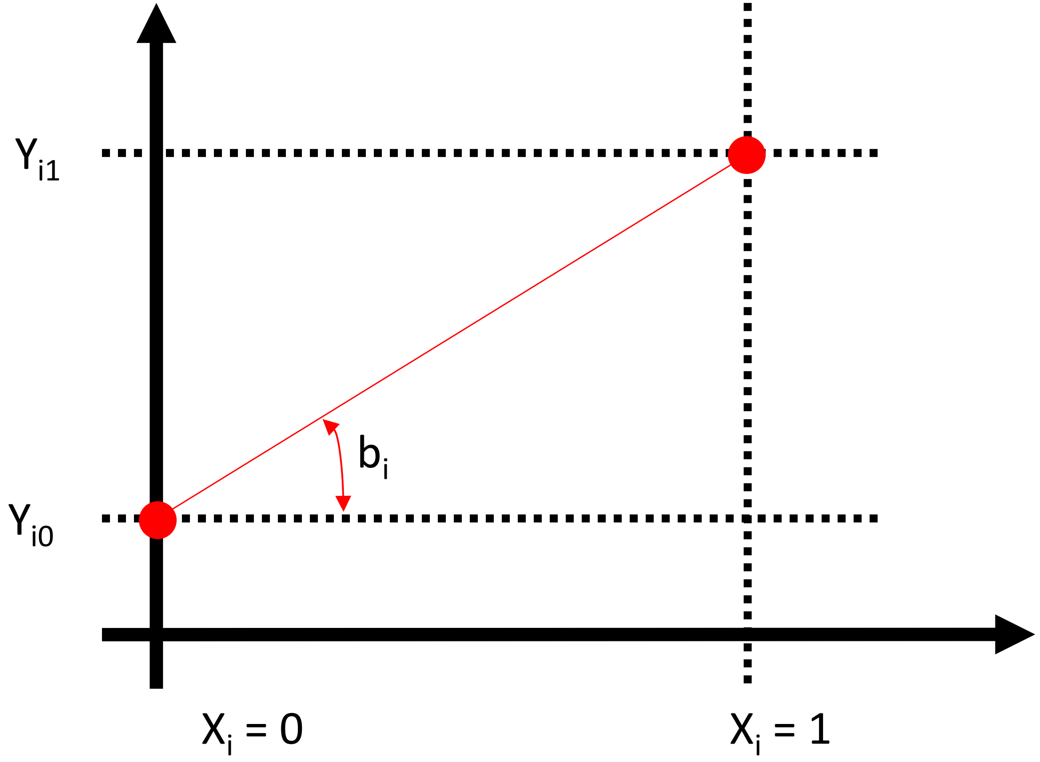
Review – Econometrics B

Ekin Ilseven – P4

What do we want?

Causal effect of **X** on **Y**: $\mathbf{X} \xrightarrow{b} \mathbf{Y}$

- b is treatment effect



For 1 person and X is binary

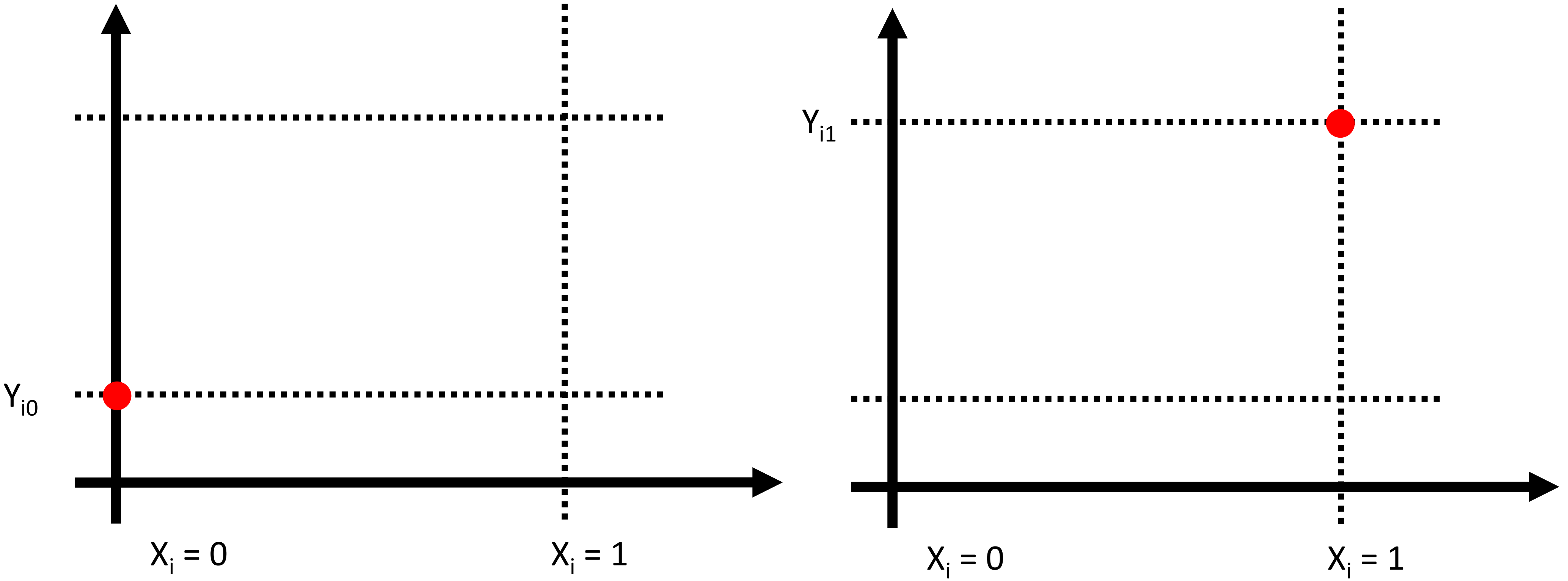
Can we have what we want?

Not easily…

1. Missing values
2. Selection bias
3. Heterogeneity in treatment effect

Missing values

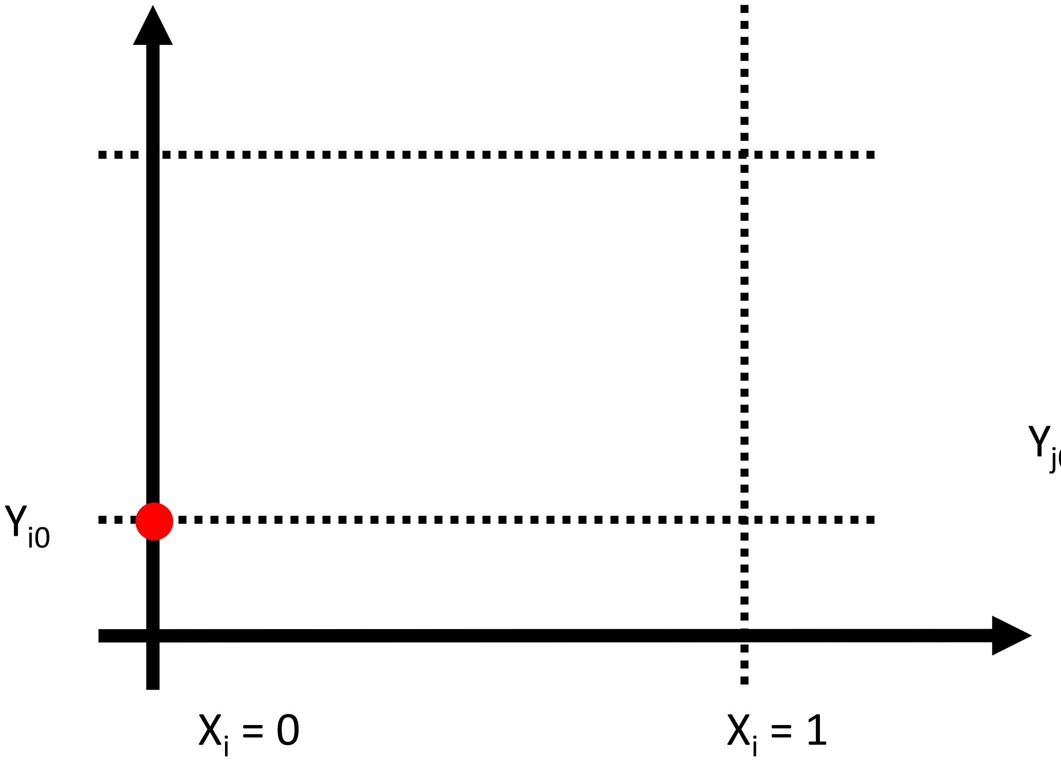
- We either have one or the other. Unit of analysis is NOT Schrödinger's cat.

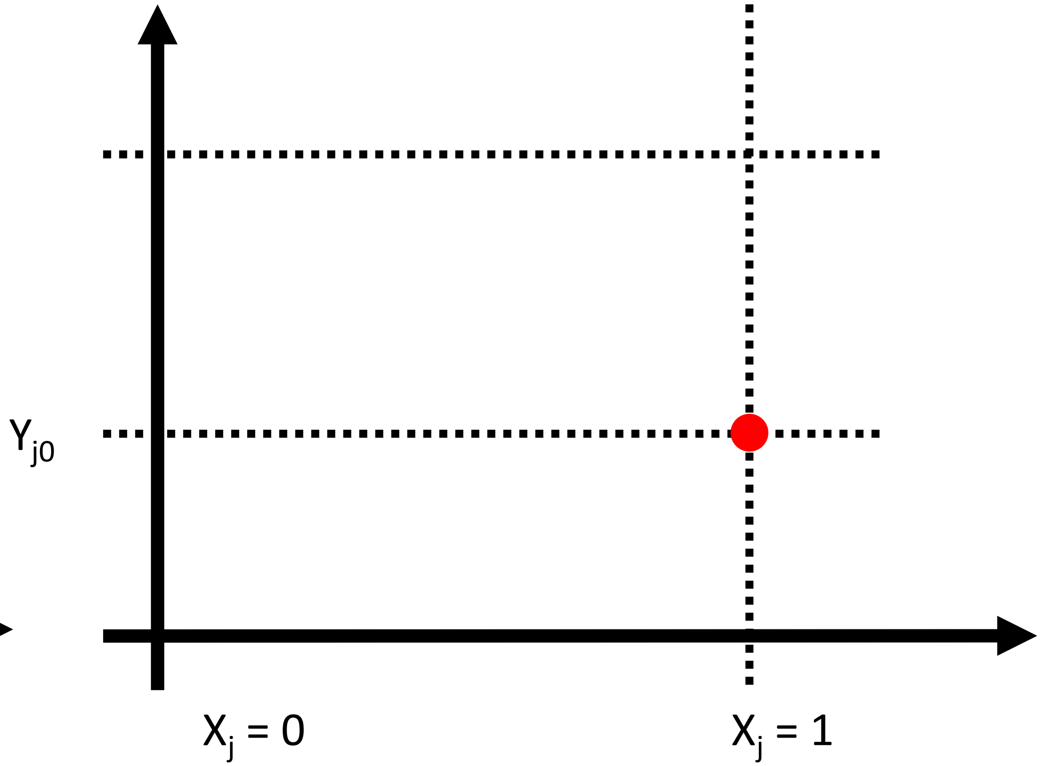


For 1 person and X is binary

Selection bias

- The Y_{i0} might be confounded with treatment X_i.

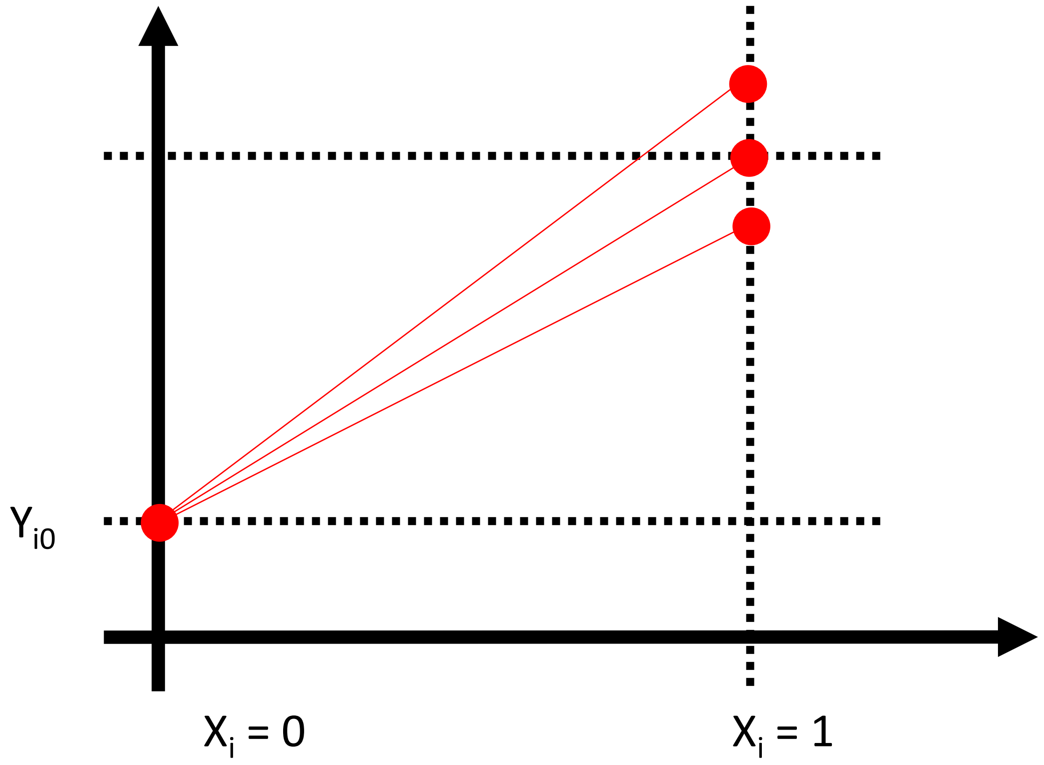




For 1 person and X is binary

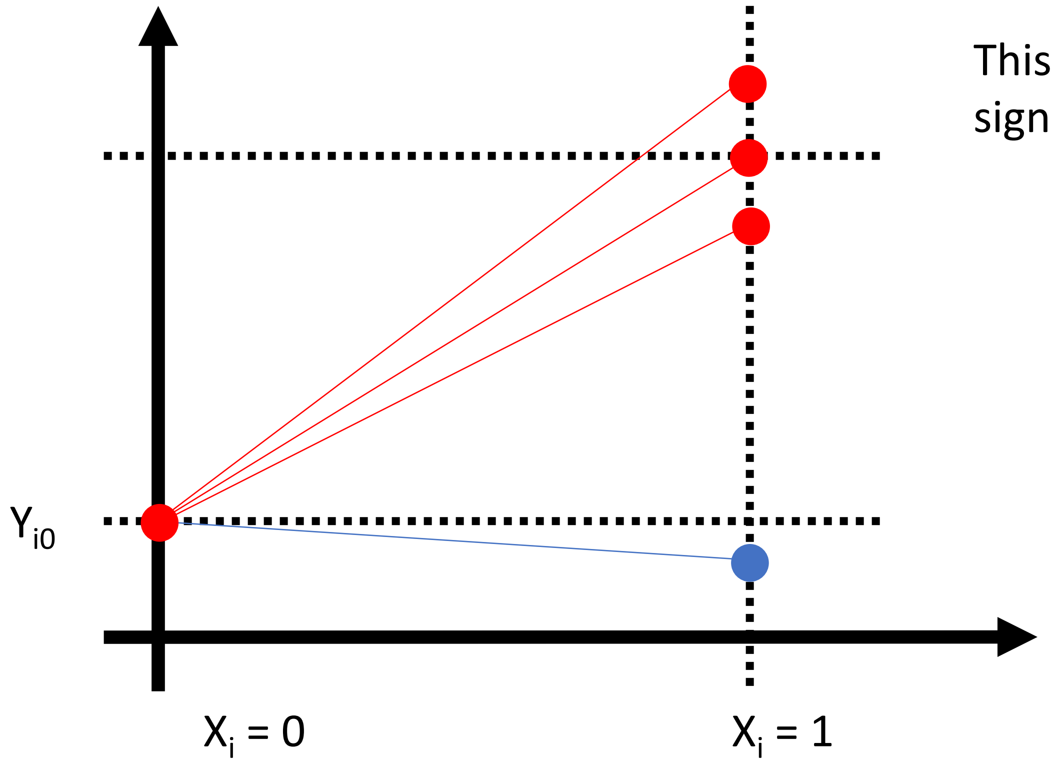
Heterogeneity in Treatment

- Even if we have counterfactuals and no selection, different slopes…



Heterogeneity in Treatment

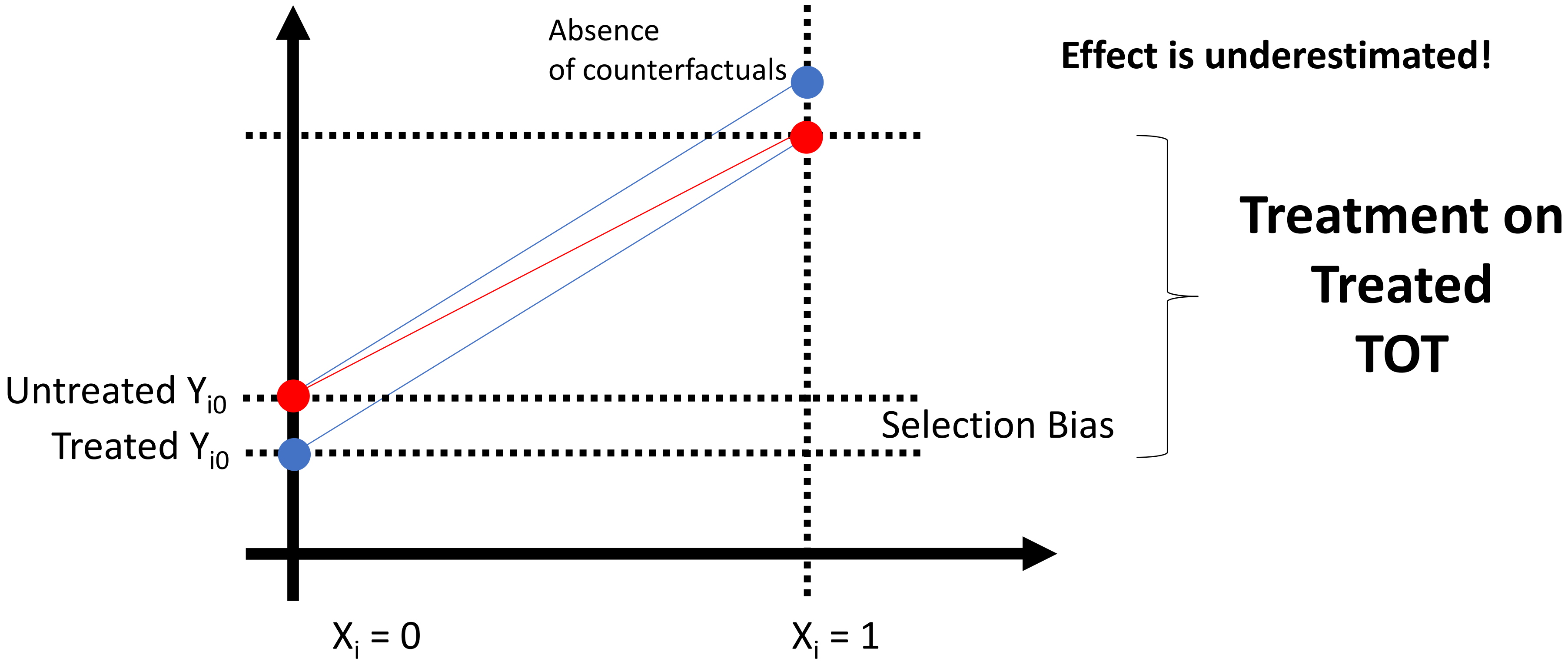
- Even if we have counterfactuals and no selection, different slopes…



This is a nice case, but what if sign flips for some people?

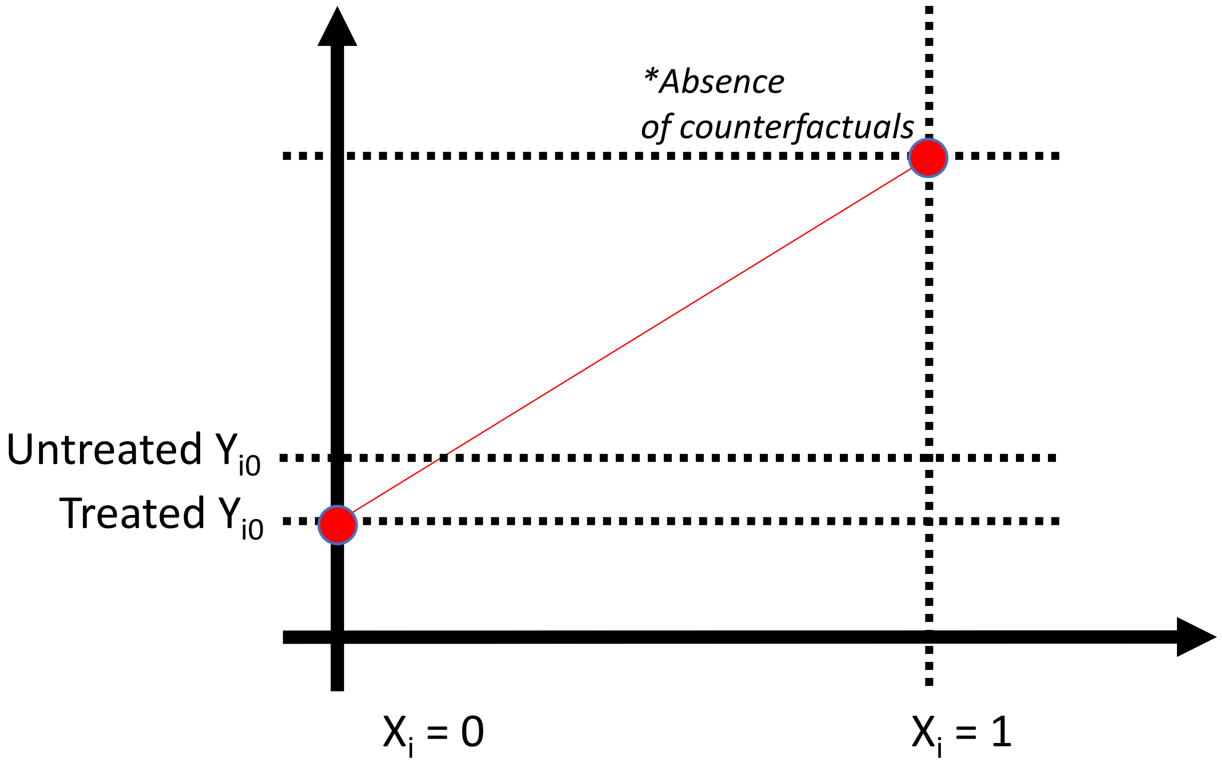
Heterogeneity in Treatment

- Assume away heterogeneity, keep counterfactual and selection:



Randomized Control Trials

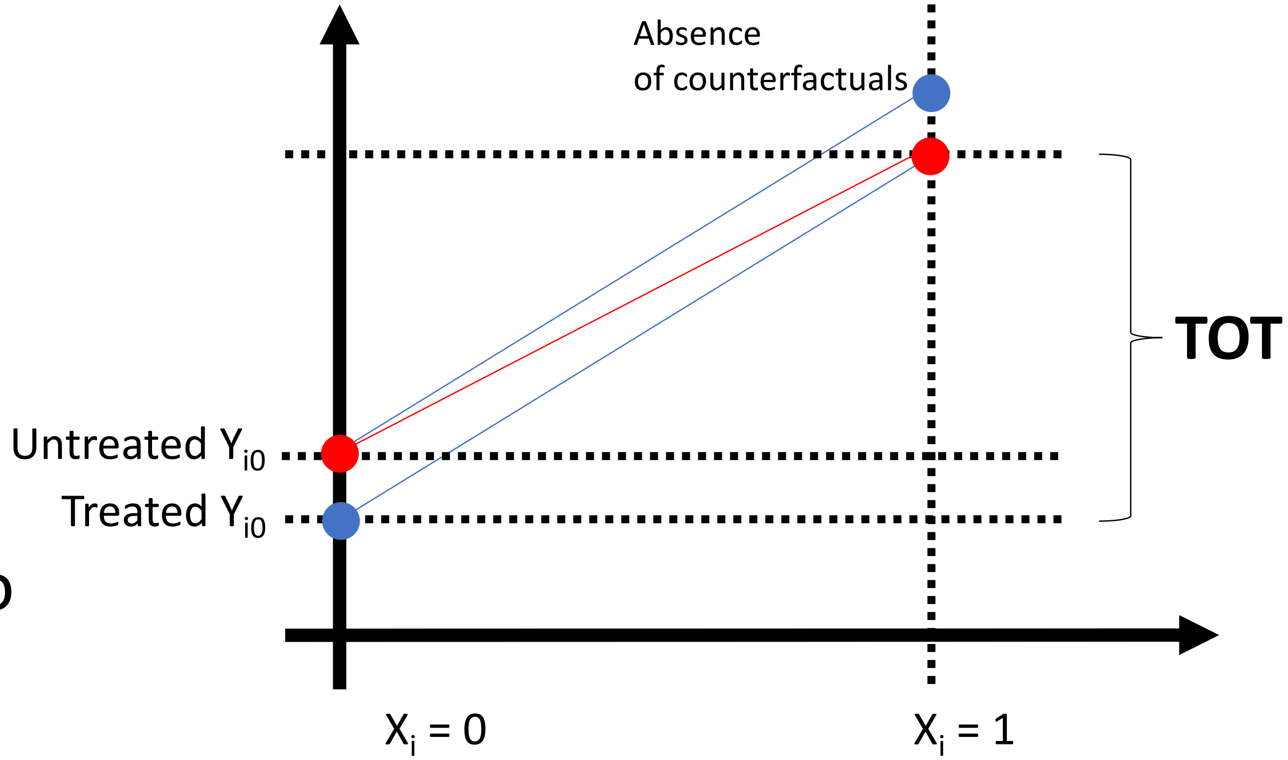
- Tackles selection bias, hence we can use treated and untreated as counterfactuals.



Matching – Session 1

Matching

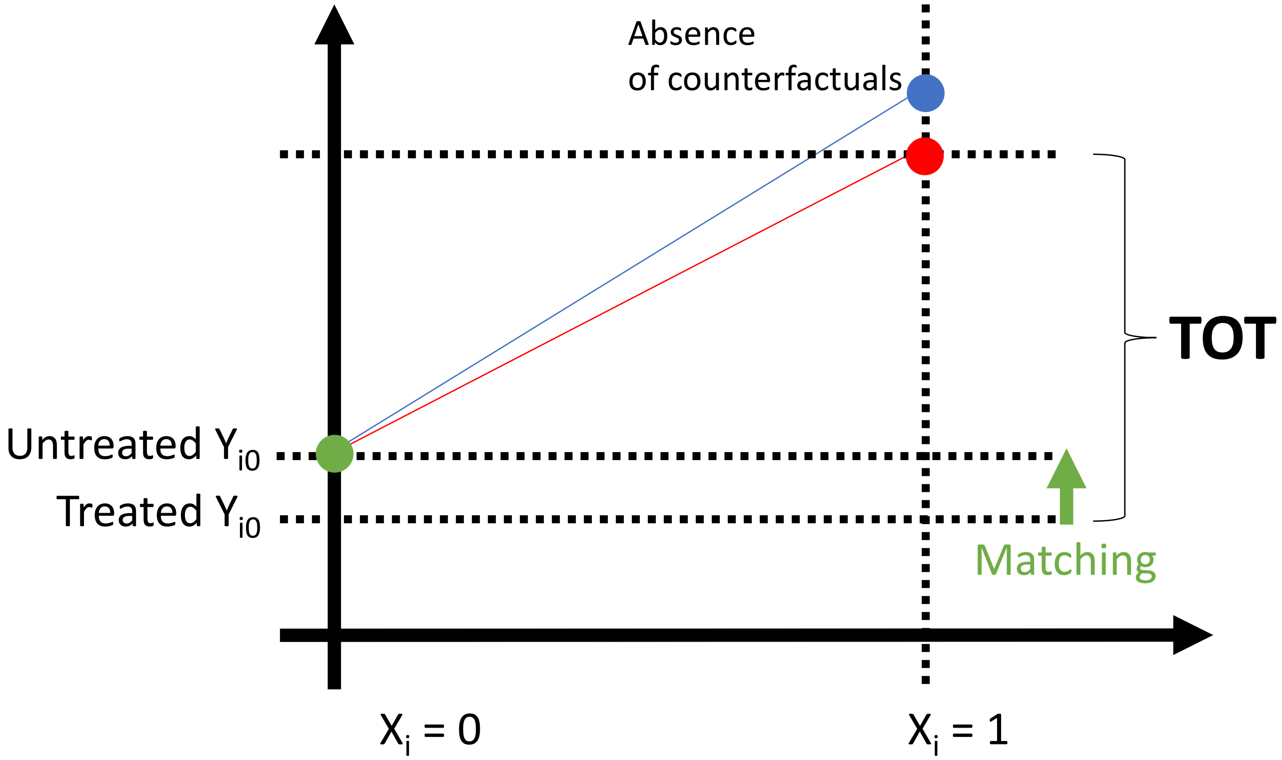
- Also tackles selection bias, but we have to build the counterfactuals.
- Assume selection bias is **detectable** by observables.
- Meaning, given distribution of observables **O**, I can find out how many people will select into treatment and how many not, and adjust the population.



How exactly?

Causal effect of **X** on **Y** now: **X** $\longrightarrow$ **O** $\longrightarrow$ **Y** (Conditional Independence Assumption)

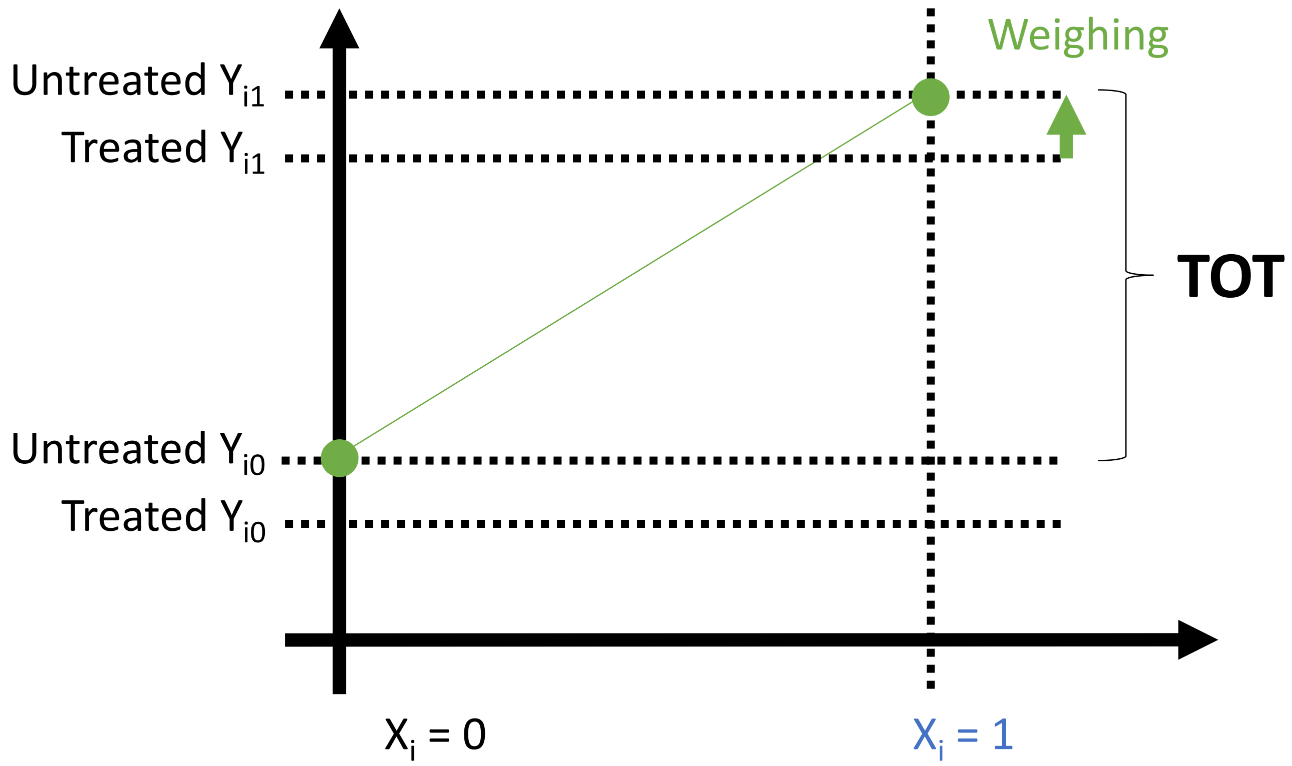
- Because we can measure **O**, now I can replace treated **Y**$_{i0}$ (not measured!) with untreated **Y**$_{i0}$ (measured).



How exactly?

Causal effect of **X** on **Y** now: **X** ⟶ **O** ⟶ **Y** (Conditional Independence Assumption)

- Because we can measure O, now I can replace treated Y_{i0} (not measured!) with untreated Y_{i0} (measured).
- For **TOT**, we are only concerned with the distribution of O for treated people.
- Take weighted average of the effect according to how O is distributed with treatment.
- Constrain O, only to values where there are both treated and untreated, otherwise no effect to identify.



Weighing: Match or saturated OLS?

- With discrete covariates: X ⟶ O ⟶ Y, D ⟶ X ⟶ Y

$$E[Y_{1i} - Y_{0i} | D_i = 1] = \sum_x \delta_x \times P(X_i = x | D_i = 1)$$

- Consider the regression framework (saturated in X_i): O ⟶ X ⟶ Y
X ⟶ D ⟶ Y

$$Y_i = \sum_x \alpha_x \times \mathbb{1}(X_i = x) + \beta \times D_i + u_i$$

$$\beta = \frac{\sum_x \delta_X \times \sigma_D^2(X_i) \times P(X_i = x)}{\sum_x \sigma_D^2(X_i) \times P(X_i = x)}$$

If too many covariates, reduce to propensity

- Propensity score is not about mechanism of X -> Y, but about the mechanism of assignment.
  - Remark: If assignment process is confounding with the outcome, you may have efficiency problems.
- Using logit/probit, propensity score determines the conditions under which individuals will be treated or not. Similar to “clustering” of people into groups that are treated or untreated.
- Pay attention to the distribution of treatment according to the observable covariates used for matching
  - OLS and matching weighs factors differently for each case.

Review of Material and Questions - I

Making Regression Make Sense 73

TABLE 3.3.1
Uncontrolled, matching, and regression estimates of the effects of voluntary military service on earnings

Race	Average Earnings in 1988–1991 (1)	Differences in Means by Veteran Status (2)	Matching Estimates (3)	Regression Estimates (4)	Regression Minus Matching (5)
Whites	14,537	1,233.4 (60.3)	−197.2 (70.5)	−88.8 (62.5)	108.4 (28.5)
Non-whites	11,664	2,449.1 (47.4)	839.7 (62.7)	1,074.4 (50.7)	234.7 (32.5)

Notes: Adapted from Angrist (1998, tables II and V). Standard errors are reported in parentheses. The table shows estimates of the effect of voluntary military service on the 1988–91 Social Security–taxable earnings of men who applied to enter the armed forces between 1979 and 1982. The matching and regression estimates control for applicants' year of birth, education at the time of application, and AFQT score. There are 128,968 whites and 175,262 nonwhites in the sample.

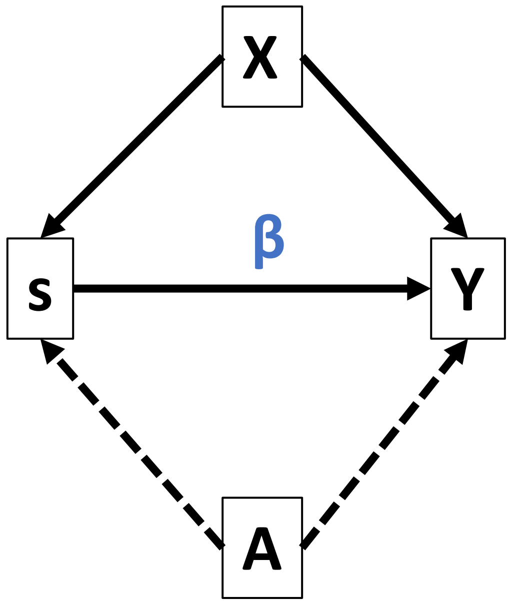
© Princeton University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/help/faq-fair-use/.

- What causes the difference between "differences in means" and "matching estimates"?
- Same for "matching estimates" and "regression estimates"?
- What other experimental design could we use to estimate the effects?

Instrumental Variable – Session 2

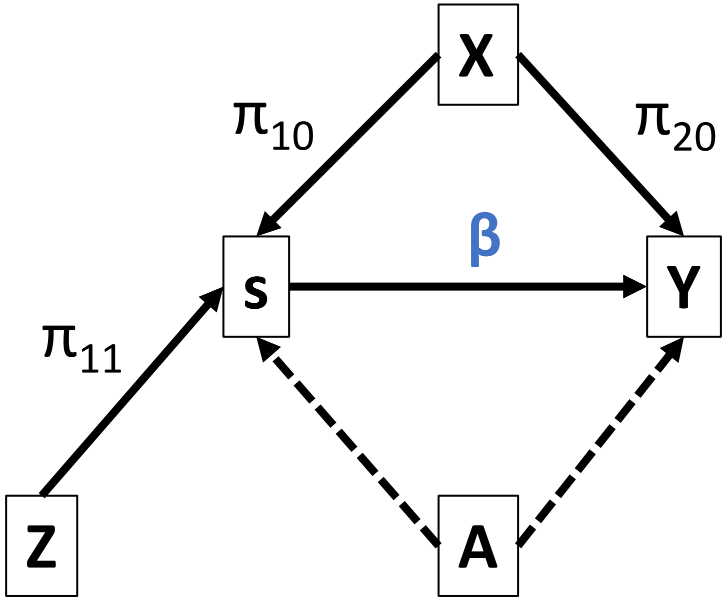
When do we need Instrumental Variable? (IV)

- We are interested in the causal effect β of **s** on **Y**.
- We have some covariates **X** which are observable.
- We have an omitted variable **A** which is unobservable.



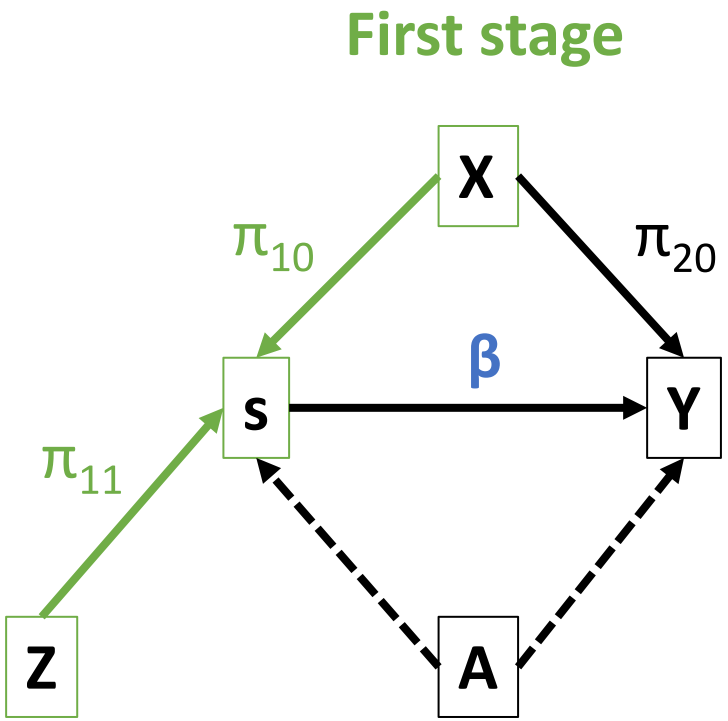
When do we need Instrumental Variable? (IV)

- We are interested in the causal effect β of **s** on **Y**.
- We have some covariates **X** which are observable.
- We have an omitted variable **A** which is unobservable.
- Let's introduce **Z** as following:
  - No correlation with **A** or **Y**.
  - Correlation with **s**.



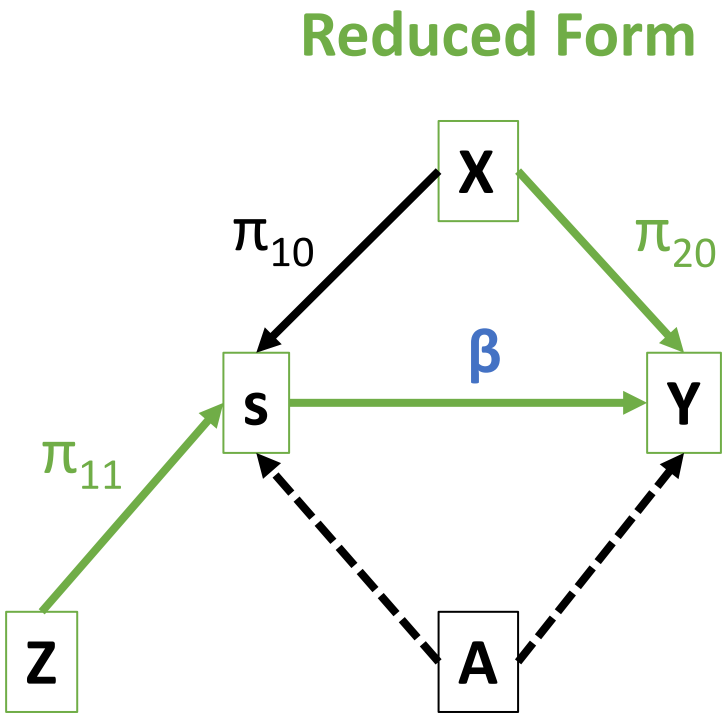
How do we get the effect?

- We are interested in the causal effect β of **s** on **Y**.
- We have some covariates **X** which are observable.
- We have an omitted variable **A** which is unobservable.
- Let's introduce **Z** as following:
    - No correlation with **A** or **Y**.
    - Correlation with **s**.



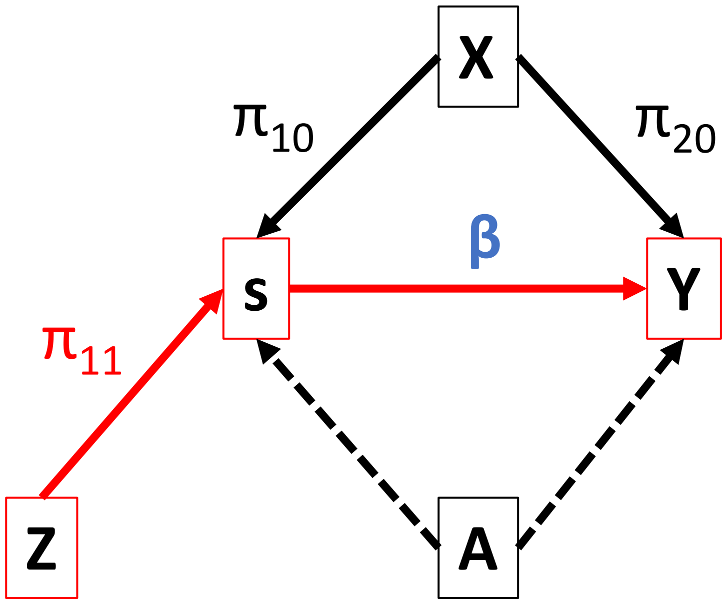
How do we get the effect?

- We are interested in the causal effect β of **s** on **Y**.
- We have some covariates **X** which are observable.
- We have an omitted variable **A** which is unobservable.
- Let's introduce **Z** as following:
  - No correlation with **A** or **Y**.
  - Correlation with **s**.



How do we get the effect?

- We are interested in the causal effect β of **s** on **Y**.
- We have some covariates **X** which are observable.
- We have an omitted variable **A** which is unobservable.
- Let's introduce **Z** as following:
  - No correlation with **A** or **Y**.
  - Correlation with **s**.

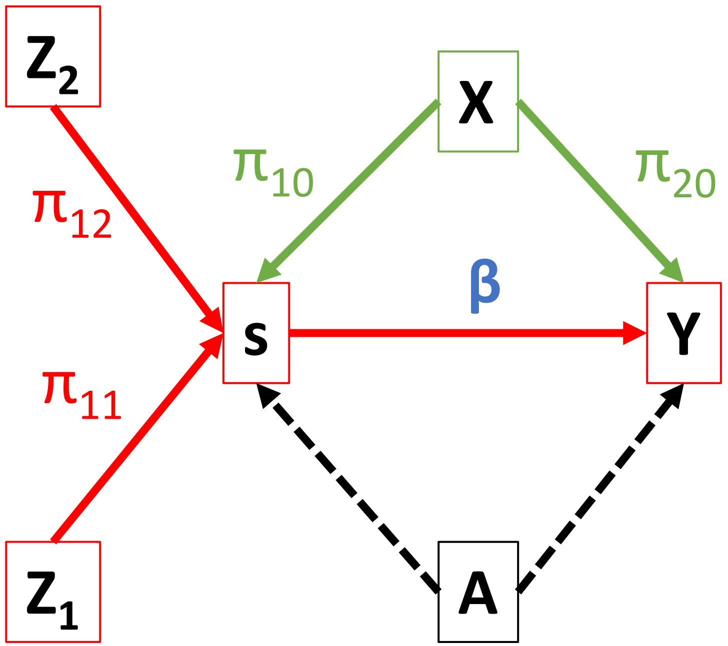


$\pi_{21} = \pi_{11} * \beta$

Caveat: The standard error of 2SLS needs to be adjusted according to the variance of **s** explained by **Z**.
Use software packages for IV estimation taking care of this instead of manual regressions.

If more than one IV? Over-identification

- If there are more than two IVs, estimate the coefficients separately.
- The resulting coefficient is a weighted average of the coefficients.
- If $\boldsymbol{\pi_{11} > \pi_{12}}$, then final $\boldsymbol{\beta}$ is closer to $\boldsymbol{\beta_1}$ and vice versa.



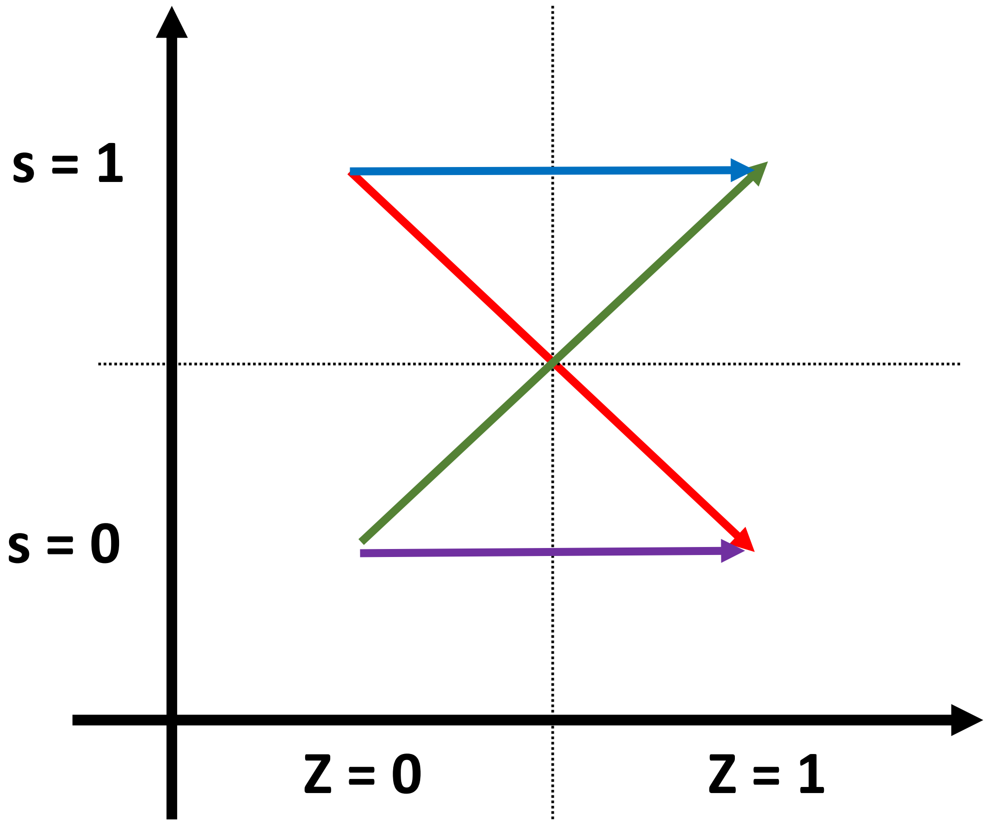
$\pi_{21} = \pi_{11} * \beta_1$

$\pi_{22} = \pi_{12} * \beta_2$

Instrument – Intention to Treat (ITT)

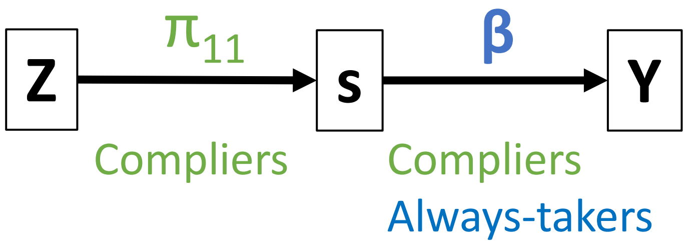
- Even if we may have intention to treat somebody, whether the person is treated or not depends on their compliance.
- We can use IV design to understand heterogeneity in intention to treat reaching success.
  - We have four types of outcomes in binary cases:

1. Compliers
2. Always-takers
3. Never-takers
4. Defiers (have to assume non-exist)

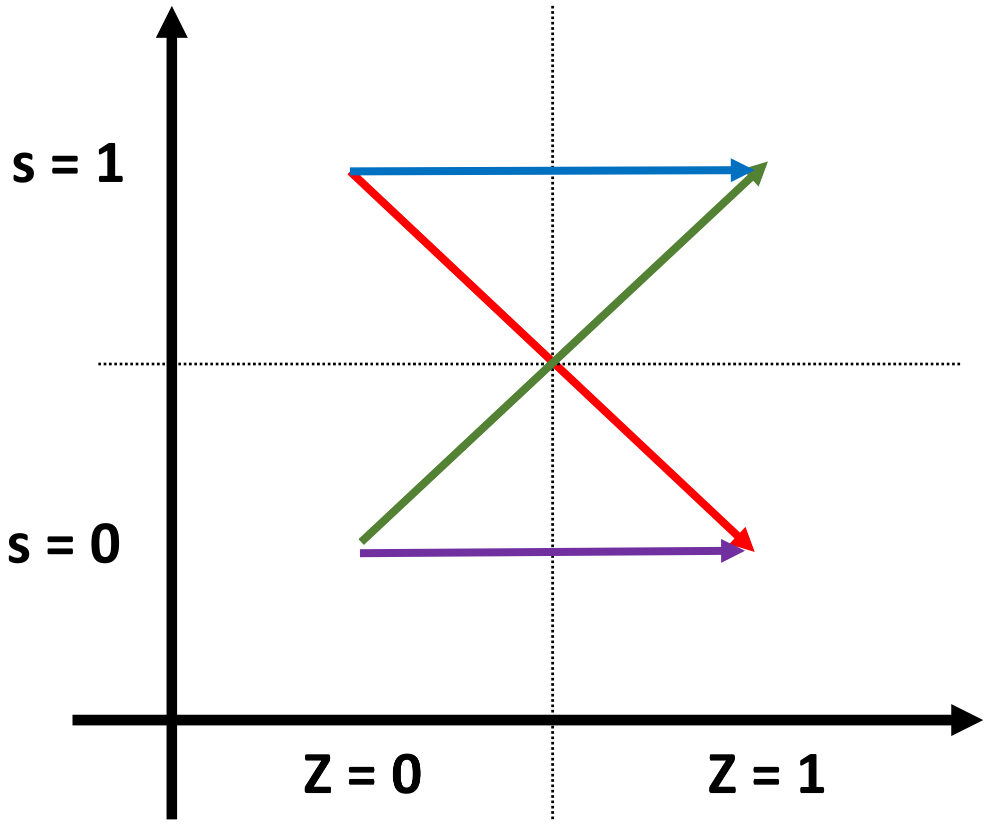


Instrument – Intention to Treat (ITT)

- Hence in an IV design, we can present this as:

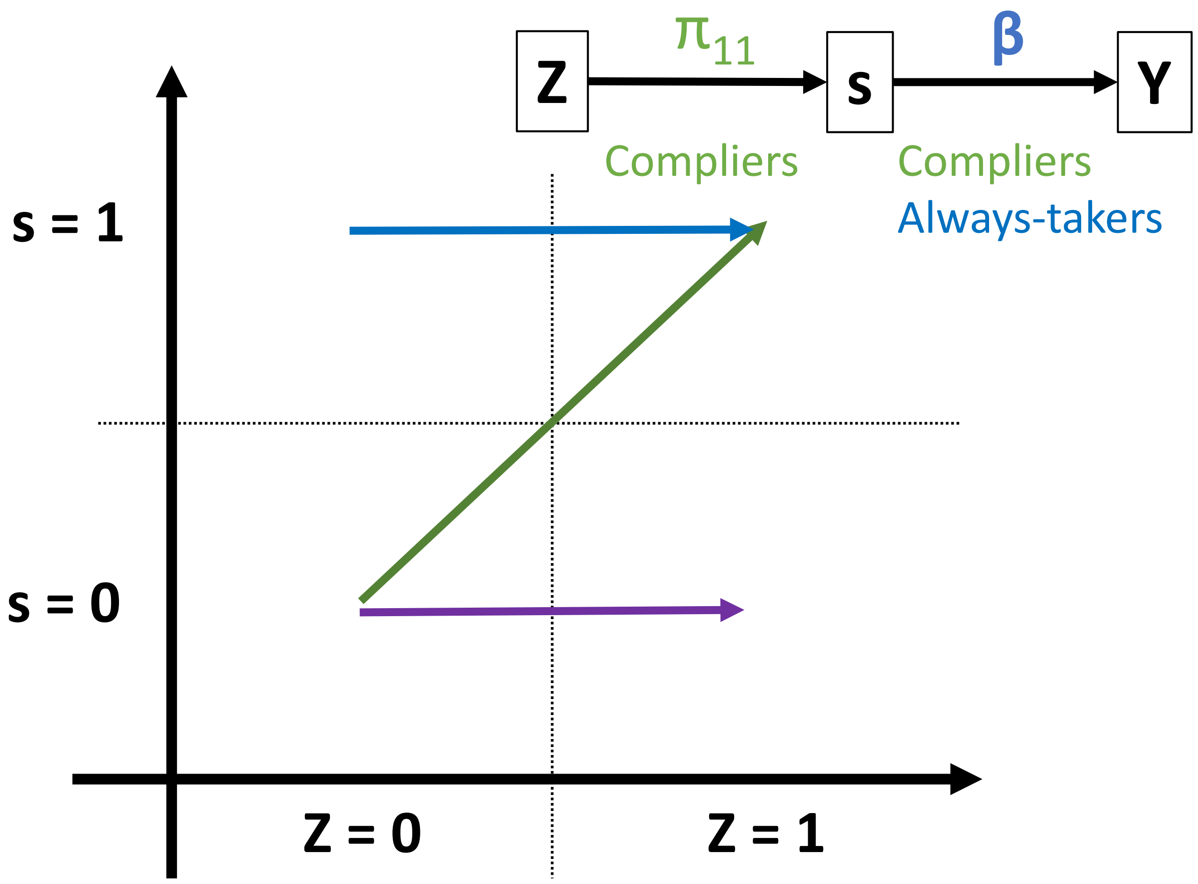


- However, effect of **β** does not have to be the same for compliers and always-takers.
- How to resolve?

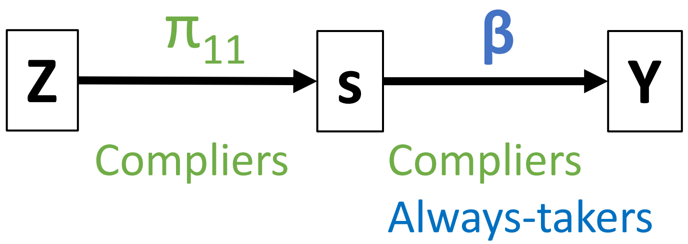


LATE Theorem approves IV design only if:

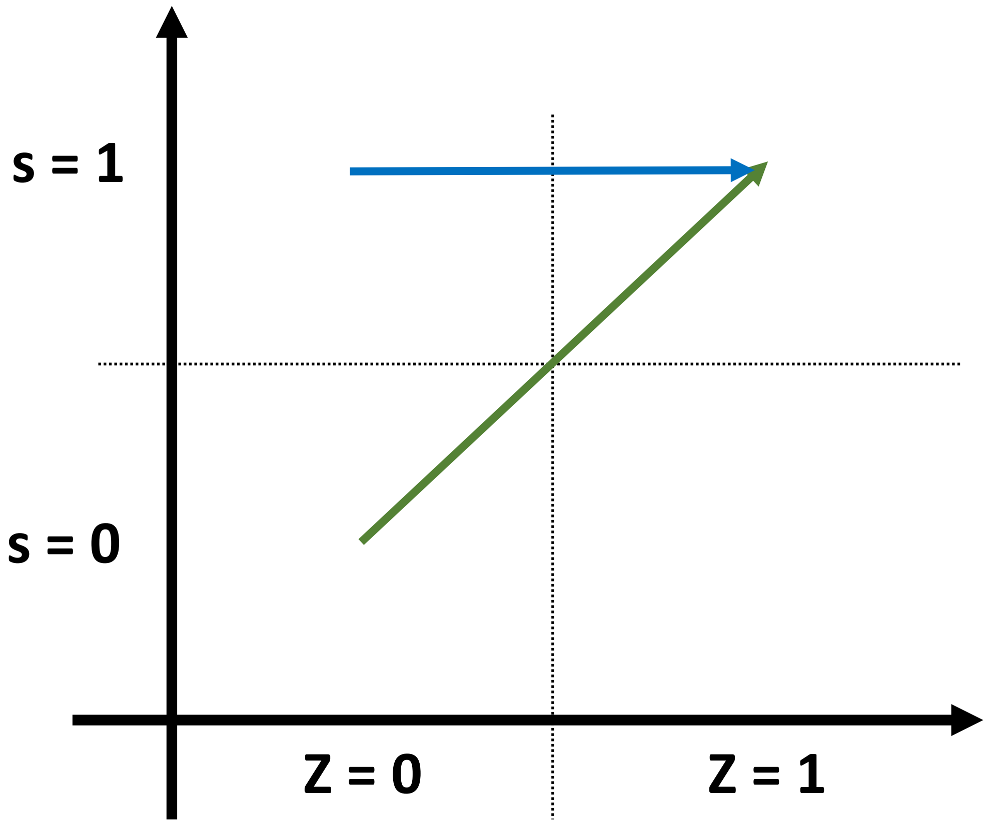
1. Independence: No any other arrows connecting **s** and **Y** with **Z**.
2. The value of treatment outcome is the same whether intended to treat or not.
3. Compliers exist.
4. No defiers.



Local Average Treatment Effect



- The effect estimated is a local average treatment effect due to heterogeneity.
- If no always-takers, then TOT.
- LATE = ITT/(Compliance Rate)
- Multiple IVs will have different profiles of compliers – increases the quality of inference.



Review of Material and Questions - II

Table 1. *Wald Estimates of the Effects of Family Size on Labor Supply*

Dependent Variable	Mean	OLS (1)	Twins Instrument: First Stage (2)	Twins Instrument: Wald Estimates (3)	Same-Sex Instrument: First Stage (4)	Same-Sex Instrument: Wald Estimates (5)	Both: 2SLS Estimates (6)
Weeks worked	20.83	−8.98	0.603	−3.28	0.060	−6.36	−3.97
		(0.072)	(0.008)	(0.634)	(0.002)	(1.18)	(0.558)
	Overid: $\chi^2(1)$	–	–	–	–	–	5.3
	(p-value)						(0.02)
Employment	0.565	−0.176		−0.076		−0.132	−0.088
		(0.002)		(0.014)		(0.026)	(0.012)
	Overid: $\chi^2(1)$	–	–	–	–	–	3.5
	(p-value)						(0.06)

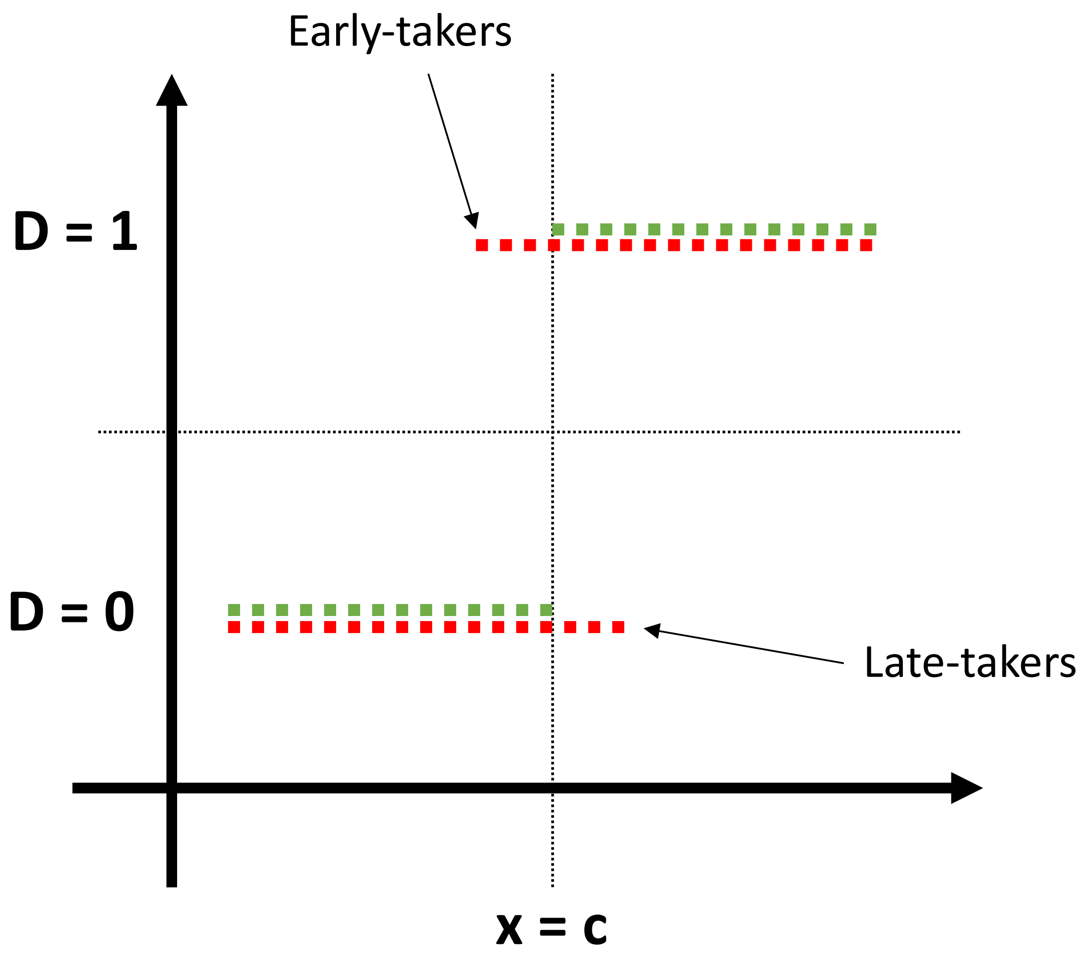
Note: The table reports OLS, Wald, and 2SLS estimates of the effects of a third birth on labor supply using twins and sex composition instruments. Data are from the Angrist and Evans (1998) extract from the 1980 U.S. census 5 percent sample, including women aged 21–35 with at least two children. OLS models include controls for mother's age, age at first birth, ages of the first two children, and dummies for race. The sample size is 394,840.

- What are the reduced form coefficients of family size on weeks worked?
- How can we interpret first stage results?
- Can you explain 2SLS estimates?
- What could be the problem with OLS?
- How about external validity?

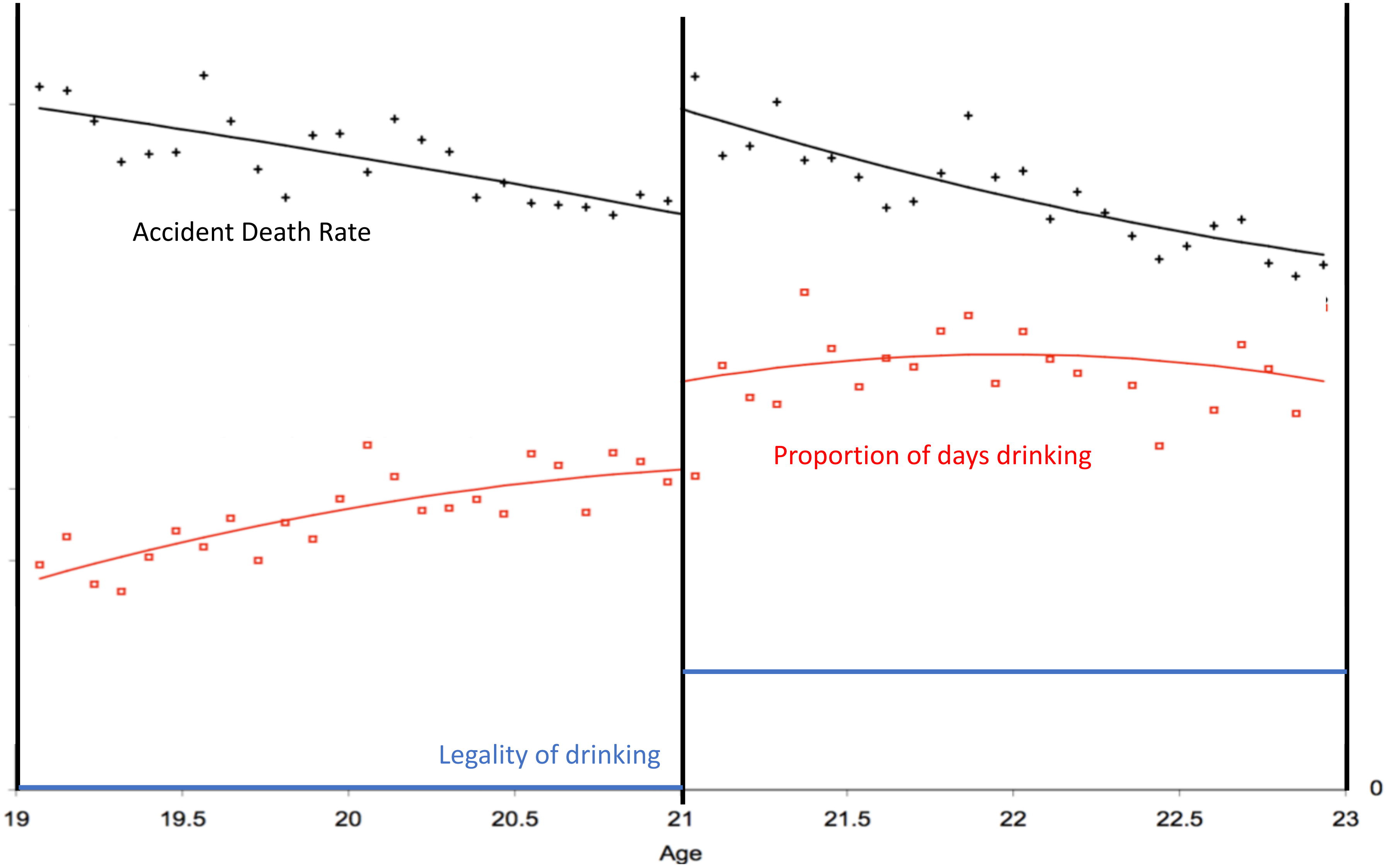
Regression Discontinuity
Session 3

Sharp and Fuzzy RD (No kink design)

- Sharp RD: The treatment status **D**, jumps at **x = c**.
- Fuzzy RD: The probability of the treatment status becoming positive, jumps at **x = c**.
- As previously, fuzzy RD is relevant for imperfect compliance with discontinuity.



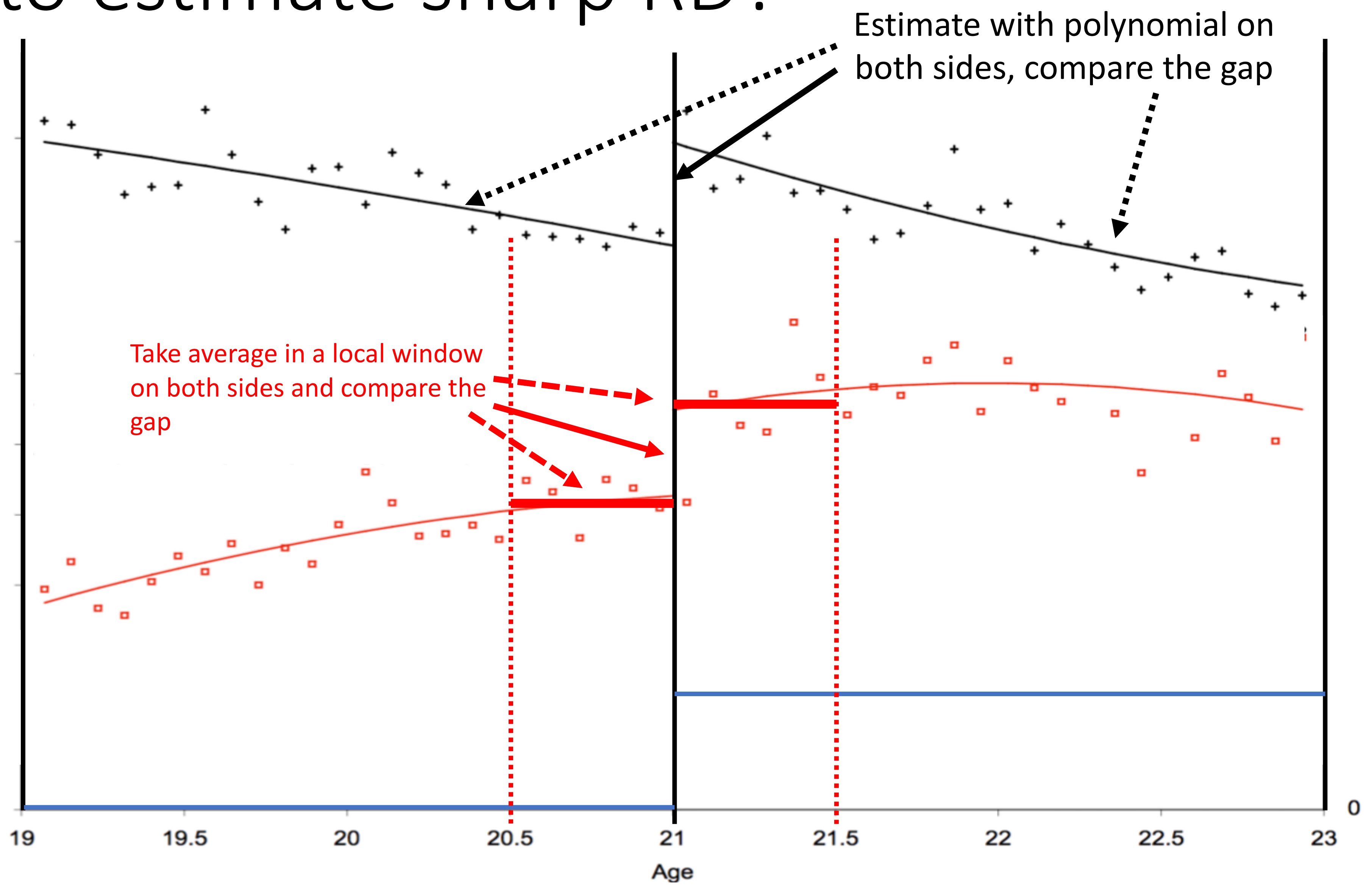
Compiled Example from Lecture



Three Conditions

- Running variable is single and continuous.
- The agents can not control running variable (often, time can't be controlled).
- Check for excess or missing mass on both sides.
- There are no other jumps except the jump in running variable of interest.
    - In some countries, driving AND drinking become legal with age of 18.

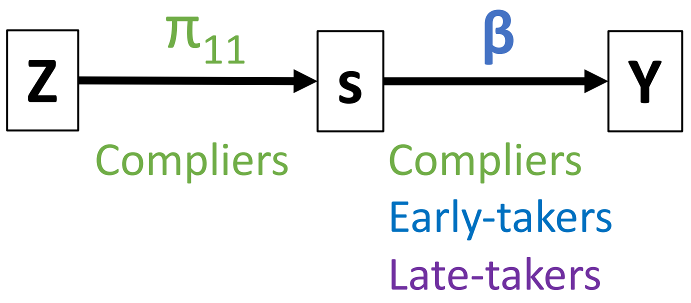
How to estimate sharp RD?



Limitations and solution?

How to do fuzzy RD?

- If there is imperfect compliance, then fuzzy RD.
- Use the threshold as IV and do sharp RD on top.
- **Z,** the dummy variable denoting threshold switch
  - Everybody complies.
- **s,** discontinuous independent variable.
- **Y,** discontinuous dependent variable.



Review of Material and Questions - III

- What are the pros and cons of RD estimation?
- What are the important checks to do before interpreting the results?
- How do we interpret the results of an RD?
- When is donut RD useful?