The Role Individual Heterogeneity in Treatment Effect Analysis

Edward Vytlacil, New York University

drawing on joint work with co-authors
including James Heckman, University of Chicago

2013 Annual Health Econometrics Workshop
October, 2013

Outline

Outline:

- Notation
- Treatment effects,
  essential heterogeneity in treatment effects.
- Evaluation Problem: Selection and sorting.
- Instrumental Variables under essential heterogeneity
- Selection models with essential heterogeneity.
- Estimation, feasibility, and practicality

Counterfactual Notation

Notation:

- D_i dummy variable for treatment for individual i, $D_i = 1$ if treated, $= 0$ otherwise.
- Y_{1i} potential outcome for individual i if treated, what would be observed if treated.
- Y_{0i} potential outcome for individual i if not treated, what would be observed if not treated.
- Observed outcome for individual i:

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \begin{cases} Y_{1i} & \text{if } D_i = 1, \\ Y_{0i} & \text{if } D_i = 0. \end{cases}$$

Counterfactual Notation (cont'd)

Define:

- X_i controls, called confounders within biostatistics, observed variables that directly affect Y_{0i}, Y_{1i}.
- Z_i, instruments, variables that affect D_i but do not directly affect Y_{0i}, Y_{1i}.
- Potential treatments

$$\{D_i(z) : z \in \mathcal{Z}\},$$

treatment choice that i would have chosen if randomly assigned $Z_i = z$, where $\mathcal{Z}$ is set of possible instrument values.

Treatment Effects and Heterogeneity

Let $\Delta_i = Y_{1i} - Y_{0i}$, treatment effect for individual i.

Central issue: how does Δ_i vary with i?

- Homogenous treatment effect:
  Treatment effects Δ_i constant across individuals (conditional on X_i)
- Heterogeneous treatment effects:
  Treatment effects Δ_i vary across individuals (even conditional on X_i)

Allowing for heterogeneous treatment effects fundamentally changes the evaluation problem if there is "Essential Heterogeneity" (Heckman, Vytlacil and Urzua, 2006).

Treatment Effects and Heterogeneity

Essential Heterogeneity
(Heckman, Vytlacil and Urzua, 2006):

- $Y_{1i} - Y_{0i} \not\perp D_i | X_i$.
- Agents select into treatment based, in part, on their own idiosyncratic effect.

Allowing for Essential Heterogeneity fundamentally changes the evaluation problem, raising questions as to what is the parameter of interest, complicating identification analysis, and changing the interpretation of results.

Evaluation Problem with Homogeneous Treatment Effect: Selection Bias

Evaluation Problem (Homogeneous Treatment Effects)

Suppose Δ a constant.

Parameter of interest: $\Delta = Y_{1i} - Y_{0i}$.

Classical evaluation problem: Selection Bias

$$\begin{aligned} \Delta_i = \Delta \Rightarrow Y_i &= Y_{0i} + D_i\Delta \\ &\Rightarrow E(Y_i|D_i = 1) - E(Y_i|D_i = 0) \\ &\quad = \Delta + \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\text{Selection Bias}}. \end{aligned}$$

Same analysis conditional on X_i if treatment effect constant conditional on X_i.

Evaluation Problem (Homogeneous Treatment Effect): Selection Bias

$E(Y_0|D=1) - E(Y_0|D=0)$ is selection bias:

- Selection on the base state

  if treated had not received treatment, would they have similar outcomes as the non treated?

- Sometimes called "Ability Bias" in labor economics.
- Common worry: omitted variable (e.g., ability), omitted variable correlated with selection into treatment.
- Common solution: instrumental variables.

Evaluation Problem with Heterogeneous Treatment Effects: Parameters

Allow Essential Heterogeneity.
Δ_i random, possibly correlated with treatment choice.

What is parameter of interest?

Most often, consider average treatment parameters:

- Average Treatment Effect, ATE $= E(Y_{1i} - Y_{0i})$,
- Treatment on the Treated, TT$= E(Y_{1i} - Y_{0i}|D_i = 1)$,
- Treatment on the Untreated, TUT$= E(Y_{1i} - Y_{0i}|D_i = 0)$.

Can also consider IV-defined parameters (e.g., LATE, Imbens and Angrist 1994), Policy Relevant Treatment Effect (Heckman and Vytlacil, 2001), Person Centered Treatment Effects (Basu, 2013), etc

Evaluation Problem with Heterogeneous Treatment Effects: Parameters (Cont'd)

Average treatment parameters:

- Average Treatment Effect, ATE $= E(Y_{1i} - Y_{0i})$.
- Treatment on the Treated, TT$= E(Y_{1i} - Y_{0i}|D_i = 1)$.
- Treatment on the Untreated, TUT$= E(Y_{1i} - Y_{0i}|D_i = 0)$.

If treatment effects are heterogeneous without Essential Heterogeneity ($Y_{1i} - Y_{0i} \perp\!\!\!\perp D_i|X_i$), then all of these mean parameters coincide (conditional on X).

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain

Allow Heterogeneous Treatment effect.

$$Y = Y_0 + D(Y_1 - Y_0).$$

$$\begin{aligned}
&E(Y|D=1) - E(Y|D=0) \\
&= \underbrace{E(Y_1 - Y_0|D=1)}_{\text{TT}} + \underbrace{E(Y_0|D=1) - E(Y_0|D=0)}_{\text{Selection Bias}} \\
&= E(Y_1 - Y_0) + \left\{ \begin{array}{c} E(Y_1 - Y_0|D=1) \\ -E(Y_1 - Y_0) \end{array} \right\} + \left\{ \begin{array}{c} E(Y_0|D=1) \\ -E(Y_0|D=0) \end{array} \right\} \\
&= \quad \text{ATE} \quad + \quad \text{Sorting Gain} \quad + \quad \text{Selection Bias}
\end{aligned}$$

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain (cont'd)

With heterogeneous effects, bias depends on parameter of interest.

- For TT, bias is selection bias, as before.
- For ATE, additional bias term: sorting gain
  - selection on the gain, benefit to those who sort into treatment versus average person.
  - Expect nonzero under essential heterogeneity
  - Positive for Roy model.
- If effects are heterogeneous but without essential heterogeneity, than analysis is the same as for homogeneous case, sorting gain is zero.
- Classical IV results do not hold

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain (cont'd)

When considering alternative methods to evaluate effects of a treatment, important to consider:

1. Essential heterogeneity?
2. What is parameter of interest?
3. What is bias of method for particular parameter?

Potential "Solution": Instrumental Variables

- Suppose for instrument Z:
  1. $\text{Cov}(D, Z) \neq 0$ (Instrument Relevance),
  2. $Z \perp\!\!\!\perp Y_0, Y_1$ (Instrument Exogeneity).
- Probability limit of IV:

$$\text{plim} IV = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)}.$$

- Will $\text{plim} IV$ correspond to an object of interest?

Potential "Solution": Instrumental Variables

$$\begin{aligned} Y &= Y_0 + D(Y_1 - Y_0) \\ &= E[Y_0] + DE[Y_1 - Y_0] + \{\varepsilon + \eta D\}, \end{aligned}$$

where

$$\begin{aligned} \varepsilon &= Y_0 - E[Y_0] \\ \eta &= (Y_1 - Y_0) - E(Y_1 - Y_0). \end{aligned}$$

Need Z to be uncorrelated with $[\varepsilon + \eta D]$ to use IV identify $E(Y_1 - Y_0)$.

Potential "Solution": Instrumental Variables

$$\begin{aligned} \varepsilon &= [Y_0 - E(Y_0)] \\ \eta &= (Y_1 - Y_0) - E(Y_1 - Y_0). \end{aligned}$$

Need Z to be uncorrelated with $[\varepsilon + \eta D]$ to use IV identify $E(Y_1 - Y_0)$.

$$Z \perp\!\!\!\perp Y_0, Y_1 \Rightarrow \begin{array}{c} \text{Cov}(Z, \varepsilon) = 0 \\ \text{Cov}(Z, \eta) = 0, \end{array}$$

$$\nRightarrow \text{Cov}(Z, \eta D) = 0.$$

IV when Δ is a constant

Suppose Δ is a constant $\Rightarrow \ \eta = 0$.

- $\text{Cov}(Z, \varepsilon + \eta D) = \text{Cov}(Z, \varepsilon) = 0$ and thus

$$\text{plim IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, D)} = E(Y_1 - Y_0).$$

- If other instruments exist, each identifies the same parameter.
- No restriction is needed on selection process.

Suppose Δ Varies Across Individuals $\Rightarrow$ η Random.

$Y = E(Y_0) + DE(Y_1 - Y_0) + \{\epsilon + \eta D\}$
where $\epsilon = [Y_0 - E(Y_0)]$, $\eta = (Y_1 - Y_0) - E(Y_1 - Y_0)$.

- $E[\eta D|Z] = E(\eta \mid D = 1, Z) \Pr(D = 1 \mid Z)$.
- Need Z to be uncorrelated with $[\varepsilon + \eta D]$ to use IV identify $E(Y_1 - Y_0)$.
- This condition will be satisfied if policy adoption is made without knowledge of $\eta = (Y_1 - Y_0) - E(Y_1 - Y_0)$.
- If decisions about D are made with partial or full knowledge of η, expect $E(\eta \mid D = 1, Z)$ to depend on Z and thus for IV not to identify $E(Y_1 - Y_0)$.

- Without more conditions, IV does not identify any ATE or any other interpretable parameter under essential heterogeneity.
- With additional conditions, IV does identify an interpretable parameter: Local Average Treatment Effect (LATE; Imbens and Angrist, 1994).

Imbens Angrist conditions (1994)

IV-1 (Independence)

$$Z \perp\!\!\!\perp \left(Y_1, Y_0, \{D(z)\}_{z \in \mathcal{Z}}\right).$$

IV-2 (Rank)

$\Pr(D = 1 \mid Z)$ *depends on* Z.

IV-3 (Monotonicity)

For all $z, z' \in \mathcal{Z}$, *either* $D_i(z) \geq D_i(z')$ *for all* i,
or $D_i(z) \leq D_i(z')$ *for all* i.

Vytlacil (2002) Equivalence

Let $\mathbf{1}[\cdot]$ denote the logical indicator function.

Vytlacil (2002) shows Imbens-Angrist conditions are equivalent to the nonparametric selection model:

SELECTION-1 (Selection Model)

$D_i = \mathbf{1}[\mu(Z_i) \geq U_i]$, $Z_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}, U_i)$, and $\mu(\cdot)$ is a nontrivial function of Z_i.

Imbens Angrist (1994)

For $Z = 0, 1$, Imbens and Angrist show that these conditions imply that

$$\begin{aligned} \text{plim} IV &= \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{\Pr(D = 1 \mid Z = 1) - \Pr(D = 1 \mid Z = 0)} \\ &= E(Y_1 - Y_0 \mid D(1) = 1, D(0) = 0) \\ &\equiv \text{LATE} \end{aligned}$$

- The mean gain to those induced to switch from into treatment by a change in Z from 0 to 1.
- Not always of a priori interest.
- More complicated expression if Z non-binary.

Imbens Angrist

- In general, LATE $\neq E(Y_1 - Y_0)$, $E(Y_1 - Y_0 \mid D = 1)$, or any other parameter of a priori interest.
- Different instruments define different parameters.
- Having a wealth of different strong instruments does not improve the precision of the estimate of any particular parameter.

Alternative Monotonicity Restrictions

- Classical IV imposes strong homogeneity assumption on outcome equation but no structure on selection equation. Does not allow for essential heterogeneity.
- In contrast, LATE analysis imposes no structure on outcome equation but imposes monotonicity on selection equation. Does allow for essential heterogeneity.
- Possible to test for presence of essential heterogeneity (Heckman, Shmierer, Urzua, 2010)
- Monotonicity assumption on selection equation is testable (Kitagawa 2008, and Machado, Shaikh and Vytlacil, 2013)

Alternative Monotonicity Restrictions

LATE analysis imposes no structure on outcome equation but imposes monotonicity on selection equation.

Alternatively, can:

- Impose monotonicity symmetrically on outcome and selection equations: Identify sign of average effect, can bound average effect. Restriction is testable. (Bhattacharya, Shaikh and Vytlacil, 2012; Shaikh and Vytlacil, 2011; Machado, Shaikh and Vytlacil, 2013)
- Impose monotonicity on outcome equation instead of selection equation: Sign of average effect not necessarily identified; Can bound average effect; Some strange implications – passible to have large positive IV imply a negative average treatment effect. Restriction is testable. (Machado, Shaikh and Vytlacil, 2013).

Selection Models

Heckman, Vytlacil and co-authors

- Impose Nonparametric Selection Model
  - By Vytlacil (2002), is equivalent to Imbens and Angrist (1994) assumptions
- Goals:
  - Unify literature with a common set of underlying parameters interpretable across studies.
  - To understand how to connect the results of various disparate IV estimands within a unified framework.
  - Consider strategies other than linear IV.

Threshold Crossing Model for D

Selection Model:

$$D = \mathbf{1}\left[\mu_D(Z) - V > 0\right]$$

with $Z \perp\!\!\!\perp V$.

- $\mu_D(Z) - V$ can be interpreted as a net utility for a person with characteristics (Z, V), where V is unobserved by the analyst.
- Vytlacil (2002) shows that this model is equivalent to the independence and monotonicity restrictions of Imbens and Angrist.
- Wider class of latent index models will have a representation in this form (Vytlacil, 2006).

Threshold Crossing Model for D: Independence, Propensity Score

We define $P(z)$ as the *propensity score*:

$$P(z) \equiv \Pr(D = 1 \mid Z = z) = \Pr(\mu_D(z) > V) = F_V(\mu(z))$$

where F_V is the distribution of V.

As normalization, can rewrite model as

$$\begin{aligned} D &= \mathbf{1}\left[\mu_D(Z) \geq V\right] \\ &= \mathbf{1}\left[F_V(\mu_D(Z)) \geq F_V(V)\right] \\ &= \mathbf{1}\left[P(Z) \geq U_D\right] \end{aligned}$$

with $U_D \equiv F_V(V) \sim \mathrm{Unif}[0,1]$

Marginal Treatment Effect: Key, unifying parameter,

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

- MTE is average effect at a given unobserved desire to participate in treatment.
- MTE and the local average treatment effect (LATE) parameter are closely related.

- Under our assumptions, all standard treatment parameters are weighted averages of MTE with weights that can be estimated.

$$\text{Parameter}_j = \int \Delta^{\text{MTE}}(u_D)\,\omega_j(u_D)du_D$$

Table 1A: treatment effects and estimands as weighted averages of the marginal treatment effect

$$\text{ATE} = E\left(Y_1 - Y_0\right) = \int_0^1 \Delta^{\text{MTE}}(u)\, \omega_{\text{ATE}}(u)\, du$$

$$\text{TT} = E\left(Y_1 - Y_0 \mid D = 1\right) = \int_0^1 \Delta^{\text{MTE}}(u)\, \omega_{\text{TT}}(u)\, du$$

$$\text{TUT} = E\left(Y_1 - Y_0 \mid D = 0\right) = \int_0^1 \Delta^{\text{MTE}}(u)\, \omega_{\text{TUT}}(u)\, du$$

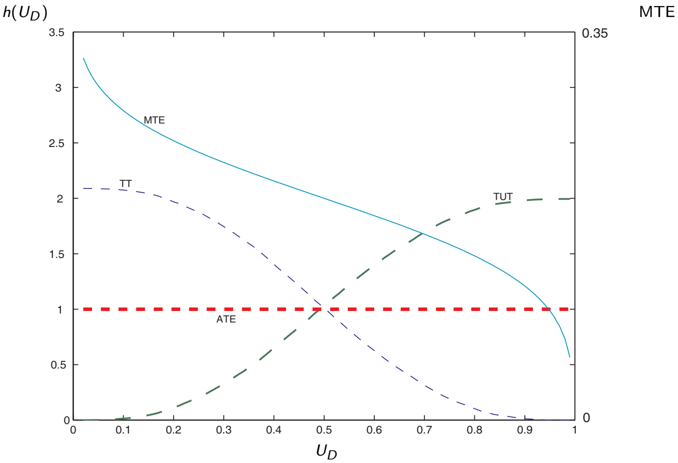
Table1B: weights

$$\omega_{\text{ATE}}(u) = 1$$

$$\omega_{\text{TT}}(u) = \frac{1 - F_P(u)}{E(P)}$$

$$\omega_{\text{TUT}}(u) = \frac{F_P(u)}{E\left((1 - P)\right)}$$

Figure 1: weights for the marginal treatment effect for different parameters



Other parameters that can be represented as weighted average of MTE include:

- Probability limit of IV
- LATE
- Policy Relevant Treatment Effect (Heckman and Vytlacil, 2001)
- Marginal Policy Relevant Treatment Effect, Average Effect of Treatment at the Margin (Carneiro, Heckman and Vytlacil, 2010, 2011).
- Person Centered Treatment Effects (Basu, 2013).

Example: Effect of Year of College on Wages (Parametric)

MTE for Effect of Year of College on Wages

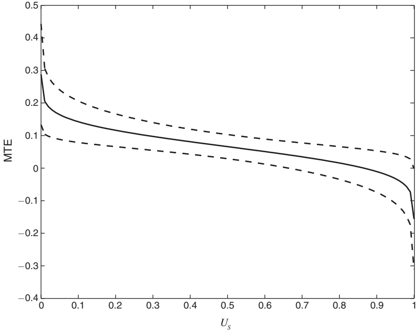


FIGURE 1. MTE ESTIMATED FROM A NORMAL SELECTION MODEL

Source: Carneiro, Heckman and Vytlacil (2011)

Example: Effect of Year of College on Wages (Semi-Parametric)

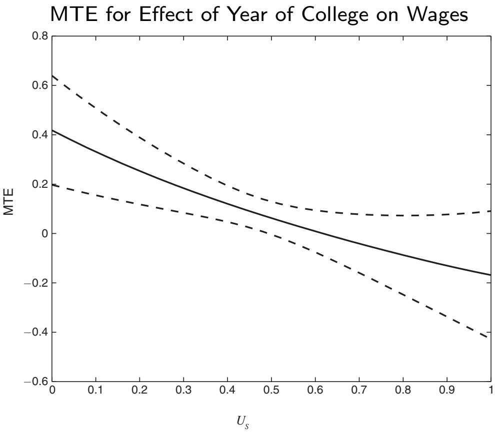


FIGURE 4. $E(Y_1 - Y_0 | \mathbf{X}, U_S)$ WITH 90 PERCENT CONFIDENCE INTERVAL—LOCALLY QUADRATIC REGRESSION ESTIMATES

Source: Carneiro, Heckman and Vytlacil (2011)

Example: Effect of Year of College on Wages

Effect of Year of College on Wages

TABLE 5—RETURNS TO A YEAR OF COLLEGE

Model		Normal	Semiparametric
$ATE = E(\beta)$		0.0670 (0.0378)	Not identified
$TT = E(\beta \mid S = 1)$		0.1433 (0.0346)	Not identified
$TUT = E(\beta \mid S = 0)$		−0.0066 (0.0707)	Not identified
MPRTE			
Policy perturbation	Metric		
$Z_\alpha^k = Z^k + \alpha$	$\lvert \mathbf{Z}\gamma - V \rvert < e$	0.0662 (0.0373)	0.0802 (0.0424)
$P_\alpha = P + \alpha$	$\lvert P - U \rvert < e$	0.0637 (0.0379)	0.0865 (0.0455)
$P_\alpha = (1 + \alpha)P$	$\lvert \frac{P}{U} - 1 \rvert < e$	0.0363 (0.0569)	0.0148 (0.0589)
Linear IV (Using $P(\mathbf{Z})$ as the instrument)		0.0951 (0.0386)	
OLS		0.0836 (0.0068)	

Source: Carneiro, Heckman and Vytlacil (2011)

Example: MTE for Effect of Vocational Rehabilitation on Employment

MTE for Effect of Vocational Rehabilitation on Employment

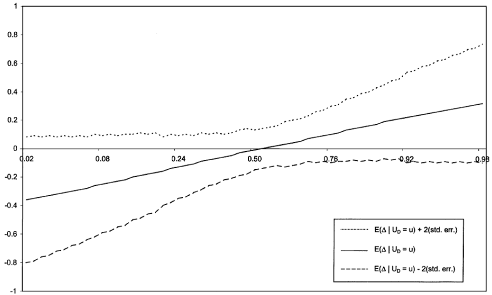


Fig. 1. Estimated marginal treatment effect.

Source: Aakvik, Heckman and Vytlacil (2005)

Example: Effect of Vocational Rehabilitation on Employment

$$ATE$$
$$\mathrm{E}(\Delta) = -0.014$$
$$(\text{standard error} = 0.08)$$

$$TT$$
$$\mathrm{E}(\Delta \mid D = 1) = -0.110$$
$$(\text{standard error} = 0.09)$$

Identification of the MTE

- Different parameters can be seen as different weighted averages of MTE, IV is a weighted average of MTE.
- If can identify MTE, can:
  1. Integrate MTE to obtain other parameters of interest
  2. Understand connection between selection into treatment and individual effects.
- How to identify MTE?

Identification of the MTE (cont'd)

Heckman-Vytlacil show that LIV (Local Instrumental Variables) identifies MTE

$$\underbrace{\frac{\partial}{\partial p} E\left(Y \mid P(Z) = p\right)}_{LIV} = \underbrace{E(Y_1 - Y_0 | U_D = p)}_{MTE}. \tag{6.1}$$

- Thus $\Delta^{\text{MTE}}(u)$ identified by LIV for $u \in \text{Supp}(P(Z))$.
- The greater the variation in $P(Z)$, the greater the range over which MTE is identified.

Using MTE for Identification of Treatment Effects

Treatment Parameter $(j) = \int_0^1 \Delta^{\text{MTE}}(u)\,\omega_j(u)\,du$,

- Identification using this relationship requires identification of $\Delta^{\text{MTE}}(u)$ for u such that $\omega_j(u) \neq 0$.
- We identify $\Delta^{\text{MTE}}(u)$ for $u \in$ Supp$(P(Z))$
- Thus, to integrate MTE to identify treatment parameter, require Supp$(P(Z)) \supseteq \{u : \omega_j(u) \neq 0\}$
- Strong requirement for traditional treatment parameters, typically "identification at infinity" requirement.

Using MTE for Identification of Treatment Effects (cont'd)

- To integrate MTE to identify treatment parameter, require $\text{Supp}(P(Z)) \supseteq \{u : \omega_j(u) \neq 0\}$
- For traditional parameters, this requirement is very strong:
  - For ATE, need $\text{Supp}(P(Z)) = [0, 1]$
  - For TT, need $\text{Supp}(P) = [0, p_u]$
  - For TUT, need $\text{Supp}(P) = [p_l, 1]$

  Can identify these parameters with an alternative identification strategy under slightly weaker conditions, but still require identification at infinity.

Using MTE for Identification of Treatment Effects (cont'd)

Without large support, can still:

1. Bound traditional parameters (Heckman and Vytlacil, 2001)
2. Understand treatment effect for some groups of individuals, and understand part of the connection between selection and individual effects, by examining MTE over identified values.
3. Identify average effect for those on margin of indifference, and effect of marginal policy changes. (Carneiro, Heckman and Vytlacil 2010, 2011).
4. Impose some functional form/parametric restrictions and extrapolate to some extent.

Nonparametric Estimation of MTE

- Making conditioning on X explicit:

$$\underbrace{\frac{\partial}{\partial p} E\left(Y \mid P(Z)=p, X=x\right)}_{LIV} = \underbrace{E(Y_1 - Y_0 | X = x, U_D = p)}_{MTE}.$$

- In theory, can non parametrically estimate $\frac{\partial}{\partial p} E\left(Y \mid P(Z)=p, X\right)$, for example, through local polynomial regression of Y on (X, P).

Nonparametric Estimation of MTE

Problem: Curse of Dimensionality

- If X contains continuous elements, especially multiple continuous elements, point wise estimation of $E(Y \mid P(Z) = p, X = x)$ will be very poor.
  - Formally: very slow rate of convergence. Expect large bias and high imprecision in finite samples. Expect asymptotics to be poor guide.
- Point-wise estimation of derivative of $E(Y \mid P(Z) = p, X = x)$ should be even more difficult.
  - All of above problems, but more so.

Nonparametric Estimation of Treatment Parameters through MTE

Additional Problem:
Support Problem, Irregular Estimation

- To estimate MTE non parametrically for all evaluation points, need support of $P(Z)$ conditional on X to be full unit interval.
  - Requires extremely powerful instrument.
- To integrate up MTE to traditional parameters, require MTE over broad support.
- Traditional treatment parameters are "non-smooth" functions MTE, expect slower than $\sqrt{N}$ estimation.

Nonparametric Estimation of MTE, Treatment Parameters

- Realistically, would need extremely large samples and extremely strong instruments to have nonparametric estimation of MTE and of traditional treatment parameters to be feasible, even if X is low dimensional.
- What is feasible?
    - Estimation of average effect for those on margin of indifference, and effect of marginal policy changes, fundamentally easier than for traditional parameters.
    - Can estimate IV, interpret.
    - Can follow bounding approach.
    - Can incorporate some parametric functional form restrictions, follow semi parametric or parametric estimation approaches.

Semiparametric Estimation of MTE

Can impose semi parametric structure, e.g., if Y is continuous, can follow Heckman, Urzua and Vytlacil (2006) and Carneiro, Heckman and Vytlacil (2010,2011):

- $Y_1 = X\beta_1 + U_1$,
- $Y_0 = X\beta_0 + U_0$

$$\Rightarrow Y = X\beta_0 + DX(\beta_1 - \beta_0) + D(U_1 - U_0) + U_0$$

$$\Rightarrow E(Y \mid X, P(Z)) = X\beta_0 + P(Z)X(\beta_1 - \beta_0) + K(P(Z))$$

Semiparametric Estimation of MTE (cont'd)

$$E\left(Y \mid X, P(Z)\right) = X\beta_0 + P(Z)X(\beta_1 - \beta_0) + K(P(Z))$$

If impose joint normality assumptions, than standard parametric problem.

Otherwise, $K(\cdot)$ unknown function.

Note dimension reduction.

Semiparametric Estimation of MTE (cont'd)

$$E\left(Y \mid X, P(Z)\right) = X\beta_0 + P(Z)X(\beta_1 - \beta_0) + K(P(Z))$$

$K(\cdot)$ unknown function, suggests semiparametric multistep estimation strategy.

1. Estimate $P(Z)$ in first step, either parametrically or semi/nonparametrically (e.g., Klein and Spady).
2. Estimate $E\left(Y \mid X, P(Z)\right)$ using estimated $P(Z)$, for example, using:
    - Partial linear regression/nonparametric double residual regression techniques, as in Heckman, Ichimura and Todd, or
    - Regress Y on $X, P(Z)X$, and a series in $P(Z)$, adapting Newey, Powell and Vella.

Semiparametric Estimation of MTE and Treatment Parameters

- Given semi-parametric partially-linear structure, support conditions now depend on support of $P(Z)$, not support of $P(Z)$ conditional on X. Less restrictive.
- If support of $P(Z)$ is limited, can:
  - Truncate integration to available support.
  - Extrapolate.
  - Follow a bounding approach.
  - Limit range of treatment parameters to be estimated.

Parametric Estimation of MTE

Alternatively, can follow a parametric approach, much less data intensive. For example:

1. If Y continuous, estimate MTE parametrically based on Heckman 2-step, generalizations of Heckman 2-step, imposing linear model with joint normality or generalizations of joint normality on error terms.
    - See, e.g., Tobias, Heckman and Vytlacil (2003), and Carneiro, Heckman and Vytlacil (2011).
2. If Y binary, estimate based on bivariate probit, or generalizations of bivariate probit.
    - See, e.g., Aakvik, Heckman and Vytlacil (2005).

Parametric Estimation of MTE

Parametric estimation:

- Much less data intensive, reasonably precise estimation feasible with smaller sample sizes.
- Naturally provides extrapolation outside of support, can estimate MTE over full unit interval and estimate all treatment parameters.
- Negative: less flexible, parametric structure might be incorrect.

Example: MTE for Effect of Vocational Rehabilitation on Employment

MTE for Effect of Vocational Rehabilitation on Employment

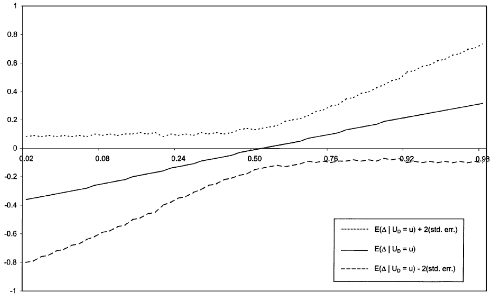


Fig. 1. Estimated marginal treatment effect.

Source: Aakvik, Heckman and Vytlacil (2005)

Example: Effect of Vocational Rehabilitation on Employment

$$ATE$$
$$\mathrm{E}(\Delta) = -0.014$$
$$(\text{standard error} = 0.08)$$

$$TT$$
$$\mathrm{E}(\Delta \mid D = 1) = -0.110$$
$$(\text{standard error} = 0.09)$$

Example: Effect of Year of College on Wages (Parametric)

MTE for Effect of Year of College on Wages

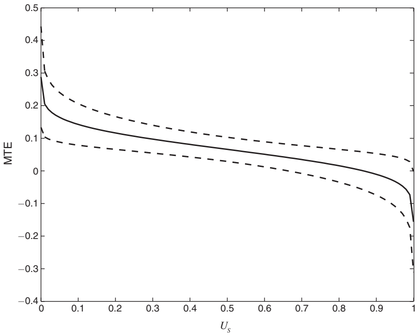


FIGURE 1. MTE ESTIMATED FROM A NORMAL SELECTION MODEL

Source: Carneiro, Heckman and Vytlacil (2011)

Example: Effect of Year of College on Wages (Semi-Parametric)

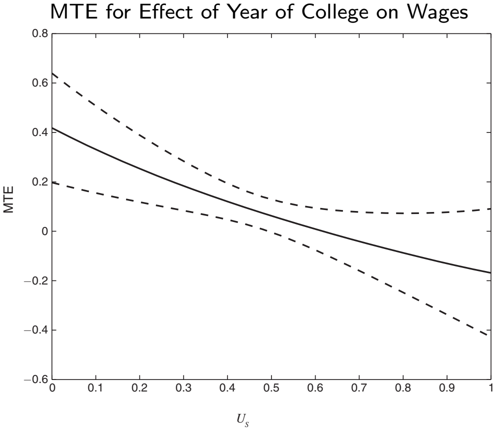


FIGURE 4. $E(Y_1 - Y_0 | \mathbf{X}, U_S)$ WITH 90 PERCENT CONFIDENCE INTERVAL—LOCALLY QUADRATIC REGRESSION ESTIMATES

Source: Carneiro, Heckman and Vytlacil (2011)

Example: Effect of Year of College on Wages

Effect of Year of College on Wages

TABLE 5—RETURNS TO A YEAR OF COLLEGE

Model		Normal	Semiparametric
$ATE = E(\beta)$		0.0670 (0.0378)	Not identified
$TT = E(\beta \mid S = 1)$		0.1433 (0.0346)	Not identified
$TUT = E(\beta \mid S = 0)$		−0.0066 (0.0707)	Not identified
MPRTE			
Policy perturbation	Metric		
$Z_\alpha^k = Z^k + \alpha$	$\lvert \mathbf{Z}\gamma - V \rvert < e$	0.0662 (0.0373)	0.0802 (0.0424)
$P_\alpha = P + \alpha$	$\lvert P - U \rvert < e$	0.0637 (0.0379)	0.0865 (0.0455)
$P_\alpha = (1 + \alpha)P$	$\lvert \frac{P}{U} - 1 \rvert < e$	0.0363 (0.0569)	0.0148 (0.0589)
Linear IV (Using $P(\mathbf{Z})$ as the instrument)		0.0951 (0.0386)	
OLS		0.0836 (0.0068)	

Source: Carneiro, Heckman and Vytlacil (2011)