IV as research design

Labor economics

Instructor: Haoran LEI

Hunan University

Review: causalities and how to find them

- Regression-based analysis generally fails to identify causal effects.
  - We economists even invented the term *"endogeneity"* specifically for this problem.

Review: causalities and how to find them

- Regression-based analysis generally fails to identify causal effects.
- In statistical practice, the ideal way to estimate the effects of D_i on Y_i is to **make the assignment of D_i random**
  - That is, making D_i independent of potential outcomes.
  - D_i is *treatment*, say **hospitalization** or **schooling**
  - Y_i is *potential outcome*, say **health status** or **wage**

Review: random assignment

Under random assignment:

- $E[Y_{1i} \mid D_i = 1] = E[Y_{1i}]$
- $E[Y_{0i} \mid D_i = 0] = E[Y_{0i}]$

And the (unconditional) *average treatment effect* is:

$$E[Y_{1i} - Y_{0i}] = E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$$

Random assignment as benchmark

While **random assignment** is generally motivated as an experimental method, it is also viewed as the benchmark for other quasi-experimental tools in economics:

- IV, (fuzzy) Regression Discontinuity, Difference-in-Differences…

IV as research design

Model: $Y_i = \beta_0 + \beta_1 X_i + U_i$

- Random assignment ensures that X and U are independent.
- Besides random assignment, we can use Z as IV if

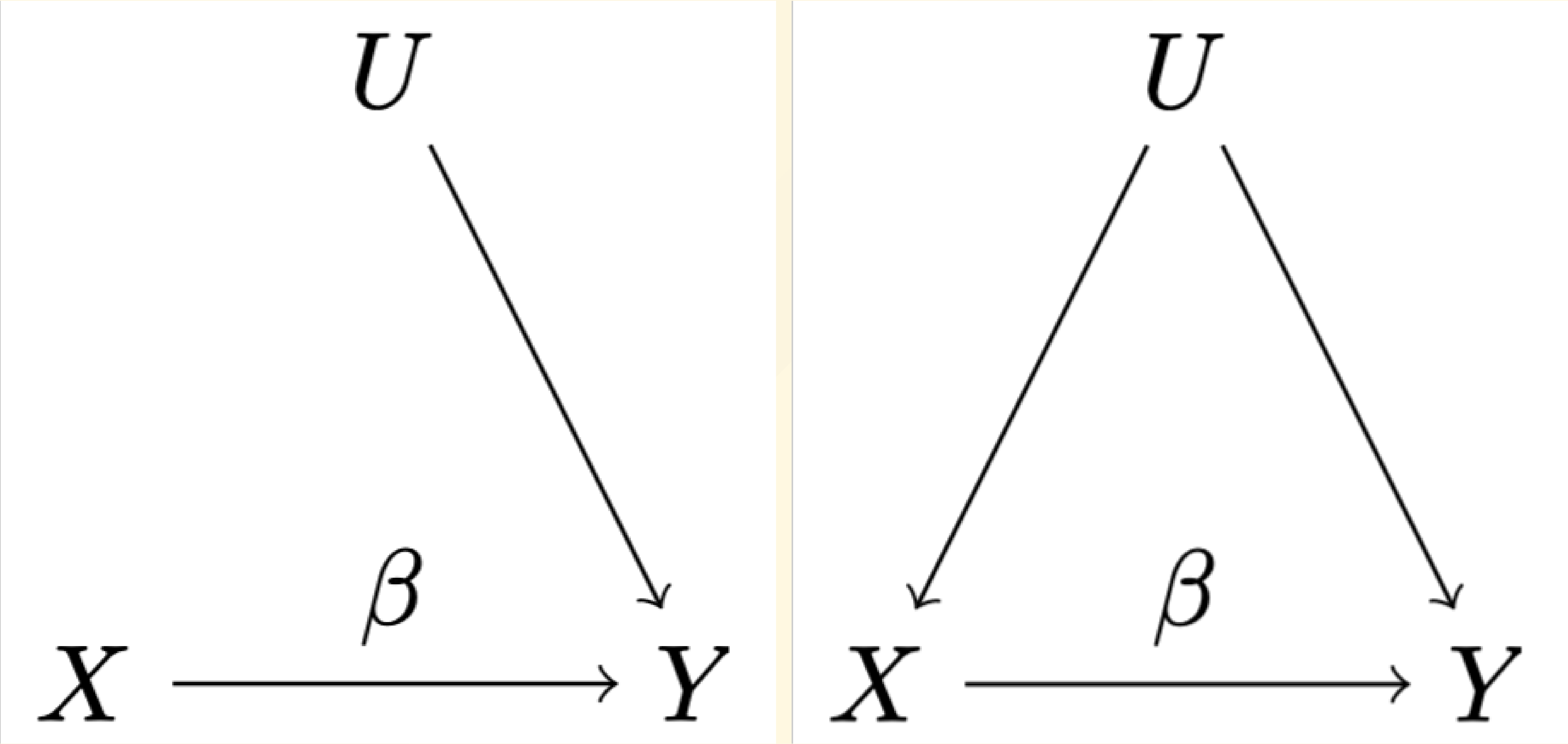
1. Z_i directly affects X_i: $Z \rightarrow X$ (**relevance**).
2. Z_i affects Y_i *only through* X_i (**exclusion**).
   Ie, the "assignment" of Z_i is independent of U_i.
3. All individuals face the same distribution of Z_i. (**As-good-as-random assignment**)
   Ie, $U \not\rightarrow Z$.

- Old texts sometimes refer to $cov(Z_i, U_i) = 0$ as the “exclusion restriction.”
- Modern IV texts distinguish between the two cases.

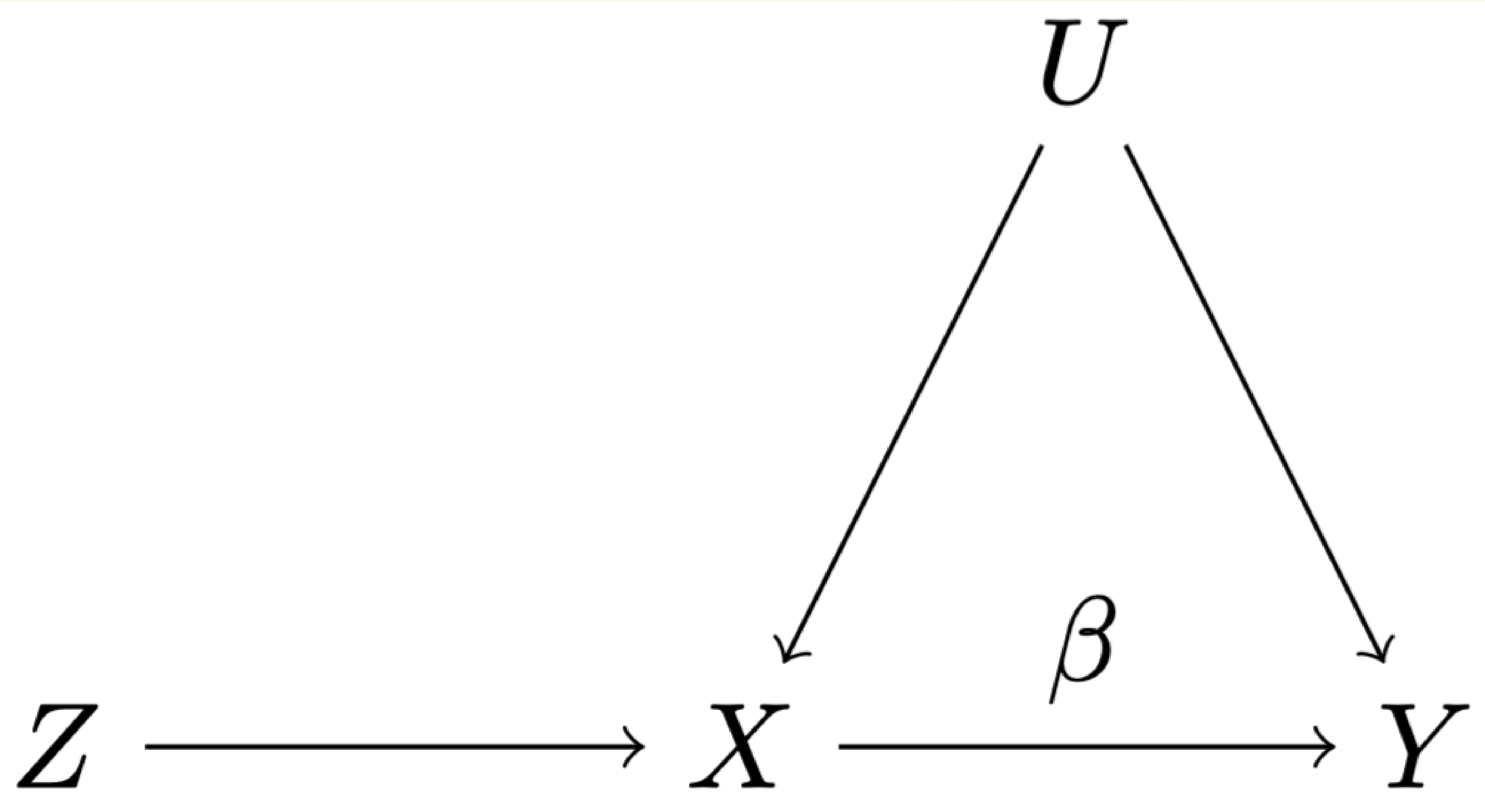
IV conditions:

1. Z_i directly affects X_i: $Z \to X$ (**relevance**).
2. Z_i affects Y_i *only through* X_i (**exclusion**).
   Ie, the "assignment" of Z_i is independent of U_i.
3. All individuals face the same distribution of Z_i. (**As-good-as-random assignment**)
   Ie, $U \not\to Z$.

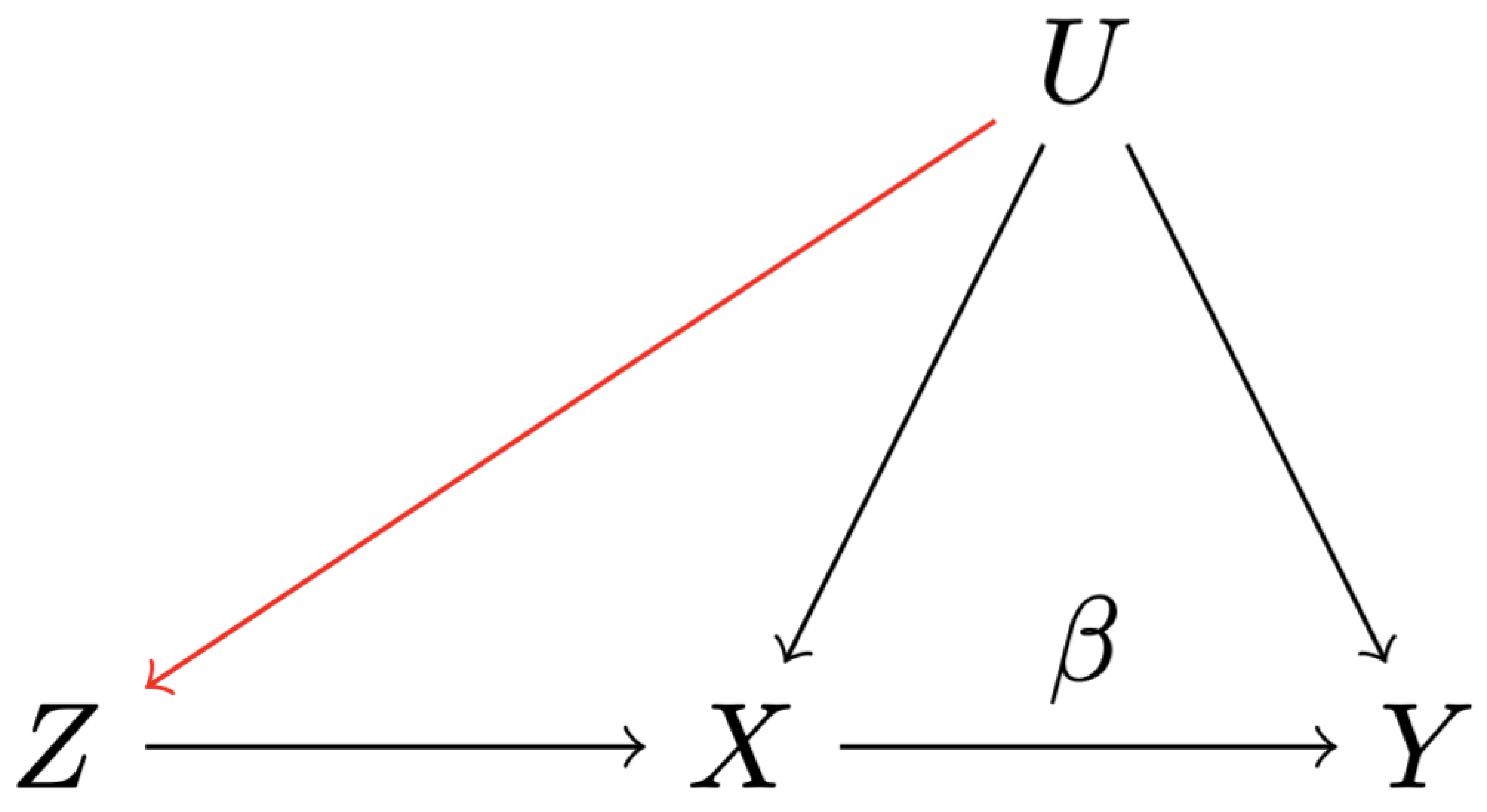
Random assignment & non-random assignment



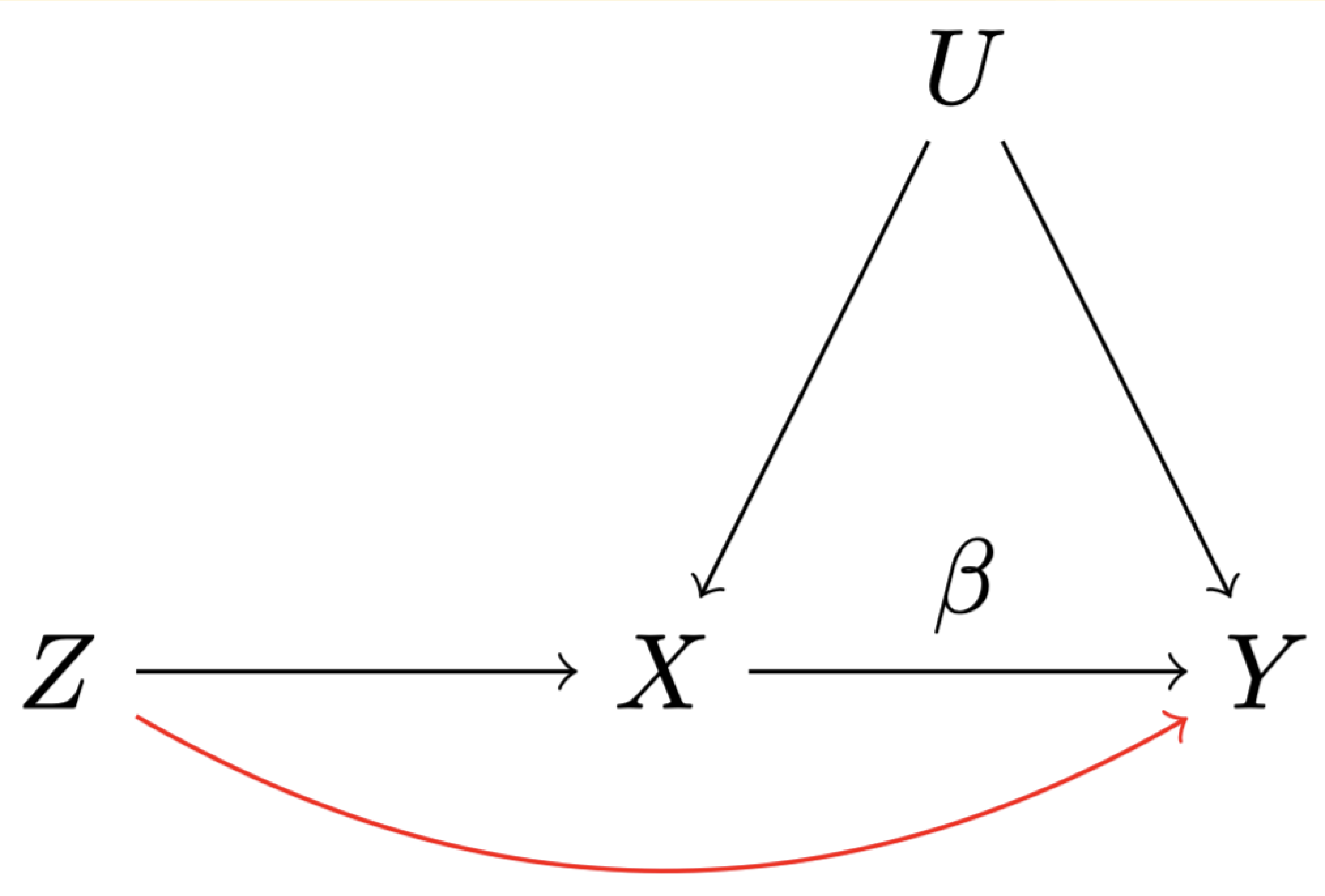
A valid instrument



Violation of "As-Good-As-Random Assignment"



Violation of "Exclusion"



Popular IVs

- Lotteries
- Natural experiments

Warnings: the taxonomy is "sloppy" and non-exhaustive.

Lotteries as IV

- (True) Lotteries guarantee that Z_i is as-good-as-randomly assigned.
- Some of the best IVs come from lotteries, either run by the researcher or "natural experiments".
- We only need to worry about the exclusion restriction.

Charter school lotteries

Abdulkadiroglu et al. (2016): whether going to a "charter" school increases students' grades

- Charter students tend to score better, but selection exists.

An **institutional feature** of charters: **admission lotteries**

- When more kids want to enroll than there are seats, admission offers $Z_i \in \{0, 1\}$ are drawn from a hat.
- Offers plausibly only affect later test scores Y_i by changing charter enrollment X_i. (*exclusion*)

"China's College Entrance Exam" Lotteries

- China's College Entrance Exam Score (*Gao Kao*).
  - A fair and controversial heritage of the *Ke Ju* system.
- Whether Albert (a high school student) attends a tier-1 or tier-2 college depends on his grades in *Gao Kao*.
- Suppose the bar for entering a tier-1 college is **600 points**. Then **those scoring 600 and 601** are given admission offers (Z_i) while **those scoring 599 and 598** are not.
  - As-good-as-randomly assigned
  - Exclusion

"China's College Entrance Exam" Lotteries

"The Value of Elite Education in China" Jia and Li (2017)

- The reality is not as ideal as what I just described. Occasionally a student scoring 599 got an offer while another scoring 601 didn't.
- Instead, Jia and Li (2017) use **(fuzzy) Regression Discontinuity Design** (**RDD**), a quasi-experimental evaluation.

Natural experiments

- Natural experiments are not literally random such as lotteries
- However, we may credibly argue Z_i is as-good-as-randomly assigned **conditional on some W_i**.
- Still need to worry about exclusion.

Quarter-of-birth

Angrist and Krueger (1991) estimate labor market returns to schooling with a creative IV: **student quarter-of-birth**

- *Compulsory schooling* requirements prevent students from dropping before *the day they turn 16*
- Fixing school start dates, students who drop out at 16 get **more or less schooling** (X) depending on their **birth date** (Z)

As-good-as-randomly assigned? Exclusion?

Angrist and Krueger (1991)

- Due to compulsory school attendance laws, quarter of birth is related to educational attainment.
- Individuals born in the **beginning of the year** start school at an older age, and can therefore drop out after completing **less schooling** than individuals born near the end of the year.
- Roughly 25 percent of potential dropouts remain in school because of compulsory schooling laws. We estimate the impact of compulsory schooling on earnings by **using quarter of birth as an instrument for education**.
- The result suggests that OVB is not severe in traditional OLS studies.

Why is the exclusion restriction challenging?

- Beware of the exclusion restriction.
- Intuitively, it feels like something (nearly) randomly assigned should satisfy this restriction, so long as it affects X.
- This is not sufficient. One needs to think critically about the IV.
  - Note that it's impossible to empirically test "exclusion".
- I'll give two examples in which the exclusion restriction fails.
  - Vietnam vet
  - Rainfall

Vietnam war lottery numbers

- Using Vietnam war lottery numbers as an IV for military service, studying the impact on mortality.
  - Y is death; X is vietnam vet; Z is lottery number
- Lottery number was randomly assigned as a function of birthdate
  - As-good-as-randomly assigned!
- Does that necessarily satisfy exclusion restriction?

Vietnam war lottery numbers

- Z (lottery number) probably fails the exclusion restriction.
- Consider one simple example: being drafted induces you to change your behavior to avoid the draft
  - Stay in school
  - Flee to Canada
- This would violate the exclusion restriction.

"Identification of causal effects using instrumental variables," Angrist, Imbens and Rubin (1996), *JASA*.

> “ But a draftee who managed to avoid military service by staying in school or moving abroad could experience an effect of Z on future life outcomes that would violate the *exclusion restriction.* For both these groups of noncompliers, the exclusion restriction requires the researcher to consider a difference in outcomes that were potentially observable, even though after the population was randomly allocated to treatment and control groups, only one of the outcomes was actually observed. ”

"Identification of causal effects using instrumental variables: Comment," Rosenbaum (1996), *JASA*.

> “ The fact that economists do not always make a clear distinction between ignorability and exclusion restrictions is evidenced by Moffitt's incorrect comment that randomization makes the draft lottery *by necessity an obvious and convincing instrument* for the effect of the military service. In fact, one contribution of our approach is to provide a framework that clearly separates ignorability and exclusion assumptions. Both statisticians and economists should find this separation useful and clarifying. ”

Second example: rainfall

- Consider rainfall as an instrument for income in agriculture environments (many crops are heavily dependent on it)
  - This is not uncommon in development papers, as Sarsons (2015) points out
  - Y is conflict; D is income; Z is rainfall.
- Exclusion restriction is that rainfall has no effect on conflict beyond income

Second example: rainfall

- While Exclusion restriction seems reasonable, Sarsons (2015) shows that places with dams (which protect against the income shocks due to rain) have similar conflict to those without dams.
- Plausible that while rain is “random”, it might have many channels besides "rain $\rightarrow$ income $\rightarrow$ conflict".

Rainfall and Conflict: A Cautionary Tale, Sarsons (2015)

> “ … there could be some unobserved variable X that is correlated with a district being dam-fed and that variable also increases the marginal effect of rainfall on rioting through a non-income channel. For example, dam-fed districts could all have dirt roads and an increase in rainfall destroys these roads, making it more difficult for people to organize and riot. While this is possible, it does not validate the use of rainfall as an instrument for income. **There is now a non-income through which rainfall is affecting rioting in damfed districts.** ”

The Local Average Treatment Effect (LATE)

Heterogeneous Effects

- We have implicitly assumed that treatment X has the same effect (β) on everyone.
- In real life treatment effects may be *heterogeneous*.
  - Eg: it's unlikely that each person would get the same benefit from an **additional year of schooling**; for some people, the effect can even be negative.
- We always say that we are estimating the **average** treatment effect. However, when treatment effect is heterogeneous, the instrumental variable approach is not valid.

LATE

- But all may not be lost. We just need a different interpretation of the estimate.
- We can interpret the estimate as the average effect for **a subset of the population**.
- The corresponding *estimand* is called the Local Average Treatment effect.
  - BTW, do you know the differences between *estimand, estimate* and *estimator*?

A brief aside: estimands, estimators and estimates

- **Estimand**: the quantity to be estimated
- **Estimate**: the approximation of the estimand using a finite data sample
- **Estimator**: the method or formula for arriving at the estimate for an estimand

A brief aside: estimands, estimators and estimates

- **Estimand**: the quantity to be estimated
- **Estimate**: the approximation of the estimand using a finite data sample
- **Estimator**: the method or formula for arriving at the estimate for an estimand

For example, we specify the data generation process $y = 1 + \beta x + \epsilon$ with $\beta = 2$. We use R to simulate the data $\{x_i, y_i\}_{i=1}^N$ and `lm(y~x)` yields $\hat{\beta} = 2.1$.
What are estimands, estimators and estimates?

Heterogeneous Effects: Model

$$Y_i = \alpha + \color{red}{\beta_i} X_i + \epsilon_i$$

- Intuitively, different "research designs" (e.g. instruments) may capture different effects of the same treatment — even when all are valid
  - Different IVs may yield very different estimates!
- This idea is formalized in the (Nobel-winning) Imbens and Angrist 1994 LATE theorem.
  - It uses a general potential outcomes framework.

Potential Outcome Setup

Let $Y_i(0)$ and $Y_i(1)$ be an individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$.

- Observed outcomes: $Y_i = (1 - D_i)Y_i(0) + D_iY_i(1)$

Potential Outcome Setup

Let $Y_i(0)$ and $Y_i(1)$ be an individual i's potential outcomes given a binary treatment $D_i \in \{0, 1\}$.

- Observed outcomes: $Y_i = (1 - D_i)Y_i(0) + D_iY_i(1) = \alpha_i + \beta_i D_i$
- (Individaul-level) treatment effects $\beta_i = Y_i(1) - Y_i(0)$

Imbens-Angrist: we can also do this for an IV first stage:

- Let $D_i(0)$ and $D_i(1)$ denote individual i's potential treatment given a binary treatment $Z_i \in \{0, 1\}$. **What is** `ivreg(Y ~ D|Z)`?

Imbens and Angrist (1994) Assumptions

1. **As-good-as-random assignment:**
    - The assignment of Z_i is independent of Y_i and D_i
2. **Exclusion:** Z_i only affects Y_i through D_i.
    - Implicit in our potential outcomes notation: $Y_i(D)$ not indexed by Z_i
3. **Relevance:** Z_i is correlated with D_i.
    - Equivalently, $E[D_i(1) - D_i(0)] \neq 0$.

Imbens and Angrist (1994) Assumptions

1. **As-good-as-random assignment:**
   - The assignment of Z_i is independent of Y_i and D_i
2. **Exclusion:** Z_i only affects Y_i through D_i.
3. **Relevance:** Z_i is correlated with D_i.
4. **Monotonicity:** $D_i(1) \geq D_i(0)$ for all i.
   - The instrument can only shift the treatment in one direction.
   - Egs: College distance, rainfall, birth-quarter,...

Local Average Treatment Effect (LATE) Identification

Imbens and Angrist showed that under these assumptions:

$$\beta^{IV} = E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)]$$

The IV estimator β^{IV} identifies a **LATE**:

- the **average treatment effect** $Y_i(1) - Y_i(0)$ among *compliers:* those with $1 = D_i(1) > D_i(0) = 0$.

Local Average Treatment Effect (LATE) Identification

Imbens and Angrist showed that under these assumptions:

$$\beta^{IV} = E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)]$$

The IV estimator β^{IV} identifies a **LATE**:

- the **average treatment effect** $Y_i(1) - Y_i(0)$ among *compliers:* those with $1 = D_i(1) > D_i(0) = 0$.

Intuitively, IV can't tell us anything about the treatment effects of *never-takers* and *always-takers*.

Complier, Always Taker, Never Taker, Defier

1. *Complier:* $\Pr(D = 1 | Z = 1, C) = \Pr(D = 0 | Z = 0, C) = 1$
2. *Always Taker:* $\Pr(D = 1 | Z = 1, A) = \Pr(D = 1 | Z = 0, A) = 1$
3. *Never Taker:* $\Pr(D = 0 | Z = 1, N) = \Pr(D = 0 | Z = 0, N) = 1$
4. *Defier:* $\Pr(D = 0 | Z = 1, De) = \Pr(D = 1 | Z = 0, De) = 1$

Complier, Always Taker, Never Taker, Defier

1. *Complier:* $\Pr(D = 1|Z = 1, C) = \Pr(D = 0|Z = 0, C) = 1$
2. *Always Taker:* $\Pr(D = 1|Z = 1, A) = \Pr(D = 1|Z = 0, A) = 1$
3. *Never Taker:* $\Pr(D = 0|Z = 1, N) = \Pr(D = 0|Z = 0, N) = 1$
4. *Defier:* $\Pr(D = 0|Z = 1, De) = \Pr(D = 1|Z = 0, De) = 1$

LATE Thm: IV only identifies the average treatment effect among *Compliers*. The monotonicity assumption rules out *Defiers*.

What Does This Mean *Practically*?

Two conceptually distinct considerations: **internal vs. external validity**

- **Context of an IV**, and who the **compliers** are, matter.
- Usual "over-identification" test logic fails: two valid IVs may have different estimands. (Kitagawa, 2015)

What Does This Mean *Practically*?

Two conceptually distinct considerations: **internal vs. external validity**

- **Context of an IV**, and who the **compliers** are, matter.
- Usual "over-identification" test logic fails: two valid IVs may have different estimands. (Kitagawa, 2015)

In addition to as-good-as-random assignment and exclusion, we may need to worry about **monotonicity** when we do IV.

- Monotonicity holds in almost all IV examples
- But it may fail. Eg, judge IV.

Judge (or examiner) IV design

Imbens and Angrist (1994) provide an example in which Monotonicity fails.

EXAMPLE 2 (Administrative Screening):[5] Suppose applicants for a social program are screened by two officials. The two officials are likely to have different admission rates, even if the stated admission criteria are identical. Since the identity of the official is probably immaterial to the response, it seems plausible that Condition 1 is satisfied. The instrument is binary so Condition 3 is trivially satisfied. However, Condition 2 requires that if official A accepts applicants with probability $P(0)$, and official B accepts people with probability $P(1) > P(0)$, official B must accept *any* applicant who would have been accepted by official A. This is unlikely to hold if admission is based on a number of criteria. Therefore, in this example we *cannot* use Theorem 1 to identify a local average treatment effect nonparametrically despite the presence of an instrument satisfying Condition 1.

[5] This example was suggested to us by Geert Ridder.

Judge (or examiner) IV design

Imbens and Angrist (1994) provide an example in which Monotonicity fails.

A judge (or examiner) IV design leverages the *idiosyncratic assignment* of individuals to a set of judges. A more tolerant judge is likely to be more merciful in deciding the prison sentence.

- Kling (2006): sentencing judges
- Doyle (2007): foster care investigators
- Maestas et al. (2013): SSDI benefit examiners
- Doyle et al. (2015): ambulance companies

Judge (or examiner) IV design

Imbens and Angrist (1994) provide an example in which Monotonicity fails.

A judge (or examiner) IV design leverages the *idiosyncratic assignment* of individuals to a set of judges. A more tolerant judge is likely to be more merciful in deciding the prison sentence.

All these works need to take special care of the monotonicity restriction.

- Eg: Frandsen et al. (2019) formalize a weaker **"average monotonicity"** condition.

Extensions of Angrist and Imbens (1994)

Angrist and Imbens worked out the original LATE theroem for binary D_i, discrete Z_i, and no included controls.

- Angrist/Imbens '95: multi-valued (ordered) D_i , saturated covariates
- Angrist/Graddy/Imbens ’00: continuous D_i (supply/demand setup)
- Heckman/Vytlicil ’05: continuous Z_i
- Multiple unordered treatments is harder (e.g. Behaghel et al. 2013)

Extensions of Angrist and Imbens (1994)

Angrist and Imbens worked out the original LATE theroem for binary D_i, discrete Z_i, and no included controls.

- Angrist/Imbens '95: multi-valued (ordered) D_i , saturated covariates
- Angrist/Graddy/Imbens ’00: continuous D_i (supply/demand setup)
- Heckman/Vytlicil ’05: continuous Z_i
- Multiple unordered treatments is harder (e.g. Behaghel et al. 2013)

Recent discussions highlight importance of including flexible controls

- Sloczynski ’20, Borusyak and Hull ’21, Mogstad et al. ’22

Better LATE than nothing

- LATE is not a new *estimator*. We are still using 2SLS as before.
- LATE is about the *estimand*. When we are doing IV, we are really just estimating the average treatment effect on the *Compliers*.
  - Besides **exclusion** and **as-good-as-random assignment**, the **monotonicity restriction** is also needed.
- LATE implies that we should always beaware of *Context of an IV* and who the *Compliers* are.
  - In other words, LATE makes it **harder to misuse the IV research design**.
- For these reasons, Imbens (2010) argues that we should always think in the LATE framework --- better LATE than nothing.

Final words

- We have spent 2-3 courses on IV. However, lots of stuff are still not covered.
  - Specifically, how to **characterize compliers**? How to use IV with **panel data** or in a **DiD** setting? **GMM version** of IV?

Final words

- We have spent 2-3 courses on IV. However, lots of stuff are still not covered.
- I hope that starting from this course, you can continue your study of IV and use it in your own research.
  - Mixtape IV
  - Pinkham's Applied Empirical Methods at Yale

Final words

- We have spent 2-3 courses on IV. However, lots of stuff are still not covered.
- I hope that starting from this course, you can continue your study of IV and use it in your own research.
- It's getting more popular to view IV (and DiD, RD...) as experimental methods, and use "causal inference" (instead of regression-based analysis) as a unifying framework.
    - We do not really talk too much about the potential outcome framework. But if you are interested, there is an online course.