Program Evaluation: Marginal Treatment Effects

Chris Conlon

April 4, 2021

Applied Econometrics

Marginal Treatment Effects

Motivation

We started with by talking about two approaches to program evaluation for $\Delta_i = Y_i(1) - Y_i(0)$.

- In the structural approach the goal was to recover the full distribution $f(\Delta_i)$, and with that recover whichever objects we wanted
- The other approach focused on characterizing particular moments of the distribution

$$ATT(x) = E[\Delta_i | X_i = x, T_i = 1]$$
$$ATU(x) = E[\Delta_i | X_i = x, T_i = 0]$$
$$LATE(x) = E[\Delta_i | X_i = x, T_i(z_i = 1) > T_i(z_i = 0)]$$

- But how exactly are these approaches linked?

The answer is through a nonparametric function known as the marginal treatment effect (MTE).

One quantity to rule them all: MTE

- Consider a treatment effect $\Delta_i = Y_i(1) - Y_i(0)$.
- Think about a single-index such that $T_i = 1(v_i \leq Z_i'\gamma)$.
- Think about the person for whom $v_i = Z_i'\gamma$
  (indifferent between treatment and non-treatment).

$$\Delta^{MTE}(x, v_i) = E[\Delta_i | X_i = x, v_i = Z_i'\gamma]$$

- Now instead of being a number like LATE, the MTE: $\Delta^{MTE}(x, v)$ is a function of v
- It is the average impact of receiving a treatment for everyone with the same $Z_i'\gamma$.

One quantity to rule them all: MTE

But what is $Z_i'\gamma$?

- For any single index model we can rewrite

$$T_i = \mathbf{1}(v_i \leq Z_i'\gamma) = \mathbf{1}(u_{is} \leq F(Z_i'\gamma)) \text{ for } u_{is} \in [0,1]$$

- F is just the cdf of v_i: (could be logit or probit, or anything else).
- Use the CDF to write things as a uniform distribution.
- Now we can write propensity score $P(Z_i) = Pr(T_i = 1|Z_i) = F(Z_i'\gamma)$.

Alternative Definitions

- Heckman (1997) also shows the relationship to LATE:

$$\Delta^{MTE}(x,z) = \lim_{z' \to z} \text{Wald}(z, z', x)$$

- A function where we evaluate the limit of LATE at each value of z
- The alternative way to define the MTE is as Local IV:

$$\Delta^{LIV}(x,p) = \frac{\partial E[Y_i | X_i = x, P(Z_i) = p]}{\partial p}$$

- How does the outcome Y_i change as we push one more person into treatment (via the Propensity Score)

MTE: Derivation

Now we can write,

$$\begin{aligned} Y(0) &= \gamma_0' X + U_0 \\ Y(1) &= \gamma_1' X + U_1 \end{aligned}$$

$P(T = 1|Z) = P(Z)$ works as our instrument with two assumptions:

1. $(U_0, U_1, u_s) \perp P(Z)|X$. (Exogeneity)
2. Conditional on X there is enough variation in Z for $P(Z)$ to take on all values $\in (0, 1)$.
    - This is much stronger than typical relevance condition. Much more like the special regressor method we will discus later.

MTE: Derivation

Now we can write,

$$
\begin{aligned}
Y &= \gamma_0' X + T(\gamma_1 - \gamma_0)' X + U_0 + T(U_1 - U_0) \\
E[Y|X, P(Z) = p] &= \gamma_0' X + p(\gamma_1 - \gamma_0)' X + E[T(U_1 - U_0)|X, P(Z) = p]
\end{aligned}
$$

Observe $T = 1$ over the interval $u_s = [0, p]$ and zero for higher values of u_s. Let $U_1 - U_0 \equiv \eta$.

$$
\begin{aligned}
E[T(U_1 - U_0)|P(Z) = p, X] &= \int_{-\infty}^{\infty} \int_0^p (U_1 - U_0) f((U_1 - U_0)|U_s = u_s) du_s d(U_1 - U_0) \\
E[T(\eta)|P(Z) = p, X] &= \int_{-\infty}^{\infty} \int_0^p \eta f(\eta|U_s = u_s) d\eta d u_s
\end{aligned}
$$

MTE: Derivation

Recall:

$$E[Y|X, P(Z) = p] = \gamma_0' X + p(\gamma_1 - \gamma_0)' X + E[T(U_1 - U_0)|X, P(Z) = p]$$

And the derivative:

$$\begin{aligned}\Delta^{MTE}(p) = \frac{\partial E[Y|X, P(Z) = p]}{\partial p} &= (\gamma_1 - \gamma_0)' X + \int_{-\infty}^{\infty} \eta f(\eta | U_s = p) d\eta \\ &= \underbrace{(\gamma_1 - \gamma_0)' X}_{ATE(X)} + E[\eta | u_s = p]\end{aligned}$$

What is $E[\eta|u_s = p]$? The expected unobserved gain from treatment of those people who are on the treatment/no-treatment margin $P(Z) = p$.

Everything is an MTE

Calculate the outcome given (X, Z) (actually X and $P(Z) = p$).

$$\Delta^{ATE}(x, T=1) = E\left(\Delta^{MTE}|X=x\right)$$
$$\Delta^{TT}(x, P(z), T=1) = E\left(\Delta^{MTE}|X=x, u_s \leq P(z)\right)$$
$$\Delta^{LATE}\left(x, P(z), P\left(z'\right)\right) = E\left(\Delta^{MTE}|X=x, P\left(z'\right) \leq u_s \leq P(z)\right)$$

ATE : This one is obvious. We treat everyone!

$$\int_{-\infty}^{\infty} \Delta^{MTE}(p) = (\gamma_1 - \gamma_0)'X + \underbrace{\int_{-\infty}^{\infty} E(\eta|u_s)d\,u_s}_{0}$$

ATT: Treat only those with a large enough propensity score $P(z) > p$:

$$TT(x) = \int_{-\infty}^{\infty} \Delta^{MTE}(p, x)\frac{Pr(P(Z|X) > p)}{E[P(Z|X)]}d\,p$$

Everything is an MTE

LATE: Integrate over the compliers:

$$LATE(x,z,z') = \frac{1}{P(z)-P(z')}\int_{P(z')}^{P(z)} \Delta^{MTE}(p,x)$$

OLS and IV are hard:

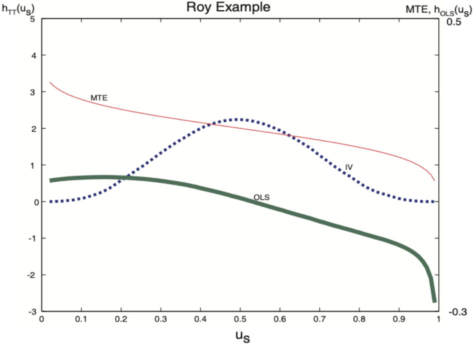
$$w^{IV}(u_s) = \left[E\left(P(Z)|P(Z)>u_s\right) - E(P(Z))\right]\frac{E(P(Z))}{\mathrm{Var}(P(Z))}$$

$$w^{OLS}(u_s) = 1 + \frac{E\left(U_1|U_S=u_s\right)h_1 - E\left(U_0|U_S=u_s\right)h_0}{\Delta^{MTE}(u_s)}$$

$$h_1 = \frac{E\left(P(Z)|P(Z)>u_s\right)}{E(P(Z))}, \quad h_0 = \frac{E\left(P(Z)|P(Z)<u_s\right)}{E(P(Z))}$$

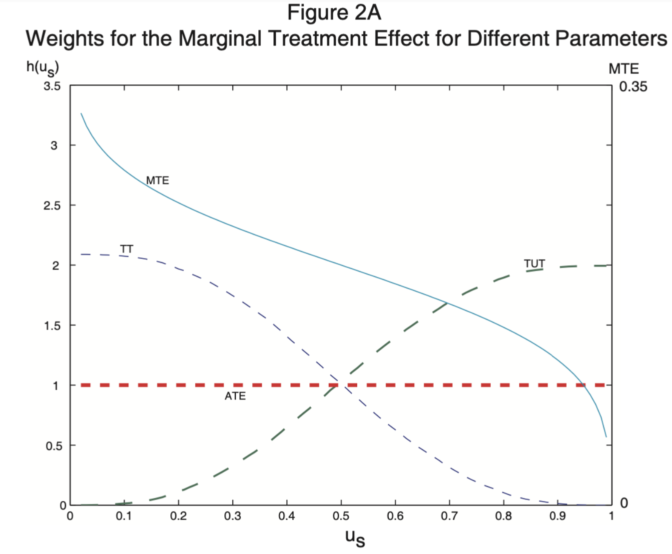
MTE: Parametric Example

Figure 2B
Marginal Treatment Effect vs Linear Instrumental Variables and Ordinary Least Squares Weights
Roy Example



$$\begin{aligned} \ln Y_1 &= \alpha + \bar{\beta} + U_1 \\ \ln Y_0 &= \alpha + U_0 \\ S &= 1 \text{ if } Z - U_S > 0 \end{aligned} \qquad \begin{aligned} U_1 &= \sigma_1 \varepsilon \\ U_0 &= \sigma_0 \varepsilon \\ U_S &= \sigma_S \varepsilon \end{aligned} \qquad \begin{aligned} \alpha &= 0.67 \\ \bar{\beta} &= 0.2 \\ & \varepsilon \tilde{\ } N(0,1) \\ & Z \tilde{\ } N(-0.0026, 0.27) \\ \sigma_1 &= 0.012 \\ \sigma_0 &= -0.05 \end{aligned}$$

MTE: Parametric Example



How to Estimate an MTE

Easy?

1. Estimate $P(Z) = Pr(T = 1|Z)$ nonparametrically (include exogenous part of X in Z).
2. Nonparametric regression of Y on X and $P(Z)$ (polynomials?)
3. Differentiate w.r.t. $P(Z)$
4. plot it for all values of $P(Z) = p$.

So long as $P(Z)$ covers $(0, 1)$ then we can trace out the full distribution of $\Delta^{MTE}(p)$.

Carneiro, Heckman and Vytlacil (AER 2010)

- Estimate returns to college (including heterogeneity of returns).
- NLSY 1979
- $Y = \log(wage)$
- Covariates X: Experience (years), Ability (AFQT Score), Mother's Education, Cohort Dummies, State Unemployment, MSA level average wage.
- Instruments Z: College in MSA at age 14, average earnings in MSA at 17 (opportunity cost), avg unemployment rate in state.

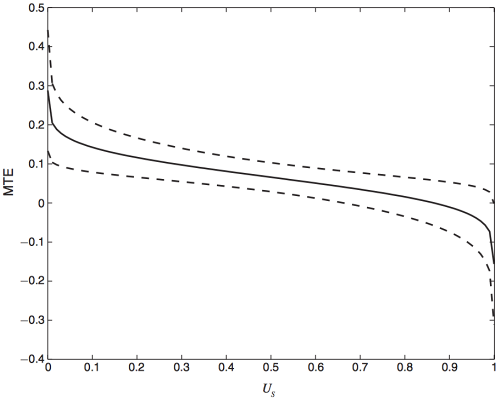


FIGURE 1. MTE ESTIMATED FROM A NORMAL SELECTION MODEL

Notes: To estimate the function plotted here, we estimate a parametric normal selection model by maximum likelihood. The figure is computed using the following formula:

$$\Delta^{\mathrm{MTE}}(\mathbf{x}, u_S) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) - (\sigma_{1V} - \sigma_{0V})\,\Phi^{-1}(u_S),$$

where σ_{1V} and σ_{0V} are the covariances between the unobservables of the college and high school equation and the unobservable in the selection equation; and **X** includes experience, current average earnings in the county of residence, current average unemployment in the state of residence, AFQT, mother's education, number of siblings,

TABLE 4— TEST OF LINEARITY OF $E(Y|\mathbf{X}, P = p)$ USING POLYNOMIALS IN P; AND TEST OF EQUALITY OF LATEs OVER DIFFERENT INTERVALS (H_0: $LATE^j\,(U_S^{Lj}, U_S^{Hj}) - LATE^{j+1}\,(U_S^{Lj+1}, U_S^{Hj+1}) = 0$)

Panel A. Test of linearity of $E(Y|\mathbf{X}, P = p)$ using models with different orders of polynomials in P[a]

Degree of polynomial for model	2	3	4	5
p-value of joint test of nonlinear terms	0.035	0.049	0.086	0.122
Adjusted critical value		0.057		
Outcome of test		Reject		

Panel B. Test of equality of LATEs (H_0: $LATE^j\,(U_S^{Lj}, U_S^{Hj}) - LATE^{j+1}\,(U_S^{Lj+1}, U_S^{Hj+1}) = 0$)[b]

Ranges of U_S for LATE^j	(0, 0.04)	(0.08, 0.12)	(0.16, 0.20)	(0.24, 0.28)	(0.32, 0.36)	(0.40, 0.44)
Ranges of U_S for LATE^{j+1}	(0.08, 0.12)	(0.16, 0.20)	(0.24, 0.28)	(0.32, 0.36)	(0.40, 0.44)	(0.48, 0.52)
Difference in LATEs	0.0689	0.0629	0.0577	0.0531	0.0492	0.0459
p-value	0.0240	0.0280	0.0280	0.0320	0.0320	0.0520
Ranges of U_S for LATE^j	(0.48, 0.52)	(0.56, 0.60)	(0.64, 0.68)	(0.72, 0.76)	(0.80, 0.84)	(0.88, 0.92)
Ranges of U_S for LATE^{j+1}	(0.56, 0.60)	(0.64, 0.68)	(0.72, 0.76)	(0.80, 0.84)	(0.88, 0.92)	(0.96, 1)
Difference in LATEs	0.0431	0.0408	0.0385	0.0364	0.0339	0.0311
p-value	0.0520	0.0760	0.0960	0.1320	0.1800	0.2400
Joint p-value				0.0520		

Carneiro, Heckman and Vytlacil

Table 5—Returns to a Year of College

Model	Metric	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		
$Z^k_\alpha = Z^k + \alpha$	$\lvert \mathbf{Z}\gamma - V \rvert < e$	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)	

Notes: This table presents estimates of various returns to college, for the semiparametric and the normal selection models: average treatment effect (ATE), treatment on the treated (TT), treatment on the untreated (TUT), and different versions of the marginal policy relevant treatment effect (MPRTE). The linear IV estimate uses P as the instrument. Standard errors are bootstrapped (250 replications). See online Appendix Table A-1 for the exact definitions of the weights. See Table 1 for the weights for MPRTE. For more discussion of MPRTE, see Carneiro, Heckman, and Vytlacil (2010).

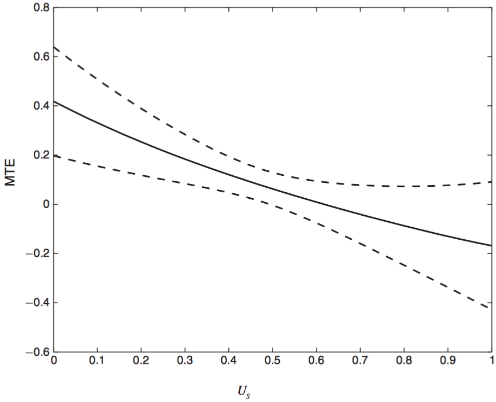


FIGURE 4. $E(Y_1 - Y_0 | \mathbf{X}, U_S)$ WITH 90 PERCENT CONFIDENCE INTERVAL—
LOCALLY QUADRATIC REGRESSION ESTIMATES

Notes: To estimate the function plotted here, we first use a partially linear regression of log wages on polynomials in $\mathbf{X}$, interactions of polynomials in $\mathbf{X}$ and P, and $K(P)$, a locally quadratic function of P (where P is the predicted probability of attending college), with a bandwidth of 0.32; $\mathbf{X}$ includes experience, current average earnings in the county of residence, current average unemployment in the state of residence, AFQT, mother's education, number of siblings, urban residence at 14, permanent local earnings in the county of residence at 17, permanent unemployment in the state of residence at 17, and cohort dummies. The figure is generated by evaluating by the derivative of (9)

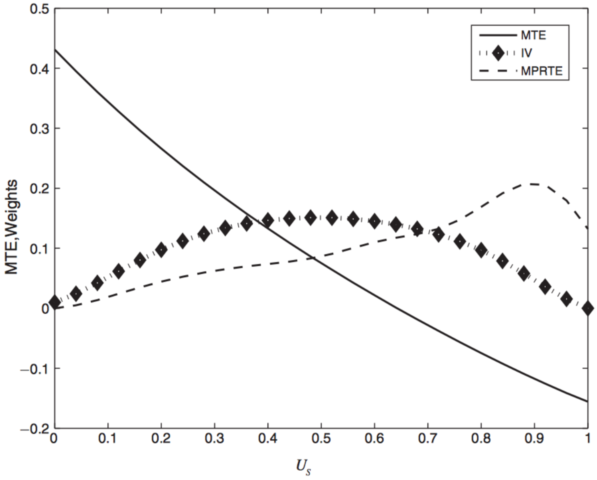


FIGURE 6. WEIGHTS FOR IV AND MPRTE

Note: The scale of the *y*-axis is the scale of the MTE, not the scale of the weights, which are scaled to fit the picture.