IV: Applications and Weak Instruments

Gov 51: Data Analysis and Politics

Scott Cunningham

Harvard University

Week 12
April 14, 2026

Science, politics, and medicine all ask causal questions — not just descriptive ones

Field	What we observe	The causal question
Medicine	Smokers have higher cancer rates	Does smoking *cause* lung cancer?
Political science	Big spenders win elections	Does money *cause* wins, or do strong candidates attract both?
Public health	Incarcerated youth commit more crimes later	Does incarceration *cause* worse outcomes?
Economics	College grads earn more	Does education *cause* higher earnings?

Left column: a fact. Right column: a *causal* claim. They require completely different methods.

Every major policy debate is ultimately a causal question

Domestic and social policy:

- Does raising the minimum wage *cause* unemployment?
- Does Medicaid expansion *cause* better health?
- Does housing assistance *cause* upward mobility?

Criminal justice:

- Does policing *reduce* crime, or follow it?
- Does juvenile detention *cause* more adult crime?
- Does incarceration *cause* people to re-offend less, or more?

Get the answer wrong and the policy **causes harm** instead of helping. Correlation is not enough.

Industry runs on causal questions too

Marketing and product:

- Does showing this ad *cause* a purchase? Or would the buyer have bought anyway?
- Does faster page load *cause* longer sessions? Or do popular pages get both?

People and management:

- Does this training program *cause* better performance?
- Does remote work *cause* lower productivity?

The problem: Optimize on correlation → invest in the wrong things.

The need: Causal answers, not just predictions.

This is why tech companies run A/B tests. But A/B tests aren't always possible.

Randomized experiments answer causal questions definitively

Recall from last week:

$$\underbrace{\bar{Y}_{D=1} - \bar{Y}_{D=0}}_{\text{SDO (what we observe)}} = \underbrace{E[Y_1 - Y_0]}_{\text{ATE}} + \underbrace{E[Y_0 \mid D=1] - E[Y_0 \mid D=0]}_{\text{selection bias}}$$

Why randomization works:

- Assign D randomly: $D \perp\!\!\!\perp (Y_0, Y_1)$
- Control group is identical to treated in expectation
- $E[Y_0 \mid D=1] = E[Y_0 \mid D=0]$
- Selection bias $= 0$

$\Rightarrow$ SDO = ATE

Randomization **deletes** selection bias by making the two groups equivalent in every way — on average.

But most important causal questions cannot be answered with a randomized experiment

Unethical	Infeasible	Unaffordable
Randomly assign children to drug-using parents	Reassign colonial histories to countries	Relocate colleges across counties
Randomly incarcerate juveniles to study the effect	Randomly assign a generation to schooling levels	Run a 30-year health study for every policy

The most important causal questions — the ones with the highest policy stakes — are the hardest to study with a randomized experiment.

When experiments are impossible, we look for the experiment that already happened

Causal question	**Natural experiment (instrument)**
Does parental drug use *cause* child maltreatment?	Policy variation in drug exposure *(Cunningham & Finlay)*
Do colonial institutions *cause* prosperity today?	Settler mortality rates *(Acemoglu et al. 2001)*
Does education *cause* higher wages?	Distance to nearest 4-year college *(Card 1995)*

Natural experiments: when the trial cannot be run, we find the trial that already happened.

A 13-year-old is arrested in Chicago — which judge they face is a coin flip

The scene:

- A minor is arrested and enters Cook County juvenile court
- Chicago has dozens of juvenile court judges
- Cases are **randomly assigned** to judges — not by merit, not by severity

Two judges in this courthouse:

- **Judge A**: strict — incarcerates 70% of cases
- **Judge B**: lenient — incarcerates 30% of cases

The coin flip: Which judge you face is chance, not character.

The variation: Lenient judge → go home. Strict judge → detention.

The instrument: Judge leniency Z randomly assigned → predicts incarceration D.

If incarceration causes harm, every conviction is making things worse

If incarceration CAUSES bad outcomes:

- Detention $\rightarrow \downarrow$ high school graduation
- Detention $\rightarrow \uparrow$ adult incarceration
- System is **creating the problem it claims to solve**

$\Rightarrow$ Case for diversion, supervision, alternatives

If it is pure selection:

- Troubled youth have bad outcomes regardless
- No causal path from detention to adult crime
- System may not be causing additional harm

$\Rightarrow$ Different policy debate entirely

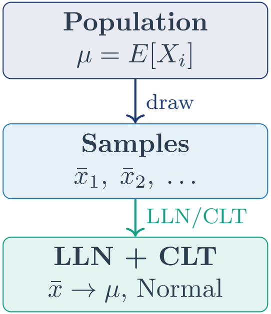
You cannot answer this with OLS. Kids who get incarcerated are already higher risk. You need the coin flip.

Part I: We Have Crossed a Line
From Statistics to Causal Inference

Statistics gave us a clean, assumption-free mapping

The setup: average income in a city.

1. Population: $\mu = E[X_i] = \$50{,}000$
2. Draw a sample: get $\bar{x}_1 = \$45{,}000$
3. Draw another: $\bar{x}_2 = \$52{,}000, \ldots$
4. By the **LLN**: $\bar{x} \xrightarrow{p} \mu$
5. By the **CLT**: $\bar{x} \sim \text{Normal}(\mu, \sigma^2/n)$



No extra assumptions needed. The math does it all.

Causality breaks the mapping — even the population estimate can be wrong

Statistics world

Estimand:	$\mu = E[X_i]$
Estimator:	$\bar{x} = \frac{1}{n}\sum X_i$
Gap:	$\bar{x} - \mu \to 0$ by LLN
Assumptions:	none

Sample converges to population.
Beautiful.

Causal world

Estimand:	$\text{ATE} = E[Y_1 - Y_0]$
Estimator:	$\text{SDO} = \bar{Y}_{D=1} - \bar{Y}_{D=0}$
Gap:	$\text{SDO} = \text{ATE} +$ **selection bias**
Assumptions:	required

Even at the population level, SDO $\neq$ ATE
unless the design guarantees it.

Identification is the assumption that makes the calculation causal

Two strategies that close the gap:

1. Random assignment

- $Y_0, Y_1 \perp\!\!\!\perp D$
- Selection bias $= 0$
- SDO $=$ ATE

Designed: you control who gets treatment.

2. Valid instrument

- $Z \perp\!\!\!\perp u, \;\; \text{Cov}(Z, D) \neq 0$
- No direct $Z \to Y$
- Wald $= \beta_1$

Found: you discover randomization that was already in the world.

Neither assumption is testable from data alone. You must *design* or *discover* them.

Part II: The Wald Estimator — A Quick Recap

Wald: the ratio of two covariances recovers the causal effect

$$\hat{\beta}_1^{\text{Wald}} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} = \frac{\text{Reduced form}}{\text{First stage}}$$

Three conditions:

1. **Relevance:** $\text{Cov}(Z, D) \neq 0$ — testable
2. **Exclusion:** $Z \perp\!\!\!\perp u$ — not testable
3. **Independence:** Z not caused by U — not testable

The coin flip again:
$Z \perp\!\!\!\perp u$ means any $\text{Cov}(Y, Z)$ runs through D only.

Worksheets 1 & 2 derive this step by step.

Stage 1: project the treatment onto the instrument, save fitted values

Regress D on Z:

$$D_i = \pi_0 + \pi_1 Z_i + v_i$$

Save the fitted values:

$$\hat{D}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

$\hat{D}_i$ contains only the variation in D that is explained by Z. The endogenous noise v_i has been stripped out.

Stage 2: regress the outcome on the cleaned-up treatment

Regress Y on $\hat{D}$:

$$Y_i = \delta_0 + \delta_1 \hat{D}_i + e_i$$

$$\hat{\delta}_1^{\text{2SLS}} = \frac{\text{Cov}(Y, \hat{D})}{\text{Var}(\hat{D})} = \frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} = \hat{\beta}_1^{\text{Wald}}$$

2SLS and the Wald estimator are the same thing. *Worksheet 2* shows the algebra.

Part III: Seeing IV With Your Eyes
First Stage and Reduced Form

Acemoglu, Johnson and Robinson asked why some former colonies are rich and others poor

The puzzle:

- North America, Australia, New Zealand: rich
- Congo, Nigeria, Haiti: poor
- All were European colonies — why?

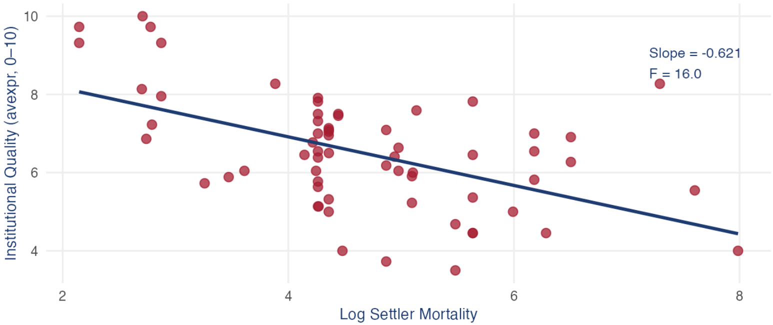
Their theory:

- Settlers built property rights and rule of law where they lived
- Extractive colonies got institutions designed only to ship resources home
- Those institutional differences persist 200 years later

The endogeneity problem: Rich countries also build good institutions — OLS cannot separate cause from effect.

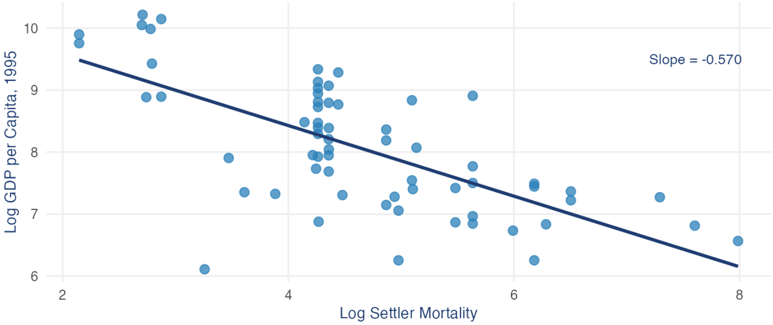
The coin flip: Malaria and yellow fever determined where Europeans could settle — before anyone chose where to colonize.

First stage: settler mortality predicts institutional quality



Slope = -0.621. Higher mortality $\rightarrow$ fewer settlers $\rightarrow$ weaker institutions.

Reduced form: settler mortality predicts income today



Slope = -0.570. Higher mortality $\rightarrow$ weaker institutions $\rightarrow$ lower income.

Wald: divide the reduced form slope by the first stage slope

$$\hat{\beta}^{\text{Wald}} = \frac{-0.570}{-0.621} = \mathbf{0.917}$$

Reduced form -0.570 $\div$ **First stage** -0.621

Wald $= 0.917$

R: iv_robust() runs both stages and fixes standard errors

```r
library(estimatr)
library(tidyverse)

ajr <- read_csv("data/ajr2001.csv")
# Outcome: log GDP per capita (logpgp95)
# Endogenous: institutional quality (avexpr)
# Instrument: log settler mortality (logem4)
fit_iv <- iv_robust(logpgp95 ~ avexpr | logem4,
                    data = ajr, se_type = "HC2")
tidy(fit_iv, conf.int = TRUE)
```

Never run Stage 2 by hand: use iv_robust() for correct SEs.

What IV does — four things to remember

1. **IV requires an assumption about the world**
   The exclusion restriction cannot be tested.
   You must argue it from institutional knowledge.
2. **2SLS uses only the clean variation in D**
   Stage 1 strips the endogenous part.
   Stage 2 identifies β_1 from what remains.
3. **AJR: a real natural experiment**
   Settler mortality was set by ecology.
   Institutions were set by settlement.
   Income was set by institutions.
4. **The Wald: RF $\div$ FS**
   No more, no less.

Thursday: What happens when the first stage is weak — and what to do about it.

Next class: when the coin is bent

Weak instruments and what goes wrong:

- What "weak" means precisely — and why it is catastrophic for 2SLS
- How to test: the F-statistic and the Olea-Pflüger effective F
- Anderson-Rubin confidence intervals: valid even when F is small
- Finding natural experiments: judge leniency (Aizer-Doyle 2015)
- How bad data can fake a strong instrument: the Albouy critique of AJR

If your F-statistic is below 10, Thursday's lecture is
the most important thing you will hear this semester.

**IV finds the randomiza-
tion that was already there.
The formula just harvests it.**