MARGINAL TREATMENT EFFECTS

Ivan A. Canay
Northwestern University[1]
ECON 481-3

1. Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. Econometrica, 73(3), 669-738.
2. Carneiro, P., Heckman, J. J., & Vytlacil, E. J. (2011). Estimating marginal returns to education. American Economic Review, 101(6), 2754-81.

[1]Today's class is based on Alex Torgovitsky's notes. I'd like to thank him for kindly sharing them.

Topics of Part I

- Lec I: Selection on Observables
  1. Potential Outcomes vs Latent Variables
  2. Causal Inference
  3. Selection Bias
  4. Selection on Observables & Selection on Prop. Score
- Lec II: Roy Models and LATE
  1. The role of heterogeneity
  2. Multiple instruments, Covariates, and Abadie's κ
- Lec III: Marginal Treatment Effect
  1. Parameters as functions of MTEs
  2. Policy Relevant Treatment Effects
- Lec IV: Extrapolations
  1. Semi-Parametrics MTEs
  2. Weights for Target Parameters

Outline

1. **Vytlacil's equivalence result**
2. Marginal Treatment Effects
3. Policy Relevant Treatment Effects
4. Empirical Application: Carneiro, Heckman, Vytlacil (11)

Selection and Monotonicity

Background

- AIR argue the interpretation of "error terms" is difficult to understand
  "Error terms" are a way to refer to latent variables (unobservables)
- They argue that assumptions about latent variables are hard to interpret
- Recall mapping from latent variable to potential outcome notation

$$D = I\{U \leq v(Z)\} \quad \Rightarrow \quad D_z = I\{U \leq v(z)\}$$

- The reverse mapping is not necessarily as easy to understand

Monotonicity and latent variables

- AIR made another argument (pg. 450) that turns out to be simply wrong:
  "Monotonicity has no explicit counterpart in the econometric formulation"
- This statement was shown to be wrong by Vytlacil (2002)
- Monotonicity is *equivalent* to a Roy model with a separable index

Vytlacil's (02) Equivalence Result

Model 1 ("IA") as in Imbens and Angrist (1994)

1. $Y = DY_1 + (1 - D)Y_0$
2. $D = \sum_{z \in \mathcal{Z}} I\{Z = z\} D_z$
3. For any $z, z' \in \mathcal{Z}$, $D_z \geq D_{z'}$ or $D_{z'} \geq D_z$ (a.s.)
4. $(Y_0, Y_1, \{D_z\}_{z \in \mathcal{Z}}) \perp\!\!\!\perp Z$

Model 2 ("Roy") is the nonparametric Roy model

1. $Y = DY_1 + (1 - D)Y_0$ (same as in Model 1)
2. $D = I\{U \leq \nu(Z)\}$ for latent variable U and unknown function ν
3. $(Y_0, Y_1, U) \perp\!\!\!\perp Z$

- Vytlacil (2002) showed that these two models are **equivalent**
- Model 2 $\Rightarrow$ Model 1 is easy (next slide)
- Model 1 $\Rightarrow$ Model 2 is more subtle

PROOF OF EQUIVALENCE

WTS $D = I\{U \leq v(Z)\}$ and $(Y_0, Y_1, U) \perp\!\!\!\perp Z$ imply conditions $1.2 - 1.4$.

Outline

1. Vytlacil's equivalence result
2. **Marginal Treatment Effects**
3. Policy Relevant Treatment Effects
4. Empirical Application: Carneiro, Heckman, Vytlacil (11)

Heckman and Vytlacil (99/05)

Motivation

- Vytlacil's result shows that nothing is lost from the Roy model
- Equivalence result means *same assumptions* as the "LATE framework"
- The Roy model may actually be easier to interpret for economists
- Interpret $v(Z) - U$ as latent utility in the choice problem for D
  Just keep in mind that the definition of U depends on Z

Setup

- The analysis uses the same non-parametric Roy model as before:

$$
\begin{aligned}
Y &= DY_1 + (1-D)Y_0 \\
D &= I\{U \leq v(Z)\} \quad \text{with } (Y_0, Y_1, U) \perp\!\!\!\perp Z
\end{aligned}
$$

- U continuously distributed, normalized to be uniform $[0,1]$
  Implies that $v(Z) = p(Z)$
- Everything is "conditional-on-X," so suppress X in the notation

Normalization: Selection Equation

The Marginal Treatment Effect

Definition

- HV define the **marginal treatment effect** (MTE) as:

$$\text{MTE}(u) \equiv E[Y_1 - Y_0 | U = u]$$

- MTE(u) is the ATE for those agents with first stage unobservable u

  Those with small u (close to 0) often choose $D = 1$

  Those with large u (close to 1) infrequently choose $D = 1$
- Unobserved treatment heterogeneity if and only if non-constant MTE

Pointwise identification of the MTE

- The MTE is point identified for all $p \in$ int supp $p(Z)$ (next slide):

$$\underbrace{\frac{\partial}{\partial p} E[Y | p(Z) = p]}_{\text{"local IV estimand"}} = \text{MTE}(p)$$

- Note that this requires Z to be continuously distributed!

Proof of identification of MTE

Proof of identification of MTE

$$E[Y|P(Z) = p] = E[Y_1|U \leq p]p + E[Y_0|U > p](1 - p).$$

Utility of the MTE

An organizing principle

- U provides a single dimension on which we can organize heterogeneity
- Many quantities can be written as weighted averages of the MTE

$$\theta = \int_a^b MTE(u)\omega(u)du \; .$$

We will see several examples ahead

- Point identification follows from LIV if support at non-zero weights
- Tool to discuss new quantities that answer specific policy questions

Implementation and extrapolation

- The MTE also gives us something we can restrict and/or parameterize
- This is useful for estimation (dimension reduction)
- It is also useful for thinking about extrapolation

ATE: Unweighted Average MTE

- The ATE is the unweighted average of the MTEs:

$$\text{ATE} = E\left[E[Y_1 - Y_0|U]\right] = \int_0^1 \text{MTE}(u) \times 1 \underbrace{du}_{U \text{ uniform}}$$

- The ATE is point identified if $\{0,1\} \in \operatorname{supp} p(Z)$
- This follows from the LIV-MTE identification argument, e.g.:

$$E[Y|p(Z) = 1] = E[Y_1|U \leq 1] = E[Y_1]$$

- Requiring $\{0,1\} \in \operatorname{supp} p(Z)$ is a **large support condition**
- It says that there exist instrument values $z_0, z_1 \in \mathcal{Z}$ such that:
  Every agent with z_0 would *never* take the treatment ($p(z_0) = 0$)
  Every agent with z_1 would *always* take the treatment ($p(z_1) = 1$)
- A *severe* demand to place on the data $\Rightarrow$ limited scope
- Basically means you have random assignment since $Z = z_d \Rightarrow D = d$

ATT/ATU: Weighted Average MTE

ATT

- The ATT can be written as (see problem set)

$$\text{ATT} = \int_0^1 \text{MTE}(u) \frac{P[p(Z) \geq u]}{P[D=1]} \, du \equiv \int_0^1 \text{MTE}(u) \omega_{\text{ATT}}(u) \, du,$$

- Those with low values of u are more highly weighted
  These are the most likely to take treatment
- The weights are known or identifiable and integrate to 1

ATU

- Analogous argument for the ATU:

$$\text{ATT} = \int_0^1 \text{MTE}(u) \frac{P[p(Z) < u]}{P[D=0]} \, du \equiv \int_0^1 \text{MTE}(u) \omega_{\text{ATU}}(u) \, du,$$

- High values of u are more highly weighted (least likely to take treatment)

LATE as a Weighted MTE

Weights

- Suppose $p(z) > p(z')$ for two values z' and z — then

$$\text{LATE}_{z'}^{z} = \int_0^1 \text{MTE}(u)\omega_{\text{LATE}}(u)du, \quad \text{where} \;\; \omega_{\text{LATE}}(u) \equiv \frac{I\{p(z') < u \leq p(z)\}}{p(z) - p(z')}$$

- $u \leq p(z')$ are always-takers for $z' \to z$ and $u > p(z)$ are never-takers
- $u \in (p(z'), p(z)]$ are $z' \to z$ compliers
- $\text{LATE}_{z'}^{z}$ puts equal weight on compliers, 0 weight on all others

LATE as a Weighted MTE

Limiting case

- Notice that as we take $p(z) \searrow p(z')$,

$$\lim_{p(z)\searrow p(z')} \mathsf{LATE}_{z'}^{z} = \lim_{p(z)\searrow p(z')} \frac{\int_{p(z')}^{p(z)} \mathsf{MTE}(u)\,du}{p(z)-p(z')} = \mathsf{MTE}(p(z')),$$

- So the MTE is a limiting (*marginal*) version of the LATE

The IV Estimand as a Weighted MTE

- Suppose we use $J(Z)$ as an instrument for D — IV estimand $\beta_{IV,J}$
- Using similar arguments it can be shown that

$$\beta_{IV,J} \equiv \frac{\text{cov}(J(Z), Y)}{\text{cov}(J(Z), D)} = \int_0^1 \text{MTE}(u)\omega_{IV,J}(u)du,$$

$$\text{with} \quad \omega_{IV,J}(u) \equiv \frac{(E[J(Z)|p(Z) \geq u] - E[J(Z)])P[p(Z) \geq u]}{\text{cov}(J(Z), D)},$$

- Weights are 0 for $u < \inf \text{supp}\, P$ and $u > \sup \text{supp}\, P$
- Weights integrate to 1
- Weights will generally be negative for some u:

$$E[J(Z)|p(Z) \geq u] - E[J(Z)] \text{ may be both positive and negative}$$

- Example of only positive is $J(Z) = p(Z)$ or a monotone transformation
- So IV/TSLS need not estimate a "causal effect" in general
- Still consistent with the IA results since they ordered Z by $p(Z)$

Outline

1. Vytlacil's equivalence result
2. Marginal Treatment Effects
3. **Policy Relevant Treatment Effects**
4. Empirical Application: Carneiro, Heckman, Vytlacil (11)

"Policy Relevant" Parameter?

- The MTE framework partitions all agents in a clear way
- Provides a foundation for thinking about "ideal" treatment effects
- The "ideal" treatment effect clearly depends on the question

- The ATE receives a lot of attention in the literature

  But not very useful for policy — can agents still *choose* D?
- The ATT is somewhat clearer in this regard

  Loss in benefit to treated group from discontinuing $D = 1$

- Perhaps more relevant is changing the agent's *choice problem*
- For example, $D \in \{0, 1\}$ is attending a four-year college
- Average effect of forcing college/no college (ATE) is not interesting
- Nor is the effect on college-goers of shutting down college (ATT)
- More interesting are the effects via D of adjusting tuition Z

Policy Relevant Treatment Effects

- HV formalize this idea as **policy relevant treatment effects** (PRTE)
- Aggregate effect on Y of a change in the propensity score/instrument
- Change corresponds to a policy that affects treatment choice

- Let $p^\star(Z^\star), Z^\star$ be the propensity score/instrument under a new policy
- Let $D^\star$ denote the treatment choice under the new policy:

$$D^\star = I\{U \leq p^\star(Z^\star)\}$$

- Letting $Y^\star = D^\star Y_1 + (1 - D^\star)Y_0$ be the outcome under the new policy,

HV define the PRTE as: $\beta_{\text{PRTE}} \equiv \dfrac{E[Y^\star] - E[Y]}{E[D^\star] - E[D]}$

- The mean effect (per net person) of the policy change
- Implicit assumption is that the policy does not affect (Y_0, Y_1, U)
  Intuitively necessary — see HV for a formalization

The PRTE as a Weighted MTE

- One can show that

$$\beta_{\text{PRTE}} \equiv \frac{E[Y^\star] - E[Y]}{E[D^\star] - E[D]} = \int_0^1 \text{MTE}(u)\omega_{\text{PRTE}}(u)\, du$$

$$\text{with} \quad \omega_{\text{PRTE}}(u) \equiv \frac{F_P^-(u) - F_{P^\star}^-(u)}{E[P^\star] - E[P]}$$

  where F_P and $F_{P^\star}$ are the distributions of $P \equiv p(Z)$ and $P^\star \equiv p^\star(Z^\star)$

  $F_P^-(u) \equiv \lim_{v \uparrow u} F_P(u)$ is the left-limit of F_P at u
- The weights show that point identifying β_{PRTE} will be difficult
- In particular, the support of $P^\star$ must be contained in that of P
- Since we can only possibly point identify $\text{MTE}(u)$ on the support of P
- Restricts to interpolating policies vs. extrapolating policies
- In addition, still need to have a continuous instrument
- Or else cannot nonparametrically point identify $\text{MTE}(u)$ for *any* u

PRTEs: Two Counterfactual Policies

- Instead of contrasting with status quo, could have two policies:

$$D^a \equiv I\{U \leq p^a(Z^a)\} \quad \text{and} \quad D^b \equiv I\{U \leq p^b(Z^b)\}$$
$$Y^a \equiv D^a Y_1 + (1 - D^a) Y_0 \quad \text{and} \quad Y^b \equiv D^b Y_1 + (1 - D^b) Y_0$$

- Then define the PRTE for b relative to a as

$$\text{PRTE}_a^b \equiv \frac{E[Y^b] - E[Y^a]}{E[D^b] - E[D^a]}$$

- Derivation of the weights just requires relabeling the previous argument

The LATE is a PRTE

- Policy a: Every agent receives $Z = z'$: $\quad p^a(\cdot) = p(\cdot), Z^a = z'$
- Policy b: Every agent receives $Z = z$: $\quad p^b(\cdot) = p(\cdot), Z^b = z$
- Then $\text{PRTE}_a^b = \text{LATE}_{z'}^z$

  If Z is a policy lever, the LATE may be intrinsically interesting

Outline

1. Vytlacil's equivalence result
2. Marginal Treatment Effects
3. Policy Relevant Treatment Effects
4. **Empirical Application: Carneiro, Heckman, Vytlacil (11)**

Empirical Application

Carneiro, Heckman & Vytlacil (2011)

- Study returns to schooling in the NLSY 79 for $N = 1{,}747$ white males
- Y is (roughly) log average hourly wages
- $D \in \{0, 1\}$ is attending some college — annualized in various ways
- X contains the usual suspects plus some controls relevant for Z
- Z are taken from a variety of other previous studies:
  1. The presence of a four-year college in the county of residence at age 14
  2. Local wage in the county of residence at age 17
  3. Local unemployment in the state of residence at age 17
  4. Average tuition in public four-year colleges in age 17 county of residence

- Analysis so far has been "conditional-on-X" — won't work in practice

  Nonparametric conditioning leads to the usual curse of dimensionality
- CHV solve this by imposing some semiparametric structure
- The way they do this also helps with limited instrument support

Variables Used in CHV (2011)

Variable	Definition
Y	Log wage in 1991 (average of all nonmissing wages between 1989 and 1993)
$S = 1$	If ever enrolled in college by 1991; zero otherwise
$\mathbf{X}$	AFQT,[a] mother's education, number of siblings, average log earnings 1979–2000 in county of residence at 17, average unemployment 1979–2000 in state of residence at 17, urban residence at 14, cohort dummies, years of experience in 1991, average local log earnings in 1991, local unemployment in 1991
$\mathbf{Z}\backslash\mathbf{X}$[b]	Presence of a college at age 14 (Card 1995; Stephen V. Cameron and Christopher Taber 2004), local earnings at 17 (Cameron and Heckman 1998; Cameron and Taber 2004), local unemployment at 17 (Cameron and Heckman 1998), local tuition in public four-year colleges at 17 (Thomas J. Kane and Cecilia E. Rouse 1995)

[a] We use a measure of this score corrected for the effect of schooling attained by the participant at the date of the test, since at the date the test was taken, in 1981, different individuals have different amounts of schooling and the effect of schooling on AFQT scores is important. We use a correction based on the method developed in Karsten T. Hansen, Heckman, and Kathleen J. Mullen (2004). We take the sample of white males, perform this correction, and then standardize the AFQT to have mean 0 and variance 1 within this sample. See Table A-2 in the online Appendix.
[b] The papers in parentheses are papers that previously used these instruments.

Notational differences

- Their S is my D
- Their Z is like my (X, Z)
- So their $Z \setminus X$ are the (excluded) instruments — what I call Z
- X are the covariates as in my notation

A Semiparametric MTE Model

- Suppose that we make the following functional form assumptions:

$$Y_d = X'\delta_d + V_d, \qquad V_d \equiv Y_d - E[Y_d|X] \quad \text{for } d = 0, 1$$

- Under these assumptions one can show that

$$E[Y|X = x, P = p] = x'\delta_0 + px'(\delta_1 - \delta_0) + K(p, x)$$

$$\text{where} \quad K(p, x) \equiv \int_0^p E[V_1 - V_0|X = x, U = u]\, du$$

$$\text{so that} \quad \text{MTE}(p, x) = x'(\delta_1 - \delta_0) + \frac{\partial}{\partial p} K(p, x)$$

- This is *almost* a **partially linear model** (Robinson 1988)

  X and PX enter linearly, and through an unknown function $K(P, X)$
- But still a dimensionality problem, because $K(P, X)$ depends on X
- Also, we wouldn't be able to separately identify (δ_0, δ_1) from $K(p, x)$

A Partially Linear MTE Model

- To address this, CHV assume that $(X, Z) \perp\!\!\!\perp (V_0, V_1, U)$, so then

$$\begin{aligned} K(p, x) &\equiv \int_0^p E[V_1 - V_0 | X = x, U = u]\, du \\ &= \int_0^p E[V_1 - V_0 | U = u]\, du \equiv K(p) \end{aligned}$$

- The nonparametric component is now a function of a scalar, i.e.

$$Y = X'\delta_0 + PX'(\delta_1 - \delta_0) + K(P) + \varepsilon \quad \text{where} \quad \varepsilon \equiv Y - E[Y|X, P] \qquad (\star)$$

- Now apply the classic argument for partially linear models:

$$\begin{aligned} (\star) \Rightarrow E[Y|P] &= E[X|P]'\delta_0 + PE[X|P]'(\delta_1 - \delta_0) + K(P) \\ \text{let} \quad \widetilde{Y} &\equiv Y - E[Y|P] \text{ and } \widetilde{X}_j \equiv X_j - E[X_j|P] \text{ for each component } j \\ \text{then} \quad \widetilde{Y} &= \underbrace{\widetilde{X}'\delta_0 + P\widetilde{X}'(\delta_1 - \delta_0)}_{\text{linear in } \widetilde{X},\, P\widetilde{X}} + \epsilon \quad \text{with} \quad E[\epsilon | \widetilde{X}, P\widetilde{X}] = 0\,, \end{aligned}$$

Semiparametric Estimation

1. Estimate $P(X)$ using a logit or probit to get $\widehat{P}(X)$
2. Construct $\widetilde{Y}$ and $\widetilde{X}$ through (1-dimensional!) nonparametric regression
3. Linear regression of $\widetilde{Y}$ on $\widetilde{X}$ and $P\widetilde{X}$ to get $(\widehat{\delta}_0, \widehat{\delta}_1)$
4. Nonparametrically regress $Y - X'\widehat{\delta}_0 - \widehat{P}X'(\widehat{\delta}_1 - \widehat{\delta}_0)$ on $\widehat{P}$,

   noting that $\quad E[Y - X'\delta_0 - PX'(\delta_1 - \delta_0)|P] = K(P)$

   Note that we want to estimate the *derivative* of K
5. Then $x'(\widehat{\delta}_1 - \widehat{\delta}_0) + \widehat{K}'(p)$ is an estimate of MTE(p, x)

- $K(p, x)$ can only possibly be point identified for $(p, x) \in \text{supp}(P, X)$
- That is, for a fixed x, only for $p \in \text{supp}(P|X = x)$
- But $K(p)$ can be point identified for any $p \in \text{supp}(P) \supseteq \text{supp}(P|X = x)$
- So assuming $(X, Z) \perp\!\!\!\perp (V_0, V_1, U)$ has addressed the support issues
- As in most applications, these support issues are a big problem ...

Joint Support of P and Index of X

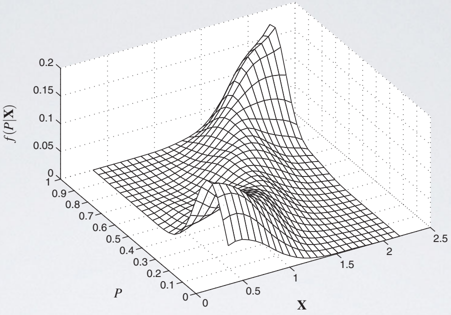

FIGURE 2. SUPPORT OF P CONDITIONAL ON $\mathbf{X}$

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). $\mathbf{X}$ corresponds to an index of variables in the outcome equation.

It is striking how small the support of P is for each value of the X index.

Marginal Support of P

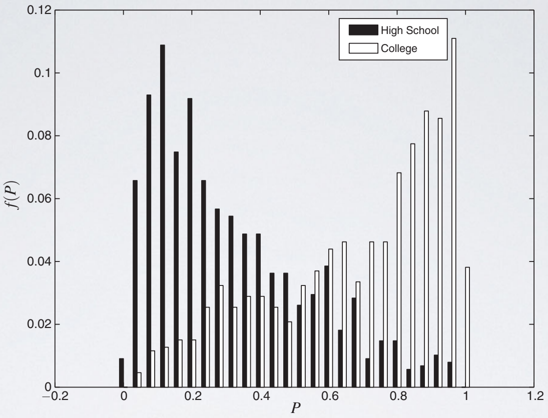

Figure 3. Support of P for $S = 0$ and $S = 1$

The support of the estimated $P(Z)$ is almost full for both $D \in \{0,1\}$

A Better Way to Separability

- The assumption used in CHV is $(V_0, V_1, U) \perp\!\!\!\perp (X, Z)$
  $X \perp\!\!\!\perp U$: unattractive given the variables usually included in X
- X is still allowed to have a direct effect on Y_d via $X'\delta_d$
- So X are completely exogenous (and correctly parameterized) covariates

A weaker condition

- In fact, all that was used in the CHV derivation was:

$$E[V_1 - V_0 | X = x, U = u] = E[V_1 - V_0 | U = u]$$

- A sufficient condition is $(V_0, V_1) \perp\!\!\!\perp X | U$ — does not require $X \perp\!\!\!\perp U$
- The benefits can be seen from:

$$\frac{\partial}{\partial p}\text{MTE}(p, x) = \frac{\partial}{\partial p}\left[x'(\delta_1 - \delta_0) + \frac{\partial}{\partial p}K(p)\right] = \frac{\partial^2}{\partial p^2}K(p)$$

- So the **slope** of the MTE does not depend on x — **separability**

Estimated MTE: Semiparametric Case

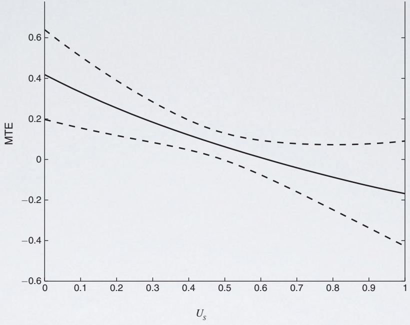

FIGURE 4. $E(Y_1 - Y_0 | \mathbf{X}, U_S)$ WITH 90 PERCENT CONFIDENCE INTERVAL—LOCALLY QUADRATIC REGRESSION ESTIMATES

Notes: To estimate the function plotted here, we first use a partially linear regression of log wages on polynomials in **X**, interactions of polynomials in **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; **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) at the average value of **X**. Ninety percent standard error bands are obtained using the bootstrap (250 replications).

- This plots $x'(\widehat{\delta}_1 - \widehat{\delta}_0) + \frac{\partial}{\partial p}\widehat{K}(p)$ evaluated at the average of x
- Presumably only for p in its unconditional support $[.032, .978]$

Estimated MTE: Normal Model

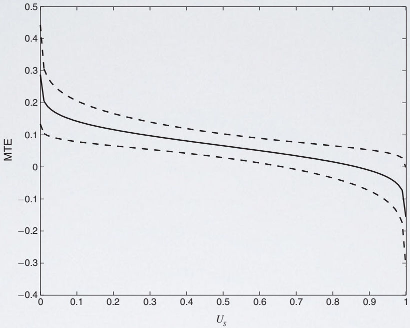

FIGURE 1. MTE ESTIMATED FROM A NORMAL SELECTION MODEL

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:

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

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, 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. We plot 90 percent confidence bands.

- Notice that the normal model restricts the MTE to be monotone
- What happens as p (u_S in their notation) tends to 0 or 1?

H_0: No Unobserved Heterogeneity

- CHV provide two tests of H_0 : **no unobserved heterogeneity**
  Important null — without heterogeneity we could use simple linear IV
- Recall that under their assumptions:

$$E[Y|X=x,P=p] = x'\delta_0 + px'(\delta_1 - \delta_0) + K(p)$$

$$\text{where} \quad K(p) \equiv \int_0^p E[V_1 - V_0|U=u]\, du$$

- No unobserved heterogeneity in treatment response if and only if

$$E[V_1 - V_0|U=u] = C \quad \text{for all } u \text{, some } C \quad \Leftrightarrow \quad K(p) = pC$$

- So test $H_0 : E[Y|X=x,P=p]$ is linear in p for each x

One way to implement this test

- Assume $K(p) = \beta_0 + \beta_1 p + \cdots + \beta_k p^k$ in the semiparametric procedure
- Then test $H_0 : \beta_2 = \cdots = \beta_k = 0$

Testing No Unobs. Heterogeneity

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			

[a]The size of the test is controlled using a critical value constructed by the bootstrap method of Romano and Wolf (2005) using a 10 percent significance level.

[b]In order to compute the numbers in this table, we construct groups of values of U_S and average the MTE within these groups, by computing $E(Y_1 - Y_0 | \mathbf{X} = \bar{\mathbf{x}}, U_S^{Lj} \leq U_S \leq U_S^{Hj})$, where U_S^{Lj} and U_S^{Hj} are the lowest and highest values of U_S defined for interval j. Then we compare the average MTE across adjacent groups and test whether the difference is equal to zero using the bootstrap with 250 replications.

- Panel B is a more direct (maybe less clean) test of the same null
- Tests the implication that MTE is constant over ranges of $[0, 1]$

QUESTIONS?