IV and Marginal Treatment Effects

Framework

- Treatment D_i, Potential outcome if treated Y_{1i}, potential outcome if not treated Y_{0i}

$$Y_{0i} = \mu_0(X_i) + U_{0i} \tag{1}$$

$$Y_{1i} = \mu_1(X_i) + U_{1i} \tag{2}$$

$$D_i^* = \mu_D(X_i, Z_i) - V_i \tag{3}$$

$$D_i = 1 \text{ if } D_i^* \geq 0, \qquad D_i = 0 \text{ otherwise}, \tag{4}$$

- D_i^* is the latent propensity to take the treatment
- V_i iid error term indicating unobserved heterogeneity in the propensity for treatment
  $D_i = 1$ if $P(X_i, Z_i) \geq U_{Di} = F_V(V_i)$

IV

- Different evaluation parameters are an average over parts of the distribution of impacts.
    - The ATE averages over the entire distribution
    - The ATT averages over the distribution of impacts for those who are somehow allocated into treatment
    - LATE averages over the distribution of impacts for those who switch into treatment as a result of a reform or more precisely, as a result of a change of the value of some instrument affecting decisions to participate.
- They all represent an aggregation over different margins
- They are not comparable and they are difficult to interpret from the general perspective
- As a unifying parameter, Heckman and Vytlacil (Econometrica, 2005) defined the MARGINAL TREATMENT EFFECT

IV

- Covariate specific IV (binary instrument)

$$\text{Wald}(x) = \frac{E[Y_i|Z_i=1,X_i=x]-E[Y_i|Z_i=0,X_i=x]}{E[D_i|Z_i=1,X_i=x]-E[D_i|Z_i=0,X_i=x]}. \tag{11}$$

  with $D_i = Z_i D_{1i} + (1 - Z_i) D_{0i}$

- IV assumptions: (i) Independence (ii) Existence of first stage (iii) Monotonicity

- Under these assumptions, IV identifies LATE

$$\begin{aligned} \text{LATE}(x) &= E[Y_{1i}-Y_{0i}|D_{1i}>D_{0i},X_i=x] \\ &= \mu_1(X_i)-\mu_0(X_i)+E[U_{1i}-U_{0i}|D_{1i}>D_{0i},X_i=x] \end{aligned} \tag{12}$$

$$IV = \sum_{x\in\mathcal{X}} \omega(x)\text{LATE}(x)$$

- Saturated model: weights are the contribution of the observations with $Xi = x$ to the variance of the first-stage fitted values

- **compliers**: individuals whose potential treatment status changes in response to the extra encouragement for treatment as the instrument changes

- Eg: if the instrument is a dummy variable for a college being located nearby an individual's place of residence, then the LATE is the treatment effect averaged over the group of individuals who attend college if living nearby a college, but who do not attend college if the college is far away
  - These might be people who are constrained in their resources to take up college far away from their place of residence
  - Or people who feel that their return from college would not warrant the cost of attending college in a faraway location

- IV is not informative on the effect for the subgroup of always-takers (defined by $D_{1i} = D_{0i} = 1$) and never-takers (defined by $D_{1i} = D_{0i} = 0$)

- In education example: always-takers could be individuals who estimate their returns as high enough in order to warrant college attendance even in a faraway location,

- And never-takers would not attend college even in a nearby location

- Important difference between LATE and other treatment effect parameters
- ATE, ATT, ATU are parameters that answer economic policy questions and are defined independently of any instrument
- LATE is defined by IV used and does not necessarily answer an economic policy question and does not necessarily represent a treatment parameter for an economically interesting group of the population
- Special cases where LATE coincides with one of the TE parameters
- Example 1: Oreopoulos (2006) uses an increase of the compulsory school-leaving age as a binary instrument
- there were no never-takers: all untreated are compliers so LATE=ATU

- Example 2: Chetty et al. (2016) who evaluate the long-run effects of the Moving To Opportunity (MTO) experiment
- Offered randomly selected families housing vouchers to move from high-poverty housing projects to lower-poverty neighborhoods
- Because nobody in the control group had access to the treatment, there were no always-takers, implying that all treated are compliers
- In general, RCTs with one-sided compliance also identify ATT or ATU

Continuous Instrument

Continuous Instrument

- z continuous: one can exploit any pair of values z and z' of Z_i as a binary instrument calculating the covariate specific IV estimator

$$\text{Wald}(z, z', x) = \frac{E[Y_i|Z_i = z, X_i = x] - E[Y_i|Z_i = z', X_i = x]}{E[D_i|Z_i = z, X_i = x] - E[D_i|Z_i = z', X_i = x]}. \tag{13}$$

- With IV assumptions it identifies LATE for compliers with a change in the instrument from z to z'
- Monotonicity needs to hold between all pairs of values z and z': All individuals whose treatment status is affected by a change of the instrument from z to z' have to either all be shifted into treatment, or all be shifted out of treatment
- latent index choice model in Eq (3) and (4) ensure this is true

Continuous Instrument

- Assuming a move from z to z' shift people into treatment

$$\text{LATE}(z, z', x) = E[Y_{1i} - Y_{0i} | D_{zi} > D_{z'i}, X_i = x]. \tag{14}$$

- where D_{zi} is a binary indicator for the potential treatment status of individual i for instrument value $Z_i = z$
- In latent index model: $D_{Zi} > D_{Z'i} \iff P(Z') < U_D < P(Z)$
- Compliers are individuals with intermediate values of the "distaste" for treatment, such that they do not choose treatment when faced with a propensity score value of $P(z')$, but they choose treatment when faced with the higher value $P(z)$

Illustration

- Assuming a move from z to z' shift people into treatment

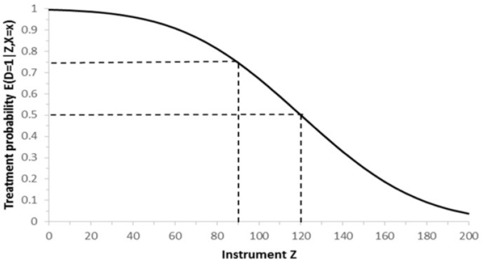


Fig. 1. Treatment probability as a function of a continuous instrument. *Notes:* Based on hypothetical data, the figure shows the effect of a continuous instrument Z on the probability of treatment in a sample with fixed covariates ($E[D = 1, Z, X = x]$). For example, the horizontal axis could represent distance to college and the vertical axis could represent the probability to attend college. *Data source:* Simulated hypothetical data.

- A reduction of the instrument from $Z_i = 120$ to $Z_i = 90$ raises the probability of treatment from $P(120) = .5$ to $P(90) = .75$

- This shifts individuals with $.5 < U_D < .75$ into treatment

- The associated LATE would thus be the treatment effect for this subgroup
- In practice: discretize the range of Z

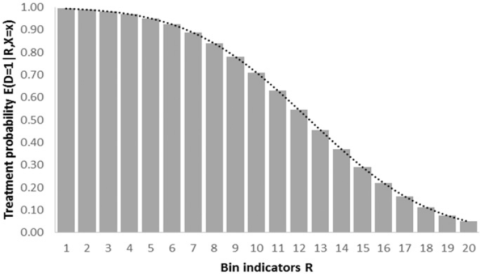


Fig. 2. Treatment probability in discrete bins of a continuous instrument. *Notes:* Based on hypothetical data, the bins in this figure show the probability of treatment in a sample with fixed covariates ($E[D=1,R,X=x]$) as a function of a discrete variable R, which has been generated by grouping the values of the continuous instrument depicted in Fig. 1 into 20 equally spaced bins. The dotted line reproduces the function depicted in Fig. 1. *Data source:* Simulated hypothetical data.

- $LATE(r, r', x)$

$$\frac{E[Y_i|R_i=r,X_i=x]-E[Y_i|R_i=r\prime,X_i=x]}{E[D_i|R_i=r,X_i=x]-E[D_i|R_i=r\prime,X_i=x]}$$

Aggregatin pairwise LATEs

- Efficient way: by 2SLS, using group indicator dummies for the values of R_i as instruments, fully saturating the first and second stage in the covariates, and interacting the instruments in the first stage with the covariates

- This gives variance-weighted average of covariate-specific LATEs

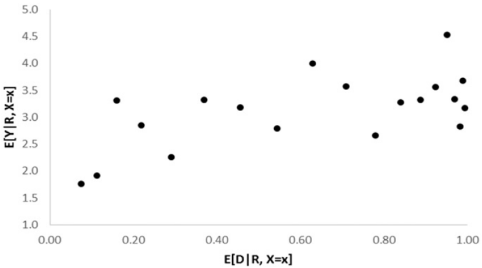


Fig. 3. Grouped data IV. *Notes:* Based on hypothetical data, the figure plots the average outcome against the average treatment probability in a sample with fixed covariates for 20 groups, which are equal to the bins depicted in Fig. 2 and correspond to 20 equally sized bins of an underlying continuous instrument. Grouped data IV can be visualized as fitting a line through these points. *Data source:* Simulated hypothetical data.

- Interpretation is complex compared to binary instrument: overall IV effect is now representative for compliers with changes between all values of the instrument, with different weights attached to groups of compliers at different pairs of values
- Aggregate IV may also hide interesting information, such as which pairs of values of the instrument shift a particularly large group of individuals or a group of individuals with particularly large treatment effects, into treatment

Margina Treatment Effect (MTE)

MTE

- MTE is the effect of a treatment on the marginal individual entering treatment
- It provides a way to interpret several evaluation parameters
- They will provide a bridge between structural and treatment effect parameters and allow us to understand the way they are related

MTE: Formal definition

- Consider a discrete treatment T
- The rule allocating to treatment may be written as: $T = 1(v_i \leq Z_i'\gamma)$
- For a particular value of $Z_i'\gamma$, the marginal individual is the one with $v_i = Z_i'\gamma$
- Individual take treatment if propensity scores exceeds quantile of distribution of of v_i at which it is located
- Now consider the effect of treatment for the ith individual $\beta_i = Y_i^1 - Y_i^0$
- MTE is defined: $MTE(Z_i'\gamma) = E(\beta_i | v_i = Z_i'\gamma)$
- Thus the MTE is the average impact for the marginal individual receiving treatment among those with value of the index equal to $Z_i'\gamma$
- It turns out that all treatment effect parameters we have looked at can be written as weighted averages of this parameter

MTE: A justification

- Suppose we think of a very simple model of College attendance choice
- The Cost of attending is $C_i = W_0 + Z_i'\gamma + u_{si}$ where W_0 is the opportunity cost in lost earnings and $Z_i'\gamma + u_{si}$ represents the direct costs
- The benefits in the simplest form are $L(W_i^1 - W_i^0)$ where L represents a lifecycle factor and depends on the discount rate
- An individual will go to College if $L(W_i^1 - W_i^0) \geq W_0 + Z_i'\gamma + u_{si}$
- The marginal individual satisfies the condition $u_{si} = L(W_i^1 - W_i^0) - W_0 - Z_i'\gamma$
- This shows how the allocation to treatment will depend on the returns and why conditioning on u_s will give us the treatment effect for the marginal individual at a given Z_i

- It's convenient to rewrite the treatment model as: $T = \mathbb{1}(u_{si} \leq F(Z_i'\gamma))$, where u_{si} is now uniform [0,1]
- This can be done by defining $U_s \equiv F(V)$ where $F(.)$ is the distribution of V
- Thus $F(Z'\gamma) = Pr(T = 1|Z) = P(Z)$
- Define

$$\begin{aligned} Y_0 &= \gamma_0' X + U_0 \\ Y_1 &= \gamma_1' X + U_1 \end{aligned}$$

- Consider the observed outcome: $Y = \gamma_0' X + T(\gamma_1 - \gamma_0)' X + U_0 + T(U_1 - U_0)$

- We are now going to think of the probability of assignment to treatment $P(Z)$ as an instrument
- Assume that $(U_0, U_1, U_s) \perp P(Z)|X$
- We will also make the assumption that given X, $P(Z)$ has continuous support in the open interval (0,1):
- This means that the excluded variables Z vary sufficiently for any fixed value of X to make the treatment assignment probability vary anywhere between 0 and 1

- Now take the expected value of the outcome given the instrument $P(Z) = p$ and X
  $E(Y|X, P(Z) = p) = \gamma_0' X + p(\gamma_1 - \gamma_0)' X + E[T(U_1 - U_0)|X, P(Z) = p]$
- $T = 1$ over the interval for $u_s = [0, p]$ and 0 for higher values of u_s and u_s is uniform

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

- We can now write the MTE as

$$
\begin{aligned}
\Delta^{MTE}(p) &= \frac{\partial E(Y|X, P(Z)=p)}{\partial p} \\
&= (\gamma_1 - \gamma_0)'X + \int_{-\infty}^{+\infty} (U_1 - U_0) f((U_1 - U_0)|U_s = u_s) d(U_1 - U_0) \\
&= (\gamma_1 - \gamma_0)'X + E(U_1 - U_0|U_s = p)
\end{aligned}
$$

- $E(U_1 - U_0|U_s = p)$ is the average unobserved gain of treatment for those whose unobserved characteristics make them indifferent between treatment or not at $P(Z) = p$

- The above process suggests an estimation procedure:
- Estimate the nonparametric regression of the outcome variable Y on X and on P(Z)
- This can be achieved by fitting Y on polynomials of X and P(Z)
- Differentiate the result with respect to P(Z)
- If P(Z) indeed varies from (0, 1), i.e. has full support, then it will be possible to estimate the marginal treatment effect

MTE and ATE

- We can now define all parameters of interest as a function of the MTE
- Averaging over all marginal individuals we obtain

$$
\begin{aligned}
ATE(X) &= \int_{-\infty}^{+\infty} \Delta^{MTE}(p)dp = (\gamma_1 - \gamma_0)'X + \int_{-\infty}^{+\infty} E(U_1 - U_0|u_s)du_s \\
&= (\gamma_1 - \gamma_0)'X
\end{aligned}
$$

- This is the effect of assigning treatment randomly to everyone of type X, assuming full compliance, and ignoring general equilibrium effects

MTE and LATE

- Now consider LATE. Here for any given X, P(Z) takes two values, say $P(Z) = b(X)$ or $P(Z) = a(X)$ with say $a > b$
- So lets think of $a(X) - b(X)$ as the policy induced change in the treatment probability for someone with characteristics X
- All those with unobserved propensity to be assigned to treatment such that $b(X) \leq u_s \leq a(X)$ will now switch into treatment under this policy

$$
\begin{aligned}
LATE(X) &= \frac{1}{a(X) - b(X)} \int_{b(X)}^{a(X)} \Delta^{MTE}(p)dp \\
&= (\gamma_1 - \gamma_0)'X + \frac{1}{a(X) - b(X)} \int_{b(X)}^{a(X)} E(U_1 - U_0 | u_s) du_s
\end{aligned}
$$

- The value of LATE will depend on the interval over which we integrate, i.e. it will depend on which margin the specific policy tend to shift into treatment
- This of course depends on the instrument used!

MTE and LATE

- Heckman and Vytlacil (2005) show that every estimator can be written as a weighted average of the MTE.
- For example LATE is the average MTE with weights $\frac{1}{a(X)-b(X)}$ in the range $(a(X), b(X))$ and zero everywhere else
- Consider now the treatment parameter on the treated. This can be written as:

$$\beta^{TT}(X) = \int_{-\infty}^{+\infty} \Delta^{MTE}(p) \left[\frac{Pr(P(Z|X) > p)}{E(P(Z|X))} \right] dp$$

- Thus the higher the probability of participating the larger the weight on the overall effect

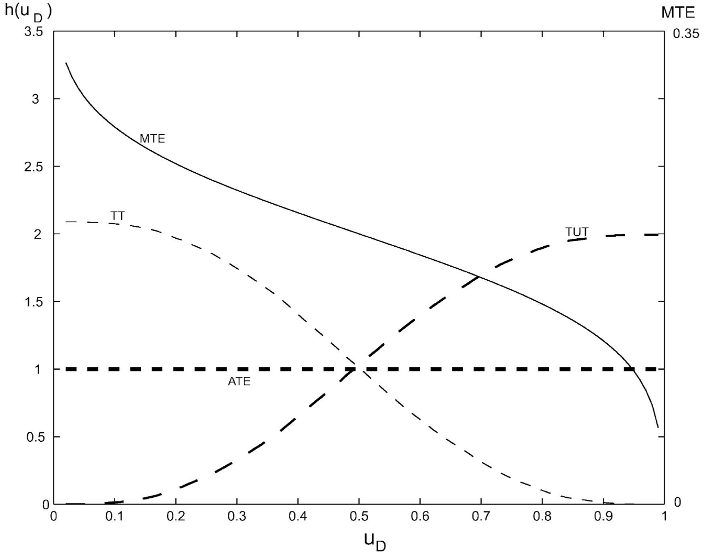


FIGURE 1A.—Weights for the marginal treatment effect for different parameters.

Application: Marginal Returns to Education

- Caneiro, Heckman and Vytlacil (AER, 2011) estimates marginal returns to college for individuals induced to enroll in college by different marginal policy changes
- Aim: analyse the heterogeneity of returns to education
- 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 rate in State

- Estimate a logit model for College participation on cohort dummies and on polynomial terms of the instruments
- The Probability of College attendance is:

$$P(Z) \equiv Pr(T = 1|Z) = \frac{1}{1 + exp(-Z'\beta)}$$

- The average derivatives are then:

$$\frac{1}{N} \sum_{sample} \left[\frac{\partial Pr(T = 1|Z)}{\partial AFQT} \right] = \frac{1}{N} \sum_{sample} \left[P(Z)(1 - P(Z)) \frac{\partial Z'\beta}{\partial AFQT} \right]$$

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	0.0149
(in percent)	(0.0100)
Tuition in 4 year public colleges at 17	−0.0027
(in $100)	(0.0017)*
Test for joint significance of instruments: p-value	0.0001

Support and identification: How does the probability of going to College differ between those who go to College and those who do not?

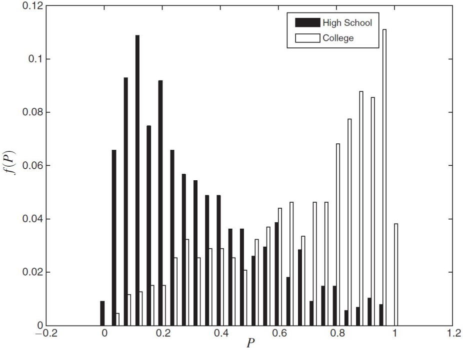


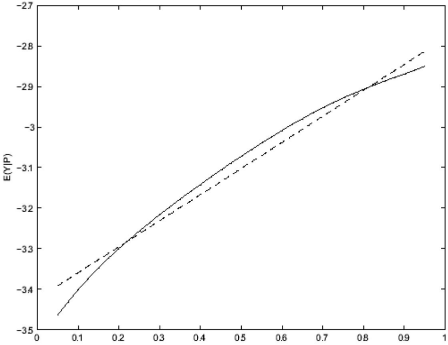
FIGURE 3. SUPPORT OF P FOR $S = 0$ AND $S = 1$

Notes: P is the estimated probability of going to college. It is estimated from a logit regression of college attendance on corrected AFQT, mother's education, number of siblings, urban residence at 14, permanent earnings in the county of residence at 17, permanent unemployment in the state of residence at 17, cohort dummies, a dummy variable indicating the presence of a college in the county of residence at age 14, average log earnings in the county of residence at age 17, and average state unemployment in the state of residence at age 17 (see Table 3).

Are returns Heterogeneous? - direct evidence

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

Figure 2: $E(Y|X,P)$ as a Function of P for Average X



The Marginal Treatment Effect by unobserved cost of College

Figure 3: $E(Y_1 - Y_0|X, U_S)$ Estimated Using Locally Quadratic Regression (Averaged Over X)

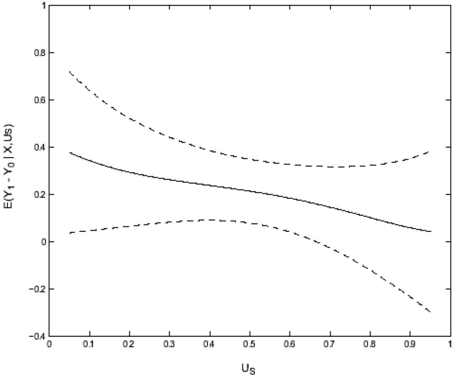


Table 1
Treatment effects parameters

	(1)
	Returns to college
ATE	0.067* (0.038)
TT	0.143*** (0.035)
TUT	−0.007 (0.071)
IV	0.095** (0.039)

- ATT: substantial returns to college because it puts most weight on low U_s individuals
- ATE is only of 6.7%
- ATUT: returns for an average person who doesn't attend college are close to zero and statistically insignificant
- Expansion of college to individuals who currently do not attend would not be effective

Next Week

Measurement Errors and Policy Evaluation

- Chen, Hong, and Nekipelov (JEL, 2011): "Nonlinear Models of Measurement Errors"
- Hu and Schennach (Econometrica, 2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models"
- Lewbel (Econometrica, 2007): "Estimation of Average Treatment Effects with Misclassification"
- Kelly (2020) "Understanding Persistence" - old version "Standard errors of persistence"