Instrumental Variables

Gov 51: Data Analysis and Politics

Scott Cunningham

Harvard University

Week 11
April 9, 2026

Part I: The Problem
OLS Fails When Treatment Is Endogenous

Selection bias makes OLS unreliable without randomization

In an experiment:

- Randomization ensures $D_i \perp Y_{0i}$
- Treated and control groups have equal baseline outcomes
- Difference in means = ATE

Randomization kills selection bias

In the world:

- We cannot randomize institutions, education, or conflict
- Who gets treatment is not random
- Those who select into $D = 1$ differ from $D = 0$ in ways that also affect Y

Observational data: selection bias contaminates every comparison

Difference in means = ATT + selection bias

Start from what we observe: $E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$

$$\underbrace{E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]}_{\text{What we observe}} = \underbrace{E[Y_{1i} - Y_{0i} \mid D_i = 1]}_{\text{ATT (causal)}} + \underbrace{E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]}_{\text{Selection bias}}$$

- **ATT:** average treatment effect for those who were treated
- **Selection bias:** the treated group would have had different baseline outcomes even without treatment
- If selection bias = 0 (randomization!), DiM = ATT $\approx$ ATE

OLS bias and selection bias are the same problem in different notation

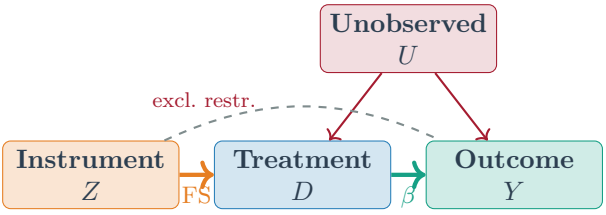
$$\hat{\beta}_{OLS} \xrightarrow{p} \beta + \frac{\text{Cov}(D_i,\, \varepsilon_i)}{\text{Var}(D_i)}$$

- ε_i: everything driving Y other than D
- If those things also drive D (ability, wealth, geography…), $\text{Cov}(D_i, \varepsilon_i) \neq 0$
- Regression model: $Y_i = \alpha + \beta D_i + \varepsilon_i$

$$E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = \frac{\text{Cov}(D_i, \varepsilon_i)}{\text{Var}(D_i)}$$

Selection bias and OVB: two names, one problem.

IV uses a third variable to break the endogeneity



Three conditions:

1. **Relevance:** Z shifts D
2. **Exclusion:** $Z \to Y$ only via D
3. **Independence:** $Z \perp U$

Only (1) testable.
(2)&(3): theory required.

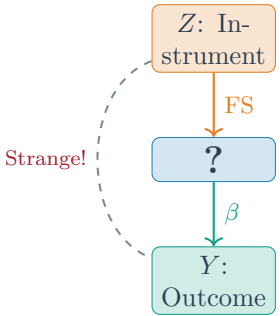
Part II: Good Instruments Are a Bit Strange

A valid instrument looks bizarre until you know the treatment story

The strangeness principle:

- A good instrument's reduced-form correlation with Y makes no sense without knowing D
- *Without* D: correlation seems random
- *With* D: it becomes inevitable

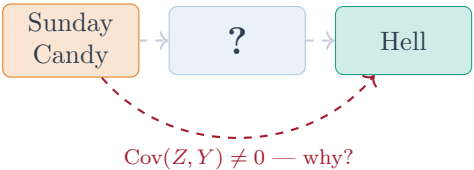
Strangeness $\Leftrightarrow$ Exclusion:
the only path from Z
to Y runs through D



How can a song affect the afterlife?

Chance the Rapper, "Ultralight Beam":

"I made Sunday Candy,
I'm never going to hell"



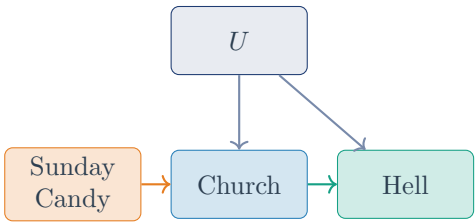
- Why would a song predict going to hell?

Sunday Candy is a good instrument for going to church

What if Chance meant:

- Made Sunday Candy → pastor heard it
- Pastor invited him to church; he went
- Went to church → won't go to hell
- Don't believe the spiritual story? Fine — the *logic* still works

Good instruments **feel strange**:
$\text{Cov}(Z, Y) \neq 0$ only
makes sense through D

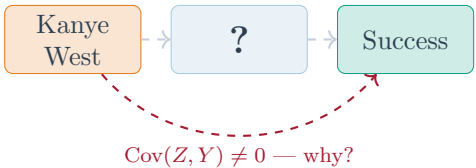


Meeting Kanye West predicts success — but why?

Chance the Rapper continues:

"I met Kanye West,
I'm never going to fail"

- Why would meeting Kanye predict success?

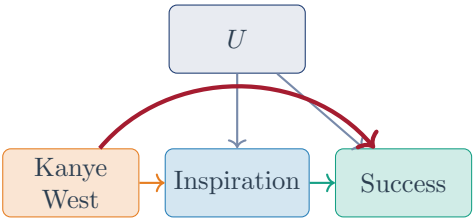


Kanye is not strange enough to be an instrument

Kanye directly launches careers:

- Meeting Kanye → inspiration: plausible
- Meeting Kanye → success **directly**: also plausible
- **Not strange** — you already know why $\text{Cov}(Z, Y) \neq 0$
- Kanye opens doors independent of D

Kanye West → Success directly: exclusion restriction violated



Part III: Acemoglu, Johnson & Robinson (2001)
"Why Nations Fail" — with Data

The central question: do institutions cause prosperity?

The hypothesis:

- Countries with secure property rights, rule of law, and protection against expropriation grow richer
- This is about *economic institutions*, not culture or geography

The problem:

- Rich countries can *afford* better institutions
- Reverse causality: wealth → institutions
- Geography, culture, disease burden all confound

2024 Nobel Prize

Daron Acemoglu
Simon Johnson
James A. Robinson

"for studies of how institutions are formed and affect prosperity"

Colonialism created two very different kinds of institutions

Settler colonies (low mortality)

- North America, Australia, New Zealand, southern Latin America
- Europeans could survive → came in large numbers
- Built inclusive institutions: property rights, rule of law, representative government
- Needed to protect their own assets

Result: high income today

Extractive colonies (high mortality)

- Much of sub-Saharan Africa, parts of South Asia
- Europeans died rapidly → few settled permanently
- Built extractive institutions: resource extraction, forced labor, no property rights for locals
- Goal was to transfer wealth to Europe

Result: lower income today

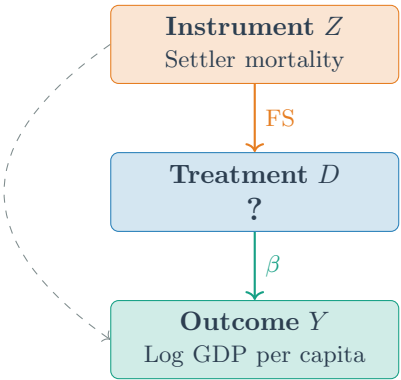
Settler mortality passes the strangeness test — it only matters through institutions

The IV strategy:

1. Settler mortality rates from military records (1817–1848)
2. Higher mortality → fewer settlers → extractive institutions
3. Instrument: log(settler mortality)

Variables:

- ▷ `logem4`: log settler mortality
- ▷ `avexpr`: avg. expropriation protection (0–10)
- ▷ `logpgp95`: log GDP per capita, 1995



Part IV: What Makes an Instrument Valid

A valid instrument must be relevant, excluded, and independent

1. Relevance

$$\text{Cov}(Z_i, D_i) \neq 0$$

Testable: check first-stage F. $F < 10$: weak instrument.

2. Exclusion

$$Z_i \not\to Y_i \text{ directly}$$

Not testable. Argued on theory.
Only path: $Z \to D \to Y$.

3. Independence

$$Z_i \perp U_i$$

Not testable. No backdoor paths from Z through unobservables.

Only condition (1) is testable. Conditions (2) and (3) require economic reasoning — this is where IV arguments are won or lost.

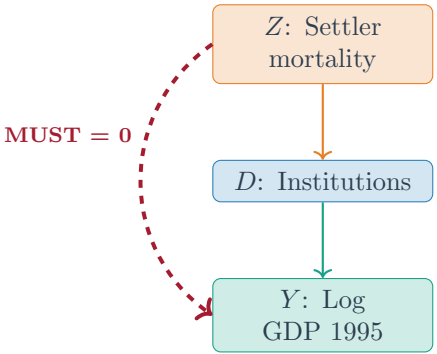
The exclusion restriction holds if mortality only matters through institutions

For exclusion:

- Mortality in 1820 cannot directly affect GDP in 1995
- Channel: mortality → settlement → institutions → income
- Geography/latitude controlled separately

Possible violations:

- Disease burden correlates with labor productivity today
- Geography that drove mortality also drives growth independently



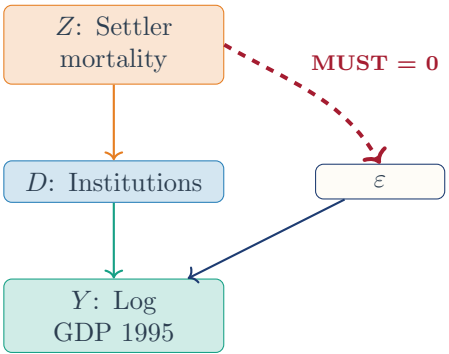
The instrument must also be independent of the error: $\text{Cov}(\varepsilon, Z) = 0$

Two conditions:

- **Exclusion:** Z has no *direct* effect on Y
- **Independence:** $\text{Cov}(\varepsilon, Z) = 0$ — Z is uncorrelated with all omitted causes of Y

For AJR:

- Exclusion: mortality in 1820 cannot directly shift GDP in 1995
- Independence: mortality must not correlate with geography's direct effect on income



Valid IV requires **both**: no direct effect *and* $\text{Cov}(\varepsilon, Z) = 0$

Part V: The Wald Estimator

The IV logic: use only the exogenous variation in D

- Z shifts D (first stage): $\alpha_1 = E[D \mid Z = 1] - E[D \mid Z = 0]$
- Z shifts Y (reduced form): $\pi_1 = E[Y \mid Z = 1] - E[Y \mid Z = 0]$
- If exclusion holds, the RF effect on Y flows entirely through D

First stage (FS)

$\alpha_1 = E[D \mid Z = 1] - E[D \mid Z = 0]$

How much does Z shift D?

Reduced form (RF)

$\pi_1 = E[Y \mid Z = 1] - E[Y \mid Z = 0]$

How much does Z shift Y?

If Z only shifts Y through D, then: $\beta = \pi_1/\alpha_1$

Dividing the reduced form by the first stage recovers the causal effect

$$\hat{\delta}_{\text{Wald}} = \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]} = \frac{\text{Reduced form}}{\text{First stage}}$$

- Numerator: total effect of instrument on outcome
- Denominator: total effect of instrument on treatment
- Ratio: how much each unit of *treatment* (driven by instrument) changes outcome

Dividing the reduced form by the first stage recovers the causal effect

$$\hat{\delta}_{\text{Wald}} = \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]} = \frac{\text{Reduced form}}{\text{First stage}}$$

- Numerator: total effect of instrument on outcome
- Denominator: total effect of instrument on treatment
- Ratio: how much each unit of *treatment* (driven by instrument) changes outcome

For AJR: RF $= -0.570$ FS $= -0.621$ $\Rightarrow$ Wald $\approx$ **0.917**

Wald = 2SLS when one instrument, one endogenous variable (just-identified)

Part VI: From Wald to 2SLS
Two-Stage Least Squares

The 2SLS intuition: use only the "clean" part of D

The problem with OLS: D_i is correlated with ε_i (omitted ability, reverse causality, ...)

The 2SLS fix: decompose D_i into two parts:

Clean part $\hat{D}_i$

Explained by Z_i
Exogenous by assumption
Use this in the structural equation

Dirty part $\hat{\nu}_i$

Unexplained residual
May be correlated with ε_i
Throw this away

$$D_i = \underbrace{\hat{D}_i}_{\text{clean}} + \underbrace{\hat{\nu}_i}_{\text{dirty}}$$

Step 1: Run the first stage — project D onto Z

First stage regression:

$$D_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i + \hat{\nu}_i$$

- Regress the endogenous variable D_i on the instrument Z_i (and all controls)
- Fitted values $\hat{D}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i$ contain only the variation in D that comes from Z
- Since Z is exogenous, $\hat{D}$ is also exogenous

Step 1: Run the first stage — project D onto Z

First stage regression:

$$D_i \quad = \quad \hat{\alpha}_0 \; + \; \hat{\alpha}_1 Z_i \; + \; \hat{\nu}_i$$

- Regress the endogenous variable D_i on the instrument Z_i (and all controls)
- Fitted values $\hat{D}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i$ contain only the variation in D that comes from Z
- Since Z is exogenous, $\hat{D}$ is also exogenous

For AJR:

$\widehat{\texttt{avexpr}}_i = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot \texttt{logem4}_i$

$\hat{\alpha}_1 = -0.621 \quad \Rightarrow \quad$ Higher mortality $\rightarrow$ weaker institutions

Always check: is the first stage strong? Compute the F-statistic.
AJR: $F \approx 16$ — passes the $F > 10$ threshold.

Step 2: Replace D_i with $\hat{D}_i$ in the structural equation

Second stage regression:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_{2SLS}\hat{D}_i + \hat{\varepsilon}_i$$

- Replace the endogenous D_i with the predicted value $\hat{D}_i$ from Stage 1
- $\hat{D}_i$ is a linear function of Z_i only $\Rightarrow$ uncorrelated with structural error ε_i
- OLS on Stage 2 is now unbiased (the endogeneity has been removed)

Step 2: Replace D_i with $\hat{D}_i$ in the structural equation

Second stage regression:

$$Y_i \quad = \quad \hat{\beta}_0 \; + \; \hat{\beta}_{2SLS}\hat{D}_i \; + \; \hat{\varepsilon}_i$$

- Replace the endogenous D_i with the predicted value $\hat{D}_i$ from Stage 1
- $\hat{D}_i$ is a linear function of Z_i only $\Rightarrow$ uncorrelated with structural error ε_i
- OLS on Stage 2 is now unbiased (the endogeneity has been removed)

For AJR: $\widehat{\log \text{GDP}}_i = \hat{\beta}_0 + 0.917 \cdot \widehat{\texttt{avexpr}}_i$

Important: Use software (`iv_robust()`) for standard errors.
Doing Stage 2 manually gives wrong SEs:
they don't account for Stage 1 estimation.

Key result: 2SLS = Wald in the just-identified case

When there is **one instrument** and **one endogenous variable**:

$$\hat{\beta}_{2SLS} = \hat{\delta}_{Wald} = \frac{\hat{\pi}_1}{\hat{\alpha}_1}$$

This is exact: the two-stage algebra produces the ratio formula.

- AJR: $-0.570/-0.621 = 0.917$
- Run `iv_robust()`: get 0.917
- **They match exactly.**

When does 2SLS $\neq$ Wald?

With multiple instruments: Wald doesn't generalize, but 2SLS does.

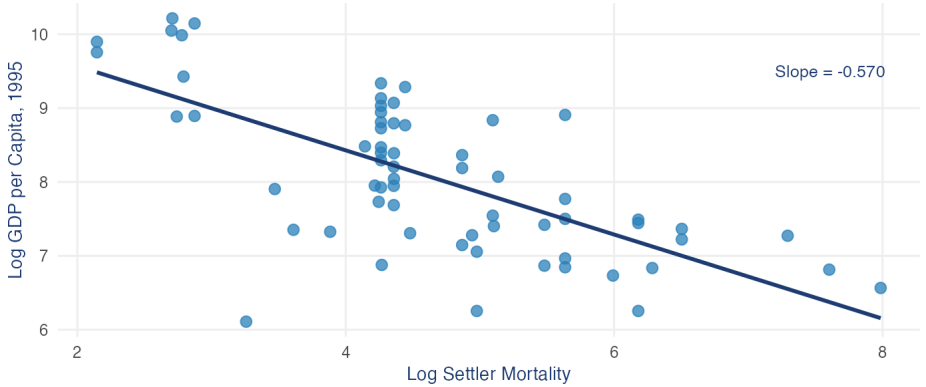
With controls: add them to both stages and the formula.

Part VII: Seeing IV Graphically

First stage: settler mortality predicts institutional quality

FS coefficient: -0.621 SE: 0.155 $t = -4.00$ $F = 16.0$

Reduced form: settler mortality predicts income today



RF coefficient: -0.570 SE: 0.074 $t = -7.72$

2SLS "pulls" the reduced form through the first stage

The graphical intuition:

1. **Reduced form:** $Z \to Y$ (the picture on the right)
2. **First stage:** $Z \to D$ (how much Z shifts D)
3. **2SLS:** rescale the RF slope by the FS slope

Imagine projecting the reduced form scatter onto the first stage scatter:

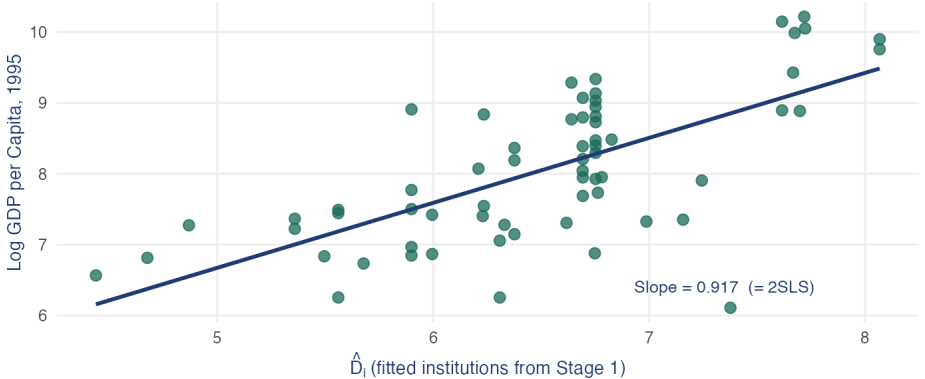
- x-axis shifts from Z to $\hat{D}$ (the fitted institutions)
- The slope steepens: $-0.570/(-0.621) = 0.917$

2SLS re-scales the RF

$$\hat{\beta}_{2SLS} = \frac{\partial Y/\partial Z}{\partial D/\partial Z}$$

"Per unit of D caused by Z"

The 2SLS scatter: log GDP vs. fitted institutions



x-axis is $\hat{D}_i$ — the part of institutions explained by settler mortality 2SLS slope $= 0.917$

Part VIII: Properties of the IV Estimator

The parameter we want and the formula we use are not the same thing

Estimand

- Feature of the data-generating process
- Fixed (not random)
- Example: β = causal effect of institutions on GDP

Estimator

- Function of the random sample
- Has a sampling distribution
- Example: $\hat{\beta}_{IV} = \hat{\pi}_1 / \hat{\alpha}_1$

Unbiased: $E[\hat{\beta}] = \beta$ (finite-sample property)
Consistent: $\hat{\beta} \xrightarrow{p} \beta$ as $n \to \infty$ (asymptotic property)

IV is biased in finite samples . . .

$$E[\hat{\beta}_{IV}] \quad \approx \quad \beta \; + \; \underbrace{\frac{\sigma_{\varepsilon\nu}}{\sigma_\nu^2}}_{\text{OLS bias}} \cdot \frac{1}{F}$$

- Large F: bias $\to 0$
- $F \approx 1$ (weak): bias $\approx$ OLS bias
- The cure is worse than the disease when the instrument is weak

. . . but IV is consistent as $n \to \infty$

$$\hat{\beta}_{IV} \xrightarrow{p} \beta + \frac{\text{Cov}(Z_i, \varepsilon_i)}{\text{Cov}(Z_i, D_i)} = \beta + \frac{0}{\text{Cov}(Z, D)} = \beta$$

- $\text{Cov}(Z_i, \varepsilon_i) = 0$ by exclusion $\Rightarrow$ numerator vanishes
- $\text{Cov}(Z_i, D_i) \neq 0$ by relevance $\Rightarrow$ denominator is non-zero
- Bias is $O(1/n)$: shrinks to zero as $n \to \infty$
- AJR: $n = 64$, so some finite-sample bias remains

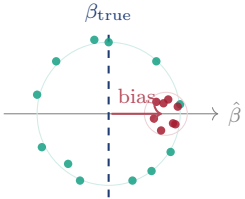
OLS: biased forever. IV: biased in small samples, consistent as $n \to \infty$.

IV variance is always larger than OLS variance

Why IV is noisier:

- OLS uses *all* variation in D
- IV uses only the slice of D caused by Z
- If Z explains 25% of D, IV discards 75%
- Less signal ⇒ estimates bounce more across samples

Each dot = one sample's $\hat{\beta}$:
IV: centered on β_{true}, wide
OLS: tight cluster, shifted (biased)



Bias-variance tradeoff: IV is unbiased but noisy; OLS is biased but precise. Neither always wins — the right measure is MSE = Variance + Bias^2.

IV variance is always larger than OLS variance

$$\text{Var}(\hat{\beta}_{OLS}) = \frac{\sigma^2_\varepsilon}{n \cdot \text{Var}(D)} \qquad \text{Var}(\hat{\beta}_{IV}) = \frac{\sigma^2_\varepsilon}{n \cdot \text{Var}(D) \cdot \rho^2_{ZD}}$$

$$\frac{\text{Var}(\hat{\beta}_{IV})}{\text{Var}(\hat{\beta}_{OLS})} = \frac{1}{\rho^2_{ZD}} \geq 1$$

- $\rho_{ZD} = 0.5$: IV SEs are **2**$\times$ larger than OLS
- $\rho_{ZD} = 0.1$: IV SEs are **10**$\times$ larger
- IV always trades precision for consistency

Today: IV logic and the AJR application — Tuesday: weak instruments

What we built today:

- The IV logic: use only the exogenous variation in D
- Relevance, exclusion, and independence — what each requires
- The Wald estimator and 2SLS: RF divided by FS
- AJR: settler mortality instruments for institutions
- IV variance is always larger than OLS — the bias-variance tradeoff

Tuesday (Lecture 11b): When instruments are weak — what goes wrong, how to detect it, and what to do about it.

IV: a valid instrument turns history into a natural experiment.