MTE as Generator of All Treatment Effects: IV and Policy Weights

James J. Heckman
University of Chicago

Extract from: Econometric Evaluation of Social Programs, Part II, Handbook of Econometrics 2007

Econ 312, Spring 2023

Index Model of Choice and Treatment Effects: Definitions and Unifying Principles

$$Y_1 = \mu_1(X, U_1) \quad (1)$$
$$Y_0 = \mu_0(X, U_0). \quad (2)$$

$$D^* = \mu_D(Z) - V\,; \qquad D = 1 \quad \text{if} \quad D^* \geq 0\,; \qquad D = 0 \quad \text{otherwise,} \tag{3}$$

- (Z, X) is observed
- (U_0, U_1, V) is unobserved
- The random variable V may be a function of (U_0, U_1)
- In the original Roy Model, μ_1 and μ_0 are additively separable in U_1 and U_0 respectively, and $V = -[U_1 - U_0]$

- In the original formulations of the Generalized Roy Model, $V = -[U_1 - U_0 - U_C]$
- U_C arises from the cost function.
- Define Z so that it includes all of the elements of X as well as any additional variables unique to the choice equation.

Assumptions

A (A-1)

*(U_0, U_1, V) are independent of Z conditional on X (**Independence**), i.e. $(U_0, U_1, V) \perp\!\!\!\perp Z | X$;*

A (A-2)

*$\mu_D(Z)$ is a non-degenerate random variable conditional on X (**Rank Condition**);*

A (A-3)

The distribution of V is continuous;

A (A-4)

*The values of $E(Y_1)$ and $E(Y_0)$ are finite (**Finite Means**);*

A (A-5)

$0 < \Pr(D = 1 \mid X) < 1$.

- (A-1) assumes that V is independent of Z given X, and is used below to generate counterfactuals.
- For the definition of treatment effects, we do not need either (A-1) or (A-2).
- Definitions of treatment effects and their identification through MTE *do not* require any elements of Z that are not elements of X or independence assumptions.
- However, analysis of instrumental variables requires that Z contain at least one element not in X.
- Assumptions (A-1) or (A-2) justify application of instrumental variables methods and nonparametric selection or control function methods.

- Assumption (A-4) is needed to satisfy standard integration conditions: guarantees that the mean treatment parameters are well defined.
- Assumption (A-5) is the assumption in the population of both a treatment and a control group for each X. Observe that there are no exogeneity requirements for X.
- This is in contrast with the assumptions commonly made in the conventional structural literature and the semiparametric selection literature.

- A counterfactual "no feedback" condition facilitates interpretability so that conditioning on X does not mask the effects of D.
- Letting X_d denote a value of X if D is set to d, a sufficient condition that rules out feedback from D to X is:

A (A-6)

Let X_0 denote the counterfactual value of X that would be observed if D is set to 0. X_1 is defined analogously. Assume $X_d = X$ for $d = 0, 1$. (The X_D are invariant to counterfactual manipulations.)

THE UNIVERSITY OF CHICAGO

- Vytlacil (2002): **assumptions (A-1)–(A-5) for the selection model and (1)–(3) are equivalent to the assumptions used to generate the LATE model of Imbens and Angrist (1994)**.
- The nonparametric selection model for treatment effects developed by Heckman and Vytlacil is implied by the assumptions of the Imbens-Angrist instrumental variable model for treatment effects.
- Approach links the IV literature to the literature on economic choice models exposited in Part I.

- The model of equations (1)-(3) and assumptions (A-1)–(A-5) impose two testable restrictions on the distribution of (Y, D, Z, X).
- First, it imposes an **index sufficiency restriction:** for any set $\mathcal{A}$ and for $j = 0, 1$,

$$\Pr(Y_j \in \mathcal{A} \mid X, Z, D = j) = \Pr(Y_j \in \mathcal{A} \mid X, P(Z), D = j).$$

- Z (given X) enters the model only through the propensity score $P(Z)$.

THE UNIVERSITY OF CHICAGO

- This restriction has empirical content when Z contains two or more variables not in X.
- Second, the model also imposes monotonicity in p for $E(YD \mid X = x, P = p)$ and $E(Y(1 - D) \mid X = x, P = p)$. Heckman and Vytlacil (2005) develop this condition further in Appendix A, and show that it is testable.

Definitions of Treatment Effects

- (ATE): $\Delta^{\text{ATE}}(x) \equiv E(\Delta \mid X = x)$ where $\Delta = Y_1 - Y_0$.
- This is the effect of assigning treatment randomly to everyone of type X assuming full compliance, and ignoring general equilibrium effects.
- The average impact of treatment on persons who actually take the treatment is Treatment on the Treated (TT): $\Delta^{\text{TT}}(x) \equiv E(\Delta \mid X = x, D = 1)$.
- This parameter can also be defined conditional on $P(Z)$: $\Delta^{\text{TT}}(x, p) \equiv E(\Delta \mid X = x, P(Z) = p, D = 1)$.

- The mean effect of treatment on those for whom $X = x$ and $U_D = u_D$: Marginal Treatment Effect (MTE)

$$\Delta^{\text{MTE}}(x, u_D) \equiv E(\Delta \mid X = x, U_D = u_D). \tag{4}$$

- Parameter defined independently of any instrument.
- Separate the definition of parameters from their identification.
- For u_D evaluation points close to zero, $\Delta^{\text{MTE}}(x, u_D)$ is the expected effect of treatment on individuals with the value of unobservables that make them most likely to participate in treatment and who would participate even if the mean scale utility $\mu_D(Z)$ is small. If U_D is large, $\mu_D(Z)$ would have to be large to induce people to participate.

- Can also interpret $E(\Delta \mid X = x,\ U_D = u_D)$ as the mean gain in terms of $Y_1 - Y_0$ for persons with observed characteristics X who would be indifferent between treatment or not if they were randomly assigned a value of Z, say z, such that $\mu_D(z) = V$.

LATE:

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

  for $u_D = \Pr(D(z) = 1) = P(z)$, $u'_D = \Pr(D(z') = 1) = P(z')$,
- Assumption (A-1)implies that
- $\Pr(D(z) = 1) = \Pr(D = 1 \mid Z = z)$ and
- $\Pr(D(z') = 1) = \Pr(D = 1 \mid Z = z')$.

THE UNIVERSITY OF CHICAGO

- *Imbens and Angrist define the LATE parameter as the probability limit of an estimator*.
- Conflates issues of definition of parameters with issues of identification.
- This representation of LATE allows us to separate these two conceptually distinct matters and to define the LATE parameter more generally.
- One can, in principle, evaluate the right hand side of the preceding equation at any u_D, u_D' points in the unit interval and not only at points in the support of the distribution of the propensity score $P(Z)$ conditional on $X = x$ where it is identified.

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

$$\text{ATE}(x) = E\left(Y_1 - Y_0 \mid X = x\right) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, du_D$$

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

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

$$\text{PRTE}(x) = E\left(Y_{a'} \mid X = x\right) - E\left(Y_a \mid X = x\right) = \int_0^1 \Delta^{\text{MTE}}\left(x, u_D\right)\, \omega_{\text{PRTE}}\left(x, u_D\right)\, du_D$$

for two policies a and a' that affect the Z but not the X

$$\text{IV}_J(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, \omega_{\text{IV}}^J(x, u_D)\, du_D, \text{ given instrument } J$$

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

Source: Heckman and Vytlacil (2005)

- All weights (except OLS) integrate up to 1

THE UNIVERSITY OF CHICAGO

Table 1B: Weights

$$\omega_{\text{ATE}}(x, u_D) = 1$$

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

$$\omega_{\text{TUT}}(x, u_D) = \left[\int_{0}^{u_D} f_{P|X}(p|X = x)\, dp\right] \frac{1}{E\left((1 - P)\,|X = x\right)}$$

$$\omega_{\text{PRTE}}(x, u_D) = \left[\frac{F_{P_{a'}|X}(u_D|x) - F_{P_a|X}(u_D|x)}{\Delta\bar{P}(x)}\right], \text{ where } \Delta\bar{P}(x) = E\left(P_a \mid X = x\right) - E\left(P_{a'} \mid X = x\right)$$

$$\omega^{J}_{\text{IV}}(x, u_D) = \left[\int_{u_D}^{1} \int \left(J(Z) - E\left(J(Z) \mid X = x\right)\right) f_{J,P|X}\left(j, t \mid X = x\right)\, dj\, dt\right] \frac{1}{\text{Cov}(J(Z), D \mid X = x)}$$

$$\omega_{\text{OLS}}(x, u_D) = 1 + \frac{E(U_1 \mid X = x, U_D = u_D)\,\omega_1(x, u_D) - E(U_0 \mid X = x, U_D = u_D)\,\omega_0(x, u_D)}{\Delta^{\text{MTE}}(x, u_D)}$$

$$\omega_1(x, u_D) = \left[\int_{u_D}^{1} f_{P|X}(p \mid X = x)\, dp\right] \frac{1}{E(P \mid X = x)}$$

$$\omega_0(x, u_D) = \left[\int_{0}^{u_D} f_{P|X}(p \mid X = x)\, dp\right] \frac{1}{E((1 - P) \mid X = x)}$$

Source: Heckman and Vytlacil (2005)

- From assumptions (A-1), (A-3), and (A-4), $\Delta^{\text{LATE}}(x, u_D, u_D')$ is continuous in u_D and u_D' and

$$\lim_{u_D' \uparrow u_D} \Delta^{\text{LATE}}(x, u_D, u_D') = \Delta^{\text{MTE}}(x, u_D).$$

$$\text{Treatment Parameter}\,(j) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, \omega_j(x, u_D)\; du_D$$

- $\omega_j(x, u_D)$ is the weighting function for the MTE.
- Integral is defined over the full support of u_D.
- Except for the OLS weights, the weights in the table all integrate to one, although in some cases the weights for IV may be negative.

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

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

where

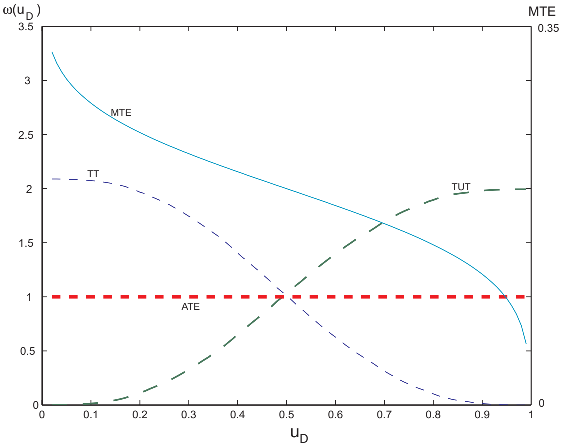
$$\omega_{TT}(x, u_D) = \frac{1 - F_{P|X}(u_D \mid x)}{\int_0^1 \left(1 - F_{P|X}(t \mid x)\right) dt} = \frac{S_{P|X}(u_D \mid x)}{E(P(Z) \mid X = x)}, \quad (5)$$

and $S_{P|X}(u_D \mid x)$ is $\Pr(P(Z) > u_D \mid X = x)$ and $\omega_{\mathrm{TT}}(x, u_D)$ is a weighted distribution.

- $\Delta^{\text{TT}}(x)$ oversamples $\Delta^{\text{MTE}}(x, u_D)$ for those individuals with low values of u_D that make them more likely to participate in the program being evaluated.
- Treatment on the untreated (TUT) is defined symmetrically with TT and oversamples those least likely to participate. The various weights are displayed in table 1AB.
- The other weights, treatment effects and estimands shown in this table are discussed later.

- Observe that if $E(Y_1 - Y_0 \mid X = x, U_D = u_D) = E(Y_1 - Y_0 \mid X = x)$ so $\Delta = Y_1 - Y_0$ is mean independent of U_D given $X = x$, then $\Delta^{\text{MTE}} = \Delta^{\text{ATE}} = \Delta^{\text{TT}} = \Delta^{\text{LATE}}$.
- Where there is no heterogeneity in terms of unobservables in MTE (Δ constant conditional on $X = x$) or agents do not act on it so that U_D drops out of the conditioning set, marginal treatment effects are average treatment effects, so that all of the evaluation parameters are the same.
- Otherwise, they are different.
- Only in the case where the marginal treatment effect is the average treatment effect will the "effect" of treatment be uniquely defined.

Figure 1A: Weights for the marginal treatment effect for different parameters



Source: Heckman and Vytlacil (2005)

- A high u_D is associated with higher cost, relative to return, and less likelihood of choosing $D = 1$.
- The decline of MTE in terms of higher values of u_D means that people with higher u_D have lower gross returns.
- TT overweights low values of u_D (i.e., it oversamples U_D that make it likely to have $D = 1$).
- ATE samples U_D uniformly.
- Treatment on the Untreated $(E(Y_1 - Y_0 \mid X = x, D = 0))$, or TUT, oversamples the values of U_D which make it unlikely to have $D = 1$.

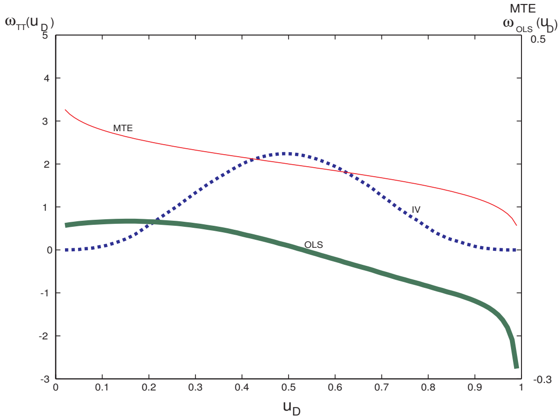
Table 3: Treatment parameters and estimands in the generalized Roy example

Treatment on the Treated	0.2353
Treatment on the Untreated	0.1574
Average Treatment Effect	0.2000
Sorting Gain[a]	0.0353
Policy Relevant Treatment Effect (PRTE)	0.1549
Selection Bias[b]	−0.0628
Linear Instrumental Variables[c]	0.2013
Ordinary Least Squares	0.1725

[a] $TT - ATE = E(Y_1 - Y_0 \mid D = 1) - E(Y_1 - Y_0)$
[b] $OLS - TT = E(Y_0 \mid D = 1) - E(Y_0 \mid D = 0)$
[c] Using Propensity Score $P(Z)$ as the instrument.
Note: The model used to create Table 3 is the same as those used to create Figures 1A and 1A. The PRTE is computed using a policy t characterized as follows: If $Z > 0$ then $D = 1$ if $Z(1+t) - V \geq 0$. If $Z \leq t$ then $D = 1$ if $Z - V \geq 0$. For this example t is set equal to 0.2.
Source: Heckman and Vytlacil (2005)

- Table 3 shows the treatment parameters produced from the different weighting schemes for the model used to generate the weights in figures 1A and 1B (next).
- Given the decline of the MTE in u_D, it is not surprising that TT>ATE>TUT. This is the generalized Roy version of the principle of diminishing returns.
- Those most likely to self select into the program benefit the most from it. The difference between TT and ATE is a sorting gain: $E(Y_1 - Y_0 \mid X, D = 1) - E(Y_1 - Y_0 \mid X)$, the average gain experienced by people who sort into treatment compared to what the average person would experience.
- Purposive selection on the basis of gains should lead to positive sorting gains of the kind found in the table. If there is negative sorting on the gains, then TUT $\geq$ ATE $\geq$ TT.

Figure 1B: Marginal Treatment Effect vs. Linear Instrumental Variables and Ordinary Least Squares Weights



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

Source: Heckman and Vytlacil (2005).

- The additively separable latent index model for D (equation (3)) and assumptions (A-1)–(A-5) are far stronger than what is required to define the parameters in terms of the MTE.
- The representations of treatment effects defined in Table 1A remain valid even if Z is not independent of U_D
- If there are no variables in Z that are not also contained in X, or if a more general nonseparable choice model generates D (so $D^* = \mu_D(Z, U_D)$).
- No instrument Z is needed to define the parameters (task 1: theory).
- Instrument needed to *identify* parameters.

Appendix: Derivations of Weights

- Given the index structure of the Generalized Roy Model, a simple relationship exists among the parameters.
- From the definitions $D = \mathbf{1}(U_D \leq P(z))$ and $\Delta = Y_1 - Y_0$ that

$$\Delta^{\mathrm{TT}}(x, P(z)) = E(\Delta | X = x, U_D \leq P(z)). \tag{6}$$

- Next consider $\Delta^{\text{LATE}}(x, P(z), P(z'))$.
- Note that $E(Y|X = x, P(Z) = P(z))$

$$= P(z)\Big[E(Y_1|X = x, P(Z) = P(z), D = 1)\Big]$$

$$+ (1 - P(z))\Big[E\left(Y_0|X = x, P(Z) = P(z), D = 0\right)\Big]$$

$$= \int_0^{P(z)} E(Y_1|X = x, U_D = u_D)du_D + \int_{P(z)}^1 E(Y_0|X = x, U_D = u_D)du_D,$$

Thus

$$E(Y|X = x, P(Z) = P(z)) - E(Y|X = x, P(Z) = P(z'))$$
$$= \int_{P(z')}^{P(z)} E(Y_1|X = x, U_D = u_D)du_D - \int_{P(z')}^{P(z)} E(Y_0|X = x, U_D = u_D)du_D.$$

THE UNIVERSITY OF CHICAGO

Thus

$$\Delta^{\text{LATE}}(x, P(z), P(z')) = E(\Delta | X = x, P(z') \leq U_D \leq P(z)).$$

- Notice that this expression could be taken as an alternative definition of LATE.
- Note that, in this expression, we could replace $P(z)$ and $P(z')$ with u_D and u'_D.
- *No instrument needs to be available to define LATE.*

THE UNIVERSITY OF CHICAGO

- We write these relationships in succinct form:

$$
\begin{aligned}
&\Delta^{\text{MTE}}(x, u_D) = E(\Delta | X = x, U_D = u_D) \\
&\Delta^{\text{ATE}}(x) = \int_0^1 E(\Delta | X = x, U_D = u_D) du_D \\
&P(z)[\Delta^{\text{TT}}(x, P(z))] = \int_0^{P(z)} E(\Delta | X = x, U_D = u_D) du_D \\
&(P(z) - P(z'))[\Delta^{\text{LATE}}(x, P(z), P(z'))] = \\
&\int_{P(z')}^{P(z)} E(\Delta | X = x, U_D = u_D) du_D.
\end{aligned}
\tag{7}
$$

- The relationship between MTE and LATE or TT conditional on $P(z)$ is analogous to the relationship between a probability density function and a cumulative distribution function.
- The probability density function and the cumulative distribution function represent the same information, but for some purposes the density function is more easily interpreted.
- Likewise, knowledge of TT for all $P(z)$ evaluation points is equivalent to knowledge of the MTE for all u_D evaluation points, so it is not the case that knowledge of one provides more information than knowledge of the other.
- However, in many choice-theoretic contexts it is often easier to interpret MTE than the TT or LATE parameters.
- It has the interpretation as a measure of willingness to pay on the part of people on a specified margin of participation in the program.

THE UNIVERSITY OF CHICAGO

- $\Delta^{\text{MTE}}(x, u_D)$ is the average effect for people who are just indifferent between participation in the program $(D = 1)$ or not $(D = 0)$ if the instrument is externally set so that $P(Z) = u_D$.
- For values of u_D close to zero, $\Delta^{\text{MTE}}(x, u_D)$ is the average effect for individuals with unobservable characteristics that make them the most inclined to participate in the program $(D = 1)$.
- For values of u_D close to one it is the average treatment effect for individuals with unobserved (by the econometrician) characteristics that make them the least inclined to participate.
- ATE integrates $\Delta^{\text{MTE}}(x, u_D)$ over the entire support of U_D (from $u_D = 0$ to $u_D = 1$). It is the average effect for an individual chosen at random from the entire population.

- $\Delta^{\text{TT}}(x, P(z))$ is the average treatment effect for persons who chose to participate at the given value of $P(Z) = P(z)$;
- It integrates $\Delta^{\text{MTE}}(x, u_D)$ up to $u_D = P(z)$.
- As a result, it is primarily determined by the MTE parameter for individuals whose unobserved characteristics make them the most inclined to participate in the program.
- LATE is the average treatment effect for someone who would not participate if $P(Z) \leq P(z')$ and would participate if $P(Z) \geq P(z)$.
- The parameter $\Delta^{\text{LATE}}(x, P(z), P(z'))$ integrates $\Delta^{\text{MTE}}(x, u_D)$ from $u_D = P(z')$ to $u_D = P(z)$.

THE UNIVERSITY OF CHICAGO

- Use the third expression in equation (7)
- Substitute into equation (6)
- We obtain an alternative expression for the TT parameter as a weighted average of MTE parameters:

$$\Delta^{\mathrm{TT}}(x)$$
$$= \int_0^1 \frac{1}{p} \left[\int_0^p E(\Delta | X = x, U_D = u_D) du_D \right] dF_{P(Z)|X,D}(p|x, D = 1).$$

THE UNIVERSITY OF CHICAGO

Using Bayes' rule, it follows that

$$dF_{P(Z)|X,D}(p|x,1) = \frac{\Pr(D=1|X=x, P(Z)=p)}{\Pr(D=1|X=x)} dF_{P(Z)|X}(p|x).$$

- Since $\Pr(D = 1|X = x, P(Z) = p) = p$, it follows that

$$\Delta^{\text{TT}}(x) = \frac{1}{\Pr(D = 1|X = x)} \int_0^1 \left(\int_0^p E(\Delta|X = x, U_D = u_D) du_D \right) dF_{P(Z)|X}(p|x). \tag{8}$$

- Note that $\Pr(D = 1|X = x) = E(P(Z)|X = x) = \int_0^1 (1 - F_{P(Z)|X}(t|x))dt$.

- Reinterpret (8) as a weighted average of local IV parameters where the weighting is similar to that obtained from a length-biased, size-biased, or P-biased sample:

$$\begin{aligned}
\Delta^{\mathrm{TT}}(x) &= \frac{1}{\Pr(D=1|X=x)} \quad x\int_0^1 \left(\int_0^1 \mathbf{1}(u_D \le p) E(\Delta|X=x, U_D=u_D) du_D\right) dF_{P(Z)|X}(p|x) \\
&= \frac{1}{\int (1-F_{P(Z)|X}(t|x))dt} \int_0^1 \left(\int_0^1 E(\Delta|X=x, U_D=u_D)\mathbf{1}(u_D \le p) dF_{P(Z)|X}(p|x)\right) du_D \\
&= \int_0^1 E(\Delta|X=x, U_D=u_D)\left(\frac{1-F_{P(Z)|X}(u_D|x)}{\int (1-F_{P(Z)|X}(t|x))dt}\right) du_D \\
&= \int_0^1 E(\Delta|X=x, U_D=u_D) g_x(u_D) du_D
\end{aligned}$$

Where

$$g_x(u_D) = \frac{1 - F_{P(Z)|X}(u_D|x)}{\int (1 - F_{P(Z)|X}(t|x))dt}$$

- $g_x(u_D)$ is a *weighted distribution*.
- Since $g_x(u_D)$ is a non-increasing function of u_D, drawings from $g_x(u_D)$ oversample persons with low values of U_D, i.e., values of unobserved characteristics that make them the most likely to participate in the program no matter what their value of $P(Z)$.

Since

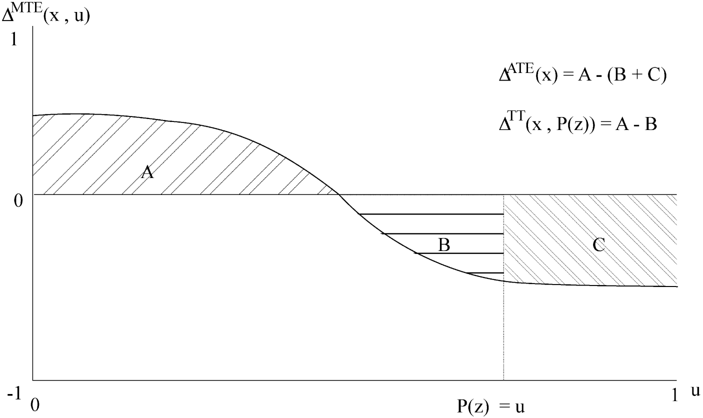
$$\Delta^{\text{MTE}}(x, u_D) = E(\Delta | X = x, U_D = u_D)$$

it follows that

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

- The TT parameter is thus a weighted version of MTE, where $\Delta^{\text{MTE}}(x, u_D)$ is given the largest weight for low u_D values and is given zero weight for $u_D \geq p_x^{max}$, where p_x^{max} is the maximum value in the support of $P(Z)$ conditional on $X = x$.

Figure 3: MTE integrates to ATE and TT under full support (for dichotomous outcome)



Source: Heckman and Vytlacil (2000).

- The curve is the MTE parameter as a function of u_D, and is drawn for the special case where the outcome variable is binary so that MTE parameter is bounded between -1 and 1.

- The ATE parameter averages $\Delta^{\text{MTE}}(u_D)$ over the full unit interval (i.e., is the area under A minus the area under B and C in the figure).
- $\Delta^{\text{TT}}(P(z))$ averages $\Delta^{\text{MTE}}(u_D)$ up to the point $P(z)$ (is the area under A minus the area under B in the figure).
- Because $\Delta^{\text{MTE}}(u_D)$ is assumed to be declining in u_D, the TT parameter for any given $P(z)$ evaluation point is larger then the ATE parameter.

- Equation (7) relates each of the other parameters to the MTE parameter. One can also relate each of the other parameters to the LATE parameter.

- MTE is the limit form of LATE:

$$\Delta^{\text{MTE}}(x, p) = \lim_{p' \to p} \Delta^{\text{LATE}}(x, p, p').$$

- Relationship between LATE and ATE is immediate:

$$\Delta^{\text{ATE}}(x) = \Delta^{\text{LATE}}(x, 0, 1).$$

- Using Bayes' rule, the relationship between LATE and TT is

$$\Delta^{\text{TT}}(x) = \int_0^1 \Delta^{\text{LATE}}(x, 0, p) \frac{p}{\Pr(D = 1|X = x)} dF_{P(Z)|X}(p|x). \tag{9}$$

THE UNIVERSITY OF CHICAGO

Policy Relevant Treatment Effect

- The conventional treatment parameters do not always answer economically interesting questions. Their link to cost-benefit analysis and interpretable economic frameworks is sometimes obscure.
- Each answers a different question. Many investigators estimate a treatment effect and hope that it answers an interesting question.
- A more promising approach for defining parameters is to postulate a policy question or decision problem of interest and to derive the treatment parameter that answers it.
- Taking this approach does not in general produce the conventional treatment parameters or the estimands produced from instrumental variables.

- Consider a class of policies that affect P, the probability of participation in a program, but do not affect Δ^{MTE}. The policies analyzed in the treatment effect literature that change the Z not in X are more restrictive than the general policies that shift X and Z analyzed in the structural literature.
- An example from the schooling literature would be policies that change tuition or distance to school but do not directly affect the gross returns to schooling (Card, 2001). Since we ignore general equilibrium effects in this chapter, the effects on (Y_0, Y_1) from changes in the overall level of education are assumed to be negligible.

- Let p and p' denote two potential policies and let D_p and $D_{p'}$ denote the choices that would be made under policies p and p'. When we discuss the Policy Relevant Treatment Effect, we use "p" to denote the policy and distinguish it from the realized value of $P(Z)$.
- Under our assumptions, the policies affect the Z given X, but not the potential outcomes.

- Let the corresponding decision rules be $D_p = \mathbf{1}[P_p(Z_p) \geq U_D]$, $D_{p'} = \mathbf{1}[P_{p'}(Z_{p'}) \geq U_D]$, where $P_p(Z_p) = \Pr(D_p = 1 \mid Z_p)$ and $P_{p'}(Z_{p'}) = \Pr(D_{p'} = 1 \mid Z_{p'})$.
- To simplify the exposition, we will suppress the arguments of these functions and write P_p and $P_{p'}$ for $P_p(Z_p)$ and $P_{p'}(Z_{p'})$. Define $(Y_{0,p}, Y_{1,p}, U_{D,p})$ as (Y_0, Y_1, U_D) under policy p, and define $(Y_{0,p'}, Y_{1,p'}, U_{D,p'})$ correspondingly under policy p'

- Assume that Z_p and $Z_{p'}$ are independent of $(Y_{0,p}, Y_{1,p}, U_{D,p})$ and $(Y_{0,p'}, Y_{1,p'}, U_{D,p'})$ respectively, conditional on X_p and $X_{p'}$. Let $Y_p = D_p Y_{1,p} + (1 - D_p) Y_{0,p}$ and $Y_{p'} = D_{p'} Y_{1,p'} + (1 - D_{p'}) Y_{0,p'}$ denote the outcomes that would be observed under policies p and p', respectively.
- Assume Δ^{MTE} is policy invariant.

- PRTE, denoted $\Delta^{\text{PRTE}}(x)$:

$$E(Y_p) \mid X = x) - E((Y_{p'}) \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\{F_{P_{p'}|X}(u_D \mid x) - F_{P_p|X}(u_D \mid x)\}\, du_D, \quad (10)$$

where $F_{P_p|X}(\cdot \mid x)$ and $F_{P_{p'}|X}(\cdot \mid x)$ are the distributions of P_p and $P_{p'}$ conditional on $X = x$

- Defined for the different policy regimes and $\Delta^{\text{MTE}}_{\Upsilon}(x, u_D) = E(\Upsilon(Y_{1,p}) - \Upsilon(Y_{0,p}) \mid U_{D,p} = u_D, X_p = x)$.

- The weights in expression (10) are derived under the assumption that the policy does not change the joint distribution of outcomes.
- To simplify the notation, throughout the rest of this chapter when we discuss PRTE, we assume that $\Upsilon(Y) = Y$. Modifications of our analysis for the more general case are straightforward.

- Define $\Delta\bar{P}(x) = E(P_p \mid X = x) - E(P_{p'} \mid X = x)$, the change in the proportion of people induced into the program due to the intervention. Assuming $\Delta\bar{P}(x)$ is positive, we may define per person affected weights as $\omega_{\text{PRTE}}(x, u_D) = \frac{F_{P_{p'}|X}(u_D|x) - F_{P_p|X}(u_D|x)}{\Delta\bar{P}(x)}$.

Derivation of PTRE

Proof.

(Equation (10)) Define $\mathbf{1}_{\mathcal{P}}(t)$ to be the indicator function for the event $t \in \mathcal{P}$. Then $E\left(Y_p \mid X\right)$

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

This derivation involves changing the order of integration. *Q.E.D.*

- This derivation involves changing the order of integration.
- Note that from (A-4),

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

so the change in the order of integration is valid by Fubini's theorem.

- Comparing policy p to policy p',

$$E\left(Y_p \mid X\right) - E\left(Y_{p'} \mid X\right)$$
$$= \int_0^1 E(\Delta \mid X, U_D = u_D)(F_{P_{p'}|X}(u_D) - F_{P_p|X}(u_D))\, du_D,$$

THE UNIVERSITY OF CHICAGO