

# Introduction to causal identification



**Seema Jayachandran**

**BREAD-IGC-ISI Summer School, July 2012**

# Outline of lecture

1. Overview of microempirical methods
2. Rubin causal model
3. Instrumental variables

# Estimating causal effects

- In policy-oriented research, we often want to estimate the causal effect of a program
- What is this effect of a performance incentive system for teachers on student test scores?
- Observe the students scores after school adopts new pay scheme
- Need to know counterfactual of what the test scores would have been had the teachers not been incentivized, *but all else had been equal*

# Challenge facing micro-empirical researchers

- Once upon a time, could run a regression and control for some variables to isolate the causal effect
  - Control for observable measures of the headmaster's quality or student characteristics
  - Regard program status, conditional on observables, as exogenous
- Today, there are high standards for ruling out unobservables biasing the results
  - Policy changes are endogenous
  - Take-up of interventions is endogenous
  - Other sources of omitted variable bias or reverse causality
- Biased estimates means we are drawing the wrong policy conclusions from the data

# Microempirical methods aimed at solving endogeneity

- Randomized experiments
- Natural experiments
- IV
- Regression discontinuity
- Difference-in-differences

# Rubin Causal Model

- Framework for thinking about causal effects
- We apply the treatment and comparison terminology liberally— to problems where we want to evaluate a causal effect
- A policy does not need to be binary (you are treated or you are not), but we consider this case for simplicity

# Notation

- Let  $T$  denote treated and  $C$  denote comparison or control
- Let  $Y_i^T$  be the test score of a student  $i$  in a school with teacher incentives
- $Y_i^C$ : test score of the *same student*  $i$  if her school did not have teacher incentives

# Potential outcomes

- Here treated means your school has the teacher incentive program in place
- $Y_i^T$  and  $Y_i^C$  are “potential outcomes”; they might be actual outcomes or hypothetical outcomes

# Causal effect of being treated

- We are interested in the difference:

$$Y_i^T - Y_i^C$$

- This is the effect of the teacher incentive program on test scores
- Problem: we don't observe student  $i$  both with and without the teacher incentive program in her school at the same time
- Fundamental Problem of Causal Inference:

It is impossible to observe the value of  $Y_i^T$  and  $Y_i^C$  on the same unit and, therefore, it is impossible to observe the effect of  $T$  for  $i$ .

# Average causal effects in a population

- Statistical solution replaces the impossible-to-observe causal effect of  $T$  on a specific unit  $i$  with the possible-to-estimate average causal effect of  $T$  over a population of units
- We can hope to learn the *average* effect of the teacher incentive program in a population of students

$$E[Y_i^T - Y_i^C]$$

- For simplicity, let's assume the treatment effect is constant across individuals:  $Y_i^T - Y_i^C = Y_j^T - Y_j^C$  for all individuals  $i$  and  $j$

# Assumptions

- Treatment status of one person does not affect others' outcomes, i.e., no spillovers
  - *SUTVA*: Stable Unit Treatment Value Assumption
- Homogenous treatment
- Neither of these are insurmountable problems if they do not hold
- Essentially means we need to think of different treatments for individual  $i$ , e.g., with and without others being treated if there are spillovers or intense versus light if treatment type varies

# Measuring average effects in a population

- Imagine we have access to data on several individuals in an area
- Some schools have the teacher incentive program, some do not
  - Policy is optional and some headmasters adopt new policy
  - Government chooses some schools in which to pilot the policy
- We can take the average in both groups and calculate the difference

$$\begin{aligned} E[Y_i | \text{program in place}] - E[Y_i | \text{program not in place}] \\ = E[Y_i^T | T] - E[Y_i^C | C] \end{aligned}$$

- Subtract and add  $E[Y_i^C | T]$

$$\begin{aligned} E[Y_i^T | T] - E[Y_i^C | T] - E[Y_i^C | C] + E[Y_i^C | T] \\ = E[Y_i^T - Y_i^C | T] + E[Y_i^C | T] - E[Y_i^C | C] \end{aligned}$$

# Comparing the treated to the untreated

$$\begin{aligned} E[Y_i^T|T] - E[Y_i^C|C] = \\ \underbrace{E[Y_i^T - Y_i^C|T]}_{\text{treatment effect}} + \underbrace{E[Y_i^C|T] - E[Y_i^C|C]}_{\text{selection bias}} \end{aligned}$$

- The first term  $E[Y_i^T - Y_i^C|T]$  is the *treatment effect* that we are trying to isolate

On average, what is the effect of the teacher incentive program on student test scores?

- In words, what is:
  - $E[Y_i^C|T]$ ?
  - $E[Y_i^C|C]$ ?
  - The difference  $E[Y_i^C|T] - E[Y_i^C|C]$ ?

# Selection bias

- Is the difference  $E[Y_i^C|T] - E[Y_i^C|C]$  likely to be positive or negative?
- The difference is the *selection bias*
- There may be systematic differences between students in schools with the program and those in schools without it
- How would pre-period data before any school had the program help?

Under what circumstances would it eliminate the selection bias?

– Next lecture: Difference-in-differences

# Analogous problem when evaluating other policies

- Suppose a microcredit organization began offering microcredit in some villages in a province but not others
- Can we measure the effect of the program by comparing villages with and without the program?
- $T$  is having microcredit in the village and let  $Y$  be average income in the village
- Is  $E[Y^C|T] - E[Y^C|C]$  likely to be positive or negative?

# Eliminating the selection bias

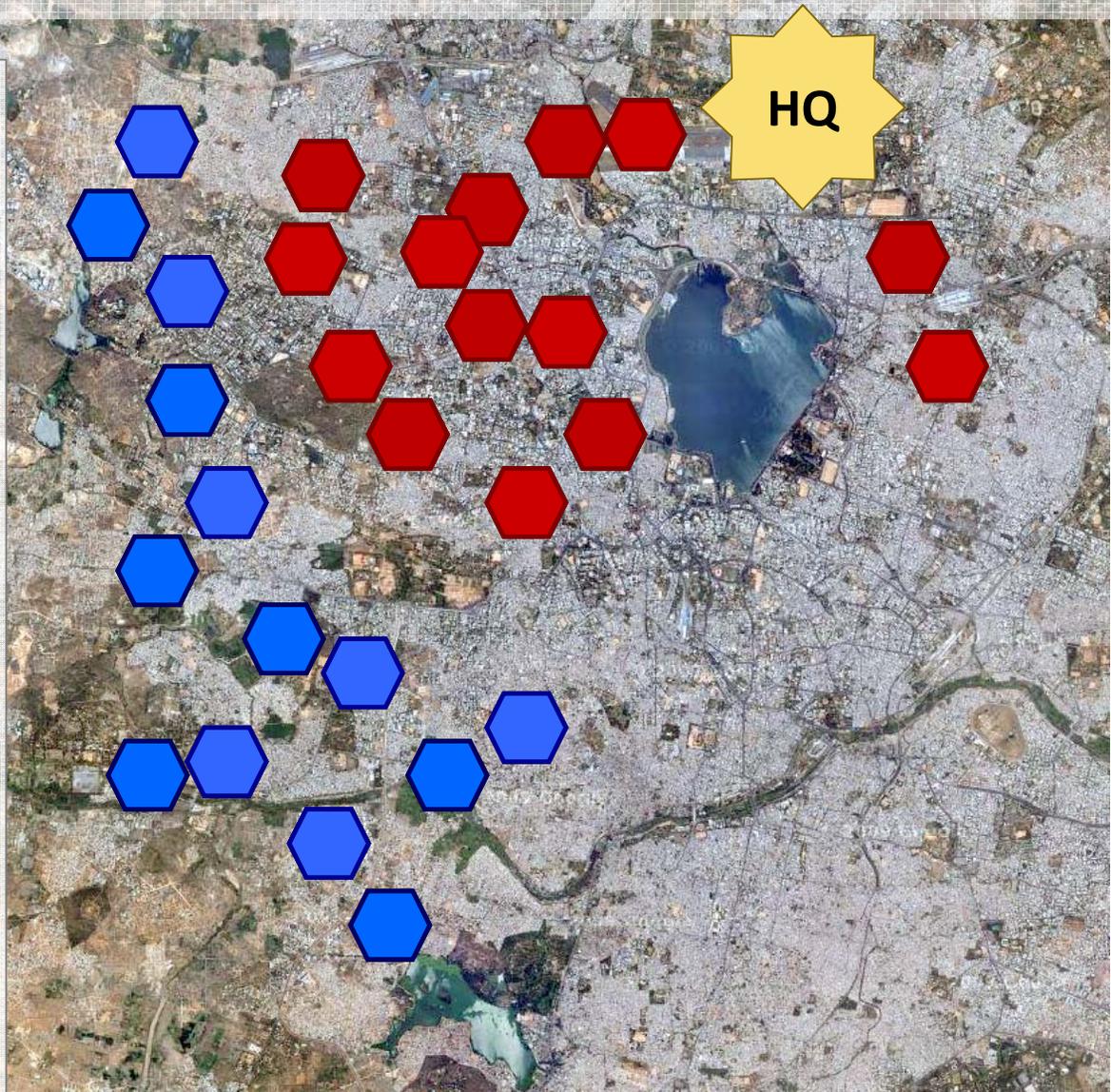
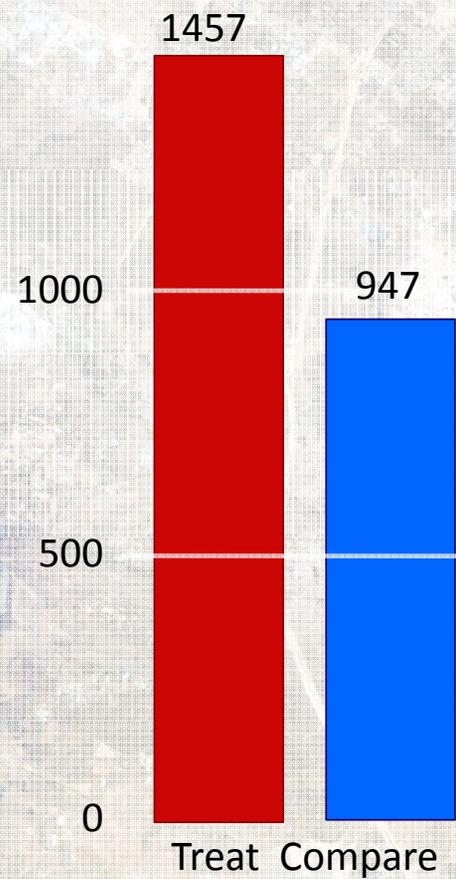
- Much of the challenge in empirical work is eliminating this selection bias
- Choosing an empirical method (“identification strategy” or “research design”) that makes selection bias less likely
- Presenting checks on data that lend credibility to the assumption of no selection bias
- Fundamental Problem of Causal Identification means that we cannot directly test the assumption of no selection bias

# Randomization as one way to solve selection bias

- If treatment and control groups are chosen randomly,  $E[Y_i^C|T] - E[Y_i^C|C] = 0$
- Eliminates systematic differences between treatment and control because no self-selection or targeting
- Can still have differences because of sampling error, of course

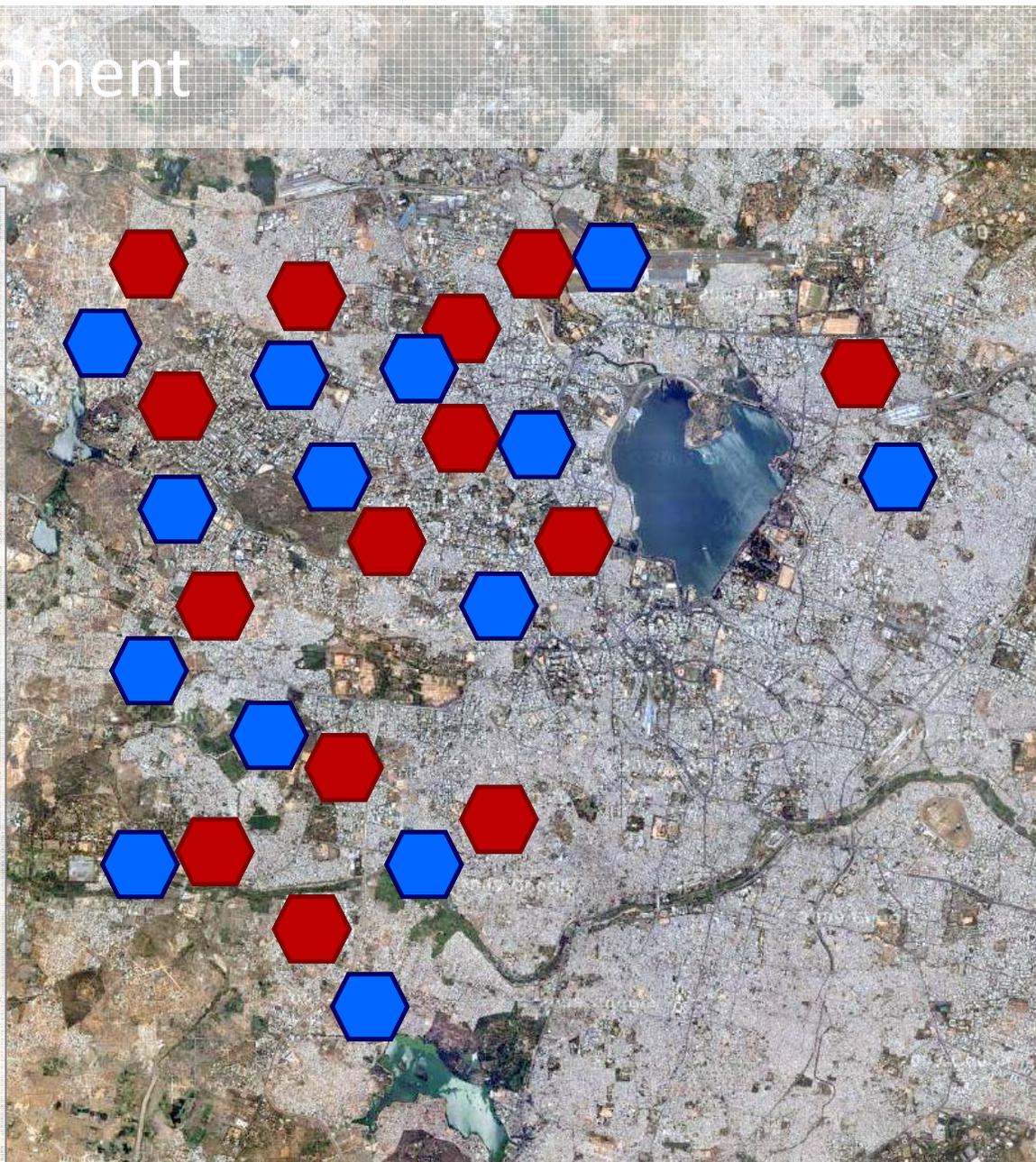
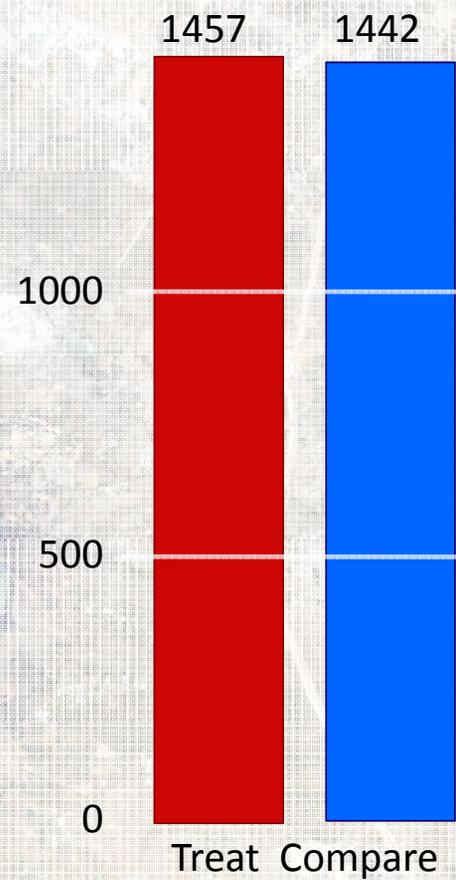
# Non random assignment

Income per person, per month



# Random assignment

Income per person, per month



## Another example

- Effect of taking dietary iron supplements on earnings
- $T$  is being offered iron supplements
- Dietary iron  $\rightarrow$  Higher hemoglobin (Hb) level in blood  $\rightarrow$  Less fatigued  $\rightarrow$  More productive worker
- This example based on Thomas et al. (2006), which used an RCT design
- Use this example to illustrate instrumental variables

# Program to distribute iron supplements

- Free iron supplement program could be interesting to us as researchers for 2 reasons
  1. We want to measure the program's effectiveness (policy question)

$$Y_i = \alpha + \beta \cdot T_i + \epsilon_i$$

2. Program enables us to measure the effect of hemoglobin on economic outcomes (more general question)

$$Y_i = \alpha + \beta \cdot Hb_i + \epsilon_i$$

## Effect of Hb on earnings

- Suppose we want to the effects of Hb on earnings

$$Y_i = \alpha + \beta \cdot Hb_i + \epsilon_i$$

- Want to regress  $Y$  on  $Hb$ , but facing omitted variable bias, i.e.  $Hb$  is an *endogenous regressor*
  - Is  $\beta$  likely to be biased upwards or downwards if we use straightforward OLS with observational data
- Use treatment  $T$  as an instrumental variable that affects  $Hb$ , and only affects  $Y$  through its effect on  $Hb$

# First and second stage

- Second stage

$$Y_i = \alpha + \beta \cdot \widehat{Hb}_i + \epsilon_i$$

- First stage

$$Hb_i = \alpha' + \beta' \cdot T_i + \epsilon'_i$$

## Requirements for IV

1. Instrument is correlated with the endogenous regressor (*strong first stage*) — testable assumption

$$E[Hb_i \cdot T_i] \neq 0$$

2. Instrument is exogenous to the outcome (*random assignment*)

$$E[\epsilon_i \cdot T_i] = 0$$

3. Instrument only affects the outcome via its effect on the endogenous regressor (*exclusion restriction*)

4. Either  $\forall$  individuals the instrument increases the endogenous regressor or  $\forall$  individuals decreases the endogenous regressor (“no defiers,” or *monotonicity condition*)

## IV estimate

- Intent to treat for Hb levels is measured by:

$$E[Hb_i|T_i = 1] - E[Hb_i|T_i = 0] \quad (1)$$

*First stage equation*

- Intent to treat estimate for labor market outcomes

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] \quad (2)$$

*Reduced form equation*

## IV estimate (cont'd)

- Using our expression for  $Y_i$ , we have:

$$E[Y_i|T_i = 1] = \alpha + \beta E[H_i|T_i = 1] + E[\epsilon_i|T_i = 1]$$

and:

$$E[Y_i|T_i = 0] = \alpha + \beta E[H_i|T_i = 0] + E[\epsilon_i|T_i = 0]$$

- Therefore,

$$\begin{aligned} & E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \\ & \beta(E[H_i|T_i = 1] - E[H_i|T_i = 0]) + E[\epsilon_i|T_i = 1] - E[\epsilon_i|T_i = 0] \end{aligned}$$

## IV estimate (cont'd)

- What can we assume about

$$E[\epsilon_i | T_i = 1] - E[\epsilon_i | T_i = 0]?$$

- Strict exogeneity assumption above implies the less strong assumption needed,  $E[\epsilon_i \cdot T_i] = 0$

## IV estimate (cont'd)

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \beta(E[H_i|T_i = 1] - E[H_i|T_i = 0])$$

- Wald estimate of  $\hat{\beta}$

$$\hat{\beta} = \frac{E[Y_i|T_i = 1] - E[Y_i|T_i = 0]}{E[H_i|T_i = 1] - E[H_i|T_i = 0]}$$

- Wald estimate: IV with binary instrument and no covariates
- Estimate the effect of health on labor market outcomes by dividing the effect of the program on labor market outcomes (reduced form, equation (2)) by the effect of the program on health (first stage, equation (1))

# Violations of exclusion restriction

- Distinction between exclusion restriction and “random assignment”
- Random assignment: instrument is exogenous in the reduced form and first stage equations
- What if the exclusion restriction were violated but  $T_i$  is still random?
- Can still interpret the first stage or reduced form results as causal (treatment on income or treatment on Hb)
- But cannot use the treatment to learn about the relationship between the endogenous regressor and the outcome of interest (Hb on income)

## Violations of monotonicity condition

- Could treatment reduce hemoglobin level for some individuals?
- Another example: Information on returns to schooling as IV for school attendance
  - Randomized information intervention telling some households what the labor-market returns to secondary school are
  - Use treatment as IV for attending secondary school
  - On average, treatment increases secondary school attendance; households had underestimated returns (first stage)
  - Random assignment of treatment and exclusion restriction probably OK
  - But what if someone had overestimated returns to schooling?

## What is the “identifying variation” with IV?

- For pedagogical reasons, will look at special case where  $H$  is also a dummy variable for high Hb
- Notation:  $Y_i(T_i, H_i(T_i))$  and  $Y_i(H_i)$  where  $T$  is treatment
- We are seeking  $E[Y_i(1) - Y_i(0)]$
- Start by decomposing effect of  $T$  on  $Y$  for person  $i$

$$\begin{aligned} Y_i(1, H_i(1)) - Y_i(0, H_i(0)) &= Y_i(H_i(1)) - Y_i(H_i(0)) \\ &= [Y_i(1) \cdot H_i(1) + Y_i(0) \cdot (1 - H_i(1))] \\ &\quad - [Y_i(1) \cdot H_i(0) + Y_i(0) \cdot (1 - H_i(0))] \\ &= (Y_i(1) - Y_i(0)) (H_i(1) - H_i(0)) \end{aligned}$$

- For the sample average of the estimand, when we use the treatment as our identifying variation in  $H$ , the only individuals who contribute to our estimate are those for whom  $H_i(1) \neq H_i(0)$

	$H_i(0) = 0$	$H_i(0) = 1$
$H_i(1) = 0$	$Y_i(1, 0) - Y_i(0, 0) = 0$ <p>Never taker</p>	$Y_i(1, 0) - Y_i(0, 1) = Y_i(0) - Y_i(1)$ <p>Defier</p>
$H_i(1) = 1$	$Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$ <p>Complier</p>	$Y_i(1, 1) - Y_i(0, 1) = 0$ <p>Always taker</p>

Notation:  $Y_i(T_i, H_i)$  and  $Y_i(H_i)$  and  $H_i(T_i)$

$Y_i(1) - Y_i(0)$  is effect of high Hb on income, the object of interest

## IV estimand for the sample

Taking the sample mean of our individual-level estimand, we get

$$\begin{aligned} E[Y_i(1 - H_i(1)) - Y_i(0, H_i(0))] &= E[(Y_i(1) - Y_i(0)) (H_i(1) - H_i(0))] \\ &= E[Y_i(1) - Y_i(0) | H_i(1) - H_i(0) = 1] \cdot \text{Prob}[H_i(1) - H_i(0) = 1] + \\ &E[Y_i(0) - Y_i(1) | H_i(1) - H_i(0) = -1] \cdot \text{Prob}[H_i(1) - H_i(0) = -1] \end{aligned}$$

First term is compliers

Second term is defiers

## Wald estimate with defiers

$$\hat{\beta}_W = \frac{E[Y_i|T_i = 1] - E[Y_i|T_i = 0]}{E[H_i|T_i = 1] - E[H_i|T_i = 0]} = \frac{\hat{\beta}_c \cdot \text{Prob}[i \text{ is a complier}] - \hat{\beta}_d \cdot \text{Prob}[i \text{ is a defier}]}{\text{Prob}[i \text{ is a complier}] - \text{Prob}[i \text{ is a defier}]}$$

- Let  $\lambda \equiv \frac{\text{Prob}[i \text{ is a defier}]}{\text{Prob}[i \text{ is a complier}] - \text{Prob}[i \text{ is a defier}]}$   $\Rightarrow \hat{\beta}_W = (1 + \lambda)\hat{\beta}_c - \lambda\hat{\beta}_d$
- If  $\beta_c = \beta_d$ , defiers are not problematic
- But with heterogeneity in treatment effects, if monotonicity condition does not hold, IV estimand is a strange weighted average (weights not restricted to  $[0,1]$ )

# LATE

- IV estimates give us a *local average treatment effect* (LATE)
- $\hat{\beta}$  is estimated for the “compliers” or those induced to have higher  $Hb$  by virtue of being assigned to the treatment group
- We cannot identify which individuals in our sample comprise the compliers

# Where do we find IVs?

- Random experiments
- “Natural experiment”: use real-world variation that we think is uncorrelated with important unobserved factors as a way to solve selection bias problem
- Often generated by a policy
  - Draft lottery number affected likelihood of serving in Vietnam
  - Construction of new schools led to an increase in education in Indonesia

## Another IV example

- Want to regress *Income* on *Military*, but those who enter military differ in other ways
- Find a factor *LotteryNo* (instrumental variable) that affects *Military*, and only affects *Income* through its effect on *Military*
- Instrument is correlated with the endogenous regressor (first stage)

$$Military_i = \alpha + \beta LotteryNo_i + \nu_i \quad (1)$$

- Only channel through which instrument affects outcome is through the endogenous regressor?
- If yes, take predicted value from (1) and run second stage

$$Income_i = \delta + \gamma \widehat{Military}_i + \varepsilon_i \quad (2)$$