Lecture 10. Roy Model, Marginal Treatment Effects

Economics 8379
George Washington University

Instructor: Prof. Ben Williams

LATE

- Let D denote a treatment variable, Z and instrument.
- Let D_z denote the (counterfactual) value of D when Z is fixed at z.

LATE

- Let D denote a treatment variable, Z and instrument.
- Let D_z denote the (counterfactual) value of D when Z is fixed at z.
- Let $Y(d, z)$ denote the counterfactual outcome.
- To simplify, suppose Z and D are both binary.

LATE assumptions

Theorem 4.4.1 in MHE.

- Assumption 1. $(Y(D_1, 1), Y(D_0, 0), D_1, D_0) \perp\!\!\!\perp Z$
- Assumption 2. $Y(d, 1) = Y(d, 0)$
- Assumption 3. $E(D_1 - D_0) \neq 0$
- Assumption 4. $D_1 - D_0 \geq 0$, or vice versa

LATE assumptions

Theorem 4.4.1 in MHE.

- Assumption 1. $(Y(D_1, 1), Y(D_0, 0), D_1, D_0) \perp\!\!\!\perp Z$
- Assumption 2. $Y(d, 1) = Y(d, 0)$
- Assumption 3. $E(D_1 - D_0) \neq 0$
- Assumption 4. $D_1 - D_0 \geq 0$, or vice versa

Then

$$\frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(D \mid Z=1) - E(D \mid Z=0)} = E(Y_1 - Y_0 \mid D_1 > D_0)$$
$$= LATE$$

Special cases

- The TT can be written as a weighted average of LATE and the average effect for the always-takers.
- In some cases, D must be equal to 0 when $Z = 0$.
  - The Bloom example – Z is a random assignment and D a treatment and there is one-way noncompliance.
  - One-way noncompliance means that some with $Z = 1$ choose $D = 0$ (refuse treatment) but no one with $Z = 0$ can have $D = 1$.
- In these cases, IV estimates TT.

Special cases

- The TUT can be written as a weighted average of LATE and the average effect for the never-takers.
- In some cases, D must be equal to 1 when $Z = 1$.
  - Suppose D indicates having a third child (as opposed to only 2) and Z indicates whether the second birth was a multiple birth.
  - Then if $Z = 1$ we must have $D = 1$.
  - There are no “never-takers”.
- In these cases, IV estimates TUT.

Compliers

- A few results:
  - $Pr(D_1 > D_0) = E(D \mid Z = 1) - E(D \mid Z = 0)$
  - for any W such that (D_1, D_0) is independent of Z conditional on W, $E(W \mid D_1 > D_0) = \frac{E(\kappa W)}{E(\kappa)}$ where

$$\kappa = 1 - \frac{D(1-Z)}{1 - Pr(Z = 1 \mid W)} - \frac{(1-D)Z}{Pr(Z = 1 \mid W)}$$

  - and, more generally, $f_{W \mid D_1 > D_0}(w)$ is equal to

$$\frac{E(D \mid Z = 1, W = w) - E(D \mid Z = 0, W = w)}{E(D \mid Z = 1) - E(D \mid Z = 0)} f_W(w)$$

LATE with covariates

- The LATE story gets quite a bit more complicated with covariates.
- Let $\lambda(x) = E(Y_1 - Y_0 \mid D_1 > D_0, X = x)$ denote the LATE conditional on X.
- We could estimate these directly using the Wald formula conditional on X.

LATE with covariates

- The LATE story gets quite a bit more complicated with covariates.
- Let $\lambda(x) = E(Y_1 - Y_0 \mid D_1 > D_0, X = x)$ denote the LATE conditional on X.
- We could estimate these directly using the Wald formula conditional on X.
- If we do 2SLS where the first stage is fully saturated and the second stage is saturated in X we get a weighted average of the $\lambda(x)$.
    - The weights are larger for values of x such that $Var(E(D \mid X = x, Z) \mid X = x)$ is larger.

LATE with covariates

- The LATE story gets quite a bit more complicated with covariates.
- Let $\lambda(x) = E(Y_1 - Y_0 \mid D_1 > D_0, X = x)$ denote the LATE conditional on X.
- We could estimate these directly using the Wald formula conditional on X.
- If we do 2SLS where the first stage is fully saturated and the second stage is saturated in X we get a weighted average of the $\lambda(x)$.
  - The weights are larger for values of x such that $Var(E(D \mid X = x, Z) \mid X = x)$ is larger.
- if $Pr(Z = 1 \mid X)$ is a linear function of X then 2SLS gives the minimum MSE approximation to $E(Y \mid D, X, D_1 > D_0)$.
  - This is useful because $E(Y \mid D = 1, X, D_1 > D_0) - E(Y \mid D = 0, X, D_1 > D_0) = \lambda(X)$.
  - Abadie (2003) proposes a way to estimate this same minimum MSE approximation when $Pr(Z = 1 \mid X)$ is not linear.

Roy model

The Roy model is a model of comparative advantage:

- Potential earnings in sectors 0 and 1: Y_0, Y_1
- Individuals choose sector 1 if and only if $Y_1 - Y_0 \geq c$ where c is a nonrandom cost.
- Heckman and Honore (1990) studied the empirical implications and identification of this model.

Roy model

The extended and generalized Roy model:

- the extended model allows for an observable cost component, $D = \mathbf{1}(Y_1 - Y_0 \geq c(Z))$ where Z is a vector of covariates and c is a possibly unknown function.
- the generalized model allows for an unobservable cost component, $D = \mathbf{1}(Y_1 - Y_0 \geq c(Z, V))$ where V is unobservable

Roy model

- Let $Y_d = \mu_d + U_d$ where $E(U_d) = 0$ for $d = 0, 1$.
- If we observe a vector of covariates X, $\mu_d = \mu_d(X)$.
  - Often $\mu_d(X) = \beta_d' X$.

Roy model

- Let $Y_d = \mu_d + U_d$ where $E(U_d) = 0$ for $d = 0, 1$.
- If we observe a vector of covariates X, $\mu_d = \mu_d(X)$.
  - Often $\mu_d(X) = \beta_d' X$.
- We can write

$$\begin{aligned} Y_i &= Y_{0i} + (Y_{1i} - Y_{0i})D_i \\ &= \mu_0 + (\mu_1 - \mu_0 + U_{1i} - U_{0i})D_i + U_{0i} \\ (&= \alpha \quad + \qquad \beta_i \qquad D_i + u_i) \end{aligned}$$

Roy model

- Let $Y_d = \mu_d + U_d$ where $E(U_d) = 0$ for $d = 0, 1$.
- If we observe a vector of covariates X, $\mu_d = \mu_d(X)$.
  - Often $\mu_d(X) = \beta_d' X$.
- We can write

$$
\begin{aligned}
Y_i &= Y_{0i} + (Y_{1i} - Y_{0i})D_i \\
&= \mu_0 + (\mu_1 - \mu_0 + U_{1i} - U_{0i})D_i + U_{0i} \\
(&= \alpha \quad + \qquad \beta_i \qquad D_i + u_i)
\end{aligned}
$$

- What the Roy model gives us is that it adds a model for D to the potential outcomes framework and demonstrates the important link between the model for D and the model for the potential outcomes.

Roy model

- Problem #1
  - if $\mu_d = \mu_d(X)$ then OLS does not identify $ATE = E(Y_{1i} - Y_{0i})$ generally because of nonlinearity ($\mu_d(X) \neq \beta_d X$) and observed heterogeneity ($\mu_1(x) - \mu_0(x)$ varies with x)

Roy model

- Problem #1
  - if $\mu_d = \mu_d(X)$ then OLS does not identify $ATE = E(Y_{1i} - Y_{0i})$ generally because of nonlinearity $(\mu_d(X) \neq \beta_d X)$ and observed heterogeneity $(\mu_1(x) - \mu_0(x)$ varies with $x)$
  - Of course, if $\mu_d = \beta_d' X$ then we solve this problem by regressing Y_i on D_i, X_i and $D_i X_i$.
  - Alternatively, we do matching to overcome these two problems.
  - Or, we simply do OLS (without the interaction) which identifies a weighted average of conditional treatment effects.

Roy model

- Problem #2
  - The above solutions only work under the conditional independence assumption, $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i \mid X_i$.
  - In the generalized Roy model, this is only satisfied if

$$(U_{0i}, U_{1i}) \perp\!\!\!\perp (U_{1i} - U_{0i}, V_i, Z_i) \mid X_i$$

Roy model

- Problem #2
  - The above solutions only work under the conditional independence assumption, $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i \mid X_i$.
  - In the generalized Roy model, this is only satisfied if

$$(U_{0i}, U_{1i}) \perp\!\!\!\perp (U_{1i} - U_{0i}, V_i, Z_i) \mid X_i$$

    - no unobserved heterogeneity and non-random or independent costs

Roy model

- Problem #2
  - A more general model is $D = \mathbf{1}(E(Y_1 - Y_0 - c(Z, V) \mid \mathcal{I}) \geq 0)$ where $E(\cdot \mid \mathcal{I})$ represents the expected value from the decision-maker's perspective, conditional on their information set.

Roy model

- Problem #2
  - A more general model is $D = \mathbf{1}(E(Y_1 - Y_0 - c(Z, V) \mid \mathcal{I}) \geq 0)$ where $E(\cdot \mid \mathcal{I})$ represents the expected value from the decision-maker's perspective, conditional on their information set.
  - In this case, conditional independence can be stated in terms of the information available to the econometrician relative to what's available to the decision-maker.
    - What if $\mathcal{I}$ consists of X and Z but not U_{1i}, U_{0i} or V_i?

Roy model

- Problem #2
  - Note also that $Y_i = \mu_0 + (\mu_1 - \mu_0)D_i + U_{0i} + (U_{1i} - U_{0i})D_i$
  - There is a selection on unobservables problem (D_i is correlated with U_{0i}) and an unobserved heterogeneity problem ($U_{1i} - U_{0i} \neq 0$).
  - An exercise for you:
    - What happens if $U_{1i} = \Delta_i + U_{0i}$ where Δ_i is independent of U_{0i}?
    - What if $U_{1i} = U_{0i}$ but U_{0i} is not independent of D_i (perhaps because U_{0i} is correlated with V_i)?

Roy model

- Problem #3
  - In the simple Roy model, no instruments are available.
  - In the extended and generalized models, Z_i is potentially a valid instrument because it is relevant but excluded from the outcome equations.
  - However, when is $E(U_{0i} + (U_{1i} - U_{0i})D_i \mid Z_i) = 0$?
    - even if $E(U_{0i} \mid Z_i) = 0$, it is unlikely that

$$\begin{aligned} 0 &= E(U_{1i} - U_{0i})D_i \mid Z_i) \\ &= E\left((U_{1i} - U_{0i})\mathbf{1}(\mu_1 - \mu_0 + U_1 - U_0 \geq c(Z, V)) \mid Z_i\right) \end{aligned}$$

unless $U_1 = U_0$.

Roy model

- The rest of this lecture
  1. What can be identified in the various versions of the Roy model if we assume normal errors?
  2. What *does* IV estimate when there is “essential heterogeneity”?
  3. How can we estimate the ATE (or other similar parameters) when we have an instrument Z_i such that $(X_i, Z_i) \perp\!\!\!\perp (U_{0i}, U_{1i}, V_i)$?
  4. Can we estimate policy counterfactuals with such a Z_i?

Roy model

- In the Roy model, $Y_d = \mu_d + U_d$ for $d = 0, 1$.
- Suppose we observe a vector of covariates X so that $\mu_d = \beta_d' X$.
- Then

$$E(Y \mid D = 1, X = x) = \beta_1' x + E(U_1 \mid U_1 - U_0 \geq -z^*, X = x)$$
$$E(Y \mid D = 0, X = x) = \beta_0' x + E(U_0 \mid U_1 - U_0 < -z^*, X = x)$$

where $z^* = (\beta_1 - \beta_0)' x - c$.

Roy model

Assumption: $(U_1, U_0) \mid X = x \sim N(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{10} \\ \sigma_{10} & \sigma_0^2 \end{pmatrix}$$

- Let $V = U_1 - U_0$
- and $\sigma_V^2 = Var(V)$

Roy model

Assumption: $(U_1, U_0) \mid X = x \sim N(0, \Sigma)$

- Let $\tilde{z} = z^* / \sigma_V$.
- Then under this assumption,

$$E(Y \mid D = 1, X = x) = \beta_1' x + \frac{\sigma_1^2 - \sigma_{10}}{\sigma_V} \lambda(-\tilde{z})$$

$$E(Y \mid D = 0, X = x) = \beta_0' x + \frac{\sigma_0^2 - \sigma_{10}}{\sigma_V} \lambda(\tilde{z})$$

and

$$Pr(D = 1 \mid X = x) = \Phi(\tilde{z})$$

Roy model

Estimation with only sector one observed.

- If we only observe Y for those with $D = 1$ (for example, a wage-LFP model) then we can
  (1) estimate the probit: $Pr(D = 1 \mid X = x) = \Phi(\gamma_0 + \gamma_1' x) = \Phi(\tilde{z})$
  (2) compute the predicted values from (1), $\hat{\tilde{Z}} = \hat{\gamma}_0 + \hat{\gamma}_1' X$ and plug into λ to get $\lambda(-\hat{\tilde{Z}})$
  (3) estimate a regression of Y on $X, \lambda(-\hat{\tilde{Z}})$ for those with $D = 1$
- This enables us to estimate β_1 but not $\beta_0, \sigma_1, \sigma_{10}, \sigma_0$.

Roy model

Estimation with both sectors observed.

- If we only observe Y, D, X for everyone.
  (1) estimate the probit:
  $Pr(D = 1 \mid X = x) = \Phi(\gamma_0 + \gamma_1' x) = \Phi(\tilde{z})$
  (2) compute the predicted values from (1), $\hat{\tilde{Z}} = \hat{\gamma}_0 + \hat{\gamma}_1' X$ and plug into λ to get $\lambda(-\hat{\tilde{Z}})$
  (3) estimate a regression of Y on $X, \lambda(-\hat{\tilde{Z}})$ for those with $D = 1$
  (4) estimate a regression of Y on $X, \lambda(\hat{\tilde{Z}})$ for those with $D = 0$
- This enables us to estimate $\beta_1, \beta_0, \frac{\beta_1 - \beta_0}{\sigma_V}, \frac{\sigma_0^2 - \sigma_{10}}{\sigma_V}$ and $\frac{\sigma_1^2 - \sigma_{10}}{\sigma_V}$.
  - Thus we get σ_V too
  - From the variance of the residuals from the two regressions we can also identify σ_1^2, σ_0^2 and σ_{10}.

Roy model

Concerns:

- If $\lambda(-\tilde{z})$ is approximately linear then we will have a serious collinearity problem.
- If U_1, U_0 is not normal then the model is misspecified and identification is not transparent.
- If there are variable costs the model is misspecified.

Roy model

- In the extended Roy model, $Y_d = \mu_d + U_d$ for $d = 0, 1$ but

$$D = \mathbf{1}(Y_1 - Y_0 \geq \gamma_1' X + \gamma_2' Z)$$

- "'Cost" of participation varies with X and also with other variables Z.
- Then

$$\begin{aligned} E(Y \mid D = 1, X = x, Z = z) &= \beta_1' x \\ &\quad + E(U_1 \mid U_1 - U_0 \geq -z^*, X = x) \\ E(Y \mid D = 0, X = x, Z = z) &= \beta_0' x \\ &\quad + E(U_0 \mid U_1 - U_0 < -z^*, X = x) \end{aligned}$$

where $z^* = (\beta_1 - \beta_0)' x - \gamma_1' x - \gamma_2' z$.

Roy model

Assumption: $(U_1, U_0) \mid X = x, Z = z \sim N(0, \Sigma)$

- Under this assumption, if $\tilde{z} = z^*/\sigma_V$,

$$E(Y \mid D = 1, X = x) = \beta_1' x + \frac{\sigma_1^2 - \sigma_{10}}{\sigma_V} \lambda(-\tilde{z})$$
$$E(Y \mid D = 0, X = x) = \beta_0' x + \frac{\sigma_0^2 - \sigma_{10}}{\sigma_V} \lambda(\tilde{z})$$

  and

$$Pr(D = 1 \mid X = x, Z = z) = \Phi(\tilde{z})$$

- β_1 and β_0 are still identified
- Σ only identified if there is an exclusion: a component of X that does not affect costs

Roy model

What do we need/want to identify?

- The ATE is

$$E(Y_1 - Y_0) = (\beta_1 - \beta_0)E(X)$$

Roy model

What do we need/want to identify?

- The ATE is

$$E(Y_1 - Y_0) = (\beta_1 - \beta_0)E(X)$$

- The distribution of gains:

$$Y_1 - Y_0 \mid X = x \sim N((\beta_1 - \beta_0)'x, \sigma_V^2)$$

- various other counterfactuals
- need Σ to go beyond mean treatment effects

Roy model

- In the generalized Roy model, $Y_d = \mu_d + U_d$ for $d = 0, 1$ but
$$D = \mathbf{1}(\gamma_1' X + \gamma_2' Z \geq V)$$
- V includes an unobservable component of cost
- Then

$$\begin{aligned} E(Y \mid D = 1, X = x, Z = z) &= \beta_1' x \\ &\quad + E(U_1 \mid V \leq z^*, X = x) \\ E(Y \mid D = 0, X = x, Z = z) &= \beta_0' x \\ &\quad + E(U_0 \mid V > z^*, X = x) \end{aligned}$$

where $z^* = \gamma_1' x + \gamma_2' z$.

Roy model

Assumption: $(U_1, U_0, V) \mid X = x, Z = z \sim N(0, \Sigma)$

- Under this assumption,
  - β_1 and β_0 are identified
  - σ_V is identified under the exclusion restriction
  - but $Var(U_1 - U_0) \neq \sigma_V^2$ (key ingredient needed for distribution of $Y_1 - Y_0$) is not identified

Generalized roy model without normality

Assumption: $(U_1, U_0, V) \perp\!\!\!\perp X, Z$ and $\lim_{z\to\infty} Pr(D = 1 \mid X = x, Z = z) = 1$

- The first assumption is essentially the same one used by Imbens and Angrist (1994)
- The second assumption is called "identification at infinity"

Roy model without normality

- Under these assumptions
  - Let $P(x,z) = Pr(D = 1 \mid X = x, Z = z)$
  - $E(Y \mid D = 1, X = x, Z = z) = \beta_1' x + K_1(P(x,z))$
  - and $\lim_{z\to\infty} E(Y \mid D = 1, X = x, Z = z) = \beta_1' x$
  - selection on unobservables goes away in the limit
  - Using the same argument for $D = 0$, we can identify $ATE(x) = (\beta_1 - \beta_0)' x$
  - We've traded normality for identification at infinity.

Some preliminaries

- Let $D = \mathbf{1}(\gamma_1' X + \gamma_2' Z \geq V)$ and assume that $(U_1, U_0, V) \perp\!\!\!\perp (X, Z)$
- Let $U_D = F_V(V)$ where V has distribution function $F_V(\cdot)$.
- Then $D = \mathbf{1}(P(X, Z) \geq U_D)$ where $P(X, Z) = F_V(\gamma_1' X + \gamma_2' Z)$ is the propensity score and $U_D \sim Uniform(0, 1)$.

Definition of the MTE

- Then the marginal treatment effect (MTE) is defined as

$$MTE(x, u) = E(Y_1 - Y_0 \mid X = x, U_D = u)$$

- This demonstrates (observable and unobservable) heterogeneity in $Y_1 - Y_0$.

Definition of the MTE

- $MTE(x, u)$ can be interpreted as the effect of participation for those individuals who would be indifferent if we assigned them a new value of $P = P(x, z)$ equal to u.
    - Someone with a large value of u (close to 1) will participate only if P is quite large; this person will be indifferent if P is equal to u. These are the "high unobservable cost" individuals.
    - Someone with a small value of u (close to 0) will participate even if P is quite small; this person will be indifferent if P is equal to u. These are the "low unobservable cost" individuals.

Identification of the MTE

- The identifying equation:

$$\frac{\partial E(Y \mid X = x, P(X,Z) = p)}{\partial p} = MTE(x,p)$$

  - This only works if Z is continuous.
  - The effect for the “high unobserved cost” individuals is identified by the effect of a marginal increase in participation probability on Y at a high participation rate.

Other treatment parameters and methods

- $ATE(x) = \int_0^1 MTE(x,u)du$

Other treatment parameters and methods

- $ATE(x) = \int_0^1 MTE(x, u)du$
- $TT(x) = \int_0^1 MTE(x, u)\omega_{TT}(x, u)du$ where $\omega_{TT}(x, u)$ disproportionately weights smaller values of u
- OLS and IV can also be written as weighted averages of $MTE(x, u)$.

Other treatment parameters and methods

- Consider the IV estimand $\Delta_{IV}(x) = \frac{Cov(J(Z),Y|X=x)}{Cov(J(Z),D|X=x)}$
- The weight here is

$$\omega_{IV}(x,u) = \frac{E(J - E(J) \mid X = x, P \geq u)Pr(P \geq u \mid X = x)}{Cov(J, P \mid X = x)}$$

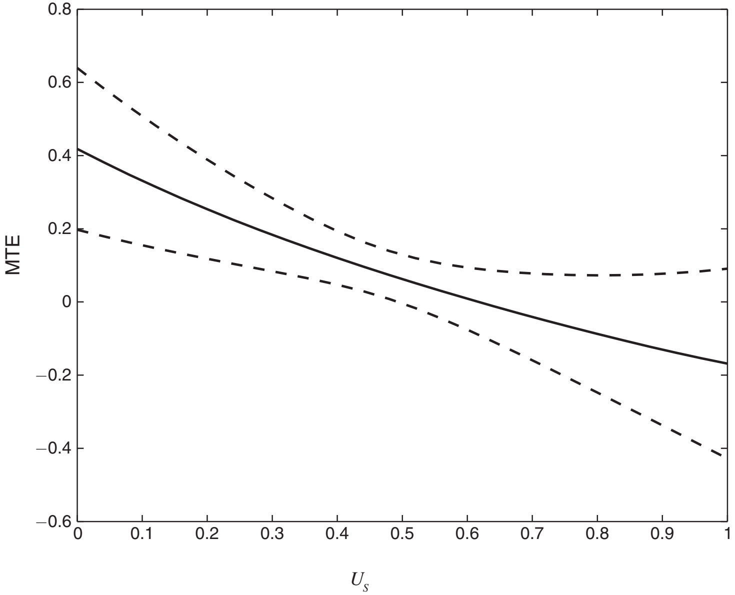
- This and many interesting implications are discussed in Heckman, Vytlacil, and Urzua (2006).

An example

Carneiro, Heckman, Vytlacil (2011)

- use data from the NLSY
- Y is log wage in 1991 (ages 28-34), D represents college attendance, X a vector of controls
- vector Z: (i) distance to college, (ii) local wage, (iii) local unemployment, (iv) average local public tuition

An example



- Estimating the MTE - normal model
  - Option 1. Estimate using MLE.
  - Option 2. Two stage estimation:
    - Probit to estimate $P(Z_i)$ (to simplify, define Z_i to include X_i *and* instrument(s))
    - Regress Y on X_i and $\hat{\lambda}_{1i} = -\frac{\phi(\Phi^{-1}(\hat{P}(Z_i)))}{\hat{P}(Z_i)}$ for $D_i = 1$
    - Regress Y on X_i and $\hat{\lambda}_{0i} = \frac{\phi(\Phi^{-1}(\hat{P}(Z_i)))}{1-\hat{P}(Z_i)}$ for $D_i = 0$
    - Then

$$MTE(x,u) = x'(\hat{\beta}_1 - \hat{\beta}_0) + (\hat{\rho}_1 - \hat{\rho}_0)\Phi^{-1}(u)$$

- Estimating the MTE – semiparametric model
  - The outcome equation can be written (see II.B in Carneiro et al. (2011)) as

$$E(Y \mid X = x, P(Z) = p) = x'\delta_0 + px'(\delta_1 - \delta_0) + K(p)$$

  - There are several ways to estimate this – perhaps the simplest is a series/spline/sieve estimator.
    - Estimate $P(Z_i)$ (logit).
    - Choose a set of basis functions (polynomials) and an order, K.
    - Run the regression:

$$Y_i = X_i'\delta_0 + \hat{P}(Z_i)X_i'(\delta_1 - \delta_0) + \gamma_1 \hat{P}(Z_i) + \ldots + \gamma_K \hat{P}(Z_i)^K + \eta_i$$

  - An important sacrifice here is that $MTE(x, u)$ is only identified for u in the support of P. (Recall identification at ∞)

- Consider policies that affect $P(Z)$ but not Y_1, Y_0, V.
- Propensity score P^* under new policy.
- It can be shown that the effect of shifting to this new policy is given by

$$\int_0^1 MTE(x,u)\left[\frac{F_{P^*|X=x}(u) - F_{P|X=x}(u)}{E(P^* \mid X=x) - E(P \mid X=x)}\right] du$$

- This will still require large support for $P(Z)$.
  - define a continuum of policies
  - consider marginal change from baseline

- MPRTE
  - Consider increasing tuition (a component of Z) by an amount α: $tuition^* = tuition + \alpha$.
  - Corresponding propensity score, P_α.
  - Define the MPRTE as

$$\lim_{\alpha \to 0} \int_0^1 MTE(x,u) \left[\frac{F_{P_\alpha | X=x}(u) - F_{P_0 | X=x}(u)}{E(P_\alpha \mid X=x) - E(P_0 \mid X=x)} \right] du$$

  - This is also equal to $\lim_{e \to 0} E(Y_1 - Y_0 \mid |\mu_D(X,Z) - V| < e)$.
  - And it can be written as $\int_0^1 MTE(x,u)\omega(x,u)$ where

$$\omega(x,u) = \frac{f_{P|X}(u) f_{V|X}(F^{-1}_{V|X}(u))}{E(f_{V|X}(\mu_D(X,Z)) \mid X)}$$