Precept 11: Unmeasured Confounding

Soc 400: Applied Social Statistics

Ziyao Tian

Princeton University

December 6, 2018

Acknowledgments: Materials are drawn from Shay O'Brien, Simone Zhang, Ryan Parsons, Xinyi Duan, Ian Lundberg, and Jeremy Cohen. Material also comes from Matt Blackwell, Adam Glynn, Jens Hainmueller and Brandon Stewart' lectures.

Today's Agenda

- Expectation, Variance, Covariance
- Instrumental Variable
  - why IV
  - assumptions
  - estimation with RStudio

Expectation

Expectation

The expected value of a random variable X is denoted by $E[X]$ and is a measure of **central tendency** of X. Roughly speaking, an expected value is like a weighted average of all of the values weighted by probability of occurrence.

The expected value of a discrete random variable X is defined as

$$E[X] = \sum_{\text{all } x} x \cdot f_X(x).$$

Expectation

The expected value of a random variable X is denoted by $E[X]$ and is a measure of **central tendency** of X. Roughly speaking, an expected value is like a weighted average of all of the values weighted by probability of occurrence.

The expected value of a discrete random variable X is defined as

$$E[X] = \sum_{\text{all } x} x \cdot f_X(x).$$

The expected value of a continuous random variable X is defined as

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

Properties of Expected Value

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X+Y] = E[X] + E[Y]$$

Properties of Expected Value

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

$$E[aX + b] = aE[X] + b$$

Properties of Expected Value

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

$$E[aX + b] = aE[X] + b$$

3) LOTUS: Law of the Unconscious Statistician

$$E[g(X)] = \sum_{x} g(x) f_X(x)$$

Properties of Expected Value

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X+Y]=E[X]+E[Y]$$

2) Homogeneity: expected value of a constant is the constant

$$E[aX+b]=aE[X]+b$$

3) LOTUS: Law of the Unconscious Statistician

$$E[g(X)]=\sum_x g(x)f_X(x)$$

However,

- $E[g(X)]\neq g(E[X])$ unless $g(\cdot)$ is a linear function
- $E[XY]\neq E[X]E[Y]$ unless X and Y are independent

Variance

The variance of a random variable X (a measure of its dispersion) is the expectation of the squared distances from the mean
$V[X] = E[(X - E[X])^2]$.
For a discrete random variable X

$$V[X] = \sum_{\text{all } x} (x - E[X])^2 f_X(x)$$

Variance

The variance of a random variable X (a measure of its dispersion) is the expectation of the squared distances from the mean $V[X] = E[(X - E[X])^2]$.
For a discrete random variable X

$$V[X] = \sum_{\text{all } x} (x - E[X])^2 f_X(x)$$

For a continuous random variable X

$$V[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

Variance

Another common form

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= E[X^2 - 2E[X]X + E[X]^2] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Properties of Variance

Property 1 of Variance: Behavior with Constants. Suppose a and b are constants and X is a random variable. Then

$$V[b] = 0$$

$$V[aX] = a^2 V[X]$$

$$V[aX + b] = a^2 V[X] + 0$$

Properties of Variance

Property 1 of Variance: Behavior with Constants. Suppose a and b are constants and X is a random variable. Then

$$V[b] = 0$$

$$V[aX] = a^2 V[X]$$

$$V[aX + b] = a^2 V[X] + 0$$

Property 2 of Variance: Additivity for Independent Random Variables. Suppose we have k independent random variables $X_1, \ldots, X_k$. If $V[X_i]$ exists for all $i = 1, \ldots, k$, then

$$V\left[\sum_{i=1}^{k} X_i\right] = V[X_1] + \cdots + V[X_k]$$

Exercise

Let i.i.d. $X_1, \ldots, X_N$ be our population. Then the population mean is the following

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

If $Var(X_1) = \sigma^2$, what is $Var[\bar{X}]$?

Exercise

Let i.i.d. $X_1, \ldots, X_N$ be our population. Then the population mean is the following

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

If $Var(X_1) = \sigma^2$, what is $Var[\bar{X}]$?

$$Var[\bar{X}] = \frac{\sigma^2}{N}$$

Wait, when do we divide by sqrt(n)?

Let X_1, X_2, ... , X_n be a random sample of size n from a distribution (population) with mean μ and variance σ^2

Estimand	**Estimator**	**Sampling Dist.**
Population Mean μ	Sample Mean $\bar{X}$	$\bar{X} \overset{\text{approx.}}{\sim} N(\mu, \frac{\sigma^2}{n})$
Population Variance σ^2	$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$	$E[S^2] = \sigma^2; S^2 \xrightarrow{n \to \infty} \sigma^2$
$SE[\bar{X}]$	$\widehat{SE}[\bar{X}] = \sqrt{\frac{S^2}{n}}$	

In this case, we adjust a population variance estimator S^2 by sqrt(n) to estimate the SE of "sample mean", which is also an estimator itself, for the population mean

Covariance

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the **linear association** between two random variables

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the **linear association** between two random variables .
- If $Cov[X, Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the **linear association** between two random variables .
- If $Cov[X, Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.
- Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$?

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the **linear association** between two random variables .
- If $Cov[X, Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.
- Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$?
- Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$?

Covariance

Definition

The **covariance** of X and Y is defined as:

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the **linear association** between two random variables .
- If $Cov[X,Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.
- Does $X \perp\!\!\!\perp Y$ imply $Cov[X,Y] = 0$?
- Does $Cov[X,Y] = 0$ imply $X \perp\!\!\!\perp Y$?
- $X \perp\!\!\!\perp Y \implies Cov[X,Y] = 0$**, but not vice versa.**

Important Identities for Variances and Covariances

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$

$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$

$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

3. Furthermore, if X and Y are independent,

$$V[X \pm Y] = V[X] + V[Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$

$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

3. Furthermore, if X and Y are independent,

$$V[X \pm Y] = V[X] + V[Y]$$

Exercise: Comparing Two Groups

Exercise: Comparing Two Groups

We will often assume the following when comparing two groups,

Exercise: Comparing Two Groups

We will often assume the following when comparing two groups,

- $X_{11}, X_{12}, ..., X_{1n_1} \sim_{i.i.d.} ?(\mu_1, \sigma_1^2)$
- $X_{21}, X_{22}, ..., X_{2n_2} \sim_{i.i.d.} ?(\mu_2, \sigma_2^2)$
- The two samples are independent of each other.

Exercise: Comparing Two Groups

We will often assume the following when comparing two groups,

- $X_{11}, X_{12}, ..., X_{1n_1} \sim_{i.i.d.} ?(\mu_1, \sigma_1^2)$
- $X_{21}, X_{22}, ..., X_{2n_2} \sim_{i.i.d.} ?(\mu_2, \sigma_2^2)$
- The two samples are independent of each other.

We will usually be interested in comparing μ_1 to μ_2, although we will sometimes need to compare σ_1^2 to σ_2^2 in order to make the first comparison.

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

$$Var[\overline{X}_1 - \overline{X}_2]$$

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

$$Var[\overline{X}_1 - \overline{X}_2] \quad = \quad Var[\overline{X}_1] + Var[\overline{X}_2]$$

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

$$
\begin{aligned}
Var[\overline{X}_1 - \overline{X}_2] &= Var[\overline{X}_1] + Var[\overline{X}_2] \\
&= \frac{1}{n_1^2}\sum Var[X_{1i}] + \frac{1}{n_2^2}\sum Var[X_{2j}]
\end{aligned}
$$

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

$$
\begin{aligned}
Var[\overline{X}_1 - \overline{X}_2] &= Var[\overline{X}_1] + Var[\overline{X}_2] \\
&= \frac{1}{n_1^2} \sum Var[X_{1i}] + \frac{1}{n_2^2} \sum Var[X_{2j}] \\
&= \frac{1}{n_1^2} \sum \sigma_1^2 + \frac{1}{n_2^2} \sum \sigma_2^2
\end{aligned}
$$

Sampling Distribution for $\overline{X}_1 - \overline{X}_2$

What is the variance of $\overline{X}_1 - \overline{X}_2$?

$$
\begin{aligned}
Var[\overline{X}_1 - \overline{X}_2] &= Var[\overline{X}_1] + Var[\overline{X}_2] \\
&= \frac{1}{n_1^2} \sum Var[X_{1i}] + \frac{1}{n_2^2} \sum Var[X_{2j}] \\
&= \frac{1}{n_1^2} \sum \sigma_1^2 + \frac{1}{n_2^2} \sum \sigma_2^2 \\
&= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}
\end{aligned}
$$

Correlation

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

The **correlation** between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

The **correlation** between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

- $Cor[X, Y]$ is a standardized measure of linear association between X and Y.

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

The **correlation** between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

- $Cor[X, Y]$ is a standardized measure of linear association between X and Y.
- Always satisfies: $-1 \leq Cor[X, Y] \leq 1$.

Equations for simple OLS $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\widehat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\
&= \frac{\frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}}{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}} \\
&= \frac{E[(X - \overline{X})(Y - \overline{Y})]}{E[(X - \overline{X})^2]} \\
&= \frac{Cov[X, Y]}{Var[X]}
\end{aligned}$$

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- What's the intuition here?

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- What's the intuition here?
- "Numerator" $\mathbf{X}'\mathbf{y}$: is approximately composed of the covariances between the columns of $\mathbf{X}$ and $\mathbf{y}$

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- What's the intuition here?
- "Numerator" $\mathbf{X}'\mathbf{y}$: is approximately composed of the covariances between the columns of $\mathbf{X}$ and $\mathbf{y}$
- "Denominator" $\mathbf{X}'\mathbf{X}$ is approximately composed of the sample variances and covariances of variables within $\mathbf{X}$

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- What's the intuition here?
- "Numerator" $\mathbf{X}'\mathbf{y}$: is approximately composed of the covariances between the columns of $\mathbf{X}$ and $\mathbf{y}$
- "Denominator" $\mathbf{X}'\mathbf{X}$ is approximately composed of the sample variances and covariances of variables within $\mathbf{X}$
- Thus, we have something like:

$$\widehat{\boldsymbol{\beta}} \approx (\text{variance of } \mathbf{X})^{-1}(\text{covariance of } \mathbf{X} \text{ \& } \mathbf{y})$$

i.e. analogous to the simple linear regression case!

Intuition for the OLS in Matrix Form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- What's the intuition here?
- "Numerator" $\mathbf{X}'\mathbf{y}$: is approximately composed of the covariances between the columns of $\mathbf{X}$ and $\mathbf{y}$
- "Denominator" $\mathbf{X}'\mathbf{X}$ is approximately composed of the sample variances and covariances of variables within $\mathbf{X}$
- Thus, we have something like:

$$\widehat{\boldsymbol{\beta}} \approx (\text{variance of } \mathbf{X})^{-1}(\text{covariance of } \mathbf{X} \text{ \& } \mathbf{y})$$

i.e. analogous to the simple linear regression case!

Disclaimer: the final equation is exactly true for all non-intercept coefficients if you remove the intercept from $\mathbf{X}$ such that $\hat{\boldsymbol{\beta}}_{-0} = \text{Var}(\mathbf{X}_{-0})^{-1}\text{Cov}(\mathbf{X}_{-0}, \mathbf{y})$. The numerator and denominator are the variances and covariances if $\mathbf{X}$ and $\mathbf{y}$ are demeaned and normalized by the sample size minus 1.

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\left(Y_i(0) \middle| D_i = 1, X_i\right)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\left(Y_i(0) \middle| D_i = 0, X_i\right)}_{\text{observed}}$$

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\left(Y_i(0) \middle| D_i = 1, X_i\right)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\left(Y_i(0) \middle| D_i = 0, X_i\right)}_{\text{observed}}$$

  - **Not testable** without the counterfactual data, which is missing by definition!

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\big(Y_i(0)\big|D_i = 1, X_i\big)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\big(Y_i(0)\big|D_i = 0, X_i\big)}_{\text{observed}}$$

  - **Not testable** without the counterfactual data, which is missing by definition!
  - With backdoor criterion, you must have the correct DAG.

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\big(Y_i(0)\big|D_i = 1, X_i\big)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\big(Y_i(0)\big|D_i = 0, X_i\big)}_{\text{observed}}$$

  - **Not testable** without the counterfactual data, which is missing by definition!
  - With backdoor criterion, you must have the correct DAG.
- Approaches to unobserved confounding

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\left(Y_i(0) \middle| D_i = 1, X_i\right)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\left(Y_i(0) \middle| D_i = 0, X_i\right)}_{\text{observed}}$$

  - **Not testable** without the counterfactual data, which is missing by definition!
  - With backdoor criterion, you must have the correct DAG.
- Approaches to unobserved confounding
  - natural experiments, IV, RD, etc.

IV in the big picture of causal inference

- Experimental ideal: unconfoundedness / "randomization"
- Selection on observables:
  - Conditioning of observed covariates X, treatment assignment D is independent of the outcome

$$\underbrace{\left(Y_i(0) \middle| D_i = 1, X_i\right)}_{\text{unobserved}} \overset{dist.}{=} \underbrace{\left(Y_i(0) \middle| D_i = 0, X_i\right)}_{\text{observed}}$$

  - **Not testable** without the counterfactual data, which is missing by definition!
  - With backdoor criterion, you must have the correct DAG.
- Approaches to unobserved confounding
  - natural experiments, IV, RD, etc.
  - goal: find plausibly exogenous variation in treatment assignment

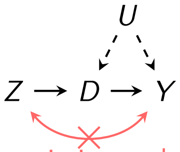
Review of Key Assumptions

1. Exogeneity of the instrument
2. Exclusion restriction
3. First-stage relationship / Relevance
4. Monotonicity

Assumptions in Graphical Model

$$
\begin{tikzcd}
 & & U \arrow[dl, dashed] \arrow[dr, dashed] & \\
Z \arrow[r] & D \arrow[rr] & & Y
\end{tikzcd}
$$

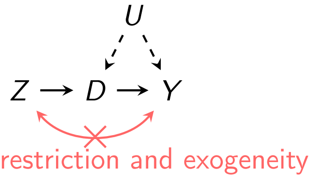
Assumptions in Graphical Model



exclusion restriction and exogeneity

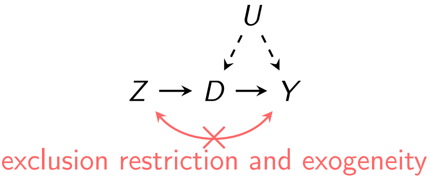
1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)

Assumptions in Graphical Model



1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)
3) Z affects D (first stage relationship / relevance)

Assumptions in Graphical Model



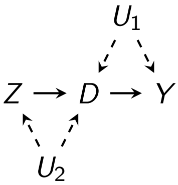
1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)
3) Z affects D (first stage relationship / relevance)
4) to allow for heterogenous treatment effect, we assume the presence of the instrument never dissuades someone from taking the treatment (monotonicity)

$$D_i(1) - D_i(0) \geq 0$$

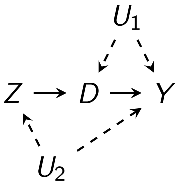
Which violation?

$$
\begin{tikzcd}
& U \arrow[dl, dashed] \arrow[d, dashed] \arrow[dr, dashed] & \\
Z \arrow[r] & D \arrow[r] & Y
\end{tikzcd}
$$

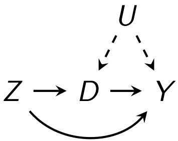
Which violation?



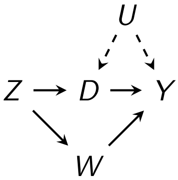
Which violation?



Which violation?



Which violation?



Remember the four subgroups of an experiment?

Name	$D_i(1)$	$D_i(0)$
Always Takers	?	?
Never Takers	?	?
Compliers	?	?
Defiers	?	?

Remember the four subgroups of an experiment?

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

A visual example

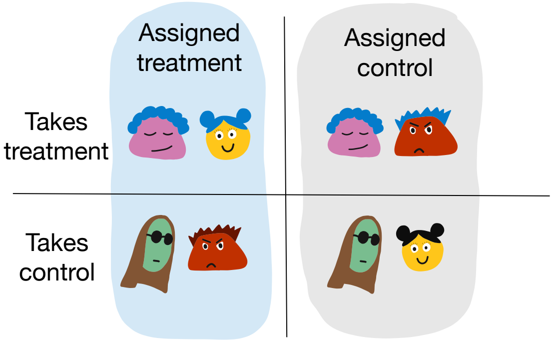
Assumption 4: Monotonicity

- To allow for heterogenous effects we need to make a new assumption about the relationship between the instrument and the treatment.
- **Monotonicity** says that the presence of the instrument **never dissuades** someone from taking the treatment:

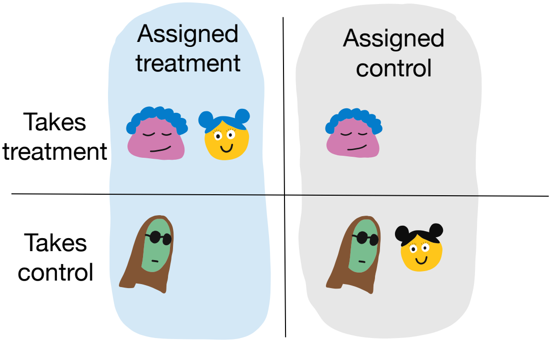
$$D_i(1) - D_i(0) \geq 0$$

- We sometimes call assumption 4 **no defiers** because the monotonicity assumption rules out the existence of defiers.
- This is not a testable assumption because no observed group is solely populated by defiers.

Monotonicity assumption



Monotonicity assumption



- If we assume there are no defiers, we can better identify the subgroups.
- Anyone with $D_i = 1$ when $Z_i = 0$ must be an **always-taker** and anyone with $D_i = 0$ when $Z_i = 1$ must be a **never-taker**.

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.
- This is the ATE among the compliers: those that take the treatment when encouraged to do so.
- That is, the LATE theorem (proof coming soon), states that:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.
- This is the ATE among the compliers: those that take the treatment when encouraged to do so.
- That is, the LATE theorem (proof coming soon), states that:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

- But doesn't $D_i(1) \geq D_i(0)$ include always-takers and never-takers?

- $\tau_{LATE} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ as a pooled effect of three groups.
- Let's focus on the numerator first.

Name	$D(1)-D(0)$	$Y(D(1))-Y(D(0))$	Fract of Pop	Avg Effect
Compliers	$1-0=1$	$Y(1)-Y(0)$	$\pi_{compliers}$	$\delta_{compliers}$
Always Takers	$1-1=0$	$Y(1)-Y(1)=0$	π_{always}	0
Never Takers	$0-0=0$	$Y(0)-Y(0)=0$	π_{never}	0
Defiers	$0-1=-1$	$Y(0)-Y(1)$	0	NA

- $\tau_{LATE} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ as a pooled effect of three groups.
- Let's focus on the numerator first.

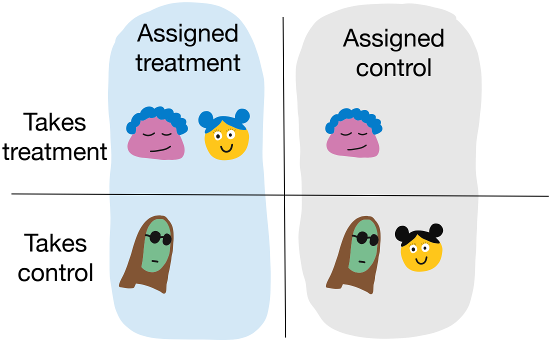
Name	$D(1) - D(0)$	$Y(D(1)) - Y(D(0))$	Fract of Pop	Avg Effect
Compliers	$1 - 0 = 1$	$Y(1) - Y(0)$	$\pi_{compliers}$	$\delta_{compliers}$
Always Takers	$1 - 1 = 0$	$Y(1) - Y(1) = 0$	π_{always}	0
Never Takers	$0 - 0 = 0$	$Y(0) - Y(0) = 0$	π_{never}	0
Defiers	$0 - 1 = -1$	$Y(0) - Y(1)$	0	NA

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

$$= E[Y(D(1)) - Y(D(0))]$$

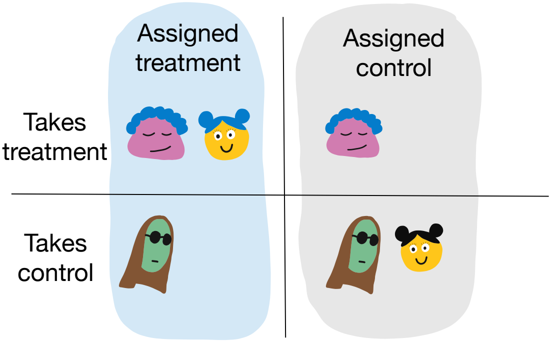
$$= \delta_{compliers} * \pi_{compliers} + 0 * \pi_{always} + 0 * \pi_{never}$$

How do we know the fraction of compliers?



- Assuming no defiers, $\pi_{compliers} = 1 - \pi_{always} - \pi_{never}$

How do we know the fraction of compliers?



- Assuming no defiers, $\pi_{compliers} = 1 - \pi_{always} - \pi_{never}$
- But, do we need to calculate the proportion of compliers?

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

$$\begin{aligned} E[Y_i|Z_i=1] &= E[Y_i(0)+(Y_i(1)-Y_i(0))D_i|Z_i=1] \\ &= E[Y_i(0)+(Y_i(1)-Y_i(0))D_i(1)] \qquad \text{(randomization)} \end{aligned}$$

- The same applies to when $Z_i = 0$, so we have

$$E[Y_i|Z_i=0] = E[Y_i(0)+(Y_i(1)-Y_i(0))D_i(0)]$$

- Thus, $E[Y_i|Z_i=1] - E[Y_i|Z_i=0] =$

$$\begin{aligned} & E[(Y_i(1)-Y_i(0))(D_i(1)-D_i(0))] \\ =& E[(Y_i(1)-Y_i(0))(1)|D_i(1)>D_i(0)]\Pr[D_i(1)>D_i(0)] \\ +& E[(Y_i(1)-Y_i(0))(-1)|D_i(1)<D_i(0)]\Pr[D_i(1)<D_i(0)] \\ =& E[Y_i(1)-Y_i(0)|D_i(1)>D_i(0)]\Pr[D_i(1)>D_i(0)] \end{aligned}$$

- The third equality comes from monotonicity: with this assumption, $D_i(1) < D_i(0)$ never occurs.

Proof (continued)

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)]$$

- We can use the same argument for the denominator:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = E[D_i(1) - D_i(0)]$$
$$= E[D_i(1) - D_i(0)] \Pr[D_i(1) > D_i(0)]$$
$$= 1 * \Pr[D_i(1) > D_i(0)]$$
$$= \Pr[D_i(1) > D_i(0)]$$

- Dividing these two expressions through gives the LATE.

Proof (continued)

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]\,\Pr[D_i(1) > D_i(0)]$$

- We can use the same argument for the denominator:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = E[D_i(1) - D_i(0)]$$

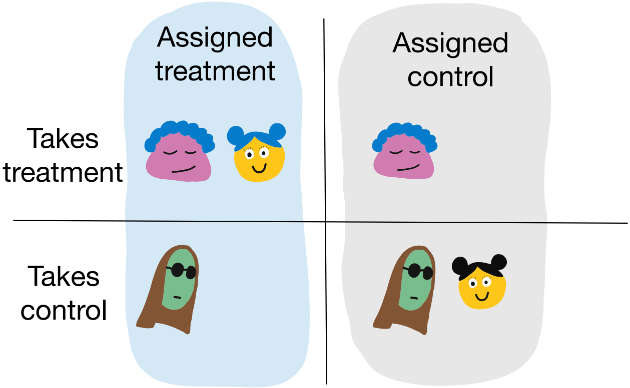
$$= E[D_i(1) - D_i(0)]\,\Pr[D_i(1) > D_i(0)]$$

$$= 1 * \Pr[D_i(1) > D_i(0)]$$

$$= \Pr[D_i(1) > D_i(0)]$$

- Dividing these two expressions through gives the LATE.
- The denominator tells us the fraction of compliers, but we don't need to calculate it to get LATE.

Monotonicity assumption



All of this allows us to interpret the LATE we identify using the instrumental variable as the average treatment effect among compliers.

IV assumptions

1) **Exogeneity**: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
2) **Exclusion**: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
3) **Relevance**: $Z \not\perp\!\!\!\perp D$

- all possibly conditional on X

IV assumptions

1) **Exogeneity**: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
2) **Exclusion**: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
3) **Relevance**: $Z \not\perp\!\!\!\perp D$

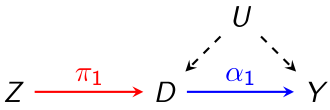
- all possibly conditional on X

4) **Monotonicity**:

$$D_i(1) - D_i(0) \geq 0$$

Assumed Models

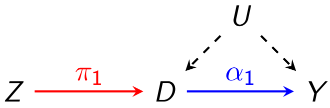
- Second Stage:
  $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage:
  $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions:
  $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and
  $Cov[u_2, Z] = 0$



IV Estimation

With our assumed model being true,

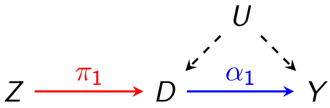
- regressing D on Z identifies π_1



IV Estimation

With our assumed model being true,

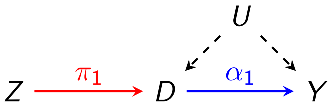
- regressing D on Z identifies π_1
- regressing Y on Z identifies $\pi_1 \cdot \alpha_1$



IV Estimation

With our assumed model being true,

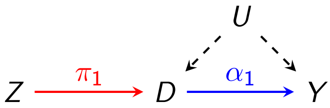
- regressing D on Z identifies π_1
- regressing Y on Z identifies $\pi_1 \cdot \alpha_1$
- $\frac{\widehat{\pi_1 \cdot \alpha_1}}{\widehat{\pi_1}}$ identifies $\frac{\pi_1 \cdot \alpha_1}{\pi_1} = \alpha_1$



IV Estimation

OLS vs. IV estimators

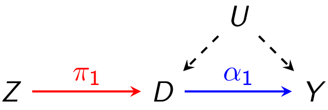
- OLS to estimate α



IV Estimation

OLS vs. IV estimators

- OLS to estimate α
- IV estimators
  - Wald estimator
  - TSLS estimator (or 2SLS)



Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$
- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} =$$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$
- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$
- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \alpha_1 + \frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

Use OLS to estimate α_1

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the **treatment variable** (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$
- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \alpha_1 + \frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

$$E[\hat{\alpha}_{1,OLS}] = \alpha_1 + E[\frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]}]$$

so bias depends on correlation between u_2 and D

Instrumental Variable Effect: Wald Estimator

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$Y = (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2)$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\alpha_1 = \frac{\gamma_1}{\pi_1}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} =$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$
\begin{aligned}
Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\
Y &= \gamma_0 + \gamma_1 Z + u_3
\end{aligned}
$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$
\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]}
$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2, Z]}{Cov[D,Z]} = \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2,Z]}{Cov[D,Z]} = \alpha_1 + \frac{Cov[u_2,Z]}{Cov[D,Z]} \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2,Z]}{Cov[D,Z]} = \alpha_1 + \frac{Cov[u_2,Z]}{Cov[D,Z]} \end{aligned}$$

$$E[\hat{\alpha}_1] =$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2,Z]}{Cov[D,Z]} = \alpha_1 + \frac{Cov[u_2,Z]}{Cov[D,Z]} \\ E[\hat{\alpha}_1] &= \alpha_1 + E\left[\frac{\widehat{Cov}[u_2,Z]}{\widehat{Cov}[D,Z]}\right] \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2,Z]}{Cov[D,Z]} = \alpha_1 + \frac{Cov[u_2,Z]}{Cov[D,Z]} \\ E[\hat{\alpha}_1] &= \alpha_1 + E\left[\frac{\widehat{Cov}[u_2,Z]}{\widehat{Cov}[D,Z]}\right] \end{aligned}$$

$\hat{\alpha}_1$ is consistent if $Cov[u_2, Z] = 0$ but has a **bias** which decreases with instrument strength.

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\ Y &= \alpha_0 + \alpha_1 (\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \\ Y &= (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1 (\hat{\pi}_1 Z) + u_2 \end{aligned}$$

- Intuition: Retain only variation in D that is induced by Z, "purged" of endogeneity

Quarter of Birth Example

Angrist, Joshua and Alan Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4).

Research Question

- Question: What is the causal effect of education on earnings?
- What are possible confounders? In what direction might those confounders bias our results?

$$\begin{tikzcd} & T \arrow[r] & Y \\ U \arrow[ur] \arrow[urr, bend right] & & \end{tikzcd}$$

"The Natural Experiment"

"The experiment stems from the fact that children born in different months of the year start school at different ages, while compulsory schooling laws generally require students to remain in schools until their sixteenth or seventeenth birthday. 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. In effect, the interaction of school-entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months."

"The Natural Experiment"

"The experiment stems from the fact that children born in different months of the year start school at different ages, while compulsory schooling laws generally require students to remain in schools until their sixteenth or seventeenth birthday. 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. In effect, the interaction of school-entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months."

- Data: Men from the 1980 Census Public Use Sample
- Note: We apply the term "natural experiment" to indicate that the "treatment" (which is instrument here) is randomized but the randomization was not controlled by the researcher

What are the key variables?

- **What's the instrument (Z)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education
- **What's the outcome (Y)?**

What are the key variables?

- **What's the instrument (Z)?**
  Quarter of birth
- **What's the treatment (T)?**
  Receiving additional education
- **What's the outcome (Y)?**
  Earnings

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**
   Does birth quarter induce variation in time spent in school?
4. **Monotonicity**

Let's evaluate the assumptions

1. **Exogeneity of the Instrument**
   Is birth quarter random?
2. **Exclusion Restriction**
   Can birth quarter affect earnings through causal channels other than education?
3. **First-stage relationship**
   Does birth quarter induce variation in time spent in school?
4. **Monotonicity**
   Are there defiers?

IV Assumption Check - First Stage Relationship

We can check by regressing treatment on the instrument. We can also gain more confidence by examining plots of the relationship:

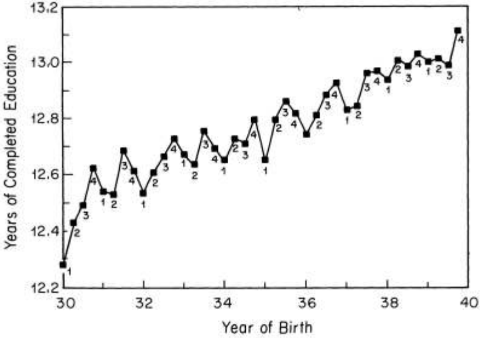


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Men born earlier in the year have less schooling

Reduced Form

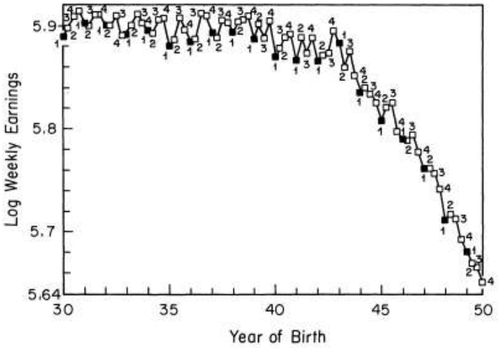


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Differences in schooling due to quarter of birth appear to translate into different earnings.

2SLS Results: White Men, 1930s Cohorts

TABLE VII

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673	0.0928	0.0673	0.0907	0.0628	0.0831	0.0628	0.0811
	(0.0003)	(0.0093)	(0.0003)	(0.0107)	(0.0003)	(0.0095)	(0.0003)	(0.0109)
Race (1 = black)	—	—	—	—	−0.2547	−0.2333	−0.2547	−0.2354
					(0.0043)	(0.0109)	(0.0043)	(0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705	0.1511	0.1705	0.1531
					(0.0029)	(0.0095)	(0.0029)	(0.0107)
Married (1 = married)	—	—	—	—	0.2487	0.2435	0.2487	0.2441
					(0.0032)	(0.0040)	(0.0032)	(0.0042)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0757	−0.0880	—	—	−0.0778	−0.0876
			(0.0617)	(0.0624)			(0.0603)	(0.0609)
Age-squared	—	—	0.0008	0.0009	—	—	0.0008	0.0009
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	163 [179]	—	161 [177]	—	164 [179]	—	162 [177]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 150 quarter-of-birth times state-of-birth interactions. Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is the same as in Table VI. Sample size is 329,509.

2SLS Results: Black Men, 1930s Cohorts

TABLE VIII

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR BLACK MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0672 (0.0013)	0.0635 (0.0185)	0.0671 (0.0003)	0.0555 (0.0199)	0.0576 (0.0013)	0.0461 (0.0187)	0.0576 (0.0013)	0.0391 (0.0199)
SMSA (1 = center city)	—	—	—	—	0.1885 (0.0142)	0.2053 (0.0308)	0.1884 (0.0142)	0.2155 (0.0324)
Married (1 = married)	—	—	—	—	0.2216 (0.0193)	0.2272 (0.0136)	0.2216 (0.0100)	0.2307 (0.0140)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
49 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0309 (0.2538)	−0.3274 (0.2560)	—	—	−0.2978 (0.0032)	−0.3237 (0.2497)
Age-squared	—	—	0.0033 (0.0028)	0.0035 (0.0028)	—	—	0.0032 (0.0027)	0.0035 (0.0028)
χ^2 [dof]	—	184 [176]	—	181 [173]	—	178 [176]	—	175 [173]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 147 quarter-of-birth times state-of-birth interactions. (There are no black men in the sample born in Hawaii.) Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is drawn from the 1980 Census. Sample size is 26,913.

Note the returns for black men appear to be smaller

Study Results

IV. Conclusion

Differences in season of birth create a natural experiment that we use to study the effect of compulsory school attendance on schooling and earnings. Because individuals born in the beginning of the year usually start school at an older age than that of their classmates, they are allowed to drop out of school after attaining less education. Our exploration of the relationship between quarter of birth and educational attainment suggests that season of birth has a small effect on the level of education men ultimately attain. To support the contention that this is a consequence of compulsory schooling laws, we have assembled evidence showing that some students leave school as soon as they attain the legal dropout age, and that season of birth has no effect on postsecondary years of schooling.

Questions?