Lecture 8: Selection Models and Policy Evaluation

Chris Walters

University of California, Berkeley and NBER

Introduction

- This lecture discusses **structural selection models** and **control function estimators** of these models
- Selection models are mathematical descriptions of how non-random samples are generated
- Control function estimators adjust for non-random selection, allowing estimation of the parameters of unselected distributions
- As we will see, this approach is intimately linked to the IV methods from previous lectures. We will emphasize the connection between these approaches
- Structural models offer the opportunity to extrapolate and predict economic parameters that are not identified by the experiment at hand, at the cost of stronger assumptions
- "Harmful" econometrics coming – tread carefully!

Selection Model Example: Labor Supply

- Simple example of a selection model: Labor supply problem

$$\max_{c,h} c - v(h) \text{ s.t. } c \leq wh + V$$

- At interior solutions:

$$v'(h^*) = w$$

- At corner solutions:

$$v'(0) > w$$

- Reservation wage is $w^* = v'(0)$; work if $w \geq w^*$

Selection Model Example

- Suppose individuals' reservation wages are described by

$$w_i^* = X_i'\theta + \eta_i$$

- Offered wages are

$$w_i = X_i'\beta + \epsilon_i$$

- Assume $E[\epsilon_i | X_i] = 0$, so $X_i'\beta$ is the population CEF
- Individual i works ($D_i = 1$) when

$$X_i'\beta + \epsilon_i \geq X_i'\theta + \eta_i$$

$$\iff X_i'(\beta - \theta) + (\epsilon_i - \eta_i) \geq 0$$

$$\iff X_i'\psi \geq v_i$$

Selection Model Example

- $D_i^* = X_i'\psi - v_i$ is a **latent index** determining D_i
- We observe outcomes in the sample with $D_i = 1$. CEF in this sample is

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

- If ϵ_i and v_i are independent, the last term is $E\left[\epsilon_i|X_i\right] = 0$ and OLS recovers β
- This is equivalent to saying we have a random sample – selection into the sample is unrelated to outcomes
- If ϵ_i and v_i aren't independent, we'll have $E\left[\epsilon_i|X_i, D_i = 1\right] \neq 0$, and OLS on observed sample is inconsistent

Selection Model Example

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

- Suppose that ϵ_i and v_i are joint normal:

$$(\epsilon_i, v_i)|X_i \sim N\left((0,0), \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon \\ \rho\sigma_\epsilon & 1 \end{bmatrix}\right)$$

- Then we can work out the expected error conditional on $D_i = 1$
- Under normality, conditional expectations are linear:

$$E\left[\epsilon_i|X_i, v_i\right] = \rho\sigma_\epsilon v_i.$$

Selection Model Example

- The CEF of w_i in the observed sample is

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

$$= X_i'\beta + \rho\sigma_\epsilon E\left[v_i|X_i, v_i < X_i'\psi\right]$$

$$= X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\delta\right)$$

- Here $\lambda(x)$ is the conditional expectation of a standard normal random variable truncated from above, also known as the **inverse Mills ratio:**

$$\lambda(x) = -\frac{\phi(x)}{\Phi(x)}.$$

Heckit

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\psi\right)$$

- ψ can be consistently estimated via a first-step probit of D_i on X_i
- Then run a second-step regression in the $D_i = 1$ sample:

$$w_i = X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\hat{\psi}\right) + u_i$$

- This two-step procedure generates consistent estimates of β; bootstrap or apply two-step correction for inference
- The Mills ratio is a **control function** or **selection correction** that accounts for selection into the observed sample
- This is Heckman's (1974, 1976, 1979) two-step selection correction ("Heckit")

Heckit Identification

- Suppose X_i is just a constant. Then the second-step regression is

$$w_i = \beta + \rho\sigma_\epsilon \cdot \lambda\left(\hat{\psi}\right) + u_i$$

$$= \delta + u_i$$

- The constant here is $\delta = (\beta + \rho\sigma_\epsilon\lambda(\psi))$, so β and $\rho\sigma_\epsilon$ are not separately identified
- More generally, if outcome and selection equations are saturated in X_i, main effects and Mills ratio term are not separately identified
- This is unattractive – there is typically no reason to believe $E[w_i|X_i]$ is linear in X_i

Heckit Identification

- Solution: Suppose there are additional variables Z_i in the selection equation, so

$$D_i = 1\{X_i'\psi + Z_i'\pi > v_i\}$$

- Assume $E[\epsilon_i|X_i, Z_i] = 0$. Then second-step CEF is

$$E[w_i|X_i, Z_i, D_i = 1] = X_i'\beta + \rho\sigma_\epsilon\lambda\left(X_i'\psi + Z_i'\pi\right)$$

- If $\pi \neq 0$ this can be estimated even if X_i is saturated since variation in Z_i separately identifies the selection term
- Identifying a Heckit without relying on functional form restrictions requires finding a Z_i that shifts the probability of selection but is excludable from the outcome equation
- Sound familiar?

Heckit with Instruments

- The requirements for a good Z_i in the Heckit model are the same as the requirements for a good instrument when we're doing IV
- This is not a coincidence. Control function and IV are methods for solving the same problem

Selection and Treatment Effects

- To see the connection between control function and IV, consider a heterogeneous treatment effects model:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- Here $\alpha_d = E\left[Y_i(d)\right]$ so $E\left[\epsilon_{id}\right] = 0$
- If we had random samples of $Y_i(1)$ and $Y_i(0)$ we could run OLS (i.e., take means) and estimate $ATE = \alpha_1 - \alpha_0$

Selection and Treatment Effects

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- But we only observe $Y_i(1)$ when $D_i = 1$, and we only observe $Y_i(0)$ when $D_i = 0$
- These are not random samples if treatment is not as good as randomly assigned
- We therefore have sample selection problems for both $Y_i(1)$ and $Y_i(0)$
- Treatment effects estimation is a two-sided sample selection problem
- An instrument is needed to solve this problem

IV and Selection Models

- We have seen that IV and control function are two methods for solving the same problem
- How should we think about the relationship between parametric sample selection models and the nonparametric LATE model of Imbens and Angrist (1994)?
- How should we think about the relationship between estimates produced by IV and control function?

IV and Selection Models

- To better understand the relationships between latent index models and the LATE model, consider a treatment effects model with a binary treatment and binary instrument:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- Suppose selection into the $D_i = 1$ sample follows the rule

$$D_i = 1\{\psi_0 + \psi_1 Z_i > v_i\}$$

$$(\epsilon_{i1}, \epsilon_{i0}, v_i) \perp\!\!\!\perp Z_i$$

$$v_i \sim F(v)$$

- $F(v)$ is some strictly increasing parametric distribution function (e.g. the normal CDF)

IV and Selection Models

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

$$D_i = 1\{\psi_0 + \psi_1 Z_i > v_i\}$$

$$(\epsilon_{i1}, \epsilon_{i0}, v_i) \perp\!\!\!\perp Z_i$$

$$v_i \sim F(v)$$

- This selection model appears to be more restrictive than the LATE model, which involves no distributional assumptions

LATE Model and Selection Model: Equivalence

- Vytlacil (2002) shows that this selection model *is* the LATE model, in the sense that
  - The selection model satisfies the LATE assumptions
  - The LATE assumptions imply that the selection model rationalizes the observed and counterfactual outcomes and treatments

LATE Model and Selection Model: Equivalence

- The first part of the proof is straightforward. Note that

$$Y_i(0) = \alpha_0 + \epsilon_{i0},\ Y_i(1) = \alpha_1 + \epsilon_{i1},$$

$$D_i(0) = 1\{\psi_0 > v_i\}\ ,\ D_i(1) = 1\{\psi_0 + \psi_1 > v_i\}$$

- $Y_i(d)$ and $D_i(z)$ are functions of $(\epsilon_{i0}, \epsilon_{i1}, v_i)$ which are independent of Z_i, so independence/exclusion are satisfied
- If $\psi_1 > 0$, then $D_i(1) \geq D_i(0)$ and monotonicity is satisfied
- $Pr\left[D_i(1) > D_i(0)\right] = Pr\left[\psi_0 + \psi_1 > v_i \geq \psi_0\right] > 0$ since $F(\cdot)$ is strictly increasing, so there is a first stage
- The selection model therefore satisfies the assumptions of the LATE framework

LATE Model and Selection Model: Equivalence

- To show that the LATE model implies the selection model representation, first note that the "parametric" assumption $v_i \sim F(v)$ is not really a restriction
- For any strictly increasing distribution function $G(\cdot)$ we can write

$$D_i = 1\left\{G^{-1}\left(F\left(\psi_0 + \psi_1 Z_i\right)\right) > G^{-1}(F(v_i))\right\}$$

$$= 1\left\{\tilde{\psi}_0 + \tilde{\psi}_1 Z_i > \tilde{v}_i\right\},$$

- where

$$\tilde{\psi}_0 = G^{-1}\left(F(\psi_0)\right),\ \tilde{\gamma}_1 = G^{-1}(F(\psi_0 + \psi_1)) - G^{-1}(F(\psi_0))$$

$$\tilde{v}_i = G^{-1}(F(v_i))$$

LATE Model and Selection Model: Equivalence

$$D_i = 1\left\{\tilde{\psi}_0 + \tilde{\psi}_1 Z_i > \tilde{v}_i\right\},$$

- The new selection error $\tilde{v}_i = G^{-1}(F(v_i))$ has CDF $G(\cdot)$
- The same selection model can be represented with any distribution function
- It is therefore sufficient to show that the LATE model implies a selection model representation for SOME distribution function

LATE Model and Selection Model: Equivalence

- Let $u_i \sim U(0,1)$ be independent of Z_i, and define

$$U_i = \begin{cases} u_i \times Pr\left[D_i(0) = 1\right], & D_i(0) = 1 \\ Pr\left[D_i(0) = 1\right] + u_i \times Pr\left[D_i(1) > D_i(0)\right], & D_i(1) > D_i(0) \\ Pr\left[D_i(1) = 1\right] + u_i \times Pr\left[D_i(1) = 0\right], & D_i(1) = 0 \end{cases}$$

- Then we can write

$$D_i = 1\left\{\psi_0 + \psi_1 Z_i > U_i\right\}$$

- Here $\psi_0 = Pr\left[D_i(0) = 1\right]$, $\psi_1 = Pr\left[D_i(1) > D_i(0)\right]$, and $U_i \sim U(0,1)$

LATE Model and Selection Model: Equivalence

- U_i is uniform on $(0, \psi_0)$ for always takers, on $(\psi_0, \psi_0 + \psi_1)$ for compliers, and on $(\psi_0 + \psi_1, 1)$ for never takers
- This model implies the same observed and counterfactual treatment choices and outcomes as the LATE model
- We can equivalently represent the selection model with the distribution $F(\cdot)$ by applying $F^{-1}(\cdot)$ to both sides of the treatment selection equation
- We have therefore shown that the LATE model and the selection model are equivalent: They are two ways of representing the same information
- Vytlacil (2002) shows that this applies to the more general LATE model with multiple instruments
- Caveat: An $F(\cdot)$ with unbounded support only works if there are always- and never-takers. Otherwise $F^{-1}(\psi_0) \to -\infty$ or $F^{-1}(\psi_0 + \psi_1) \to \infty$.

IV and Control Function

- Selection model with uniform representation of selection error:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

$$D_i = 1\{\psi_0 + \psi_1 Z_i > U_i\}$$

$$U_i \sim U(0,1)$$

$$(\epsilon_{i1}, \epsilon_{i0}, U_i) \perp\!\!\!\perp Z_i$$

- We've shown that this is the LATE model
- Does this mean that IV and control function estimates of treatment effects are also equivalent?

IV and Control Function

- No. In fact, we cannot estimate this model by control function without further assumptions
- To form control functions we need to specify $E[\epsilon_{id}|U_i]$, which we haven't done
- Control function yields estimates of α_1 and α_0, and therefore the *ATE* $\alpha_1 - \alpha_0$
- The *ATE* is not identified in the LATE model – we can only get the *LATE*
- We have to assume more if we want to extrapolate from *LATE* to *ATE*

IV and Control Function

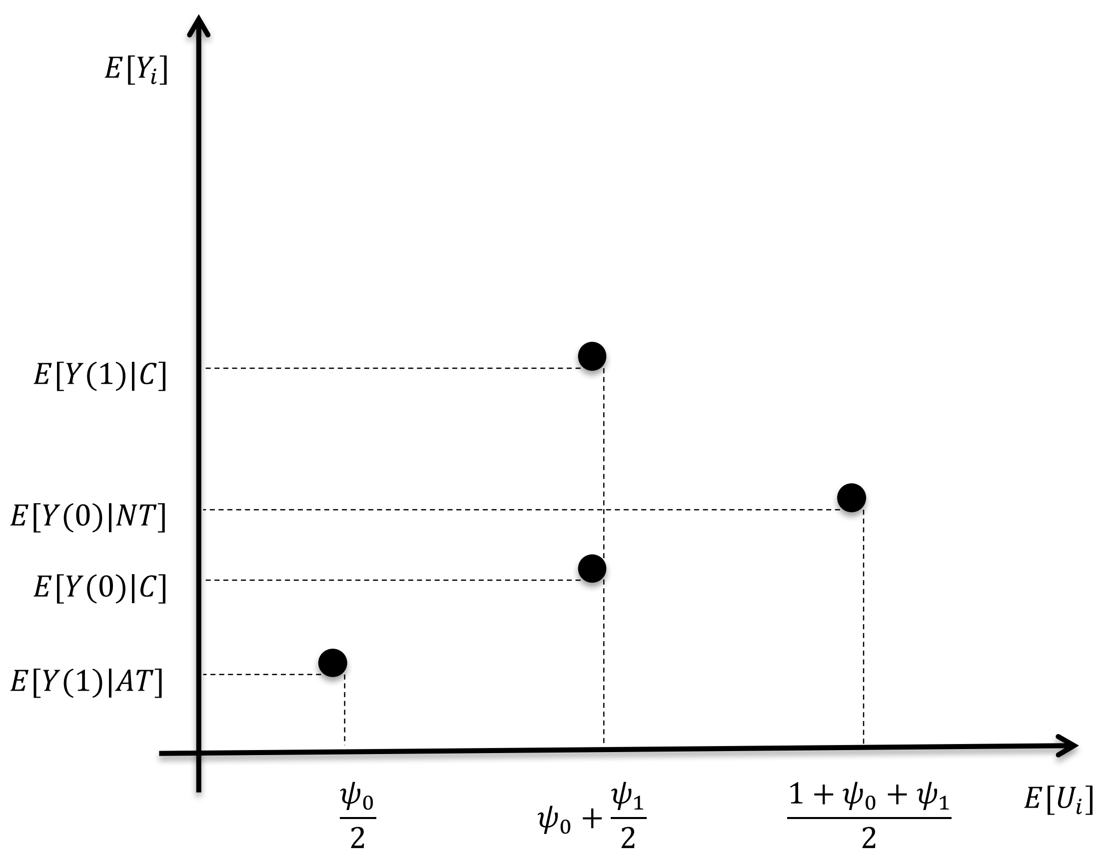
- In selection model notation, our three subgroups are defined:
  - Always takers: $U_i < \psi_0$
  - Compliers: $\psi_0 \leq U_i < \psi_0 + \psi_1$
  - Never takers: $U_i \geq \psi_0 + \psi_1$
- Then
  - $E[U_i|AT] = \frac{\psi_0}{2}$
  - $E[U_i|C] = \psi_0 + \frac{\psi_1}{2}$
  - $E[U_i|NT] = \frac{1+\psi_0+\psi_1}{2}$

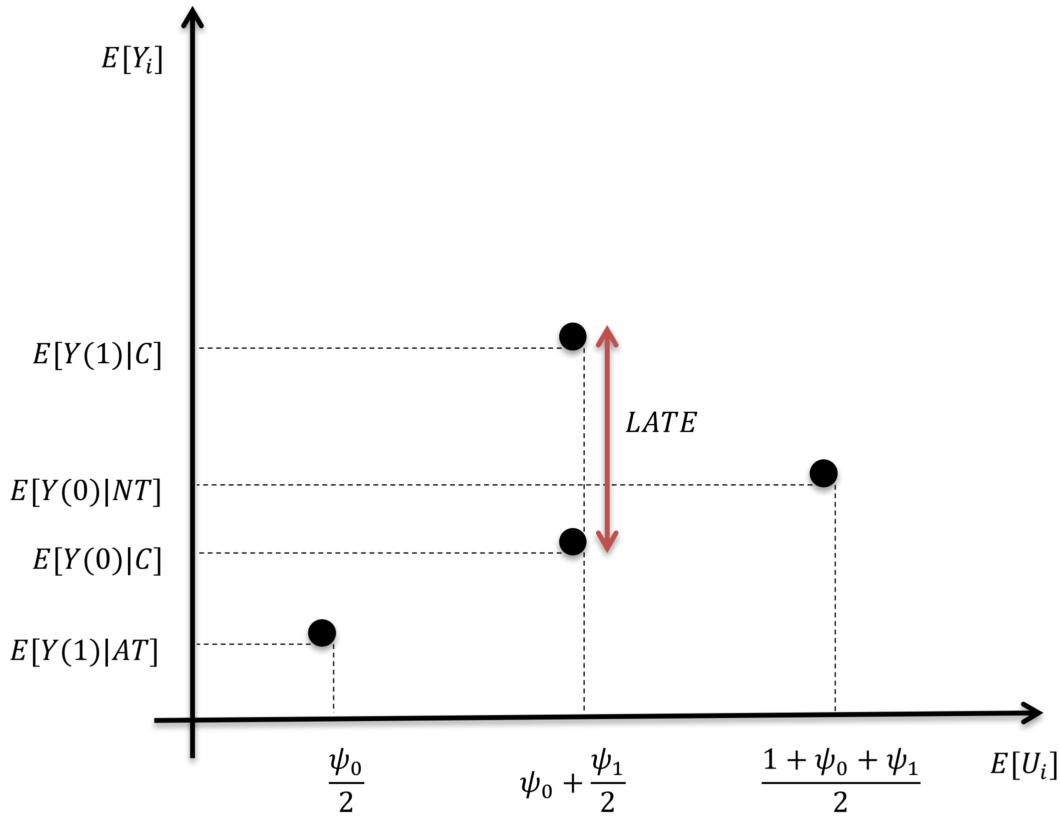
IV and Control Function

- Recall that in the LATE framework we can identify:
  - $E[Y_i(1)|AT]$
  - $E[Y_i(0)|NT]$
  - $E[Y_i(1)|C]$
  - $E[Y_i(0)|C]$
- Mean $Y_i(1)$ for always takers is observable in the $(D_i = 1, Z_i = 0)$ group
- Mean $Y_i(0)$ for never takers is observable in the $(D_i = 0, Z_i = 1)$ group
- Mean $Y_i(1)$ for compliers is obtained by removing the AT mean from the $D_i = Z_i = 1$ mix
- Mean $Y_i(0)$ for compliers is obtained by removing the NT mean from the $D_i = Z_i = 0$ mix

IV and Control Function

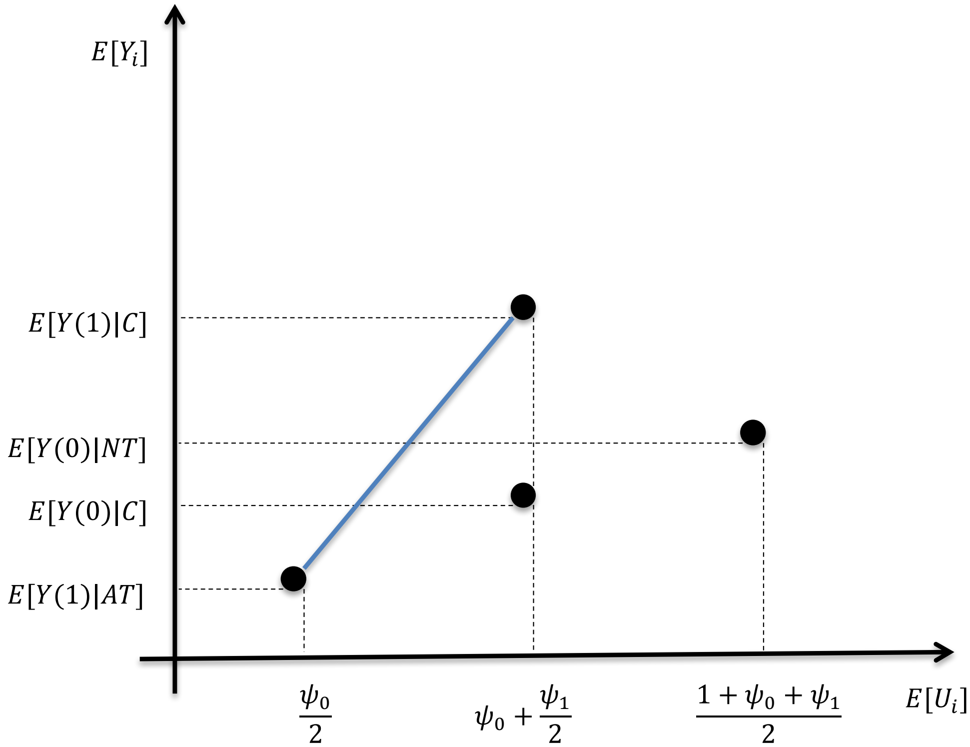
- We can therefore identify means of $Y_i(1)$ and $Y_i(0)$ for two groups each
- This yields two points on the curve $E[Y_i(d)|U_i]$ for each potential outcome





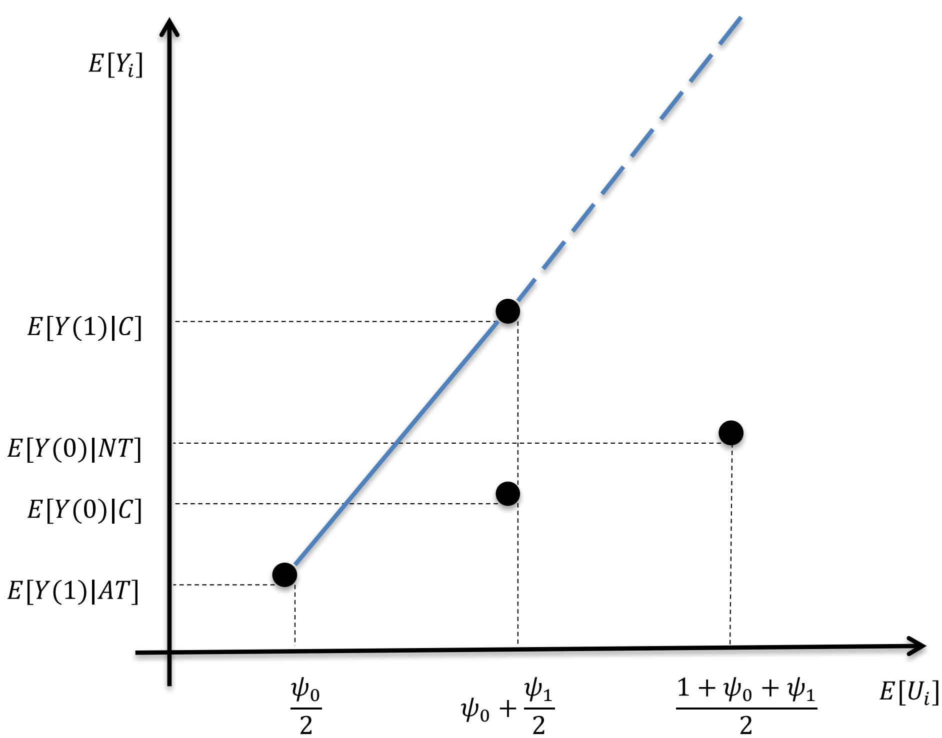
Extrapolation from LATE

- Without further assumptions we cannot identify any other treatment effects
- But by specifying a functional form for $E\left[Y_i(d)|U_i\right]$, we can "connect the dots" and extrapolate to predict effects for always takers and never takers
- This allows us to predict the effects of policies that affect different subpopulations than the instrument at hand



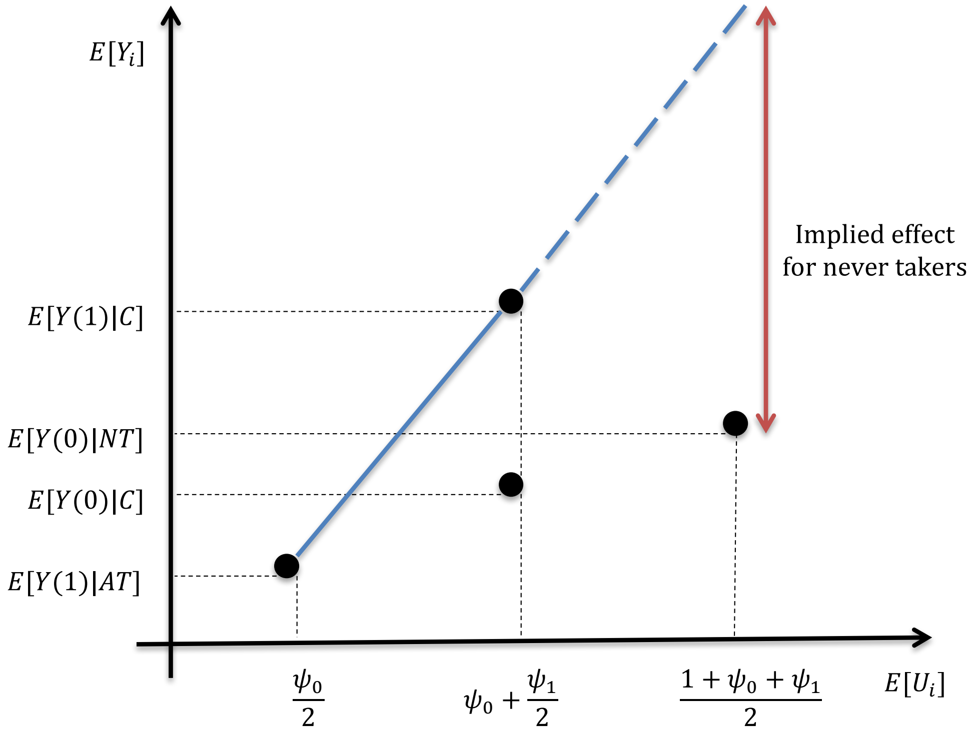
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



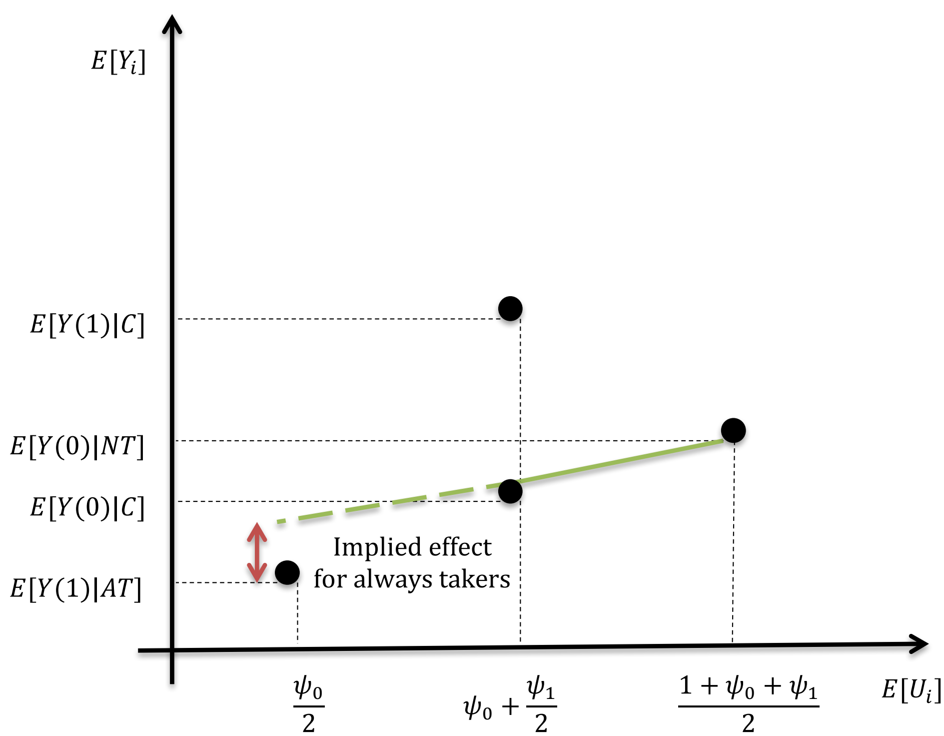
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



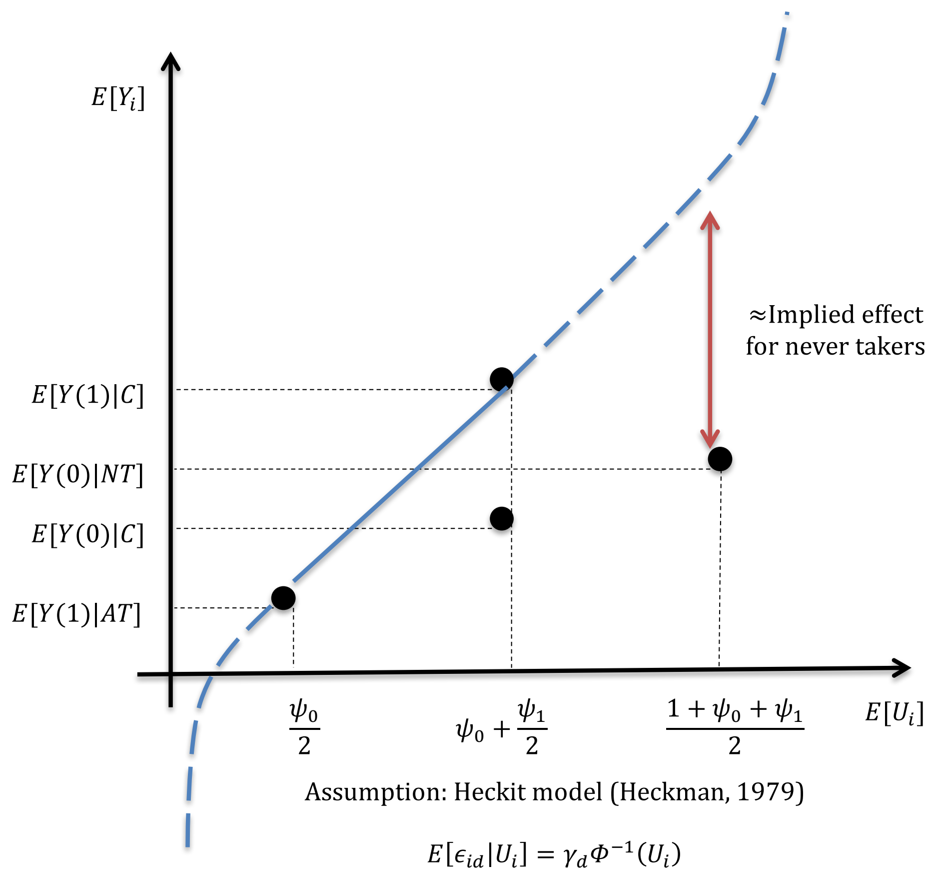
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



Extrapolation

- We can maintain the uniform representation of the selection error, $U_i \sim (0,1)$, and choose different functional forms for $E\left[Y_i(d)|U_i\right]$
  - $E[Y_i(d)|U_i] = \alpha_d + \gamma_d U_i$: Linear selection model
  - $E\left[Y_i(d)|U_i\right] = \alpha_d + \gamma_d \Phi^{-1}(U_i)$: Heckit model
- Equivalently, we can maintain the linearity restriction $E\left[Y_i(d)|U_i\right] = \alpha_d + \gamma_d U_i$, and choose different distribution functions for U_i
  - $U_i \sim U(0,1)$: Linear selection model
  - $U_i \sim N(0,1)$: Heckit model
- Specifying both a distribution for U_i and a functional form for $E\left[Y_i(d)|U_i\right]$ pins down the missing potential outcomes for ATs/NTs, allowing extrapolation from LATE

Marginal Treatment Effects

- Letting $U_i \sim U(0,1)$, choosing $E[Y_i(d)|U_i]$ implies a functional form for **marginal treatment effects** (MTE):

$$MTE(u) = E[Y_i(1) - Y_i(0)|U_i = u]$$

- MTEs are average treatment effects for individuals at a particular percentile of the unobserved cost of taking treatment (Heckman et al., 1999, 2005, 2006; Carneiro et al., 2009, 2010)
- $MTE(u)$ can be thought of as the $LATE$ associated with a hypothetical instrument that shifts the probability of treatment from u to $u + \Delta$ for small Δ
- With a continuous instrument, MTEs can be estimated as derivatives of average Y_i with respect to the conditional probability of treatment (local IV; Heckman and Vytlacil, 1999)
- With a discrete instrument, estimation requires parametric assumptions on $E[Y_i(d)|U_i]$ (Brinch et al., 2017)

Marginal Treatment Effects

- Many treatment effects of interest can be defined as weighted averages of MTEs – useful for thinking about external validity:

$$\int_0^1 \omega(u) MTE(u) du$$

- Let $\pi(z) = Pr[D_i = 1|Z_i = z]$, and $p = Pr[Z_i = 1]$
- Weights for notable treatment effects:

$$ATE: \ \omega(u) = 1$$

$$TOT: \ \omega(u) = \frac{p1\{u < \pi(1)\} + (1-p)1\{u < \pi(0)\}}{\pi(1)p + \pi(0)(1-p)}$$

$$TNT: \ \omega(u) = \frac{p1\{u \geq \pi(1)\} + (1-p)1\{u \geq \pi(0)\}}{(1-\pi(1))p + (1-\pi(0))(1-p)}$$

$$LATE: \ \omega(u) = \frac{1\{\pi(0) \leq u < \pi(1)\}}{\pi(1) - \pi(0)}$$

MTE and Policy Counterfactuals

- Models for MTE can be used to predict the effects of policies that have not been implemented
- Example: Suppose an experiment reduces the price of purchasing health insurance from p_0 to p_1, and the probability of purchase rises from π_0 to π_1
- Individuals with $U_i = \pi_1$ are on the margin between purchasing and not purchasing – we might expect them to purchase in response to a further price cut
- Heckit prediction of effect for marginal population:

$$\widehat{MTE(\pi_1)} = \hat{\alpha}_1 - \hat{\alpha}_0 + (\hat{\gamma}_1 - \hat{\gamma}_0)\,\Phi^{-1}(\hat{\pi}_1)$$

- More generally, we can use estimates of MTEs to predict TOT, TNT, ATE, or effects of other hypothetical policies

Through the Looking Glass

- CF estimate of LATE:

$$\hat{LATE} = \hat{\alpha}_1 - \hat{\alpha}_0 + \hat{E}\left[\epsilon_{i1} - \epsilon_{i0} | \gamma_0 \leq U_i < \gamma_0 + \gamma_1\right]$$

- In the binary treatment/binary instrument case with two-sided non-compliance, the two-step estimate of LATE produced by any parametric selection model is algebraically equal to the IV estimate (Kline and Walters, 2019)
- The CF estimator exactly fits the IV estimates of mean potential outcomes regardless of functional form – it connects the dots in sample
- In binary/binary case IV and CF coincide when both are used to estimate LATE
    - Equivalence serves as a natural benchmark for assessing overidentified selection models
- The assumption for $E\left[\epsilon_{it}|U_i\right]$ only matters when it is used to predict treatment effects for other subpopulations

When to Extrapolate?

- When is it reasonable to extrapolate from LATE and predict the effects of new policies?
- It depends on the interpretation of U_i, and hence on the instrument
- Equivalent to asking: when is the relationship between always taker/complier $Y_i(1)$'s likely to be a reliable guide to the relationship between complier/never taker $Y_i(1)$'s?
- If Z_i is a price shift, U_i may be viewed as (minus) willingness to pay and extrapolation may be sensible
- What would extrapolation mean in other IV examples?

Application: Kline and Walters (2016)

- Selection model example: Kline and Walters (2016) investigate effect heterogeneity with respect to counterfactual treatment choices
- Setting: Randomized evaluation of Head Start program
  - Public preschool for disadvantaged children
  - Largest preschool program in the US
  - Basic experimental impacts less impressive than earlier non-experimental analyses of HS
  - But alternative publicly subsidized preschools are now widely available for HS-eligible children. Are effects larger for kids who would otherwise stay home?

TABLE II

EXPERIMENTAL IMPACTS ON TEST SCORES

	Three-year-old cohort			Four-year-old cohort			Cohorts pooled		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Time period	Reduced form	First stage	IV	Reduced form	First stage	IV	Reduced form	First stage	IV
Year 1	0.194	0.699	0.278	0.141	0.663	0.213	0.168	0.682	0.247
	(0.029)	(0.025)	(0.041)	(0.029)	(0.022)	(0.044)	(0.021)	(0.018)	(0.031)
N		1,970			1,601			3,571	

TABLE III

PRESCHOOL CHOICES BY YEAR, COHORT, AND OFFER STATUS

		Offered			Not offered			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time period	Cohort	Head Start	Other centers	No preschool	Head Start	Other centers	No preschool	*C*-complier share
Year 1	3-year-olds	0.851	0.058	0.092	0.147	0.256	0.597	0.282
	4-year-olds	0.787	0.114	0.099	0.122	0.386	0.492	0.410
	Pooled	0.822	0.083	0.095	0.136	0.315	0.550	0.338

Kline and Walters (2016): Notation

- $Z_i \in \{0, 1\}$: Randomized experimental offer
- $D_i(z)$: Potential preschool choice.
  - h: Head Start
  - c: Other preschool center
  - n: No preschool
- Monotonicity restriction:

$$D_i(1) \neq D_i(0) \implies D_i(1) = h$$

- People only respond to a Head Start offer by enrolling in Head Start

Kline and Walters (2016): Compliance Groups

- Monotonicity implies that the population can be partitioned into five groups:
  - n-compliers: $D_i(1) = h$, $D_i(0) = n$
  - c-compliers: $D_i(1) = h$, $D_i(0) = c$
  - n-never takers: $D_i(1) = D_i(0) = n$
  - c-never takers: $D_i(1) = D_i(0) = c$
  - Always takers: $D_i(1) = D_i(0) = h$

Kline and Walters (2016): LATE

- The Head Start experiment identifies a LATE:

$$\frac{E\left[Y_i|Z_i=1\right]-E\left[Y_i|Z_i=0\right]}{E\left[1\left\{D_i=h\right\}|Z_i=1\right]-E\left[1\left\{D_i=h\right\}|Z_i=0\right]}$$

$$= E\left[Y_i(h)-Y_i(D_i(0))|D_i(1)\neq D_i(0)\right]$$

$$\equiv LATE_h$$

- This is an effect relative to a mix of counterfactuals:

$$LATE_h = S_c LATE_{ch} + (1-S_c)LATE_{nh}$$

- $LATE_{nh}$ and $LATE_{ch}$ are effects for n and c compliers relative to specific counterfactuals
- S_c is the share of c-compliers among all compliers

Kline and Walters (2016): Selection Model

- "SubLATEs" $LATE_{nh}$ and $LATE_{ch}$ aren't nonparametrically identified
- Estimate via 3-alternative selection model:

$$U_i(h) = \psi_h(X_i, Z_i) + v_{ih}$$

$$U_i(c) = \psi_c(X_i) + v_{ic}$$

$$U_i(n) = 0$$

$$(v_{ih}, v_{ic})|X_i, Z_i \sim N\left(0, \left[\begin{array}{cc} 1 & \rho(X_i) \\ \rho(X_i) & 1 \end{array}\right]\right)$$

- X_i is a vector of covariates, including demographics and experimental sites

Kline and Walters (2016): Control Functions

- Restrictions on potential outcome CEFs:

$$E\left[Y_i(d)|X_i, Z_i, v_{ih}, v_{ic}\right] = \mu_d(X_i) + \gamma_{dh} v_{ih} + \gamma_{dc} v_{ic}$$

- Averaging over individuals in a particular care alternative gives

$$E\left[Y_i(d)|X_i, Z_i, D_i = d\right] = \mu_d(X_i) + \gamma_{dh}\lambda_h\left(X_i, Z_i, d\right) + \gamma_{dc}\lambda_c\left(X_i, Z_i, d\right)$$

- $\lambda_d(X_i, Z_i, D_i)$ are bivariate versions of the Heckit Mills ratio
- Additive separability between observables and unobservables is key
- Estimates of $\mu_d(x)$, γ_{dh}, and γ_{dc} are used to construct model-based estimates of subLATEs

TABLE VIII

TREATMENT EFFECTS FOR SUBPOPULATIONS

Parameter	(1) IV	Control function (2) Covariates	Control function (3) Sites	Control function (4) Full model
$LATE_h$	0.247	0.261	0.190	0.214
	(0.031)	(0.032)	(0.076)	(0.042)
$LATE_{nh}$		0.386	0.341	0.370
		(0.143)	(0.219)	(0.088)
$LATE_{ch}$		0.023	–0.122	–0.093
		(0.251)	(0.469)	(0.154)

References

- Brinch, C., Mogstad, M., and Wiswall, M. (2017). "Beyond LATE with a discrete instrument." *Journal of Political Economy* 125(4).
- Carneiro, P., Heckman, J., and Vytlacil, E. (2010). "Evaluating marginal policy changes and the average effect of treatment for individuals at the margin." *Econometrica* 78(1).
- Carneiro, P., and Lee, S. (2009). "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality." *Journal of Econometrics* 149(2).
- Heckman, J. (1974). "Shadow prices, market wages, and labor supply." *Econometrica* 42(4).
- Heckman, J. (1976). "The common structure of statistical models of truncation, sample selection and lmiited dependent variables and a simple estimator for such models." *Annals of Economic and Social Measurement* 5(4).
- Heckman, J. (1979). "Sample selection bias as a specification error." *Econometrica* 47(1).
- Heckman, J., Urzua, S., and Vytlacil, E. (2006). "Understanding instrumental variables in models with essential heterogeneity." *Review of Economics and Statistics* 88(3).

References

- Heckman, J., and Vytlacil, E. (1999). "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the National Academy of Sciences* 96(8).
- Heckman, J., and Vytlacil, E. (2005). "Structural equations, treatment effects, and econometric policy evaluation." *Econometrica* 73(3).
- Imbens, G., and Angrist, J. (1994). "Identification and estimation of local average treatment effects." *Econometrica* 62(2).
- Kline, P., and Walters, C. (2016). "Evaluating public programs with close substitutes: the case of Head Start." *Quarterly Journal of Economics* 131(4).
- Kline and Walters (2019). "On Heckits, LATE, and numerical equivalence." *Econometrica* 87(2).
- Olsen, R. (1980). "A least squares correction for selectivity bias." *Econometrica* 48(7).
- Vytlacil, E. (2002). "Independence, monotonicity and latent index models: an equivalence result." *Econometrica* 70(1).