The Key Issue with Matching

- If individuals select into treatment based on unobservables, *CIA* will fail.

- The simple Roy Model illustrates this well:

$$\begin{aligned} Y_{id} &= \mu_d(X_i) + U_{id}, \quad d = 0,1 \\ D_i &= 1[Y_{i1} > Y_{i0}] \\ &= 1[U_{i1} - U_{i0} > \mu_0(X_i) - \mu_1(X_i)] \end{aligned}$$

- Those with high U_0 relative to U_1 (and consequently high Y_0 relative to Y_1) will tend not to be treated.

- In this case, it is very possible that $Pr[D_i|X_i, Y_{i0}, Y_{i1}] \neq Pr[D_i|X_i]$ (i.e. *CIA* fails).

IV

- Omit X for simplicity and remember the regression representation:

$$Y_i = \alpha + \Delta_i D_i + U_{0i} = \alpha + ATET\, D_i + \eta_i$$

where $\alpha = \mu_0$, $\eta_i = (U_{1i} - U_{0i} - E(U_{1i} - U_{0i}|D_i = 1))D_i$ $+U_{0i}$.

- A "natural" solution would be to look for an Instrumental Variable Z_i and rely on

$$\Delta^{IV} = \frac{cov(Y_i, Z_i)}{cov(D_i, Z_i)}$$

- If Z_i is binary, we get the *Wald estimator*:

$$\Delta^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

IV

When does IV estimate ATET?

1. $U_{0i} = U_{1i}$ ($\Rightarrow ATET = ATE$) and $E(U_{0i}|Z_i) = E(U_{0i})$. Then

$$\begin{aligned} & E(Y_i|Z_i = z) \\ = \; & \alpha + E(D_i|Z_i = z)ATET + E(U_{0i}|Z_i = z) \end{aligned}$$

   for $z = 0, 1$.

2. $U_{0i} \neq U_{1i}$. Then it is necessary that and

$$\begin{aligned} & E[U_{0i} + \{U_{1i} - U_{0i} - E(U_{1i} - U_{0i}|D_i = 1)\}D_i|Z_i] \\ = \; & E[U_{0i} + \{U_{1i} - U_{0i} - E(U_{1i} - U_{0i}|D_i = 1)\}D_i] \end{aligned}$$

   which essentially requires that the IV does not help predict the gain from treatment.

IV

- Angrist (AER, 1990) uses draft lottery as instrument for veteran status. Could be invalid for ATET if
  - Firms take into account lottery numbers in hiring
  - Workers take actions to avoid draft

- Moffitt (JASA, 1996) explores the validity of cross-section variation in welfare benefits as an instrument for participation in job training program
  - Could be invalid if anticipated benefits are correlated with welfare benefits

LATE

- Imbens and Angrist (ECMA, 1994) show that, even if the assumptions that would justify the IV for the estimation of the *ATET* are not valid, the IV estimator still identifies the Local Average Treatment Effect (LATE).

- The key assumption is *monotonicity*. Define the treatment allocation as a random variable indexed by $z \in \text{supp}(Z_i)$: $D_i = D_i(z)$. The **monotonicity** assumption holds if for any two values of the instrument $z_1 \neq z_2$ we have

$$D_i(z_1) \geq D_i(z_2) \text{ for all } i \text{ or } D_i(z_1) \leq D_i(z_2) \text{ for all } i.$$

  If people are more likely to participate when $Z_i = z_1$ then when $Z = z_2$, then anyone who would participate when $Z_i = z_1$ should also participate when $Z_i = z_2$. (Individuals are either *compliers* or *defiers*.)

LATE

- The other assumption employed by IA requires that (i) for every $z \in \text{supp}(Z_i)$, $(Y_{i1}, Y_{i0}, D_i(z)) \perp\!\!\!\perp Z_i$ and (ii) $E(D_i|Z_i = z)$ be a nontrivial function of z.
- (i) is an exclusion restriction: Z_i affects the outcomes only via D_i.
- (ii) means that the IV is relevant.

LATE

Then,

$$
\begin{aligned}
& E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = \\
= \quad & E[D_i(1)Y_{i1} + (1 - D_i(1))Y_{i0}|Z_i = 1] - E[D_i(0)Y_{i1} + (1 - D_i(0))Y_{i0}|Z_i = 0] \\
= \quad & E[D_i(1)Y_{i1} + (1 - D_i(1))Y_{i0}] - E[D_i(0)Y_{i1} + (1 - D_i(0))Y_{i0}] \text{ (by (i))} \\
= \quad & E[(D_i(1) - D_i(0))(Y_{i1} - Y_{i0})] \\
= \quad & E[Y_{i1} - Y_{i0}|D_i(1) - D_i(0) = 1]Pr(D_i(1) - D_i(0) = 1) - \\
& E[Y_{i1} - Y_{i0}|D_i(1) - D_i(0) = -1]Pr(D_i(1) - D_i(0) = -1)
\end{aligned}
$$

The IV estimates a weighted difference of the effect for those induced into treatment by Z and those induced out of treatment by Z.

If treatment effects are homogeneous this is the *ATE*, but in general it is rather meaningless.

LATE

Under monotonicity though

$$Pr(D_i(1) - D_i(0) = -1) = 0$$

(or $Pr(D_i(1) - D_i(0) = 1) = 0$) which then implies that

$$\begin{aligned} & E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] \\ = \; & E[Y_{i1} - Y_{i0}|D_i(1) - D_i(0) = 1]Pr(D_i(1) - D_i(0) = 1) \end{aligned}$$

and

$$\begin{aligned} & Pr(D_i(1) - D_i(0) = 1) \\ = \; & Pr(D_i(1) = 1 \cap D_i(0) = 0) \\ = \; & 1 - Pr(D_i(1) = 0) - Pr(D_i(0) = 1) \\ = \; & Pr(D_i(1) = 1) - Pr(D_i(0) = 1) \end{aligned}$$

LATE

- And we obtain that

$$\Delta^{IV} = E[Y_{i1} - Y_{i0} | D_i(1) - D_i(0) = 1]$$

- The IV estimator gives the average effect of the treatment for those induced into treatment by the instrument Z.
- If monotonicity can be justified, IV estimates the effect for a particular subset of the population (i.e. the compliers).
- This subset depends on the intrument and different instruments will esitmate different LATEs.

LATE

- In Angrist (AER, 1990) example, we get the effect of veteran status for the subset induced to enter the military by the draft lottery (excluding those who always join or never join, despite the draft).
- Angrist and Krueger (1994) study the effect of schooling on earnings using compulsory laws as instrument. This gives the treatment effect for the subset induced to enter school by the law (probably the “low level” schooling types).

Connecting it all: MTE

Consider the potential outcomes representation:

$$\begin{aligned} Y_i &= Y_{i1} D_i + Y_{i0}(1 - D_i) \\ Y_{id} &= \mu_d + U_{id}, \quad d = 0, 1 \\ D_i &= 1[\mu_D(Z_i) - U_{iD} \geq 0] \end{aligned}$$

where we explicity write how D_i depends on Z_i.

Assume that shocks are independent of Z_i and that U_{iD} is continuous.

Connecting it all: MTE

Write the propensity score as $P(z) = Pr(D_i = 1|Z_i = z)$ and assume there is full support $(0 < Pr(D_i = 1|Z_i) < 1)$.

If U_{iD} is continuous, we can assume without loss of generality that

$$D_i = 1[P(Z_i) - U_{iD} \geq 0],$$

where $U_{iD} \sim U[0, 1]$.

This is because

$$\underbrace{F_{U_D}(\mu_D(Z_i))}_{=P(Z_i)} \geq \underbrace{F_{U_D}(U_{iD})}_{\sim U[0,1]} \Leftrightarrow \mu_D(Z_i) \geq U_{iD}$$

Connecting it all: MTE

Define the conditional treatment effect parameters $ATE(P(Z_i))$ and $ATET(P(Z_i))$ as before (but conditioning now on the propensity score $P(Z_i)$). (Remember this is equivalent to conditioning on Z_i!)

Since Z_i may take more than two values, define $LATE(P(z), P(z'))$ is estimated by

$$\frac{E[Y_i|P(Z_i) = P(z)] - E[Y_i|P(Z_i) = P(z')]}{P(z) - P(z')}$$

The monotonicity assumption is hidden in the definition of D_i. (Can you see that?)

Connecting it all: MTE

Define the Marginal Treatment Effect as

$$MTE(u) \equiv E(\Delta_i | U_{Di} = u).$$

1. If $U_{Di} = P(z)$ the person has unobservables that make him or her indifferent between participating and not participating.
2. Those with U_{Di} close to zero, unobservables that make them most inclined to participate. The *MTE* at low values give the treatment effect on those people.
3. Those with U_{Di} close to one, unobservables that make them least inclined to participate. The *MTE* at low values give the treatment effect on those people.

Connecting it all: MTE

Note that

$$\begin{aligned} & E(Y_i|P(Z_i) = P(z)) \\ = \quad & P(z)E(Y_{i1}|P(Z_i) = P(z), D_i = 1) + (1 - P(z))E(Y_{i1}|P(Z_i) = P(z), D_i = 0) \\ = \quad & P(z)\frac{\int_0^{P(z)} E(Y_{i1}|U_{Di} = u)du}{P(z)} + (1 - P(z))\frac{\int_{P(z)}^1 E(Y_{i0}|U_{Di} = u)du}{1 - P(z)} \end{aligned}$$

and the $LATE(P(z), P(z'))$ estimator becomes

$$\begin{aligned} & \frac{\int_{P(z')}^{P(z)} E(Y_{i1}|U_{Di} = u)du - \int_{P(z')}^{P(z)} E(Y_{i0}|U_{Di} = u)du}{P(z) - P(z')} \\ = \quad & E(\Delta_i|P(z') \leq U_{Di} \leq P(z)) \\ = \quad & E(MTE(U_{Di})|P(z') \leq U_{Di} \leq P(z)) \end{aligned}$$

which is the average treatment effect those that would not participate if $Z_i = z'$ but would participate if $Z_i = z$ (i.e., the compliers).

Connecting it all: MTE

Using similar manipulations, we can show that

$$
\begin{aligned}
ATE &= \int_0^1 MTE(u)du \\
ATET &= \int_0^1 MTE(u)h_{ATET}(u)du \\
LATE(u', u'') &= \frac{\int_{u'}^{u''} MTE(s)ds}{u'' - u'}
\end{aligned}
$$

where $h_{ATET}(u) = 1 - F_{P(Z)}(u)/\int[1 - F_{P(Z)}(t)]dt$, a function that gives more weight to those more inclined to participate.

Connecting it all: MTE

Heckman and Vytlacil (2005, EMA) show also that

$$IV = \int_0^1 MTE(u) h_{IV}(u) du$$
$$OLS = \int_0^1 MTE(u) h_{OLS}(u) du$$

where

$$h_{IV}(u) = Pr(P(Z) > u) \times \frac{E[P(Z)|P(Z) > u] - E[P(Z)]}{Var(P(Z))}$$
$$h_{OLS} = 1 + \frac{E[U_1|U_D = u]h_1 - E[U_0|U_D = u]h_0}{\Delta^{MTE}(u)}$$

where $h_1 = E(P(Z)|P(Z) > u)/E(P(Z))$ and $h_0 = E(P(Z)|P(Z) < u)/E(P(Z))$.

Connecting it all: MTE

If we know the $MTE(\cdot)$, we can reconstruct all of the above parameters (plus others). How can we do that?

Notice that

$$
\begin{aligned}
& MTE(P(z)) \\
= & \lim_{P(z')\to P(z)} LATE(P(z), P(z')) \\
= & \lim_{P(z')\to P(z)} \frac{E[Y_i|P(Z_i) = P(z)] - E[Y_i|P(Z_i) = P(z')]}{P(z) - P(z')} \\
= & \frac{\partial E[Y_i|P(Z_i) = P(z)]}{\partial P(z)} \\
\equiv & \Delta^{LIV}(P(z))
\end{aligned}
$$

Connecting it all: MTE

In practice:

1. Estimate the propensity score (parametrically, semi-parametrically or non-parametrically);
2. Estimate $\partial E(Y_i|P(Z_i) = \cdot)/\partial P(z)$ non-parametrically (e.g., via slope of local linear regression estimate of $E(Y_i|P(Z_i) = \cdot)$).
3. Evaluating this function for different values of $P(z)$ traces out the *MTE* on the support of $P(Z_i)$.
4. The different estimands $ATET$, ATE and $LATE$ can be obtained by integrating under different regions of the *MTE* function.

Returns to College, CHV (AER, 2011)

- Object of study: Estimate the returns to college and analyse the heterogeneity of these returns.
- Data: NLSY 1979
- Outcome variable: log wages
- Conditioning variables (X): years of experience, cognitive ability (AFQT), maternal education, cohort dummies, log average earnings in SMSA, and average unemployment rate in state.
- Instruments (Z): Presence of a four year public college in SMSA at age 14, log average earnings in the SMSA when 17 (opportunity cost), average unemployment in state

Returns to College, CHV (AER, 2011)

Table A-3
Sample Statistics

	$S = 0$ ($N = 882$)	$S = 1$ ($N = 865$)
Log of Average Hourly Wage 1989-1993	2.2089	2.5496
	(0.4412)	(0.4959)
Years of Actual Experience	10.1042	6.8404
	(3.1260)	(3.2522)
Corrected AFQT	-0.0446	0.9515
	(0.8673)	(0.7498)
Mother's Years of Schooling	11.3083	12.9121
	(2.1056)	(2.2789)
Number of Siblings	3.2630	2.5849
	(2.0842)	(1.6450)
Urban Residence at 14	0.6995	0.7895
	(0.4587)	(0.4078)
Local Log Earnings in 1991	10.2645	10.3220
	(0.1597)	(0.1660)
Local Unemployment in 1991	6.7971	6.8226
(in %)	(1.3310)	(1.1983)
Presence of a 4 Year College at 14	0.4625	0.5884
	(0.4988)	(0.4924)
Local Log Earnings at 17	10.2780	10.2736
	(0.1619)	(0.1651)
Local Unemployment Rate at 17	7.0804	7.0846
(in %)	(1.7846)	(1.8449)
Tuition in 4 Year Public Colleges at 17	22.0164	21.1105
(in $100)	(7.8730)	8.0683
"Permanent" Local Log Earnings at 17	10.2673	10.2991
	(0.1798)	(0.1945)
"Permanent" Local Unemployment Rate at 17	6.2942	6.2077
	(1.0156)	(0.9536)

Returns to College, CHV (AER, 2011)

- Estimate a logit model for college participation on cohort dummies and polynomial terms of the instruments.
- The probability of college attendance is

$$P(Z) = \frac{1}{1 + exp(-Z^{\top}\beta)}$$

- Average derivatives are

$$N^{-1} \sum \frac{\partial \hat{P}(Z)}{\partial Z_j} = N^{-1} \sum \left[\hat{P}(Z)(1 - \hat{P}(Z)) \frac{\partial Z^{\top} \hat{\beta}}{\partial Z_j} \right]$$

Returns to College, CHV (AER, 2011)

Table 3—College Decision Model: Average Marginal Derivatives

	Average derivative
Controls (**X**)	
Corrected AFQT	0.2826
	(0.0114)***
Mother's years of schooling	0.0441
	(0.0059)***
Number of siblings	–0.0233
	(0.0068)***
Urban residence at 14	0.0340
	(0.0274)
"Permanent" local log earnings at 17	0.1820
	(0.0941)**
"Permanent" state unemployment rate at 17	0.0058
	(0.0165)
Instruments (**Z**)	
Presence of a college at 14	0.0529
	(0.0273)**
Local log earnings at 17	–0.2687
	(0.1008)***
Local unemployment rate at 17 (in percent)	0.0149
	(0.0100)
Tuition in 4 year public colleges at 17 (in $100)	–0.0027
	(0.0017)*

Returns to College, CHV (AER, 2011)

Table A-8 - OLS and IV Estimates of the Return to a Year of College

	OLS	IV					
		Presence of a college	Local earnings of unskilled workers	Local unemployment	Average Tuition in 4 year colleges	Two stage least squares using all instruments	$P(Z)$ as the instrument
Return to College	0.0836 (0.0068)	0.0576 (0.0727)	0.1736 (0.0788)	0.1582 (0.1897)	0.1211 (0.0909)	0.1253 (0.0433)	0.0951 (0.0386)

Notice:

1. How the results vary by instrumental variable and
2. How IV is larger than OLS.

Returns to College, CHV (AER, 2011)

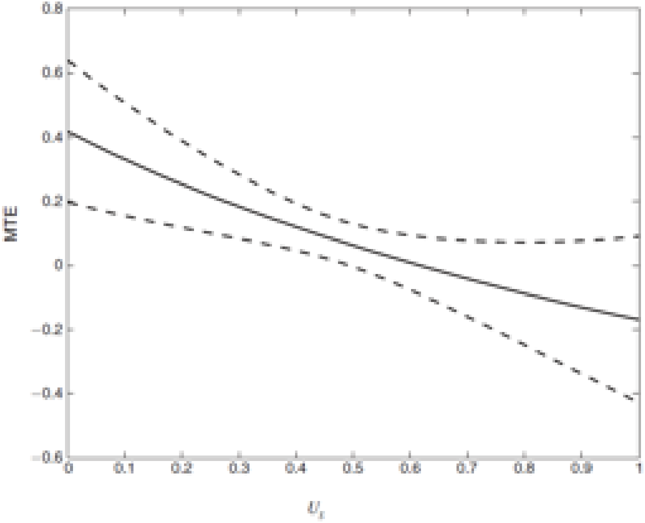

Figure 4. $E(Y_1 - Y_0 | X, U_S)$ with 90 Percent Confidence Interval—
Locally Quadratic Regression Estimates