the variation of an instrument.

LATE

- LATE is defined by the variation of an instrument.
- The instrument in LATE plays the role of a randomized assignment.

LATE

- LATE is defined by the variation of an instrument.
- The instrument in LATE plays the role of a randomized assignment.
- Randomized assignment is an instrument.

LATE

- LATE is defined by the variation of an instrument.
- The instrument in LATE plays the role of a randomized assignment.
- Randomized assignment is an instrument.
- Y_0 and Y_1 are potential *ex post* outcomes.

LATE

- LATE is defined by the variation of an instrument.
- The instrument in LATE plays the role of a randomized assignment.
- Randomized assignment is an instrument.
- 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 .

- $D(z)$ 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.

- $D(z)$ 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.

- $D(z)$ 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) .

- $D(z)$ 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 .

- $D(z)$ 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 .

- Three assumptions define LATE.

(IA-1)

$$(Y_0, Y_1, \{D(z)\}_{z \in \mathcal{Z}}) \perp\!\!\!\perp Z \mid X$$

(IA-2)

$\Pr(D = 1 \mid Z = z)$ is a nontrivial function of z conditional on X .

(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.

(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.
- 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 $D(z^2)$.

(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.
- 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*.

(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.
- 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 .

- Under LATE conditions, for two distinct values of Z , z^1 and z^2 , IV applied to (7) identifies

$$\text{LATE}(z^2, z^1) = E(Y_1 - Y_0 \mid D(z^2) = 1, D(z^1) = 0),$$

if the change from z^1 to z^2 induces people into the program ($D(z^2) \geq D(z^1)$).

- Under LATE conditions, for two distinct values of Z , z^1 and z^2 , IV applied to (7) identifies

$$\text{LATE}(z^2, z^1) = E(Y_1 - Y_0 \mid D(z^2) = 1, D(z^1) = 0),$$

if the change from z^1 to z^2 induces people into the program ($D(z^2) \geq D(z^1)$).

- This is the mean return to participation in the program for people induced to switch treatment status by the change from z^1 to z^2 .

- LATE does not identify which people are induced to change their treatment status by the change in the instrument.

- LATE does not identify which people are induced to change their treatment status by the change in the instrument.
- It leaves unanswered many policy questions.

- LATE does not identify which people are induced to change their treatment status by the change in the instrument.
- It leaves unanswered many policy questions.
- 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.

- 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**.

- 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 a variety of criteria.

- 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 a variety of criteria.
- But using the implicit economics of the problem one can do better.

Making Explicit the Implicit Economics of LATE

- Vytlačil (2002): LATE is **equivalent** to a nonparametric version of the generalized Roy model.

Making Explicit the Implicit Economics of LATE

- Vytlacil (2002): LATE 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 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.

- 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 continuous latent variable discrete choice model, which represents the individual's decision to enroll in the program being studied.

- 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 .

- 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 .
- V may depend on U_0 and U_1 in a general way.

- 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 .
- V may depend on U_0 and U_1 in a general way.
- The counterfactual choice indicator is generated by choice equation (6): $D(z) = \mathbf{1}(\mu_D(z) > V)$.

- 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 .
- V may depend on U_0 and U_1 in a general way.
- The counterfactual choice indicator is generated by choice equation (6): $D(z) = \mathbf{1}(\mu_D(z) > V)$.
- Additive separability between $\mu_D(Z)$ and V in the latent index model (6) plays an essential role in LATE.

- $P(z) \equiv \Pr(D = 1|Z = z)$.

- $P(z) \equiv \Pr(D = 1|Z = z)$.
- $P(z) = \Pr(\mu_D(z) > V) = F_V(\mu_D(z))$.

- Define: $U_D = F_V(V)$

- Define: $U_D = F_V(V)$
- Uniformly distributed over the interval $[0, 1]$

- Define: $U_D = F_V(V)$
- Uniformly distributed over the interval $[0, 1]$
- p^{th} quantile of U_D is p , i.e., the proportion of U_D that is p or lower.

- Define: $U_D = F_V(V)$
- Uniformly distributed over the interval $[0, 1]$
- p^{th} quantile of U_D is p , i.e., the proportion of U_D that is p or lower.

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

- From $P(z)$, one can identify the *ex ante* net benefit I_D up to scale and 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.

- From $P(z)$, one can identify the *ex ante* net benefit I_D up to scale and 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.
- From agent choices, one can supplement the information in LATE and ascertain *ex ante* subjective evaluations.

- The LATE assumptions imply the selection model representation (i.e., the generalized Roy model) and using the selection model representation, one can establish that $E(Y \mid Z = z) = E(Y \mid P(Z) = P(z))$: index sufficiency.

- As a consequence of Vytlacil's theorem, one can define $\text{LATE}(z^2, z^1)$ using the latent variable U_D and the values taken by $P(Z)$ when $Z = z^1$ and $Z = z^2$.

- As a consequence of Vytlačil's theorem, one can define $\text{LATE}(z^2, z^1)$ using the latent variable U_D and the values taken by $P(Z)$ when $Z = z^1$ and $Z = z^2$.
- Use the property that the Z enter the model only through $P(Z)$.

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

- As a consequence of Vytlacil's theorem, one can define $\text{LATE}(z^2, z^1)$ using the latent variable U_D and the values taken by $P(Z)$ when $Z = z^1$ and $Z = z^2$.
- Use the property that the Z enter the model only through $P(Z)$.

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

- The mean gross return to persons whose $U_D \in [P(z^1), P(z^2)]$.

- LATE can be defined within the generalized Roy model, without reference to an instrument.

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

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

the mean gross return to persons whose $U_D \in [\underline{u}_D, \bar{u}_D]$.

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

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

the mean gross return to persons whose $U_D \in [\underline{u}_D, \bar{u}_D]$.

- This separates Task 1 of Table 1 from Task 2.

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

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

the mean gross return to persons whose $U_D \in [\underline{u}_D, \bar{u}_D]$.

- This separates 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 $\Pr(D = 1 \mid Z = z^1) = p_1 = \underline{u}_D$ and $\Pr(D = 1 \mid Z = z^2) = p_2 = \bar{u}_D$).

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.

The Surplus From Treatment and the Marginal Treatment Effect

- Using Vytlacil's theorem, it is possible to understand more deeply what economic questions LATE answers.
- 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

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

- 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 and the proportion of people whose U_D is at or below p : $E(Y_1 - Y_0 \mid p \geq U_D)p = S(p)$.

- 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 and the proportion of people whose U_D is at or below p : $E(Y_1 - Y_0 \mid p \geq U_D)p = S(p)$.
- $$\begin{aligned} E(Y \mid P(Z) = p) &= E(Y_0 + \mathbf{1}(p \geq U_D)(Y_1 - Y_0)) \quad (13) \\ &= E(Y_0) + \underbrace{E(Y_1 - Y_0 \mid p \geq U_D)p}_{S(p)}. \end{aligned}$$

- 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 and the proportion of people whose U_D is at or below p : $E(Y_1 - Y_0 \mid p \geq U_D)p = S(p)$.
- $$E(Y \mid P(Z) = p) = E(Y_0 + \mathbf{1}(p \geq U_D)(Y_1 - Y_0)) \quad (13)$$
$$= E(Y_0) + \underbrace{E(Y_1 - Y_0 \mid p \geq U_D)p}_{S(p)}.$$
- Can identify the left-hand side of (13) for all values of p in the support of $P(Z)$.

- 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 and the proportion of people whose U_D is at or below p : $E(Y_1 - Y_0 \mid p \geq U_D)p = S(p)$.
- $$E(Y \mid P(Z) = p) = E(Y_0 + \mathbf{1}(p \geq U_D)(Y_1 - Y_0)) \quad (13)$$
$$= E(Y_0) + \underbrace{E(Y_1 - Y_0 \mid p \geq U_D)p}_{S(p)}.$$
- Can identify the left-hand side of (13) for all values of p in the support of $P(Z)$.
- 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.

- 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.

- Marginal increment in outcomes is

$$\frac{\partial E(Y \mid P(Z) = p)}{\partial p} = E(Y_1 - Y_0 \mid U_D = p) = \frac{\partial S(p)}{\partial p}. \quad (14)$$

- Marginal increment in outcomes is

$$\frac{\partial E(Y \mid P(Z) = p)}{\partial p} = E(Y_1 - Y_0 \mid U_D = p) = \frac{\partial S(p)}{\partial p}. \quad (14)$$

- The sample analogue of (14) is the local instrumental variable (LIV) estimator of Heckman and Vytlacil (1999, 2005).

- Marginal increment in outcomes is

$$\frac{\partial E(Y \mid P(Z) = p)}{\partial p} = E(Y_1 - Y_0 \mid U_D = p) = \frac{\partial S(p)}{\partial p}. \quad (14)$$

- The sample analogue of (14) is the local instrumental variable (LIV) estimator of Heckman and Vytlacil (1999, 2005).
- Adopting a nonparametric approach to estimating $E(Y \mid P(Z) = p)$ 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) is

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

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

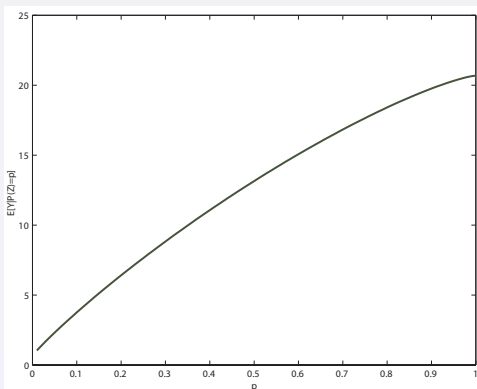
$$\text{MTE}(u_D) \equiv E(Y_1 - Y_0 \mid U_D = u_D).$$

- Expression (13) can be simplified to

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

$$\frac{\partial S(p)}{\partial p} = \text{MTE}(p).$$

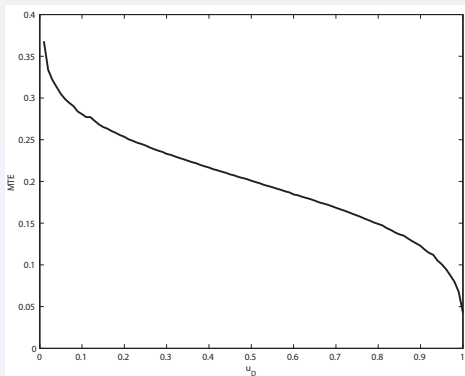
Figure 1: Plots of $E(Y|P(Z) = p)$ and the MTE derived from $E(Y|P(Z) = p)$



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

Source: Heckman and Vytlačil (2005)

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



(b) Plot of $MTE(u_D)$: The derivative of $E(Y | P(Z) = p)$ evaluated at points $p = u_D$

Source: Heckman and Vytlačil (2005)

- 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 go to college, U_D) are induced to go into college ($D = 1$) 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 $P(z)$.

- 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 go to college, U_D) are induced to go into college ($D = 1$) 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.

The Fundamental Role of the Choice Probability in Understanding What Instrumental Variables Estimate When β Depends on D

- For $p_2 > p_1$,

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

note that $\Pr(p_1 \leq U_D \leq p_2) = p_2 - p_1$.

The Fundamental Role of the Choice Probability in Understanding What Instrumental Variables Estimate When β Depends on D

- For $p_2 > p_1$,

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

note that $\Pr(p_1 \leq U_D \leq p_2) = p_2 - p_1$.

- Thus,

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

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

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

- 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 $\Pr(D = 1 \mid Z = \tilde{z}) = p_1$ and $\Pr(D = 1 \mid Z = \tilde{\tilde{z}}) = p_2$.

- The model-generated LATE can be identified if there are values of Z , say $\tilde{z}$ and $\tilde{\tilde{z}}$, such that $\Pr(D = 1 \mid Z = \tilde{z}) = p_1$ and $\Pr(D = 1 \mid Z = \tilde{\tilde{z}}) = p_2$.
- Under standard regularity conditions

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

- 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 $u_{D,M} = 1$,

- 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 $u_{D,M} = 1$,

- $$E(Y \mid U_D \leq u_{D,k}) = E(Y_0) + \sum_{j=1}^k \text{LATE}(u_{D,j}, u_{D,j-1})\eta_j,$$

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

- 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 $u_{D,M} = 1$,

- $$E(Y \mid U_D \leq u_{D,k}) = E(Y_0) + \sum_{j=1}^k \text{LATE}(u_{D,j}, u_{D,j-1})\eta_j,$$

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

- Thus

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

Counterpart to expression (15) when $p = 1$.

- 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 $u_{D,M} = 1$,

- $$E(Y \mid U_D \leq u_{D,k}) = E(Y_0) + \sum_{j=1}^k \text{LATE}(u_{D,j}, u_{D,j-1})\eta_j,$$

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

- Thus

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

Counterpart to expression (15) when $p = 1$.

- It shows how mean income can be represented as a sum of incremental gross surpluses above $E(Y_0)$.

- If $\Pr(D = 1 \mid Z = z)$ 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 $u_{D,\ell} = p_\ell, \quad \ell = 1, \dots, L.$

- $\text{MTE}(u_D)$ 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.

- $MTE(u_D)$ and the model-generated LATE (10) are structural parameters in the sense that changes in Z (conditional on X) do not affect $MTE(u_D)$ or theoretical LATE.
- They are invariant with respect to all policy changes that operate through Z .

- $MTE(u_D)$ and the model-generated LATE (10) are structural parameters in the sense that changes in Z (conditional on X) do not affect $MTE(u_D)$ 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.

- $MTE(u_D)$ and the model-generated LATE (10) are structural parameters in the sense that changes in Z (conditional on X) do not affect $MTE(u_D)$ 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.

- This deeper understanding of LATE facilitates its use in answering out of sample policy questions **P2** and **P3** for policies that operate through changing Z .

- This deeper understanding of LATE facilitates its use in answering out of sample policy questions **P2** and **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 $\Pr(D = 1 \mid Z = z^1) = P(z^1) = p_1$ and $\Pr(D = 1 \mid Z = z^2) = P(z^2) = p_2$, one can use it to evaluate any other pair of policies $\tilde{z}$ and $\tilde{\tilde{z}}$ such that

$$\Pr(D = 1 \mid Z = z^1) = \Pr(D = 1 \mid Z = \tilde{z}) = p_1$$

and

$$\Pr(D = 1 \mid Z = z^2) = \Pr(D = 1 \mid Z = \tilde{\tilde{z}}) = p_2.$$

- 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., we can compare any policy described by $\tilde{z} \in \{z \mid P(z) = p_1\}$ with any policy $\tilde{\tilde{z}} \in \{z \mid P(z) = p_2\}$ and not just the policies associated with z^1 and z^2 that identify the sample LATE.

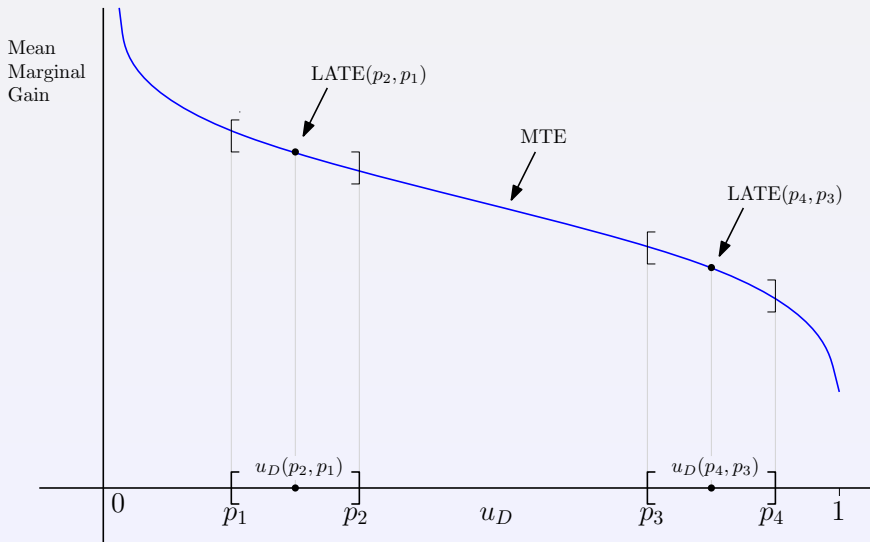
- Thus, one can use an empirical LATE determined for one set of instrument configurations to identify outcomes for other sets of instrument configurations that produce the same p_1 and p_2 , i.e., we can compare any policy described by $\tilde{z} \in \{z \mid P(z) = p_1\}$ with any policy $\tilde{\tilde{z}} \in \{z \mid P(z) = p_2\}$ and not just the policies associated with z^1 and z^2 that identify the sample LATE.
- This is a useful 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 .

- Variation in different components of Z produce variation in $P(Z)$.

- 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 .

- 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: MTE as a function of u_D : What sections of the MTE different values of the instruments and different instruments approximate.



- 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 .

- 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 .
- 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 $P(Z)$).

- If the MTE does not depend on u_D ,
 $E(Y \mid P(Z) = p) = E(Y_0) + (\mu_1 - \mu_0)p$, and all instruments
identify the same parameter: $\bar{\beta} = \mu_1 - \mu_0$.

- If the MTE does not depend on u_D ,
 $E(Y \mid P(Z) = p) = E(Y_0) + (\mu_1 - \mu_0)p$, 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.

- A test of whether $MTE(u_D)$ depends on u_D , or a test of nonlinearity of $E(Y \mid P(Z) = p)$ in p , is a test of the whether different instruments estimate the same parameter.

- A test of whether $MTE(u_D)$ depends on u_D , or a test of nonlinearity of $E(Y \mid P(Z) = p)$ in p , is a test of the whether different instruments estimate the same parameter.
- The LATE model and its extensions overturn the logic of the Durbin (1954)–Wu (1973)–Hausman (1978) test for overidentification.

- A test of whether $MTE(u_D)$ depends on u_D , or a test of nonlinearity of $E(Y \mid P(Z) = p)$ in p , is a test of the whether different instruments estimate the same parameter.
- The LATE model and its extensions overturn the logic of the Durbin (1954)–Wu (1973)–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.

All Treatment Effects Are Weighted Averages of the MTE

- For treatment effect e ,

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

where $h^e(u_D)$ is a weighting function.

- For ATE, $h^e(u_D) = 1$ and $ATE = \int_0^1 MTE(u_D) du_D$.

- For ATE, $h^e(u_D) = 1$ and $ATE = \int_0^1 MTE(u_D) du_D$.
- 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

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

where $u_{D,j} > u_{D,j-1} > u_{D,j-2} > \dots$.

- By the mean value theorem, may express each of the integrals in the sum as

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

for some $u_{j,j-1}^* \in [u_{D,j-1}, u_{D,j})$.

- By the mean value theorem, may express each of the integrals in the sum as

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

for some $u_{j,j-1}^* \in [u_{D,j-1}, u_{D,j})$.

- From (16) 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 may write expression (18) as

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

where $\tilde{h}^e(u_{j,j-1}^*) = h^e(u_{j,j-1}^*)(u_{D,j} - u_{D,j-1})$.

- Thus may write expression (18) as

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

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) is

$$\begin{aligned} & E(Y \mid \text{Alternative Policy } a) - E(Y \mid \text{Baseline Policy } b) \\ &= E(Y^a) - E(Y^b) \\ &= \int_0^1 \text{MTE}(u_D) h^{\text{PRTE}}(u_D) du_D, \end{aligned}$$

- The Policy Relevant Treatment Effect (PRTE) is

$$\begin{aligned} & E(Y \mid \text{Alternative Policy } a) - E(Y \mid \text{Baseline Policy } b) \\ &= E(Y^a) - E(Y^b) \\ &= \int_0^1 \text{MTE}(u_D) h^{\text{PRTE}}(u_D) du_D, \end{aligned}$$

- $h^{\text{PRTE}}(u_D) = F_P^b(u_D) - F_P^a(u_D)$; 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 .

- The Policy Relevant Treatment Effect (PRTE) is

$$\begin{aligned} & E(Y \mid \text{Alternative Policy } a) - E(Y \mid \text{Baseline Policy } b) \\ &= E(Y^a) - E(Y^b) \\ &= \int_0^1 \text{MTE}(u_D) h^{\text{PRTE}}(u_D) du_D, \end{aligned}$$

- $h^{\text{PRTE}}(u_D) = F_P^b(u_D) - F_P^a(u_D)$; 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 .
- 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 .

- The PRTE weights $MTE(u_D)$ by the change in the distribution of the probabilities of participation at different values of u_D .
- 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.

- Notice that the same $MTE(u_D)$ can be used to evaluate the impacts of a variety of different policies.
- It 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.

- Notice that the same $MTE(u_D)$ can be used to evaluate the impacts of a variety of different policies.
- It 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: (a) MTE Weights For Different Treatment Parameter and IVs.
(F_P is the distribution of P . f_P is its density.)

Treatment Parameter Weights

$$h^{ATE}(u_D) = 1$$

$$h^{TT}(u_D) = \left[\int_{u_D}^1 f_P(p) dp \right] \frac{1}{E(P)}$$

$$h^{TUT}(u_D) = \left[\int_0^{u_D} f_P(p) dp \right] \frac{1}{E(1-P)}$$

$$h^{PRTE}(u_D) = F_P^a(u_D) - F_P^b(u_D)$$

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

Table 3: (b) MTE Weights For Different Treatment Parameter and IVs.
 (F_P is the distribution of P . f_P is its density.)

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 Vytlacil (1999, 2005, 2007).

- Figure 3 plots an MTE taken from the analysis of Heckman and Vytlacil (2005) and the weights for MTE associated with ATE, TT, and TUT for a case where β is not independent of D .

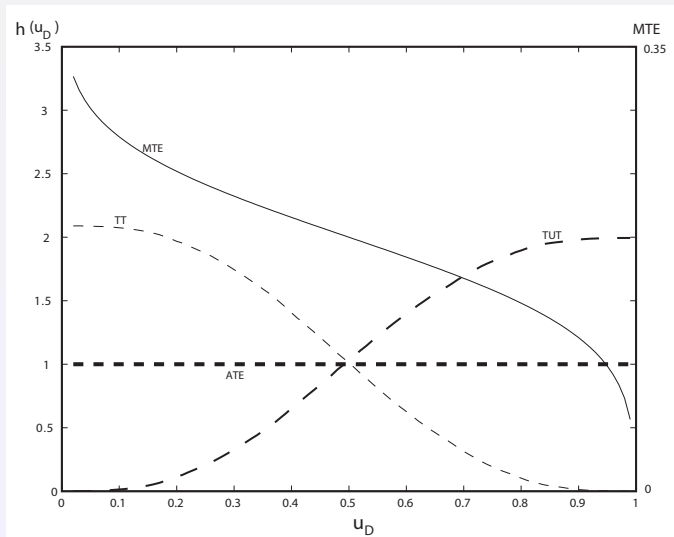
- Figure 3 plots an MTE taken from the analysis of Heckman and Vytlacil (2005) and the weights for MTE associated with ATE, TT, and TUT for a case where β is not independent of D .
- ATE weights u_D evenly.

- Figure 3 plots an MTE taken from the analysis of Heckman and Vytlacil (2005) and the weights for MTE associated with ATE, TT, and TUT for a case where β 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).

- Figure 3 plots an MTE taken from the analysis of Heckman and Vytlacil (2005) and the weights for MTE associated with ATE, TT, and TUT for a case where β 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 .

- Figure 3 plots an MTE taken from the analysis of Heckman and Vytlacil (2005) and the weights for MTE associated with ATE, TT, and TUT for a case where β 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 $\text{MTE}(u_D)$ is decreasing in u_D , $\text{TT} > \text{ATE} > \text{TUT}$.

Figure 3: Weights for the marginal treatment effect for different parameters for the model graphed in Figure 1



Source: Heckman and Vytlacil (2005)

What Does Conventional IV estimate?

- In most empirical studies, more than the two values of Z are used to construct IV estimates.

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 Yitzhaki (1989) to express IV as a weighted average of component LATEs defined for different values of the instruments.

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 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 $\text{MTE}(u_D)$ may change sign over the interval $u_D \in [0, 1]$, the IV may be negative even if some portion of the $\text{MTE}(u_D)$ is positive.

- Since $\text{MTE}(u_D)$ may change sign over the interval $u_D \in [0, 1]$, the IV may be negative even if some portion of the $\text{MTE}(u_D)$ is positive.
- For general instruments that are not monotonic functions of $P(Z)$, the output of IV is even more ambiguous.

- Since $\text{MTE}(u_D)$ may change sign over the interval $u_D \in [0, 1]$, the IV may be negative even if some portion of the $\text{MTE}(u_D)$ 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 $u_D \in [0, 1]$.

- Since $\text{MTE}(u_D)$ may change sign over the interval $u_D \in [0, 1]$, the IV may be negative even if some portion of the $\text{MTE}(u_D)$ 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.

- 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 LATE is positive, in the general case IV can be negative.

- 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 LATE is positive, in the general case IV can be negative.
- Negative components of $MTE(u_D)$ weighted by negative weights can generate a positive IV.

- This analysis is constructive because the weights can be identified from the data.

- 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 .

- 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 .
- They can identify what portion of the u_D distribution a LATE identifies (i.e. it answers the question “which people are affected by the instrument?”)

What Does IV Estimate?

- Consider a linear regression approximation of $E(Y | P(Z) = p)$:

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

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

What Does IV Estimate?

- Consider a linear regression approximation of $E(Y \mid P(Z) = p)$:

$$E^*(Y \mid P(Z) = p) = a + bp,$$

$$b = \frac{\text{Cov}(Y, P(Z))}{\text{Var}(P(Z))} = \frac{\text{Cov}(E(Y \mid P(Z)), P(Z))}{\text{Var}(P(Z))}.$$

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

- Using condition (IA-1) and (15),

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

- When $\text{MTE}(u_D)$ is constant in u_D ($\text{MTE}(u_D) = \mu_1 - \mu_0 = \bar{\beta}$) β is independent of D , the numerator 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 $b = \mu_1 - \mu_0 = \bar{\beta}$.

- When $\text{MTE}(u_D)$ is constant in u_D ($\text{MTE}(u_D) = \mu_1 - \mu_0 = \bar{\beta}$) β is independent of D , the numerator 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 $b = \mu_1 - \mu_0 = \bar{\beta}$.

- Traditional result for IV.

- In this case, the marginal gross surplus is the same as the average gross surplus for all values of p .

- In this case, the marginal gross surplus is the same as the average gross 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.

- In this case, the marginal gross surplus is the same as the average gross 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

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

- 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

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

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

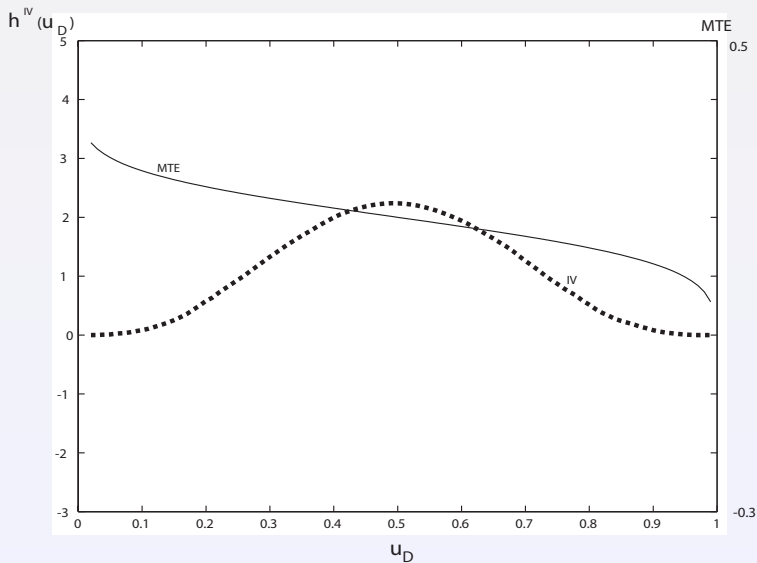
$$h_{P(Z)}^{\text{IV}}(u_D) = \frac{E(P(Z) - E(P(Z)) \mid P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Var}(P(Z))}$$

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

- Weights can be estimated from the sample distribution of $P(Z)$.

- Weights can be estimated from the sample distribution of $P(Z)$.
- Weights for $P(Z)$ as an instrument have a distinctive profile.

Figure 4: IV Weights as a Function of u_D .



Source: Heckman and Vytlacil (2005).

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

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

where $\lambda_{\ell} = \frac{1}{\text{Var}(P(Z))} \sum_{t>\ell}^L (p_t - E(P)) f_P(p_t)$ and $f_P(p_t)$ is the probability that $P(Z) = p_t$.

The Problem of Limited Support

- While the various treatment parameters can be defined from the generalized Roy model, they may not necessarily be identified from the data.

The Problem of Limited Support

- While the various treatment parameters can be defined from the generalized Roy model, they may not necessarily be identified from the data.
- $P(Z)$ may not be identified over the full unit interval.

- $P(Z)$ may only assume discrete values.

- $P(Z)$ may only assume discrete values.
- This limits the identifiability of MTE.

- $P(Z)$ 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 of $P(Z) = P(z)$ associated with the discrete instruments.

- One approach to this problem developed by Manski (1990, 1995, 2003) is to produce bounds on the treatment effects.

- One approach to this problem developed by Manski (1990, 1995, 2003) is to produce bounds on the treatment effects.
- Heckman and Vytlacil (1999, 2000, 2001a,b, 2007) develop specific bounds for the generalized Roy model that underlies the LATE model.

- One approach to this problem developed by Manski (1990, 1995, 2003) is to produce bounds on the treatment effects.
- Heckman and Vytlacil (1999, 2000, 2001a,b, 2007) 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 Vytlacil (2010) consider an alternative approach based on marginal policy changes.

- Carneiro, Heckman, and Vytlacil (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.

- Carneiro, Heckman, and Vytlacil (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).

- MP RTE 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.

- MP RTE 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.
- 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.

More General Instruments

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

More General Instruments

- Typically, economists use a variety of instruments one at a time and not just $P(Z)$ as an instrument and compare the resulting estimates (see, e.g, Card, 1999, 2001).
- When there is selection on the basis of gross gains ($\beta \not\propto D$) so that the marginal gross surplus is not the same as the average gross surplus, different instruments identify different parameters.

More General Instruments

- Typically, economists use a variety of instruments one at a time and not just $P(Z)$ as an instrument and compare the resulting estimates (see, e.g, Card, 1999, 2001).
- When there is selection on the basis of gross gains ($\beta \not\propto D$) 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.

More General Instruments

- Typically, economists use a variety of instruments one at a time and not just $P(Z)$ as an instrument and compare the resulting estimates (see, e.g, Card, 1999, 2001).
- When there is selection on the basis of gross gains ($\beta \not\propto D$) 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).

- 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 ($Z = (Z_1, \dots, Z_K)$, $K \geq 2$).

- 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 $P(Z)$.

- 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

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

- The IV estimator based on Z_1 is

$$IV_{Z_1} = \frac{\text{Cov}(Y, Z_1)}{\text{Cov}(D, Z_1)} = \frac{\text{Cov}(E(Y \mid Z_1), Z_1)}{\text{Cov}(D, Z_1)}.$$

- Note, however, that choices (and hence Y) are generated by the *full vector of Z* operating through $P(Z)$.

- The IV estimator based on Z_1 is

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

- 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 .

- 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 $\beta (= Y_1 - Y_0)$.

- 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.

- A linear regression analogy helps to fix ideas.
- Suppose that outcome Q can be expressed 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 $E(\varepsilon \mid W) = 0$.

- 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

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

where ϕ_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_ℓ , $\ell \neq 1$).

- 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

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

where ϕ_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_{ℓ} , $\ell \neq 1$).

- 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.

- 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 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.

- 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 it covaries with tuition in the sample used by the analyst.

- 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 it covaries with tuition in the sample used by the analyst.
- This is not the *ceteris paribus* effect of a variation in tuition.

- 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 it 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 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 it 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 β , the traditional instrumental variable argument that analysts do not need a model for D and can ignore other possible determinants of D besides 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 $P(Z)$.

- 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 will disagree about the interpretation of the margin identified by variation in Z_1 .

- To establish these points, recall that as a consequence of Vytlačil's theorem, Z enters the distribution of Y only through $P(Z)$.

- To establish these points, recall 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 $P(Z)$.

- To establish these points, recall that as a consequence of Vytlacil'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 Vytlacil's theorem.

- To establish these points, recall that as a consequence of Vytlacil'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 Vytlacil's theorem.
- Thus

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

where $g_{P(Z)|Z_1}(p, z_1)$ is the conditional density of $P(Z)$ given $Z_1 = z_1$.

- Putting all of these ingredients together, using (15) we obtain

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

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

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

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

$$IV_{Z_1} = \frac{\int_{-\infty}^{\infty} (z_1 - E(Z_1)) \int_0^1 S(p) g_{P(Z)|Z_1}(p, z_1) dp dz_1}{\text{Cov}(Z_1, D)}$$

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

- 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

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

- 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

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

- The weight integrates to 1 but can be negative over stretches of u_D .

- 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

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

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

- An illuminating way to represent this weight is

$$h_{Z_1}^{IV}(u_D) = \frac{E(Z_1 - E(Z_1) \mid P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Cov}(Z_1, D)}.$$

- An illuminating way to represent this weight is

$$h_{Z_1}^{IV}(u_D) = \frac{E(Z_1 - E(Z_1) \mid P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Cov}(Z_1, D)}.$$

- 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.

- An illuminating way to represent this weight is

$$h_{Z_1}^{IV}(u_D) = \frac{E(Z_1 - E(Z_1) \mid P(Z) > u_D) \Pr(P(Z) > u_D)}{\text{Cov}(Z_1, D)}.$$

- 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) for each value of u_D .

Example

Figure 5: Joint Distribution of Instruments $Z = (Z_1, Z_2)$

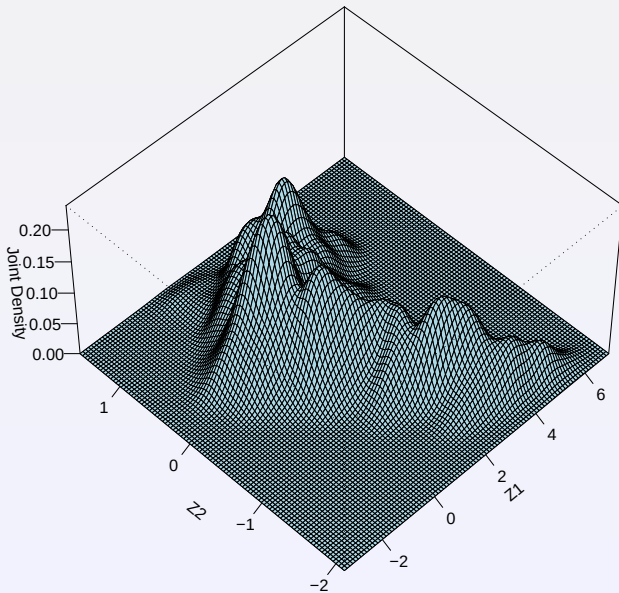
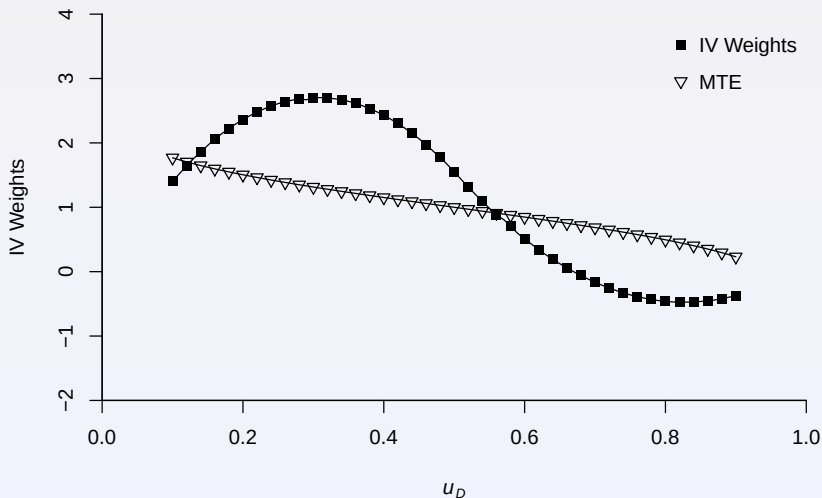


Figure 6: MTE and IV weights for a general instrument Z_1 , a component of $Z = (Z_1, Z_2)$.



- Table 4, taken from Heckman, Urzua, and Vytlacil (2006) shows how three different distributions of Z for the same underlying policy-invariant model with the same ATE can produce very different IV estimates.

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).

How Experiments Improve on LATE

- 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 .

How Experiments Improve on LATE

- 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.

- 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 $P(Z)$.

- 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 those effects directly identified by the experiment.

- Using the implicit economic theory underlying LATE, economists can do better than just report an IV estimate.

- 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.

- 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 $P(Z)$.

- 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.

- 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.

- 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.
- 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).

Policy Effects, Treatment Effects, and IV

- A main lesson of this analysis is that policy effects are not generally the same as treatment effects and, in general, neither are produced from IV estimators.

Policy Effects, Treatment Effects, and IV

- A main lesson of this analysis 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.

Policy Effects, Treatment Effects, and IV

- A main lesson of this analysis 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.

Policy Effects, Treatment Effects, and IV

- A main lesson of this analysis 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 these questions.

- Policy questions and not statistical methods drive analyses.

- Policy questions and not statistical methods drive analyses.
- Well-posed economic models are scarce in the program evaluation approach.

- 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 methods over economic content.

- 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 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.

- 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 questions that arise in addressing policy problems.

- 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 questions that arise in addressing policy problems.
- Constructing the PRTE is an example of this approach.

- 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 questions that arise in addressing policy problems.
- Constructing the PRTE is an example of this approach.
- An alternative approach developed in Heckman and Vytlačil (2005) constructs combinations of instruments using sample data on Z that address specific policy questions.

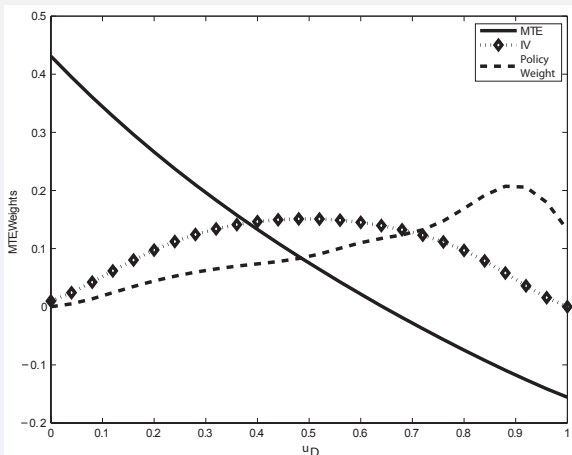
- Figure 7, taken from an analysis of the returns to attending college by Carneiro, Heckman, and Vytlačil (2009), plots the estimated weights for MTE from a marginal change in policy that proportionally expands the probability of attending college for everyone.

- Figure 7, taken from an analysis of the returns to attending college by Carneiro, Heckman, and Vytlacil (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.

- Figure 7, taken from an analysis of the returns to attending college by Carneiro, Heckman, and Vytlacil (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.

- Figure 7, taken from an analysis of the returns to attending college by Carneiro, Heckman, and Vytlacil (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.

Figure 7: Weights for IV and MP RTE in the Carneiro-Heckman-Vytlacil (2009) Analysis of the Returns to College Going.



Source: Carneiro et al. (2009).

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

Multiple Choices

- Imbens and Angrist analyze a two choice model.

Multiple Choices

- Imbens and Angrist analyze a two choice model.
- Heckman, Urzua, and Vytlacil (2006, 2010) and Heckman and Vytlacil (2007) extend their analysis to an ordered choice model and to general unordered choice models.

- 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, 2010; Heckman and Vytlačil, 2007).

- 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, 2010; Heckman and Vytlačil, 2007).
- If, however, one is interested in identifying the mean returns to any pair of outcomes, unaided IV will not do the job.

- 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 Vytlacil, 2006, 2010; Heckman and Vytlacil, 2007).
- 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.

- 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.

- 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.

- 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.

- 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.

Conclusions

- This paper compares the structural approach to empirical policy analysis with the program evaluation approach.

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.

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.

- This paper advocates placing the economic and policy questions being addressed front and center.

- This paper advocates placing the economic and policy questions being addressed front and center.
- Economic theory helps to sharpen statements of policy questions.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

- 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.

Appendix

Example: The Generalized Roy Model for the Normal Case

$$Y_1 = \mu_1(X) + U_1$$

$$Y_0 = \mu_0(X) + U_0$$

$$C = \mu_C(Z) + U_C$$

Net Benefit: $I = Y_1 - Y_0 - C$

$$I = \underbrace{\mu_1(X) - \mu_0(X) - \mu_C(Z)}_{\mu_D(Z)} + \underbrace{U_1 - U_0 - U_C}_{-V}$$

$$(U_0, U_1, U_C) \perp\!\!\!\perp (X, Z)$$

$$E(U_0, U_1, U_C) = (0, 0, 0)$$

$$V \perp\!\!\!\perp (X, Z)$$

- Assume Normally Distributed Errors.

- Assume Normally Distributed Errors.
- Assume Z contains X but may contain other variables (exclusions)

$$Y = DY_1 + (1 - D)Y_0 \quad \text{observed } Y$$
$$D = 1(I \geq 0) = 1(\mu_D(Z) \geq V)$$

- Assume Normally Distributed Errors.
- Assume Z contains X but may contain other variables (exclusions)

$$Y = DY_1 + (1 - D)Y_0 \quad \text{observed } Y$$
$$D = 1(I \geq 0) = 1(\mu_D(Z) \geq V)$$

- Assume $V \sim N(0, \sigma_V^2)$

- Propensity Score:

$$\Pr(D = 1 \mid Z = z) = \Phi \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

$$E(Y \mid D = 1, X = x, Z = z) = \mu_1(X) + \underbrace{E(U_1 \mid \mu_D(z) \geq V)}_{K_1(P(z))}$$

because $(X, Z) \perp\!\!\!\perp (U_1, V)$.

- Under normality we obtain

$$E \left(U_1 \mid \frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V} \right) = \frac{\text{Cov}(U_1, \frac{V}{\sigma_V})}{\text{Var}(\frac{V}{\sigma_V})} \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

- Why?

$$U_1 = \text{Cov} \left(U_1, \frac{V}{\sigma_V} \right) \frac{V}{\sigma_V} + \varepsilon_1$$

$$\varepsilon_1 \perp\!\!\!\perp V$$

$$\begin{aligned} E \left(\frac{V}{\sigma_V} \mid \frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V} \right) &= \frac{\int_{-\infty}^{\frac{\mu_D(z)}{\sigma_V}} t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt}{\int_{-\infty}^{\frac{\mu_D(z)}{\sigma_V}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt} = \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) \\ &= \frac{\frac{-1}{\sqrt{2\pi}} e^{(-\frac{1}{2}) \left(\frac{\mu_D(z)}{\sigma_V} \right)^2}}{\Phi \left(\frac{\mu_D(z)}{\sigma_V} \right)} = \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) = \frac{-\phi \left(\frac{\mu_D(z)}{\sigma_V} \right)}{\Phi \left(\frac{\mu_D(z)}{\sigma_V} \right)} \end{aligned}$$

- Notice

$$\lim_{\mu_D(z) \rightarrow \infty} \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) = 0$$

$$\lim_{\mu_D(z) \rightarrow -\infty} \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) = -\infty$$

- Notice

$$\lim_{\mu_D(z) \rightarrow \infty} \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) = 0$$
$$\lim_{\mu_D(z) \rightarrow -\infty} \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right) = -\infty$$

- Propensity score:

$$P(z) = \Pr(D = 1 \mid Z = z) = \Phi \left(\frac{\mu_D(z)}{\sigma_V} \right)$$
$$\therefore \left(\frac{\mu_D(z)}{\sigma_V} \right) = \Phi^{-1} (\Pr(D = 1 \mid Z = z))$$

- Thus we can replace $\frac{\mu_D(z)}{\sigma_V}$ with a known function of $P(z)$

- Notice that because $(X, Z) \perp\!\!\!\perp (U, V)$, Z enters the model (conditional on X) only through $P(Z)$.

- Notice that because $(X, Z) \perp\!\!\!\perp (U, V)$, Z enters the model (conditional on X) only through $P(Z)$.
- This is called index sufficiency.

- Put all of these results together to obtain

$$E(Y | D = 1, X = x, Z = z) = \mu_1(x) + \left(\frac{\text{Cov}(U_1, \frac{V}{\sigma_V})}{\text{Var}(\frac{V}{\sigma_V})} \right) \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

$$= E(Y_1 | D = 1, X = x, Z = z) = \mu_1(x) + \left(\frac{\text{Cov}(U_1, \frac{V}{\sigma_V})}{\text{Var}(\frac{V}{\sigma_V})} \right) \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

$$\tilde{\lambda}(z) = E \left(\frac{V}{\sigma_V} \mid \frac{V}{\sigma_V} < \frac{\mu_D(z)}{\sigma_V} \right) < 0$$

$$\lambda(z) = E \left(\frac{V}{\sigma_V} \mid \frac{V}{\sigma_V} \geq \frac{\mu_D(z)}{\sigma_V} \right) > 0$$

$$E(Y | D = 0, X = x, Z = z) = \mu_0(x) + \left(\frac{\text{Cov}(U_0, \frac{V}{\sigma_V})}{\text{Var}(\frac{V}{\sigma_V})} \right) \lambda \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

$$\text{Var} \left(\frac{V}{\sigma_V} \right) = 1$$

$$\frac{V}{\sigma_V} = -\frac{(U_1 - U_0 - U_C)}{\sigma_V}$$

$$\text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) = -\text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) + \text{Cov}\left(U_0, \frac{V}{\sigma_V}\right) + \text{Cov}\left(U_C, \frac{V}{\sigma_V}\right)$$

In Roy model case ($U_C = 0$),

$$\begin{aligned}\text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) &= -\text{Cov}\left(U_1, \frac{U_1 - U_0}{\sigma_V}\right) \\ &= -\frac{\text{Cov}(U_1 - U_0, U_1)}{\sqrt{\text{Var}(U_1 - U_0)}}\end{aligned}$$

- We can identify $\mu_1(x), \mu_0(x)$

- We can identify $\mu_1(x), \mu_0(x)$
- From Discrete Choice model we can identify

$$\frac{\mu_D(z)}{\sigma_V} = \frac{\mu_1(x) - \mu_0(x) - \mu_C(z)}{\sigma_V}$$

- We can identify $\mu_1(x), \mu_0(x)$
- From Discrete Choice model we can identify

$$\frac{\mu_D(z)}{\sigma_V} = \frac{\mu_1(x) - \mu_0(x) - \mu_C(z)}{\sigma_V}$$

- If we have a regressor in X that does not affect $\mu_C(z)$ (say regressor x_j , so $\frac{\partial \mu_C(z)}{\partial x_j} = 0$), we can identify σ_V and $\mu_C(z)$.

- We can identify $\mu_1(x), \mu_0(x)$
- From Discrete Choice model we can identify

$$\frac{\mu_D(z)}{\sigma_V} = \frac{\mu_1(x) - \mu_0(x) - \mu_C(z)}{\sigma_V}$$

- If we have a regressor in X that does not affect $\mu_C(z)$ (say regressor x_j , so $\frac{\partial \mu_C(z)}{\partial x_j} = 0$), we can identify σ_V and $\mu_C(z)$.
- $\therefore$ We can identify the net benefit function and the cost function up to scale.

- We can identify $\mu_1(x), \mu_0(x)$
- From Discrete Choice model we can identify

$$\frac{\mu_D(z)}{\sigma_V} = \frac{\mu_1(x) - \mu_0(x) - \mu_C(z)}{\sigma_V}$$

- If we have a regressor in X that does not affect $\mu_C(z)$ (say regressor x_j , so $\frac{\partial \mu_C(z)}{\partial x_j} = 0$), we can identify σ_V and $\mu_C(z)$.
- $\therefore$ We can identify the net benefit function and the cost function up to scale.
- $\therefore$ We can compute *ex ante* subjective net gains.

- Method generalizes:
Don't need normality

- Method generalizes:
Don't need normality
- Need “Identification at infinity” to identify ATE and TT

- Method generalizes:
Don't need normality
- Need “Identification at infinity” to identify ATE and TT
-

$$E(Y \mid D = 1, X = x, Z = z) = \mu_1(x) + \overbrace{K_1(P(z))}^{\text{control function}}$$

$$E(Y \mid D = 0, X = x, Z = z) = \mu_0(x) + \underbrace{K_0(P(z))}_{\text{control function}}$$

$$\lim_{P(z) \rightarrow 1} E(Y \mid D = 1, X = x, Z = z) = \mu_1(x)$$

$$\lim_{P(z) \rightarrow 0} E(Y \mid D = 0, X = x, Z = z) = \mu_0(x)$$

- If we have this condition satisfied, we can identify ATE

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- If we have this condition satisfied, we can identify ATE

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- ATE is defined in a limit set. This is true for any model with selection on unobservables (IV; selection models)

- What about treatment on the treated?

$$E(Y_1 - Y_0 \mid D = 1, X = x, Z = z)$$

(a) From the data, we observe

$$E(Y_1 \mid D = 1, X = x, Z = z)$$

(b) Can also create it from the model

(c) $E(Y_0 \mid D = 1, X = x, Z = z)$ is a counterfactual

We know

$$E(Y_0 \mid D = 0, X = x, Z = z) = \mu_0(x) + \text{Cov}\left(U_0, \frac{V}{\sigma_V}\right) \lambda\left(\frac{\mu_D(Z)}{\sigma_V}\right)$$

(this is data)

(d) We seek

$$E(Y_0 \mid D = 1, X = x, Z = z) = \mu_0(x) + \text{Cov} \left(U_0, \frac{V}{\sigma_V} \right) \tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right)$$

But under normality, we know $\text{Cov} \left(U_0, \frac{V}{\sigma_V} \right)$

We know $\frac{\mu_D(z)}{\sigma_V}$

$\tilde{\lambda}(\cdot)$ is a known function.

Can form $\tilde{\lambda} \left(\frac{\mu_D(z)}{\sigma_V} \right)$ and can construct counterfactual.

- More generally, without normality (but with $(X, Z) \perp\!\!\!\perp (U, V)$)

$$E(Y_1 \mid D = 1, X, Z) = E(Y \mid D = 1, X = x, Z = z) = \mu_1(x) + K_1(P(z))$$

$$E(Y_0 \mid D = 0, X, Z) = E(Y \mid D = 0, X = x, Z = z) = \mu_0(x) + \tilde{K}_0(P(z))$$

$$\text{where } K_1(P(z)) = E(U_1 \mid D = 1, X = x, Z = z) = E\left(U_1 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$$

$$\tilde{K}_1(P(z)) = E\left(U_1 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$$

$$\tilde{K}_0(P(z)) = E\left(U_0 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$$

- Use the transformation

$$\frac{F_V}{\sigma_V} \left(\frac{\mu_D(z)}{\sigma_V} \right) = P(z)$$

$$\frac{F_V}{\sigma_V} \left(\frac{V}{\sigma_V} \right) = U_D \quad (\text{a uniform random variable})$$

$$D = 1 \left(\frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V} \right) = 1 (P(z) \geq U_D)$$

$$K_1(P(z)) = E(U_1 \mid P(z) > U_D)$$

$$K_1(P(z))P(z) + \tilde{K}_1(P(z))(1 - P(z)) = 0$$

$\therefore$ we can construct $\tilde{K}_1(P(z))$

- Symmetrically

$$\tilde{K}_0(P(z)) = E(U_0 \mid P(z) \leq U_D)$$

$$K_0(P(z)) = E(U_0 \mid P(z) > U_D)$$

$$(1 - P(z))\tilde{K}_0(P(z)) + P(z)K_0(P(z)) = 0$$

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) = \\ = \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) &= \\ &= \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- But we can form $\mu_1(x) + K_1(P(z))$ from data

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) = \\ = \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- But we can form $\mu_1(x) + K_1(P(z))$ from data
- We get $\mu_0(x)$ from limit set $P(z) \rightarrow 0$ identifies $\mu_0(x)$

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) = \\ = \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- But we can form $\mu_1(x) + K_1(P(z))$ from data
- We get $\mu_0(x)$ from limit set $P(z) \rightarrow 0$ identifies $\mu_0(x)$
- We can form $K_0(P(z)) = -\tilde{K}_0(P(z)) \frac{P(z)}{1-P(z)}$

- $\therefore$ If we have “identification at infinity,” we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) = \\ = \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- But we can form $\mu_1(x) + K_1(P(z))$ from data
- We get $\mu_0(x)$ from limit set $P(z) \rightarrow 0$ identifies $\mu_0(x)$
- We can form $K_0(P(z)) = -\tilde{K}_0(P(z)) \frac{P(z)}{1-P(z)}$
- $\therefore$ Can construct the desired counterfactual mean.

- Notice how we can get Effect of Treatment for People at the Margin:

$$E(Y_1 - Y_0 \mid I = 0, X = x, Z = z)$$

- Notice how we can get Effect of Treatment for People at the Margin:

$$E(Y_1 - Y_0 \mid I = 0, X = x, Z = z)$$

- Under normality we have (as a result of independence and normality)

$$\begin{aligned} E(Y_1 - Y_0 \mid I = 0, X = x, Z = z) \\ &= \mu_1(x) - \mu_0(x) + E\left(U_1 - U_0 \mid \frac{\mu_D(z)}{\sigma_V} = \frac{V}{\sigma_V}, X = x, Z = z\right) \\ &= \mu_1(x) - \mu_0(x) + \text{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right) \frac{\mu_D(z)}{\sigma_V} \end{aligned}$$

In the Roy model case where $U_C = 0$ but $\mu_C(z) \neq 0$

$$\begin{aligned} &= \mu_1(x) - \mu_0(x) - \sigma_V \left(\frac{\mu_D(z)}{\sigma_V} \right) \\ &= \mu_1(x) - \mu_0(x) - \mu_D(z) \\ &= \mu_C(z) \end{aligned}$$

(marginal gain = marginal cost)

- MTE is

$$\begin{aligned} E(Y_1 - Y_0 \mid V = v, X = x, Z = z) = \\ = \mu_1(x) - \mu_0(x) + \text{Cov} \left(U_1 - U_0, \frac{V}{\sigma_V} \right) v \end{aligned}$$

- MTE is

$$\begin{aligned} E(Y_1 - Y_0 \mid V = v, X = x, Z = z) = \\ = \mu_1(x) - \mu_0(x) + \text{Cov} \left(U_1 - U_0, \frac{V}{\sigma_V} \right) v \end{aligned}$$

- Effect of Treatment for People at the Margin picks $v = \frac{\mu_D(z)}{\sigma_V}$

- MTE is

$$\begin{aligned} E(Y_1 - Y_0 \mid V = v, X = x, Z = z) = \\ = \mu_1(x) - \mu_0(x) + \text{Cov} \left(U_1 - U_0, \frac{V}{\sigma_V} \right) v \end{aligned}$$

- Effect of Treatment for People at the Margin picks $v = \frac{\mu_D(z)}{\sigma_V}$
- Notice we can use the result that

$$\begin{aligned} \frac{\mu_D(z)}{\sigma_V} &= F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(P(z)) \\ V &= F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(U_D) \end{aligned}$$

- Effect of Treatment for People at Margin of Indifference Between Taking Treatment and Not:

$$\begin{aligned} E(Y_1 - Y_0 \mid I = 0, X = x, Z = z) &= \\ &= \mu_1(x) - \mu_0(x) + \text{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right) F_{\left(\frac{V}{\sigma_V}\right)}^{-1}(P(z)) \end{aligned}$$

- Effect of Treatment for People at Margin of Indifference Between Taking Treatment and Not:

$$\begin{aligned} E(Y_1 - Y_0 \mid I = 0, X = x, Z = z) &= \\ &= \mu_1(x) - \mu_0(x) + \text{Cov} \left(U_1 - U_0, \frac{V}{\sigma_V} \right) F_{\left(\frac{V}{\sigma_V}\right)}^{-1}(P(z)) \end{aligned}$$

- MTE:

$$\begin{aligned} E(Y_1 - Y_0 \mid U_D = u_D, X = x, Z = z) &= \\ &= \mu_1(x) - \mu_0(x) + \text{Cov} \left(U_1 - U_0, \frac{V}{\sigma_V} \right) F_{\left(\frac{V}{\sigma_V}\right)}^{-1}(U_D) \end{aligned}$$

- Recent Advances in Econometrics:

- Recent Advances in Econometrics:
 - (a) Relax normality

- Recent Advances in Econometrics:

- (a) Relax normality

- (b) Do not assume linearity of $\mu_1(X)$ and $\mu_0(X)$ in terms of X

- Recent Advances in Econometrics:

- (a) Relax normality
- (b) Do not assume linearity of $\mu_1(X)$ and $\mu_0(X)$ in terms of X
- (c) Do not require identification at infinity but only because they abandon pursuit of ATE, TT, TUT or else assume that $(Y_1, Y_0) \perp\!\!\!\perp D \mid X$ (matching assumption)

- Recent Advances in Econometrics:

- (a) Relax normality
- (b) Do not assume linearity of $\mu_1(X)$ and $\mu_0(X)$ in terms of X
- (c) Do not require identification at infinity but only because they abandon pursuit of ATE, TT, TUT or else assume that $(Y_1, Y_0) \perp\!\!\!\perp D \mid X$ (matching assumption)
- (d) Identification at infinity in some version or the other is required for ATE, TT, TUT as long as there is selection on unobservables (i.e., $(Y_1, Y_0) \not\perp\!\!\!\perp D \mid X$)

- Recent Advances in Econometrics:

- (a) Relax normality
- (b) Do not assume linearity of $\mu_1(X)$ and $\mu_0(X)$ in terms of X
- (c) Do not require identification at infinity but only because they abandon pursuit of ATE, TT, TUT or else assume that $(Y_1, Y_0) \perp\!\!\!\perp D \mid X$ (matching assumption)
- (d) Identification at infinity in some version or the other is required for ATE, TT, TUT as long as there is selection on unobservables (i.e., $(Y_1, Y_0) \not\perp\!\!\!\perp D \mid X$)

- End of Example of Normal Model**