ECON 3510: Poverty and Economic Development
Lecture 1: Causality and IV

Instructor: Weizheng Lai

Bowdoin College

Fall 2025

Overview

- In this lecture, we formally define **causality**, which involves comparison to a counterfactual scenario.
- We will discuss causality in linear regressions. As it turns out, simple linear regression results cannot be interpreted as causal effects, unless additional assumptions are imposed.
- Then, we discuss the **instrumental variable (IV) strategy**, an approach economists use to infer causality. We will discuss more causal inference methods in future lectures.

Outline

1. Causality
2. Instrumental Variable Basics
3. Local Average Treatment Effect

Potential Outcomes Framework

- A simple framework to discuss causality.
  - Also called the Rubin Causal Model, named after statistician Donald Rubin.
- Consider a binary treatment $T_i \in \{0, 1\}$ and and an outcome Y_i.
  - Think of $T_i =$ holding a college degree, and $Y_i =$ earnings.
- An individual, indexed by i, has two potential outcomes:
  - Outcome under treatment, i.e., when $T_i = 1$: Y_{i1};
  - Outcome under control (non-treatment), i.e., when $T_i = 0$: Y_{i0}.
- E.g., Y_{i1} is the income individual i will get if they get a college degree.

  Y_{i0} is the wage individual i will get if they do not get a college degree.
- Note that $Y_i = T_i Y_{i1} + (1 - T_i) Y_{i0}$. (Why?)

Causal Effect

- What is the causal effect of the treatment on the outcome for individual i?

$$Y_{i1} - Y_{i0}$$

- This is the difference in outcome for individual i if they get treated versus if they do not get treated.
  - E.g., difference in income with a college degree versus not.

- Can we compute this treatment effect?

- We can't! For a given individual i, we never observe both of Y_{i1} and Y_{i0} simultaneously.
  - If i has a college degree, we observe Y_{i1} but we can't observe the **counterfactual** Y_{i0}.
  - Likewise, if i doesn't have a college degree, we observe Y_{i0} but we can't observe the **counterfactual** Y_{i1}.

- In reality, we are unable to do comparisons of Y for different hypothetical values of T—the thought experiment that defines causal effects.

Selection Bias

- Why don't we just compare a treated individual to a control individual?
  - I.e., compare an individual with a college degree to one without.
- Then, we compare individual i in treatment ($T_i = 1$) to individual j in control ($T_j = 0$):

$$Y_{i1} - Y_{j0}.$$

  Note that both Y_{i1} and Y_{j0} can be observed.
- Can we get the causal effect from this? Not necessarily. Why?

Selection Bias (Cont'd)

- We can rewrite $Y_{i1} - Y_{j0}$ as follows:

$$Y_{i1} - Y_{j0} = Y_{i1} - Y_{i0} + Y_{i0} - Y_{j0}$$

- The first component, $Y_{i1} - Y_{i0}$, is the **treatment effect** for individual i (what we want).
- However, there is a second term, $Y_{i0} - Y_{j0}$, which is the difference in potential outcomes under control between i and j. This term is called **selection bias**.
- Therefore, in general, $Y_{i1} - Y_{j0}$ is not equal to the causal effect due to the selection bias.
  - Positive selection $\Rightarrow$ upward bias, $Y_{i1} - Y_{j0} > Y_{i1} - Y_{i0}$.
  - Negative selection $\Rightarrow$ downward bias, $Y_{i1} - Y_{j0} < Y_{i1} - Y_{i0}$.
- Whether the selection bias is positive or negative is case-dependent.

Average Treatment Effect

- We usually want to estimate the average treatment effect (ATE):

$$ATE = E(Y_{i1} - Y_{i0}).$$

- ATE is over the entire population. There are other versions of average effects.

$$CATE(x) = E(Y_{i1} - Y_{i0} \mid X_i = x)$$

  where X_i is some characteristic. $CATE(x)$ is the average treatment effect among those with $X_i = x$.

- Estimating average effects also faces the challenge of selection bias

Average Treatment Effect and Selection Bias

- $ATE = E(Y_{i1} - Y_{i0})$. Because for all individuals, we only observe one of Y_{i1} and Y_{i0}, not both, we can't directly compute ATE. But we can compute the mean difference between the treated and untreated.

$$E(Y_{i1} \mid T_i = 1) - E(Y_{i0} \mid T_i = 0)$$

where both terms are observable. In the above equation, substract and add $E(Y_{i0} \mid T_i = 1)$:

$$\underbrace{E(Y_{i1} \mid T_i = 1) - E(Y_{i0} \mid T_i = 1)}_{=E(Y_{i1} - Y_{i0} \mid T_i = 1)} + E(Y_{i0} \mid T_i = 1) - E(Y_{i0} \mid T_i = 0).$$

- $E(Y_{i1} - Y_{i0} \mid T_i = 1)$: average treatment effect on the treated (ATT).
  - Not exactly ATE. But it's still a useful causal parameter—policy makers may care more about the impact on the population targeted by an intervention.
- $E(Y_{i0} \mid T_i = 1) - E(Y_{i0} \mid T_i = 0)$: **selection bias**—avg difference in Y_{i0} btw the treated and untreated.
  - E.g., the avg difference in earnings btw college grads who, *hypothetically*, did not obtain a degree and individuals who never had a college degree.
  - Positive (negative) selection would lead to upward (downward) bias for ATT.

Average Treatment Effect and Selection Bias (Cont'd)

- A sufficient condition to purge selection bias is **random assignment of the treatment**.

$$T_i \perp\!\!\!\perp Y_{i1}, Y_{i0}.$$

  $\Rightarrow E(Y_{i0} \mid T_i = 1) - E(Y_{i0} \mid T_i = 0) = E(Y_{i0}) - E(Y_{i0}) = 0$. (In fact, only $T_i \perp\!\!\!\perp Y_{i0}$ is used here, but randomization should break T_i's dependence on both potential outcomes.)

  E.g., (unrealistically) college degrees are randomly awarded.

- At least, need an assumption such that individuals would have similar average potential outcomes if they were not treated:

$$E(Y_{i0} \mid T_i = 1) = E(Y_{i0} \mid T_i = 0)$$

- The assumption that purges selection bias is called **"identifying assumption."**

Causality and Regression

- The identifying assumption under the potential outcomes model coincides with the assumption we need for consistent OLS regression.
- Note

$$\begin{aligned} Y_i &= T_i Y_{i1} + (1 - T_i) Y_{i0} \\ &= (Y_{i1} - Y_{i0}) T_i + Y_{i0} \\ &= E(Y_{i0}) + (Y_{i1} - Y_{i0}) T_i + Y_{i0} - E(Y_{i0}) \end{aligned}$$

- For simplicity, let $Y_{i1} - Y_{i0} = \beta$, i.e., constant treatment effect (this assumption is not necessary). Then,

$$Y_i = \alpha + \beta T_i + u_i$$

where $\alpha = E(Y_{i0})$, $\beta = Y_{i1} - Y_{i0}$, and $u_i = Y_{i0} - E(Y_{i0})$.

Causality and Regression (Cont'd)

- If we run an OLS regression of Y_i on T_i in the population,

$$\beta_{\text{OLS}} = \frac{Cov(T_i, Y_i)}{Var(T_i)} = \underbrace{\beta}_{\text{treatment effect}} + \underbrace{\frac{Cov(T_i, u_i)}{Var(T_i)}}_{\text{selection bias}}.$$

- OLS identifies the treatment effect, i.e., $\beta_{\text{OLS}} = \beta$ when $Cov(T_i, u_i) = 0$.

- Note $u_i = Y_{i0} - E(Y_{i0})$, a random assignment assumption ($T_i \perp\!\!\!\perp Y_{i1}, Y_{i0}$) suffices.

- **Take-away:** We can always run regressions so long as data are available, but interpreting a coefficient as a parameter of interest requires additional **assumptions**.

Conditional Independence Assumption

- The **conditional independence assumption** (CIA; also called conditional unconfoundedness):

$$D_i \perp\!\!\!\perp Y_{i1}, Y_{i0} \mid \mathbf{X}_i = \mathbf{x}.$$

$\mathbf{X}_i$: covariates/control variables/confounders.

- How does this help us?

$$\begin{aligned}
& E(Y_i \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_i \mid D_i = 0, \mathbf{X}_i = \mathbf{x}) \\
&= E(Y_{i1} \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_{i0} \mid D_i = 0, \mathbf{X}_i = \mathbf{x}) \\
&= E(Y_{i1} \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_{i0} \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) \\
&\qquad + E(Y_{i0} \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_{i0} \mid D_i = 0, \mathbf{X}_i = \mathbf{x}) \\
&= CATT(\mathbf{x}) + \text{selection bias}
\end{aligned}$$

CIA implies that

$$\text{selection bias} = E(Y_{i0} \mid \cancel{D_i = 1}, \mathbf{X}_i = \mathbf{x}) - E(Y_{i0} \mid \cancel{D_i = 0}, \mathbf{X}_i = \mathbf{x}) = 0.$$

In addition, CIA implies:

$$CATT(\mathbf{x}) = E(Y_{i1} \mid \cancel{D_i = 1}, \mathbf{X}_i = \mathbf{x}) - E(Y_{i0} \mid \cancel{D_i = 0}, \mathbf{X}_i = \mathbf{x}) = CATE(\mathbf{x}).$$

CIA and Regression

- We have shown that with CIA, $CATE(\mathbf{x})$ can be identified by observable conditional means:

$$CATE(\mathbf{x}) = E(Y_i \mid D_i = 1, \mathbf{X}_i = \mathbf{x}) - E(Y_i \mid D_i = 0, \mathbf{X}_i = \mathbf{x}).$$

- Common to approximate the CEF linearly, as

$$E[Y_i|D_i, \mathbf{X}_i] \approx D_i\beta + \mathbf{X}_i'\boldsymbol{\gamma}$$

- Then CIA implies that $CATE(\mathbf{x}) \approx \beta$.

  Doesn't depend on $\mathbf{x}$, so also have $\beta \approx ATE$.

- So if we estimate the multivariate regression

$$Y_i = D_i\beta + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i,$$

  we can interpret $\hat{\boldsymbol{\beta}}$ as an estimate of the ATE (under the CIA).

- This can be extended to multivalued and continuous treatment, where $\hat{\boldsymbol{\beta}}$ can be interpreted as an estimate of the average marginal effect of the treatment.

Choice of Covairates

- Invoking the CIA was a popular approach to causality.
- For the CIA to be valid, the choice of covariates is key. Some ideas:
  - Economic theory;
  - Institutional knowledge.
- It is important to validate the choice of covariates, i.e., conditioning upon them can purge selection bias.
- After fixing a set of baseline, most important controls, researchers will often show that their results are robust to controlling for a larger set of covariates to address various concerns.
- **Limitation of CIA:** We can only control for observable variables. The estimator can still be biased if there are important *unobservable* factors that cannot be proxied by observables.

  Modern causal inference seeks to design "good comparisons" to tease out the influence of unobservables, without relying on the "correct" choice of covariates.

Solutions to Selection Bias

- Simply assuming random assignment or zero selection bias won't give us convincing results.
- What can be done with selection bias?

1. Randomized controlled trials (RCTs): assignment is artificially randomized.
   - RCTs are costly. Moreover, many treatments cannot be randomized for feasibility and ethical reasons. E.g., it won't be possible and appropriate to randomly assign democracy to a country.
2. Control for selection: include many variables in a regression; hopefully, conditional upon them, the treatment is as *good as randomly assigned*.
   - In practice, not very promising because we rarely know what variables determine the assignment process.
3. **Natural experiments:** look for "events" that *exogenously* change the treatment.
   - This has become the state-of-the-art approach in empirical economics.
   - We will cover many methods in this course. Next, we start with **instrumental variables**.

Outline

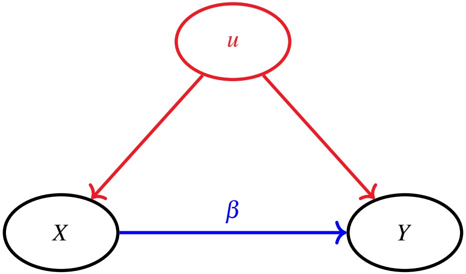
1. Causality

2. Instrumental Variable Basics

3. Local Average Treatment Effect

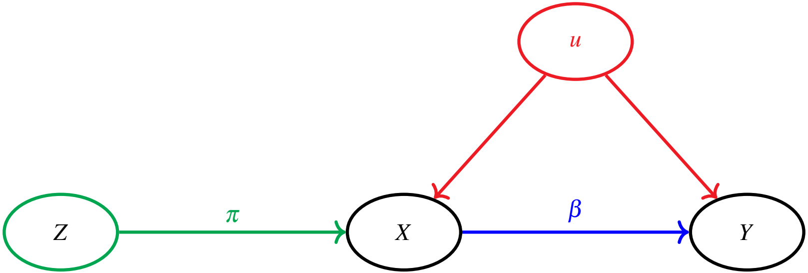
Endogeneity/Selection Bias

- $Y = \alpha + \beta X + u$ where $Cov(X, u) \neq 0$. Regression doesn't identify the causal effect β.



Idea of IV

- $Cov(X,u) \neq 0$.
- However, Z can shock X without shifting u, and it doesn't *directly* affect Y. Z is an *instrument* for X.



Basic IV Model

- Suppose Y_i = earnings, X_i = education, and u_i = unobserved ability.
- Consider a constant-effect population model (often called the structural equation):

$$Y_i = \alpha + \beta X_i + u_i.$$

$Cov(X_i, u_i) \neq 0$, thus regression coefficient $\beta_{\text{OLS}} = \frac{Cov(X_i, Y_i)}{Var(X_i)} \neq \beta$.

- Suppose we observe the outcome of a college scholarship lottery, Z_i.
  - E.g., $Z_i = 1$ if i is a scholarship lottery winner, $Z_i = 0$ otherwise.
- Assume that Z_i only affects Y_i through its effect on X_i. Z_i is said to be an *instrument* for X_i. Randomization implies $Cov(Z_i, u_i) = 0$.
  - If Z_i has a direct effect on Y_i, $u_i = \gamma Z_i + v_i$. Randomization implies $Cov(Z_i, v_i) = 0$ but it is possible $Cov(Z_i, u_i) \neq 0$.
- How is this useful?

IV Mechanics

- By $Cov(Z_i, u_i) = 0$,

$$\begin{aligned} Cov(Z_i, Y_i - \alpha - \beta X_i) &= 0 \\ \beta Cov(Z_i, X_i) &= Cov(Z_i, Y_i) \\ \beta &= \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)}. \end{aligned}$$

  Thus, β is identified by $\frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$.

- $\frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$ is said to be the instrumental variable estimand, denoted by β_{IV}.

IV Mechanics (Cont'd)

- The IV estimand can be written as

$$\beta_{\text{IV}} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, X_i)/Var(Z_i)} = \frac{\rho}{\pi}.$$

  ρ and π are coefficients from regressions:

$$Y_i = \kappa + \rho Z_i + \nu_i \qquad \text{(reduced-form regression)},$$
$$X_i = \mu + \pi Z_i + \eta_i \qquad \text{(first-stage regression)}.$$

  - Note that ρ and π identify causal effects if Z_i is as-good-as-randomly assigned.
  - ρ identifies the **intent-to-treat (ITT) effect**—the effect of an assignment regardless of actual treatment take-up.

- IV divides the "reduced-form effect" of the instrument on the outcome by the "first-stage effect" of the instrument on the treatment.
  - Intuitively, $\Delta Z \to \Delta X \to \Delta Y$.
  - Effect of Z on Y (ρ) = Effect of Z on X (π) $\times$ Effect of X on Y (β).

- If Z_i is binary, β_{IV} reduces to

$$\beta_{\text{IV}} = \frac{\rho}{\pi} = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(X_i \mid Z_i = 1) - E(X_i \mid Z_i = 0)}.$$

Validity of the IV Estimand

- When is the IV estimand $\beta_{\text{IV}} = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$ valid, i.e., $\beta_{\text{IV}} = \beta$?

1. **(Relevance)** $Cov(Z_i, X_i) \neq 0$.
   - Instrument Z_i is predictive of treatment X_i.
   - This implies the first-stage regression $X_i = \mu + \pi Z_i + \eta_i$, $\pi \neq 0$.

Validity of the IV Estimand (Cont'd)

- The key "validity" condition $Cov(Z_i, u_i) = 0$ (so that $\beta_{\text{IV}} = \beta$) requires **two distinct assumptions**. For illustration, write u_i in the structural model as $u_i = \gamma Z_i + v_i$.

② **(Independence)** Individuals with higher/lower potential outcomes (i.e., hypothetical values of Y_i conditional upon X_i and Z_i) do not face systematically different values of Z_i.

- Potential outcome $Y_i(x,z) = \alpha + \beta x + \gamma z + v_i$. Independence means $Z_i \perp\!\!\!\perp v_i$.

③ **(Exclusion Restriction)** Z_i does not directly affect Y_i through X_i.

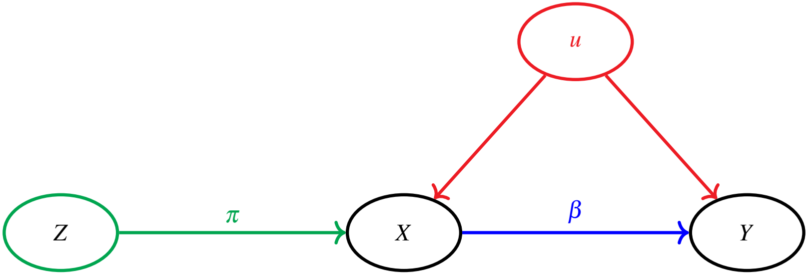
- This means $\gamma = 0$. That said, the structural equation "correctly" excludes Z_i; Z_i does not hide in u_i.

- **Independence + Exclusion Restriction = Exogeneity**. Z_i affects Y_i *only* through X_i.

- Confusingly, some old-school econometrics texts sometimes refer to $Cov(Z_i, u_i) = 0$ as the "exclusion restriction." We shall distinguish them.
  - We may adopt this terminology as it has been widely accepted.

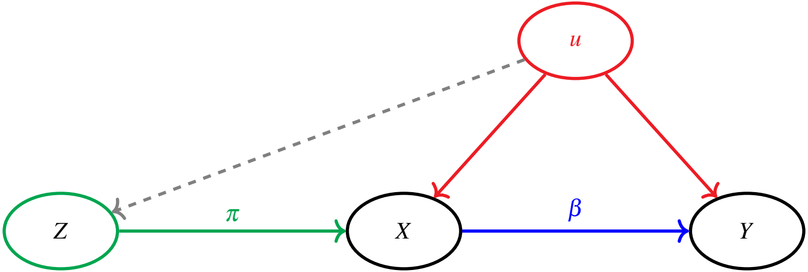
Valid IV Graph

- We can re-present $Cov(Z_i, u_i) = 0$ graphically.



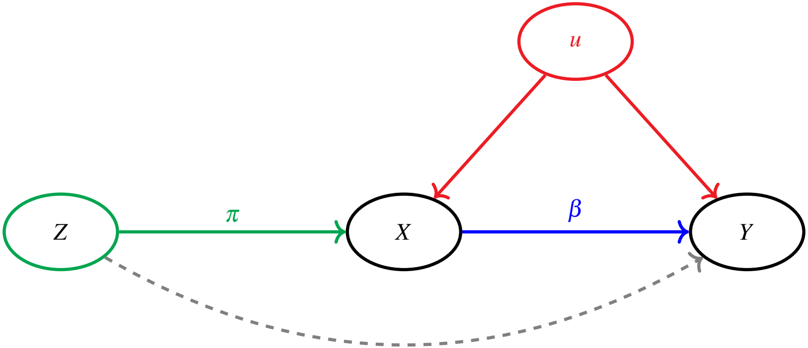
Independence Violated

- Some unobserved factors in u can also affect Z.



Exclusion Restriction Violated

- Z directly affects Y.



IV Estimation

- Use the MoM to estimate $\beta_{\text{IV}} = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,X_i)}$: replacing population moments with sample means

$$\hat{\beta}_{\text{IV}} = \frac{\sum_i (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_i (Z_i - \bar{Z})(X_i - \bar{X}_i)} = \frac{\frac{\sum_i (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_i (Z_i - \bar{Z}_i)^2}}{\frac{\sum_i (Z_i - \bar{Z})(X_i - \bar{X})}{\sum_i (Z_i - \bar{Z}_i)^2}} = \frac{\hat{\rho}}{\hat{\pi}}.$$

  $\hat{\rho}$ and $\hat{\pi}$ are from OLS reduced-form and first-stage regressions:

$$\begin{aligned} Y_i &= \hat{\kappa} + \hat{\rho} Z_i + \hat{v}_i && \text{(reduced-form regression)}, \\ X_i &= \hat{\mu} + \hat{\pi} Z_i + \hat{\eta}_i && \text{(first-stage regression)}. \end{aligned}$$

- $\hat{\beta}_{\text{IV}}$ can also be derived from regression

$$Y_i = \alpha + \beta \hat{X}_i + \varepsilon_i,$$

  where $\hat{X}_i$ is the predicted value of X_i for the first-stage regression. The above regression is often called the "second-stage regression," and the IV estimator $\hat{\beta}_{\text{IV}}$ is also called the two-stage least squares (2SLS) estimator.

- However, in Stata, one should never literally run two OLS regressions to obtain $\hat{\beta}_{\text{IV}}$. The "second stage" wouldn't take into account the uncertainty of predicting $\hat{X}_i$, so the SEs would be incorrect.
  - `ivreg2` in Stata does the SE adjustment automatically: `ivreg2 y (x = z)`.

Evaluating IV Validity

1. **(Relevance)** $Cov(Z_i, X_i) \neq 0$.
   - It implies the first-stage regression $X_i = \mu + \pi Z_i + \eta_i$, $\pi \neq 0$.
   - We can test $H_0 : \pi = 0$. If $\hat{\pi}$ is highly significant, that suggests a strong IV. A rule of thumb is that the F-stat should be at least 10 (Staiger and Stock, 1997).
   - An economic argument for relevance is appreciated.

2. **(Independence)** Individuals with higher/lower potential outcomes do not face systematically different values of Z_i.
   - This is untestable because we do not observe potential outcomes.
   - We may leverage institutional knowledge: Is the assignment of the IV random or near random condition upon some observable characteristics?
   - We may examine correlations between the IV and some observable determinants of the outcome (balance test).

3. **(Exclusion Restriction)** Z_i *only* affects Y_i through X_i.
   - This is also untestable, because alternative channels, if any, are in u_i, which we cannot observe.
   - Economic theory can help us think about whether an alternative channel is possible.
   - We can test if Z_i has an impact on variables that measure alternative channels. This can increase our confidence in the IV, however, we cannot exhaust/measure all possibilities. A strong economic argument is necessary.

Outline

1. Causality

2. Instrumental Variable Basics

3. Local Average Treatment Effect

When Treatment Effects Are Heterogeneous...

- We have assumed the treatment effects are homogeneous. That is, we consider a population model:

$$Y_i = \alpha + \beta X_i + u_i$$

  where β is constant.

- In this framework, by construction, IV identifies a causal effect that applies to everyone, though perhaps only a small group of individuals react to the IV.

- We will allow for effect heterogeneity, and show that under certain assumptions, IV lets us identify a **local average treatment effect (LATE)**—an average effect for **compliers**, i.e., people who are induced by the IV to take up the treatment.
  - Unlike in the constant-effect model, now the IV estimand is a "local" effect in the sense that it doesn't tell us about the treatment effect for people whose treatment status is not affected by the IV.

- For simplicity, we consider binary treatment $D_i \in \{0,1\}$ and binary IV $Z_i \in \{0,1\}$.
  - E.g, D_i = indicator of high/low education; Z_i = indicator of winning the scholarship lottery.

Four Types of People

- Introduce **potential treatments**: $D_i(z)$, a function of the IV value (z); analogous to the potential outcomes.
  - $D_i(1)$ is treatment status if $Z_i = 1$ (win the lottery).
  - $D_i(0)$ is treatment status if $Z_i = 0$ (lost the lottery).
- **Always takers:** people who will have high educational attainment regardless of the outcome of the lottery. Always takers have $D_i(1) = D_i(0) = 1$.
- **Never takers:** people who will never have high educational attainment regardless of the lottery. Never takers have $D_i(1) = D_i(0) = 0$.
- **Compliers:** people who have high educational attainment *only* if they win the lottery. Compliers have $D_i(1) = 1$ and $D_i(0) = 0$.
- **Defiers:** people who have high educational attainment *only* if they lose the lottery. Defiers have $D_i(1) = 0$ and $D_i(0) = 1$.

Potential Outcomes

- **Potential outcomes:** $Y_i(d,z)$, which is a function of hypothetical values of D_i and Z_i, d and z respectively.
- For notational simplicity, a general expression for $Y_i(d,z)$ is

$$Y_i(d,z) = \beta_i d + \gamma_i z + u_i,$$

where β_i and γ_i can be non-constant.

Assumptions

1. **(Relevance)** IV should be predictive of D_i. $Cov(Z_i, D_i) \neq 0$.
   - In the language of potential treatments, $D_i(1) \neq D_i(0)$.

2. **(Independence)** Individuals with higher/lower potential outcomes/treatments do not face systematically different values of Z_i. $Z_i \perp\!\!\!\perp Y_i(d,z)$ for all (d,z) and $Z_i \perp\!\!\!\perp D_i(z)$ for all z.
   - Given $Y_i(d,z) = \beta_i d + \gamma_i z + u_i$, this means $Z_i \perp\!\!\!\perp \beta_i, \gamma_i, u_i, D_i(1), D_i(0)$.

3. **(Exclusion Restriction)** Z_i *only* affects Y_i through X_i.
   - Given $Y_i(d,z) = \beta_i d + \gamma_i z + u_i$, this means $\gamma_i = 0$.
   - $Y_i(d,z)$ is not a function of z and reduces to $Y_i(d) = \beta_i d + u_i$.

4. *[New]* **(Monotonicity)** There are no defiers, i.e., $D_i(1) \geq D_i(0)$ for all i.
   - IV shifts treatment status in a single direction.
   - This is untestable because we couldn't observe both $D_i(1)$ and $D_i(0)$. It needs to hold for all individuals, but we may look at the signs of first-stage effects by subgroup.

LATE Theorem

- **Theorem:** If the four assumptions hold, then

$$\beta_{\text{IV}} = E(\beta_i \mid i \in \text{Compliers}) = E[\beta_i \mid D_i(1) = 1, D_i(0) = 0] \equiv LATE.$$

(This result won Joshua Angrist and Guido Imbens the 2021 Nobel Prize in Economics!)

- *Proof:* (You won't be responsible for this.)

$$\beta_{\text{IV}} = \frac{\rho}{\pi} = \frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(D_i \mid Z_i = 1) - E(D_i \mid Z_i = 0)}.$$

First-stage effect π:

$$\begin{aligned}
\pi &\overset{\text{A2}}{=} E[D_i(1) - D_i(0)] \\
&= 1 \times \Pr[D_i(1) > D_i(0)] + (-1) \times \Pr[D_i(1) > D_i(0)] \\
&= \Pr(\text{Compliers}) - \Pr(\text{Defiers}) \\
&\overset{\text{A4}}{=} \Pr(\text{Compliers}).
\end{aligned}$$

LATE Theorem Proof

- Reduced-form (ITT) effect ρ:

$$
\begin{aligned}
\rho &= E[\beta_i D_i(1) + u_i \mid Z_i = 1] - E[\beta_i D_i(0) + u_i \mid Z_i = 0] \\
&\overset{\text{A2}}{=} E[\beta_i (D_i(1) - D_i(0))] \\
&= \Pr[D_i(1) > D_i(0)] \times E[\beta_i \mid D_i(1) > D_i(0)] + \Pr[D_i(1) < D_i(0)] \times E[-\beta_i \mid D_i(1) < D_i(0)] \\
&\overset{\text{A4}}{=} \Pr[D_i(1) > D_i(0)] \times E[\beta_i \mid D_i(1) > D_i(0)] \\
&= \Pr(\text{Compliers}) \times E[\beta_i \mid \text{Compliers}].
\end{aligned}
$$

- Taken together,

$$
\beta_{\text{IV}} = \frac{\rho}{\pi} = E[\beta_i \mid \text{Compliers}].
$$

Is LATE more useful than ITT?

- Reduced-form (ITT) is always causal if independence holds.
- But to get LATE, we further need exclusion and monotonicity to hold.
- Why ever do IV? IV answers more interesting questions than ITT.
  - Wage effects of education vs. of winning a scholarship lottery.
- And hopefully it's more externally valid...
  - Important to think about the scope of what IV can speak to...