The Econometric Evaluation of Policy Design: Part I: Heterogeneity in Program Impacts, Modeling Self-Selection, and Parameters of Interest

Edward Vytlacil,
Yale University

Renmin University, Department of Economics
June 2017

Lectures primarily drawing upon:

- Heckman and Smith (1998), "Evaluating the Welfare State"
- Heckman, Smith with Clements (1997) "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts"
- Heckman and Vytlacil (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation"
- Heckman, Vytlacil, and Urzua (2006), "Understanding Instrumental Variables in Models with Essential Heterogeneity"

while also drawing on other work to a lesser extent.

Outline

Outline:

- Counterfactual notation, relationship to structural models.
  - Mapping back and forth between restrictions stated on counterfactual notation and restrictions on structural models.
  - Different perspectives: what can be a treatment, etc
- Treatment effects, issue of heterogeneity in treatment effects.
  - Key issue: do treatment effects vary across individuals, and do individuals know and act upon some knowledge of their own idiosyncratic effect?

Outline (cont'd)

Outline Continued:

- Treatment parameters of potential interest
  - Average treatment effect parameters
  - Distributional treatment effect parameters
  - Treatment effects on the distribution
- The Evaluation Problem:
  Selection problem, sorting problem, complications that arise if treatment effects vary and individuals know and act upon some knowledge of their own idiosyncratic effect.

Next set of lecture notes will summarize some possible approaches for the evaluation problem.

Counterfactual Notation

Notation:

- D_i dummy variable for treatment for individual i, $D_i = 1$ if treated, $= 0$ otherwise.
- Y_{1i} potential outcome for individual i if treated, what would be observed if treated.
- Y_{0i} potential outcome for individual i if not treated, what would be observed if not treated.

Counterfactual Notation (cont'd)

- Observed outcome for individual i:

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- Observe either Y_{0i} (if not treated) or Y_{1i} (if treated), never both. This problem is referred to as "the fundamental problem of causal inference" by Holland (1986)

Counterfactual Notation (cont'd)

Implicit in notation: No interaction across units

- Called "Stable Unit Treatment Value Assumption (SUTVA)" in biostatistics.
- Rules out general equilibrium effects, peer-effects, etc., in economics.
- Assumption not always appropriate, for example:
  - A large scale vaccination program, where an individual being vaccinated may prevent him from infecting non-vaccinated individuals.
  - A large scale increase in college-attendance, causing an increase in the supply of skilled labor and thus a decrease in the price of skilled labor.

Counterfactual Notation (cont'd)

Interesting work relaxing the assumption of no interactions across units, e.g., allowing peer-effect models in education (see e.g. Brock and Durlauf, 2001; Graham, 2008, 2015).

Often redefine definition of unit.

Recent work on designing randomized control trials to measure spillover effects (see, e.g., Baird, Bohren, Macintosh, and Ozler, 2016).

Counterfactual Notation (cont'd)

Can extend notation to multivalued or continuous treatment,

$$\{Y_{d,i} : d \in \mathcal{D}\}, \qquad Y_i = Y_{d,i} \text{ if } D_i = d$$

where $\mathcal{D}$ set of possible treatments. Most work focuses on D_i binary.

Examples of some exceptions:

- Matching with discrete treatment (Lechner, 2001, Imbens, 2003), or with continuous treatment (Behrman, Cheng, and Todd, 2004, Hirano and Imbens 2004).
- Instrumental variables with discrete ordered treatment (Angrist and Imbens,1995, Heckman, Vytlacil and Urzua, 2006), or with discrete unordered treatment (Heckman, Vytlacil and Urzua, 2008; Pinto 2016)
- Instrumental variables and control function with continuous treatment (Florens, Heckman, Meghir and Vytlacil, 2008)

Treatment models with continuous treatment closely related to recent work on endogenous regressors in nonseparable models, see e.g. Altonji, Ichimura and Otsu (2008), Altonji and Matzkin (2005), Blundell and Powell (2004) and Imbens and Newey (2009).

Counterfactual Notation (cont'd)

Define:

- X_i controls, called confounders within biostatistics, observed variables that directly affect Y_{0i}, Y_{1i}.
- Z_i, instruments, variables that affect D_i but do not directly affect Y_{0i}, Y_{1i}.
- Will sometimes define potential treatments

$$\{D_{zi} : z \in \mathcal{Z}\},$$

treatment choice that i would have chosen if randomly assigned $Z_i = z$, where $\mathcal{Z}$ is set of possible instrument values.

Counterfactual versus Econometric Structural Notation

Can map back and forth between structural notation and counterfactual/potential outcome notation.

For example,

$$Y = g(X_i, D_i, \epsilon_i) \Rightarrow \begin{cases} Y_{1i} &= g(X_i, 1, \epsilon_i) \\ Y_{0i} &= g(X_i, 0, \epsilon_i) \end{cases}$$

For example,

$$Y = X_i\beta + \alpha D_i + \epsilon_i \Rightarrow \begin{cases} Y_{1i} &= X_i\beta + \alpha + \epsilon_i \\ Y_{0i} &= X_i\beta + \epsilon_i \end{cases}$$

Counterfactual versus Econometric Structural Notation (equivalences)

Can express restrictions on counterfactual notation, and then translate into restrictions on structural equations, or vice versa.

For example, Vytlacil (2002) establishes equivalence between Imbens and Angrist (1994) LATE conditions and the nonparametric selection model.

Imbens Angrist conditions (1994)

Imbens and Angrist (1994) impose the following conditions:

IV-1 (Independence)

$Z \perp\!\!\!\perp \left(Y_1, Y_0, \{D(z)\}_{z \in \mathcal{Z}}\right)$.

IV-2 (Rank)

$\Pr(D = 1 \mid Z)$ *depends on* Z.

IV-3 (Monotonicity)

For all $z, z' \in \mathcal{Z}$, *either* $D_i(z) \geq D_i(z')$ *for all* i, *or* $D_i(z) \leq D_i(z')$ *for all* i.

Vytlacil (2002) Equivalence

Let $\mathbf{1}[\cdot]$ denote the logical indicator function.

Vytlacil (2002) shows that conditions (IV-1) - (IV-3) are equivalent to a nonparametric selection model of the following form:

SELECTION-1 (Selection Model)

$D_i = \mathbf{1}[\mu(Z_i) \geq U_i]$, $Z_i \perp\!\!\!\perp (Y_{0i}, Y_{1i}, U_i)$, *and* $\mu(\cdot)$ *is a nontrivial function of* Z_i.

See also Pinto (2016).

Vytlacil and Yildiz (2007) Equivalence

As another example of equating restrictions on counterfactual notation and restrictions on structural equations, Vytlacil and Yildiz (2007) establishes equivalence between weak separability and a monotonicity restriction that generalizes (IV-3).

Let $W_i = (D_i, X_i)$, D not necessarily binary.

Let $Y_i(w)$ denote potential outcome setting $W_i = w$.

Vytlacil and Yildiz (2007) Equivalence

Consider:

M-1 (Monotonicity)

For all $w, w' \in \mathcal{W}$, either $Y_i(w) \geq Y_i(w')$ for all i, or $Y_i(w) \leq Y_i(w')$ for all i.

and

M-2 (Weak Separabiltiy)

$Y_i = g(\theta(X_i, D_i), \epsilon_i)$, where θ is scalar valued and g is weakly increasing in θ.

Vytlacil and Yildiz establish that (M-1) is equivalent to (M-2).

Equating Potential Outcomes and Structural Models (cont'd)

There are some philosophical differences between those who use potential outcome frameworks from biostatistics and those who use structural models in econometrics.

For example, see Holland (1986), and Heckman and Vytlacil (2007),

- "Effect of Causes" versus "Causes of Effects"
  - "statistical solution" versus "scientific solution"
- What can be a cause?
  - Can gender be a cause? Can studying be a cause?

Treatment Effects

Suppose $D_i = 0, 1$.

$\Delta_i = Y_{1i} - Y_{0i}$ is treatment effect for individual i.

Key issue: how does Δ_i vary with i?

Four cases:

- A Homogeneous treatment effects
- B Homogeneous treatment effects conditional on X
- C1 Heterogeneous Treatment Effects without "Essential Heterogeneity"
- C2 Heterogeneous Treatment Effects with "Essential Heterogeneity"

Homogeneous Treatment Effects

(A) Homogeneous Treatment Effects:
$Y_{1i} - Y_{0i} = \Delta$, does not vary with i.

- All individuals have the same effect of treatment.
- For example: $Y_i = X_i\beta + \alpha D_i + \epsilon_i$, then $Y_{1i} - Y_{0i} = \alpha$.
- More generally, restriction is equivalent to $Y_i = g(X_i) + \alpha D_i + \epsilon_i$, imposes additive separability between D_i and (X_i, ϵ_i).
- If Y_i is discrete, then typically this restriction is logically incoherent.
- If Y_i is continuous, then this restriction is logically coherent, but is it reasonable?

Homogeneous Treatment Effects Conditional on X

(B) Homogeneous Treatment Effects Conditional on X_i:

$$X_i = X_{i'} \Rightarrow Y_{1i} - Y_{0i} = Y_{1i'} - Y_{0i'}$$

- Equivalent to $\Delta_i = \Delta(X_i)$.
- All individuals with the same X characteristics have the same effect of treatment.
- For example: $Y_i = X_i\beta + D_iX_i\alpha + \epsilon_i$, then $Y_{1i} - Y_{0i} = X_i\alpha$.
- More generally, restriction is equivalent to $Y_i = g(X_i, D_i) + \epsilon_i \Rightarrow \Delta_i = g(X_i, 1) - g(X_i, 0)$, imposes additive separability between (D_i, X_i) and ϵ_i.
- Is restriction logically coherent? Is restriction plausible?

Heterogeneous Effects

(C) Heterogeneous Treatment Effects:
Δ_i varies freely with i, even conditional on X.

- Even individuals with the same X characteristics may have different effects of the treatment.
- For example: $Y_i = X_i\beta + \alpha_i D_i + \epsilon_i$, with α_i random coefficient, then $Y_{1i} - Y_{0i} = \alpha_i$.
- More generally, restriction is equivalent to $Y_i = g(X_i, D_i, \epsilon_i)$, so that $Y_{1i} - Y_{0i} = g(X_i, 1, \epsilon_i) - g(X_i, 0, \epsilon_i)$, does not impose additive separability.
- Key distinction: "essential heterogeneity" or not (Heckman, Vytlacil, and Urzua (2006)

Heterogeneous Effects (cont'd)

Key question with heterogeneous effects:

- Do agents select into treatment based, in part, on their own idiosyncratic effect?

Two cases:

C1 $Y_{1i} - Y_{0i} \perp\!\!\!\perp D_i | X_i$, agents are not selecting into treatment based on $Y_{1i} - Y_{0i}$ (conditional on X_i).

C2 $Y_{1i} - Y_{0i} \not\perp\!\!\!\perp D_i | X_i$, agents do select into treatment based on $Y_{1i} - Y_{0i}$ (conditional on X_i).

Case [C2] is called "Essential Heterogeneity" by Heckman, Vytlacil and Urzua (2006).

Heterogeneous Effects (cont'd)

What is parameter of interest, and how hard to identify and estimate parameter of interest?

A If homogeneous effects, then only one scalar parameter of effect, Δ. Identification and estimation issues are standard and relatively easy.

Heterogeneous Effects (cont'd)

What is parameter of interest, and how hard to identify and estimate parameter of interest?

B If homogeneous treatment effect conditional on X,
- Can be interested in effect for different subgroups (defined by X),
  - Identification issues same as for [A], though estimation will be less precise, possible multiple hypothesis testing issues and curse of dimensionality issues..
- Can can be interested in average effect, averaging over X.
  - Raises question, if average over X, which average over X?
  - Often support/extrapolation issues.
  - Some new issues with estimation, inference.

Heterogeneous Effects (cont'd)

C1 If effect heterogeneous, but individuals do not select into treatment based upon own effect, then almost exactly the same as [B] if desire average effect. Only new issue is if desire distribution of treatment effects.

C2 If effect heterogeneous, and individuals do select into treatment based upon own effect, than question of what is parameter of interest is less simple. Identification and estimation is much harder.

Mean Treatment Parameters

What are parameters of interest?

Most often, consider average treatment parameters:

- Average Treatment Effect, ATE $= E(Y_{1i} - Y_{0i})$.
- Treatment on the Treated, TT$= E(Y_{1i} - Y_{0i}|D_i = 1)$.
- Treatment on the Untreated, TUT$= E(Y_{1i} - Y_{0i}|D_i = 0)$.

Mean Treatment Parameters

What are parameters of interest?

Or same parameters conditional on X.
$\text{ATE}(X_i) = E(Y_{1i} - Y_{0i}|X_i)$
$\text{TT}(X_i) = E(Y_{1i} - Y_{0i}|X_i, D_i = 1)$,
$\text{TUT}(X_i) = E(Y_{1i} - Y_{0i}|X_i, D_i = 0)$.
(Sometimes referred to as $\text{CATE}(X_i)$, etc.)

By law of iterated expectations:

- $E(Y_{1i} - Y_{0i}) = \int E(Y_{1i} - Y_{0i}|X_i)dF_X$.
- $E(Y_{1i} - Y_{0i}|D_i = 1) = \int E(Y_{1i} - Y_{0i}|X_i, D_i = 1)dF_{X|D=1}$.
- $E(Y_{1i} - Y_{0i}|D_i = 0) = \int E(Y_{1i} - Y_{0i}|X_i, D_i = 0)dF_{X|D=0}$.

Mean Treatment Parameters

What are parameters of interest?

Above parameters average over population distribution of X. Sometimes ATE is referred to as Population Average Treatment Effect (PATE).

Sometimes, especially in bio-statistics, define parameters averaging over sample distribution of X, e.g., Sample Average Treatment Effect (SATE). For my lectures, will exclusively discuss population parameters.

Mean Treatment Parameters (cont'd)

- In case [A], homogeneous treatment effects, all parameters coincide,
  $\text{ATE}=\text{TT}=\text{TUT}=\text{ATE}(X)=\text{TT}(X)=\text{TUT}(X)$
- In case [B], homogeneous effects conditional on X, all mean parameters conditional on X coincide,
  $\text{ATE}(X)=\text{TT}(X)=\text{TUT}(X)$,
  but possible to have $\text{ATE}\neq\text{TT}\neq\text{TUT}$ if distribution of X different for treated versus untreated.
- In case [C1], heterogeneous effects but not essential heterogeneity, same as in case [B].
- In case [C2], essential heterogeneity, all parameters may be different:
  $\text{ATE}\neq\text{TT}\neq\text{TUT}\neq\text{ATE}(X)\neq\text{TT}(X)\neq\text{TUT}(X)$

Mean Treatment Parameters (cont'd)

- Which parameter is recovered by idealized randomized experiment? Depends on how implemented.
- What parameter is recovered by instrumental variables estimation? by selection models?
- Important issue when comparing results across studies, different numbers from different studies might not be in contradiction, might be estimating different parameters.

Distributional Treatment Parameters

Also possible to be interested in parameters describing the distribution of treatment effects, in particular, in parameters that depend on the joint distribution of (Y_0, Y_1) including on the dependence between Y_0 and Y_1. For example:

- $\Pr[Y_{1i} > Y_{0i}]$, probability of benefiting from the treatment.
- $\Pr[Y_{1i} > Y_{0i} | Y_{0i} < c]$, probability of benefiting from the treatment among those worst off without the treatment.
- $\Pr[Y_{1i} > c | Y_{0i} < c]$, e.g., probability of being above the poverty line with treatment among those who would be below the poverty line without treatment.

Distributional Treatment Parameters (cont'd)

Additional examples of distributional treatment effect parameters:

- Median$(Y_{1i} - Y_{0i})$, or other quantiles of $Y_{1i} - Y_{0i}$.
- $E[Y_{1i} - Y_{0i} | Y_{0i} < c]$, expected benefit among those who would be worst off without the treatment (e.g., expected benefit among those who would be below the poverty line without the treatment).
  - $E[Y_{1i} - Y_{0i} | Y_{0i} < c]$ is a conditional expectation/mean, but depends on dependence between Y_0 and Y_1.

Distributional Treatment Parameters (cont'd)

- Most ambitious parameter: The joint cumulative distribution function of (Y_0, Y_1):
  $\Pr[Y_{1i} \leq t_1, Y_{0i} \leq t_0]$ for all t_0, t_1
  - Knowledge of $\Pr[Y_{1i} \leq t_1, Y_{0i} \leq t_0]$ for all t_0, t_1 would allow one to calculate all of the above parameters, and any other distributional parameter. Would provide detailed knowledge of effects of the treatment.
- Can also define treatment on the treated and treatment on the untreated versions of the distributional parameters, for example,
  - $\Pr[Y_{1i} > Y_{0i} | D = 1]$, probability of benefiting from the treatment among those selecting to receive treatment.
  - $\text{Median}(Y_{1i} - Y_{0i} | D = 1)$, median benefit among those selecting to receive treatment

Distributional Treatment Parameters (cont'd)

- It is possible to have average treatment parameters and distributional treatment parameters in opposite directions.
    - For example, it is possible to have $\Pr[Y_{1i} > Y_{0i}] < .5$ and $\text{Median}(Y_1 - Y_0) < 0$ so that most individuals are hurt by the treatment, but $E(Y_1 - Y_0) > 0$ so that the average effect of the treatment is positive, if some small fraction of individuals have a large positive treatment effect.
- Choice of parameter of interest related to social welfare function. Can also be related to voting rules. See Heckman and Smith (1998).

Distributional Treatment Parameters (cont'd)

Note:

- If Y is binary, the distributional parameters have a particularly simple form.
  - For example, if $Y = 1$ if live, $Y = 0$ die, then
    - $\Pr[Y_{1i} > Y_{0i}] = \Pr[Y_{1i} = 1, Y_{0i} = 0]$ is probability life is saved because of treatment.
    - $\Pr[Y_{1i} < Y_{0i}] = \Pr[Y_{1i} = 0, Y_{0i} = 1]$ is probability die because of treatment.

- In case (A), homogeneous treatment effects, distributional parameters have a trivial form. For example, if treatment effects are homogeneous, then $\Pr[Y_{1i} > Y_{0i}]$ equals 1 or 0, and $E(Y_1 - Y_0) = \text{Median}(Y_1 - Y_0)$, etc. Distributional parameters are only non-trivial and of real interest if effects are heterogeneous (case C).

Distributional Treatment Parameters (cont'd)

Key Issue:

Mean treatment parameters are fundamentally easier to identify than distributional treatment parameters.

- Linearity of expectations, $E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$, average treatment parameters depend only on marginal distributions of Y_0 and Y_1, not on full joint distribution.
- Above distributional parameters depend on joint distribution of (Y_0, Y_1), including on the dependence between Y_0 and Y_1, not just on the marginal distributions of Y_0, Y_1.

Distributional Treatment Parameters (cont'd)

Even with an idealized randomized experiment, one cannot identify distributional treatment parameters without more assumptions/more structure.

Reason:

- Idealized random experiments allow one to identify distribution of Y_0 from those subjects randomly denied treatment, and to identify the distribution of Y_1 from those randomly assigned to treatment, and thus to identify parameters that depend only on the marginal distributions of Y_0, Y_1 such as the average treatment effect.
- However, one never observes Y_0 and Y_1 for the same individual, and thus randomized experiments do not identify the dependence between Y_0, Y_1, and thus do not identify distributional treatment parameters.

Distributional Treatment Parameters (cont'd)

How to recover distributional treatment parameters?

Trivial if homogeneous effect or homogeneous conditional on X, but difficult otherwise.

Possible approaches include imposing rank ordering assumption between Y_{0i} and Y_{1i}.

- Could impose perfect positive dependence.
- Could impose perfect negative dependence.
- Could impose any other level of dependence. which one to impose?
- Not testable.

Table: Imposing Rank Ordering to Identify Distribution of $Y_1 - Y_0$

From Heckman, Smith with Clements (1997)
Based on JTPA Randomized Experiment

Estimated parameters of the impact distribution; perfect positive dependence and perfect negative dependence cases
(National JTPA Study 18 month impact sample; adult females)

Statistic	Perfect positive dependence	Perfect negative dependence
5th Percentile	0·00 (47·50)	−22350·00 (547·17)
25th Percentile	572·00 (232·90)	−11755·00 (411·83)
50th Percentile	864·00 (269·26)	580·00 (389·51)
75th Percentile	966·00 (305·74)	12791·00 (253·18)
95th Percentile	2003·00 (543·03)	23351·00 (341·41)
Percent positive	100·00 (1·60)	52·00 (0·81)
Impact standard deviation	1857·75 (480·17)	16432·43 (265·88)
Outcome correlation	0·9903 (0·0048)	−0·6592 (0·0184)

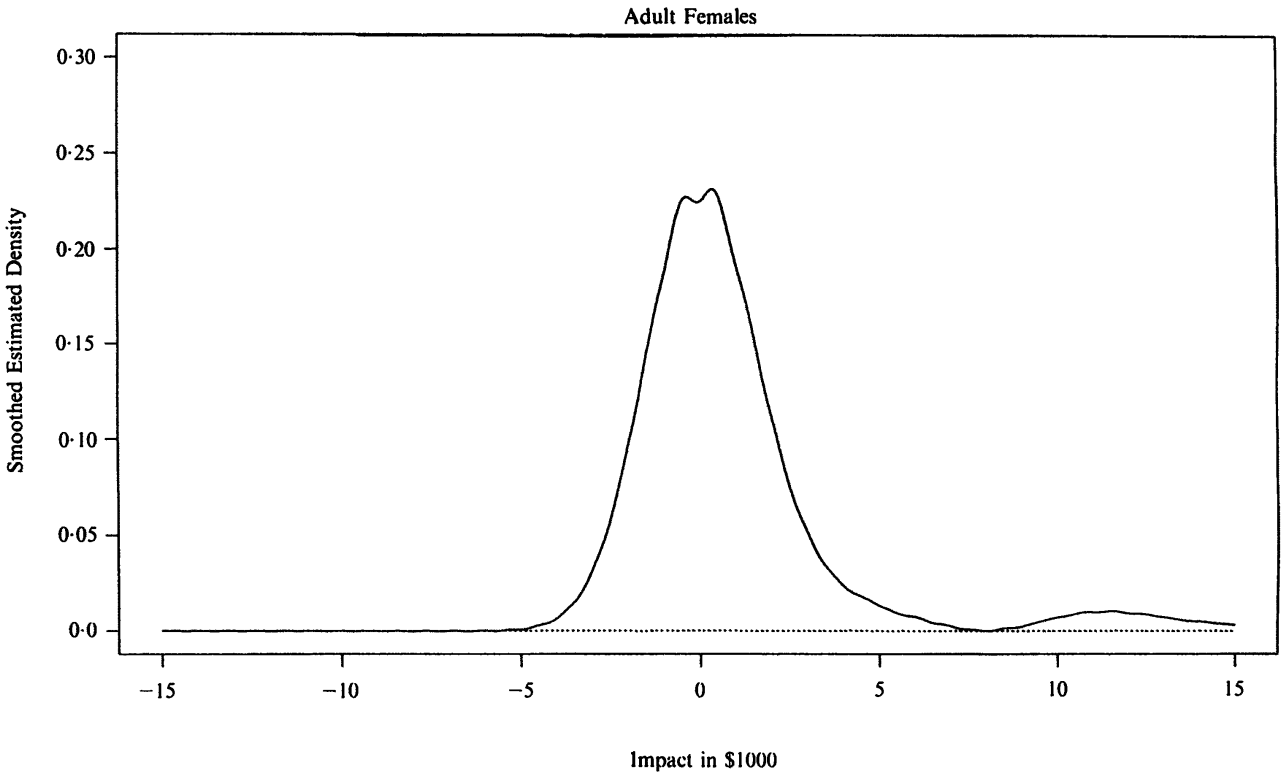
Distributional Treatment Parameters (cont'd)

Another possible approach:
Deconvolution, imposing that $Y_{1i} - Y_{0i} \perp\!\!\!\perp Y_{0i}$.

- How to justify?
- Note: difficult estimation, inference issues for deconvolution.

Figure: Deconvolution Estimator of Density of $Y_1 - Y_0$

From Heckman, Smith with Clements (1997)
Based on JTPA Randomized Experiment



Distributional Treatment Parameters (cont'd)

Possible approaches continued:

- Factor model approach, with dependence between (D, Y_0, Y_1) generated by scalar factor.
  - With factor model structure, can use dependence between D, Y_0 and between D, Y_1 to infer dependence between Y_0, Y_1
- Impose Roy model, $D_i = 1[Y_{1i} \geq Y_{0i}]$.
  - Roy model assumption allows one to use individuals self-selection into treatment to infer dependence between Y_0, Y_1.

Distributional Treatment Parameters (cont'd)

Another Approach: Do not impose more structure, instead bound distribution of (Y_{0i}, Y_{1i}) and thus bound distributional treatment effect parameters.

For example, following Heckman, Smith with Clements (1997), for Y_0, Y_1 binary, identify from randomized experiment:

1. $\Pr[Y_1 = 1]$
2. $\Pr[Y_0 = 1]$

Do not identify $\Pr[Y_1 = 1, Y_0 = 0]$, $\Pr[Y_1 = 1, Y_0 = 1]$, etc, but can use Frechet bounds on joint distribution to bound these terms.

Figure: Frechet Bounds

From Heckman, Smith with Clements (1997)

		Untreated E	Untreated N	
Treated	E	P_{EE}	P_{EN}	$P_{E\bullet}$
	N	P_{NE}	P_{NN}	$P_{N\bullet}$
		$P_{\bullet E}$	$P_{\bullet N}$	

FIGURE 2

2×2 Table representation

Table: Frechet Bounds

From Heckman, Smith with Clements (1997)

TABLE 4

Fraction employed in the 16th, 17th or 18th months after random assignment and Fréchet–Hoeffding bounds on the probabilities P_{NE} and P_{EN}

(National JTPA study 18 month impact sample; adult females)

Parameter	Estimate
Fraction employed in the treatment group	0·64 (0·01)
Fraction employed in the control group	0·61 (0·01)
Bounds on P_{EN}	[0·03, 0·39] (0·01), (0·01)
Bounds on P_{NE}	[0·00, 0·36] (0·00), (0·01)

1. Employment percentages are based on self-reported employment in months 16, 17 and 18 after random assignment. A person is coded as employed if the sum of their self-reported earnings over these three months is positive.
2. P_{ij} is the probability of having employment status i in the treated state and employment status j in the untreated state, where i and j take on the values E for employed and N for not employed. The Fréchet–Hoeffding bounds are given in the text.

Distributional Treatment Parameters (cont'd)

I will not further discuss distributional treatment parameters in these lectures. For more information on these parameters and their identification, see, e.g.,

- Heckman and Smith (1998), Heckman and Smith with Clements (1997)
- Aakvik, Heckman and Vytlacil (2005), Caneiro, Hansen and Heckman (2003)
- recent work by Yanqin Fan and her co-authors.

Treatment Effects on Distribution

Treatment effects on distribution, different from distribution of treatment effects.

For example:

- $\Pr[Y_{1i} \geq c] - \Pr[Y_{0i} \geq c]$, e.g., c might be poverty line.
- $\text{Med}(Y_{1i}) - \text{Med}(Y_{0i})$.
- More generally, difference in any quantile of Y_{1i} and corresponding quantile of Y_{0i}.

Note:
$\text{Med}(Y_{1i}) - \text{Med}(Y_{0i})$ does NOT in general equal $\text{Med}(Y_{1i} - Y_{0i})$!

Treatment Effects on Distribution (cont'd)

- Effect on distribution, not distribution of effects.
- Only requires knowledge of marginal distributions of Y_{0i}, Y_{1i}, not full joint distribution.
- Can be recovered by randomized experiments.
- Treatment effect estimators for average treatment effect parameters have corresponding estimators for corresponding quantile treatment effects (QTE) parameters.

Treatment Effects on Distribution (cont'd)

- Treatment effect estimators for average treatment effect parameters have corresponding estimators for quantile treatment effects (QTE) parameters, e.g.,
  - matching methods for QTE (see, e.g., Firpo 2007),
  - instrumental variable methods for QTE (see, e.g., Chernozhukov and Hansen, 2001, Abadie, Angrist and Imbens, 2002),
  - regression discontinuity methods for QTE (see, e.g., Frandsen, Frolich, and Melly, 2012)
  - MTE/selection model methods for quantile treatment effects (Carneiro and Lee, 2009).
- I will not discuss further in these lectures.

Other Treatment Parameters

Other parameters recently developed:

- Local Average Treatment Effect (LATE), Imbens and Angrist (1994)

  Average effect for those induced to change treatment status because of a change in the instrument. An instrument dependent parameter.

- Marginal Treatment Effect (MTE), Bjorklund and Moffitt (1987), Heckman and Vytlacil (various)

  Average effect for those individuals with a given unobserved desire to receive treatment.

Other Treatment Parameters

None of the above treatment parameters will generally answer policy questions except in extreme cases. For example, if treatment is going to college:

- Treatment on the treated answers extreme policy question "what is the effect of current regime versus eliminating all college education"
- Treatment on the untreated answers extreme policy question "what is the effect of current regime versus forcing everyone to go to college?"
- Average treatment effects answers extreme policy question "what is the effect of forcing everyone to go to college versus eliminating all college education?"

Other Treatment Parameters

Can also define treatment parameters to answer particular policy questions.

- For example, what is effect of increasing or decreasing incentives to go to college? For example, what is the effect of increasing tuition subsidies?
- See, e.g., Ichimura and Taber (2000, 2002) and Carneiro, Heckman and Vytlacil (2010, 2011).

Evaluation Problem

Now consider identification of the average treatment effect parameters

- Classical identification problem: selection bias
- New identification problem under essential heterogeneity: sorting gain.

Evaluation Problem with Homogeneous Treatment Effect: Selection Bias

Suppose Homogeneous Treatment Effect, Δ a constant.

Classical evaluation problem: Selection Bias

$$\begin{aligned}\Delta_i = \Delta \Rightarrow Y &= Y_0 + D\Delta \\ &\Rightarrow E(Y|D=1) - E(Y|D=0) \\ &= \Delta + \underbrace{E(Y_0|D=1) - E(Y_0|D=0)}_{\text{Selection Bias}}\end{aligned}$$

Same analysis conditional on X if homogeneous treatment effects conditional on X.

Selection Bias (cont'd)

$E(Y_0|D=1) - E(Y_0|D=0)$ is selection bias:

- Selection on the base state

  if treated had not received treatment, would they have similar outcomes as the non treated?

- Sometimes called "Ability Bias" in labor economics.
- Common worry: omitted variable (e.g., ability), omitted variable correlated with selection into treatment.

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain

Now allow heterogeneous treatment effects.

$$Y = Y_0 + D(Y_1 - Y_0).$$

$$\begin{aligned}
&E(Y|D=1) - E(Y|D=0) \\
&\quad = \underbrace{E(Y_1 - Y_0|D=1)}_{\text{TT}} + \underbrace{E(Y_0|D=1) - E(Y_0|D=0)}_{\text{Selection Bias}} \\
&\quad = E(Y_1 - Y_0) + \left\{ \begin{array}{c} E(Y_1 - Y_0|D=1) \\ -E(Y_1 - Y_0) \end{array} \right\} + \left\{ \begin{array}{c} E(Y_0|D=1) \\ -E(Y_0|D=0) \end{array} \right\} \\
&\quad = \quad \text{ATE} \quad + \quad \text{Sorting Gain} \quad + \quad \text{Selection Bias}
\end{aligned}$$

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain (cont'd)

With heterogeneous effects, bias depends on parameter of interest.

- For TT, bias is selection bias, exactly as before.
- For ATE, we now have additional bias term: sorting gain
  - selection on the gain, benefit to those who sort into treatment versus average person.
  - Expect sorting gain term to be nonzero under essential heterogeneity
  - Sorting gain term is positive for Roy model.
- If effects are heterogeneous but without essential heterogeneity, than analysis is the same as for homogeneous case, sorting gain term is equal to zero.

Table: Selection Bias, Sorting Gain, and Comparative Advantage

Parameter	*Estimation*
OLS	0.2929
IV*	0.5609
ATE	0.4336
TT	0.5149
TUT	0.3630
Bias**	–0.1407
Selection Bias***	–0.2220
Sorting Gain****	0.0813

*Using the propensity score as the instrument.
**$Bias = \text{OLS} - \text{ATE}$.
***$Selection\ bias = \text{OLS} - \text{TT}$.
****$Sorting\ gain = \text{TT} - \text{ATE}$.

Source: Heckman and Li (2004)
"Selection Bias, Comparative Advantage, and Heterogeneous Returns to Education: Evidence from China in 2000 ."

Evaluation Problem with Heterogeneous Treatment Effect: Selection Bias and Sorting Gain (cont'd)

When considering alternative methods to evaluate effects of a treatment, important to consider:

1. Essential heterogeneity? Do effects vary across individuals, and do individuals self-select into treatment based on their own idiosyncratic effect?
2. What is parameter of interest?
3. What is bias of econometric method for particular parameter of interest?

Different approaches may give different results but be equally valid if estimating different parameters.