

Applied Economics Issues

Session 3: Instrumental Variables

Ewen Gallic

Master 2 DADEE – 2026

Today's session

- 1 **Reading debrief** : [Angrist and Krueger \(1991\)](#), explained by you
- 2 **Instrumental variables** : a reminder of Session 2
- 3 **Practice in R** (1/2) (Wald ratio)
- 4 **Instrumental variables** : the estimator and its precision
- 5 **Practice in R** (2/2): IV by hand, standard errors, sample size
- 6 **Instrumental variables** : 2SLS, what IV estimates, what can go wrong
- 7 **Application** : [Angrist and Krueger \(1991\)](#), checking your answers
- 8 **In class exercise** : IV detective
- 9 **Next session** : reading assignment

The course: two questions

A. Does X cause Y ?

Measuring causal impact

- ▶ **S2**: the endogeneity problem; OLS under endogeneity (practice in R).
- ▶ **S3**: instrumental variables: practice in R, method, what IV estimates; application: [Angrist and Krueger \(1991\)](#); IV detective.
- ▶ **S4**: the credibility revolution; randomized trials, natural and quasi-natural experiments, regression discontinuity; panel data and difference-in-differences.

B. How much is X worth?

Assigning value in the absence of a market

- ▶ **S5**: valuation methods (stated, revealed, experimental); climate change and the price of carbon (IAMs, social cost of carbon).
- ▶ **S6**: measuring and valuing climate impacts (agriculture, health, sea-level rise, conflict).
- ▶ **S7**: behavioral economics: beliefs, social norms, nudges; wrap-up.

Where we stopped last session

- ▶ Recall that we want to estimate a causal effect β in $y_i = \mathbf{x}_i' \beta + \varepsilon_i$.
- ▶ However, OLS is consistent only if $E[\mathbf{x}_i \varepsilon_i] = \mathbf{0}$. Omitted variables, simultaneity and measurement error break this condition.
- ▶ With the R practice (simulated wages, true return to schooling $\beta = 0.08$), you saw that with omitted ability (ovb) or noisy schooling (measerr), OLS **converges to the wrong value**, and more data only narrows the confidence interval **around the wrong number**.
- ▶ We introduced the idea of an instrument: a variable z that moves x for reasons unrelated to u ; $\beta = \text{Cov}(z, y) / \text{Cov}(z, x)$; with a binary z , the Wald estimator.
- ▶ Today:
 - Part 4 of the notebook, in two steps, alternating with the method,
 - what IV actually estimates, and what can go wrong,
 - a real instrument: the quarter of birth ([Angrist and Krueger, 1991](#)),
 - then it is your turn: find the instrument in four papers.

1. Reading debrief: Angrist and Krueger (1991)

Your turn: Angrist and Krueger (1991)

Those who read Angrist and Krueger (1991, QJE) in depth: explain it to the class before we look at the slides.

- 1 Why can't we simply regress wages on years of schooling?
- 2 Why would the quarter of birth affect years of schooling?
- 3 What must be true about the quarter of birth for the estimate to be causal?
- 4 What do the authors find, compared with OLS?

We write your answers on the board, and check them against the paper at the end of the session.

2. Instrumental variables: a reminder of Session 2

The Equation of Interest, and the Problem

We want to estimate the causal return to one year of schooling, β , in

$$\ln w_i = \alpha + \beta \text{educ}_i + \varepsilon_i, \quad \varepsilon_i = \gamma u_i + v_i,$$

where u_i is **ability** (unobserved) and v_i is a wage shock.

But OLS converges to

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{\text{Cov}(\text{educ}_i, \varepsilon_i)}{\text{Var}(\text{educ}_i)} = \beta + \overbrace{\gamma \frac{\text{Cov}(\text{educ}_i, u_i)}{\text{Var}(\text{educ}_i)}}^{\text{ability bias}}$$

- ▶ More able people earn more ($\gamma > 0$) and study longer ($\text{Cov}(\text{educ}_i, u_i) > 0$): consequently, OLS **overstates** β .
- ▶ In the practice (ovb): true $\beta = 0.08$, OLS $\rightarrow 0.109$, and more data does not help.
- ▶ The same happens with measurement error in schooling (bias towards zero), or with simultaneity.
- ▶ **We need variation in schooling that has nothing to do with ability.**

One way out: an instrument

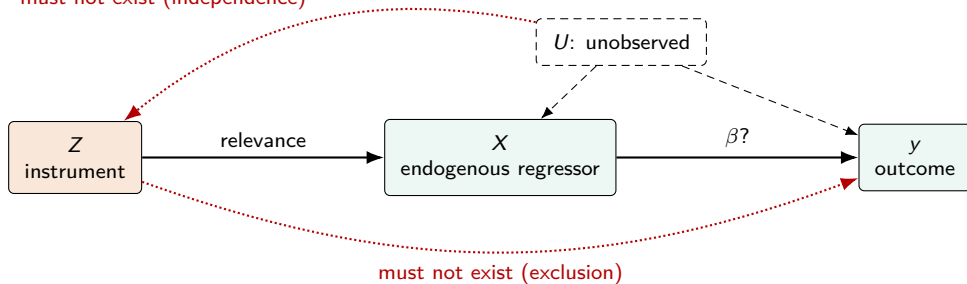
- ▶ The idea consists in finding a variable z_i that **moves schooling for reasons unrelated to ability**, and using only that part of the variation.
- ▶ Example in the practice: a **scholarship offer drawn by a coin toss**.
- ▶ Two conditions are required for the instrument:
 - **relevance**: $\text{Cov}(z_i, \text{educ}_i) \neq 0$ (the offer changes schooling);
 - **exogeneity**: $\text{Cov}(z_i, \varepsilon_i) = 0$ (the offer is unrelated to ability and affects wages only through schooling).
- ▶ Take the covariance of both sides of the wage equation with z_i :

$$\text{Cov}(z_i, \ln w_i) = \beta \text{Cov}(z_i, \text{educ}_i) + \underbrace{\text{Cov}(z_i, \varepsilon_i)}_{=0} \implies \beta = \frac{\text{Cov}(z_i, \ln w_i)}{\text{Cov}(z_i, \text{educ}_i)}.$$

How much the offer moves wages, relative to how much it moves schooling.

What Makes a Valid Instrument?

must not exist (independence)



1. Relevance : Z moves X .

► **Testable**: regress X on Z .

2. Exogeneity : Z is uncorrelated with ε .

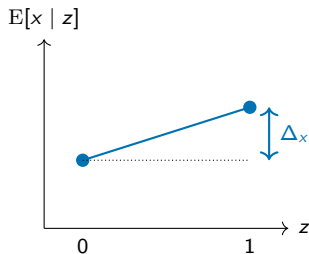
- **Independence**: Z as good as randomly assigned.
- **Exclusion**: Z affects y only through X .
- **Not testable** (ε is unobserved): it is an **argument**.

Binary Instrument: the Wald Estimator

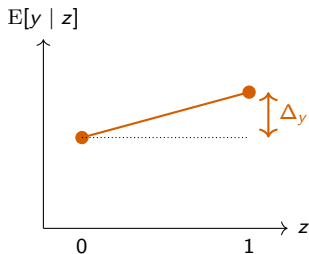
With a binary instrument $z \in \{0, 1\}$, the ratio $\text{Cov}(z, y) / \text{Cov}(z, x)$ becomes a ratio of two differences in means:

$$\hat{\beta}_{IV} = \frac{E[y | z = 1] - E[y | z = 0]}{E[x | z = 1] - E[x | z = 0]} = \frac{\text{reduced form}}{\text{first stage}}$$

“By how much does z move y , relative to how much it moves x ?”



first stage



reduced form

3. Practice (1/2): computing the Wald ratio

Practice (in class, 1/2): computing the Wald ratio

Back to our simulated data

- ▶ In `ovb` and `measerr`, OLS is inconsistent.
- ▶ The files also contain `z`: a scholarship offer, drawn by a **coin toss**.
- ▶ Is `z` a valid instrument for `educ`?
 - **relevance**: it shifts schooling (by 1 year on average);
 - **exogeneity**: it is random, so independent of ability, of the wage shock and of the noise;
 - **exclusion**: it affects wages only through schooling.

Part 4 of the notebook (≈ 15 min)

- ▶ 4.1 first stage and reduced form (`ovb`, $n = 10\,000$)
- ▶ 4.2 the Wald ratio

Predict first

- ▶ Will the first stage be close to 1? Why?
- ▶ Will the Wald ratio be closer to 0.08 or to the OLS estimate?

Leave the question on the F statistic for later.

Practice debrief (1/2): the Wald ratio, for real

Case ovb, $n = 10\,000$

- ▶ **First stage**: the offer raises schooling by ≈ 0.997 years (the true effect in the simulation is 1).
 - ▶ **Reduced form**: the offer raises log wages by ≈ 0.076 .
 - ▶ **Wald ratio**: $0.0764/0.997 \approx 0.077$ close to the true return to schooling. OLS gives 0.109.
- ⇒ Using only the variation in schooling driven by the coin toss removes the ability bias.
- ⇒ Open questions, answered next:
- how do we write this estimator in general, with controls and several regressors?
 - why does it work, and how precise is it?

4. Instrumental variables: the estimator

Moment Conditions

Let us start from something we know: how OLS is built.

- ▶ The model is $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$. If the regressors are exogenous, the true $\boldsymbol{\beta}$ satisfies K **moment conditions**:

$$E[\mathbf{x}_i \varepsilon_i] = \mathbf{0} \iff E[\mathbf{x}_i (y_i - \mathbf{x}_i' \boldsymbol{\beta})] = \mathbf{0}$$

- ▶ Replace the expectation by the sample mean, and solve for the unknown $\mathbf{b}$:

$$\frac{1}{n} \sum_i \mathbf{x}_i (y_i - \mathbf{x}_i' \mathbf{b}) = \mathbf{0} \iff \frac{1}{n} \mathbf{X}' (\mathbf{y} - \mathbf{X} \mathbf{b}) = \mathbf{0} \implies \mathbf{b} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$$

This is the OLS estimator.

- ▶ If a regressor is endogenous, $E[\mathbf{x}_i \varepsilon_i] \neq \mathbf{0}$: the condition OLS solves is **false**, so OLS is inconsistent.
- ▶ Instruments give us a condition that **is** true: $E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}$. Same recipe, with $\mathbf{z}_i$ instead of $\mathbf{x}_i$.

Setting Up the Notation: Endogenous Regressors and Instruments

The model is $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, with $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$ and $K = K_1 + K_2$:

- ▶ $\mathbf{X}_1$ ($n \times K_1$): the **endogenous** regressors.
- ▶ $\mathbf{X}_2$ ($n \times K_2$): the **exogenous** regressors (the constant, the controls).
- ▶ $\mathbf{Z}_1$ ($n \times L_1$): the **excluded instruments** (they do not appear in the equation of $\mathbf{y}$).
- ▶ $\mathbf{Z} = [\mathbf{Z}_1 \ \mathbf{X}_2]$ ($n \times L$, $L = L_1 + K_2$): the full set of instruments. Exogenous regressors are their *own* instruments.

Exogeneity

$$E[\mathbf{z}_i \varepsilon_i] = \mathbf{0} \quad (L \text{ conditions})$$

Relevance (rank condition)

$$\text{rank } E[\mathbf{z}_i \mathbf{x}_i'] = K$$

Order condition: $L \geq K$, i.e., $L_1 \geq K_1$ (at least as many excluded instruments as endogenous regressors). $L = K$: *just-identified*; $L > K$: *over-identified*.

Solving the Moment Conditions: the Just-Identified Case ($L = K$)

Suppose $L = K$ (*just-identified*): exogeneity is a list of equations

- ▶ $\mathbf{z}_i$ is the $L \times 1$ vector of instruments for observation i : the excluded instruments plus the exogenous regressors, which instrument themselves.
- ▶ Exogeneity says that each instrument is uncorrelated with the error:

$$E[\mathbf{z}_i \varepsilon_i] = \begin{cases} E[z_{i1} \varepsilon_i] = 0 \\ \vdots \\ E[z_{iL} \varepsilon_i] = 0 \end{cases}$$

These equations are called **moment conditions** because they are statements about population moments (expectations).

- ▶ Replacing ε_i turns them into equations in β :

$$E[\mathbf{z}_i(y_i - \mathbf{x}_i' \beta)] = \mathbf{0} \quad \Longleftrightarrow \quad \underbrace{E[\mathbf{z}_i \mathbf{x}_i']}_{L \times K} \beta = \underbrace{E[\mathbf{z}_i y_i]}_{L \times 1}.$$

Solving the Moment Conditions: the Just-Identified Case ($L = K$)

The sample version, step by step:

$$\frac{1}{n} \mathbf{Z}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{0}$$

$$\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\mathbf{b} = \mathbf{0}$$

$$\mathbf{Z}'\mathbf{X}\mathbf{b} = \mathbf{Z}'\mathbf{y}$$

$$\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}$$

- The last step requires $\mathbf{Z}'\mathbf{X}$ to be **invertible** (**relevance** (rank) condition).

Plug in the Model

Let us have a look at the properties of the IV estimator. Replace $\mathbf{y}$ by the true model:

$$\begin{aligned}\hat{\beta}_{IV} &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\beta + \varepsilon) \\ &= \underbrace{(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{X}}_{=\mathbf{I}}\beta + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\varepsilon \\ &= \beta + \underbrace{(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\varepsilon}_{\text{sampling error}}\end{aligned}$$

Divide each factor by n and let $n \rightarrow \infty$:

$$\hat{\beta}_{IV} = \beta + \left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)^{-1} \left(\frac{1}{n}\mathbf{Z}'\varepsilon\right) \xrightarrow{p} \beta + \mathbb{E}[\mathbf{z}_i\mathbf{x}'_i]^{-1} \underbrace{\mathbb{E}[\mathbf{z}_i\varepsilon_i]}_{=0} = \beta$$

- ▶ **Relevance**: $\mathbf{Z}'\mathbf{X}$ is invertible.
- ▶ **Exogeneity**: the second term vanishes.
- ▶ Hence $\hat{\beta}_{IV} \xrightarrow{p} \beta$: IV is **consistent**. The sampling error converges to $\mathbf{0}$ because $\frac{1}{n}\mathbf{Z}'\varepsilon \xrightarrow{p} \mathbf{0}$.

IV: Biased in Finite Samples (1/2)

Write the first stage as $\mathbf{X} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}$: the part of $\mathbf{X}$ driven by the instruments, and the rest. $\mathbf{Z}$ is unrelated to ε (exogeneity), but $\mathbf{V}$ is correlated with ε (endogeneity: ability is in $\mathbf{V}$).

$$\hat{\beta}_{IV} - \beta = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\varepsilon, \quad \mathbf{Z}'\mathbf{X} = \underbrace{\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}}_{\text{signal}} + \underbrace{\mathbf{Z}'\mathbf{V}}_{\text{first-stage noise}}$$

OLS: condition on $\mathbf{X}$

- ▶ $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is then fixed,
- ▶ and $E[\varepsilon \mid \mathbf{X}] = \mathbf{0}$ (exogeneity),
- ▶ so $E[\hat{\beta} \mid \mathbf{X}] = \beta$: **unbiased**.

IV: what can we condition on?

- ▶ On $\mathbf{X}$? Then $E[\varepsilon \mid \mathbf{Z}, \mathbf{X}] \neq \mathbf{0}$: $\mathbf{X}$ contains $\mathbf{V}$, which is informative about ε .
- ▶ On $\mathbf{Z}$ only? Then $(\mathbf{Z}'\mathbf{X})^{-1}$ is still random (through $\mathbf{Z}'\mathbf{V}$) and moves with $\mathbf{Z}'\varepsilon$.

Thought experiment: if $\mathbf{Z}'\mathbf{V} = \mathbf{0}$, then $\mathbf{Z}'\mathbf{X} = \mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi}$ is fixed given $\mathbf{Z}$, and $E[(\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi})^{-1}\mathbf{Z}'\varepsilon \mid \mathbf{Z}] = (\mathbf{Z}'\mathbf{Z}\boldsymbol{\Pi})^{-1}\mathbf{Z}' E[\varepsilon \mid \mathbf{Z}] = \mathbf{0}$.

The bias comes from the first-stage noise $\mathbf{Z}'\mathbf{V}$, which is correlated with ε .

IV: Biased in Finite Samples (2/2)

As $n \rightarrow \infty$: consistent. Divide by n :

$$\frac{1}{n} \mathbf{Z}' \mathbf{X} = \underbrace{\frac{1}{n} \mathbf{Z}' \mathbf{Z}}_{\rightarrow E[\mathbf{z}_i \mathbf{z}_i']} \boldsymbol{\Pi} + \underbrace{\frac{1}{n} \mathbf{Z}' \mathbf{V}}_{\rightarrow 0}, \quad \frac{1}{n} \mathbf{Z}' \boldsymbol{\varepsilon} \rightarrow 0 \quad \Rightarrow \quad \hat{\beta}_{IV} \xrightarrow{p} \beta.$$

The first-stage noise averages out: biased for any fixed n , but consistent.

Weak instrument ($\boldsymbol{\Pi} \approx 0$): worse. The signal $\mathbf{Z}' \mathbf{Z} \boldsymbol{\Pi}$ is small, the noise $\mathbf{Z}' \mathbf{V}$ dominates, and $(\mathbf{Z}' \mathbf{X})^{-1}$ is large: the correlation between $\mathbf{Z}' \mathbf{V}$ and $\mathbf{Z}' \boldsymbol{\varepsilon}$ is **amplified**. **Extreme case:** one x , one z , $\boldsymbol{\Pi} = 0$ (so $\mathbf{x} = \mathbf{v}$). Write $\boldsymbol{\varepsilon} = \delta \mathbf{v} + \mathbf{e}$, with $\mathbf{e}$ unrelated to $\mathbf{v}$ and $\mathbf{z}$:

$$\hat{\beta}_{IV} - \beta = \frac{\mathbf{z}' \boldsymbol{\varepsilon}}{\mathbf{z}' \mathbf{v}} = \delta + \frac{\mathbf{z}' \mathbf{e}}{\mathbf{z}' \mathbf{v}}, \quad \delta = \frac{\text{Cov}(x, \varepsilon)}{\text{Var}(x)} = \text{the OLS bias.}$$

- ▶ A useless instrument does not give a neutral answer: it is centred on the **OLS** value, with a lot of noise.
- ▶ In between, IV is biased **towards OLS**, roughly by a fraction $1/F$ of the OLS bias (F : first-stage F statistic).

How Precise Is IV?

Just-identified case, homoskedastic errors :

$$\widehat{\text{Var}}(\hat{\beta}_{IV}) = \hat{\sigma}^2 (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_i \hat{\varepsilon}_i^2, \quad \hat{\varepsilon}_i = y_i - \mathbf{x}_i' \hat{\beta}_{IV}$$

(residuals computed with the **observed** $\mathbf{x}_i$; robust versions exist).

One x , one z : the variance is approximately

$$\text{Var}(\hat{\beta}_{IV}) \approx \frac{\sigma^2}{n \text{Var}(x)} \times \frac{1}{\rho_{xz}^2}, \quad \rho_{xz} = \text{Corr}(x, z)$$

- ▶ The first factor is the OLS variance: IV is **less precise** than OLS, by a factor $1/\rho_{xz}^2$.
- ▶ In the simulation, $\rho_{xz} \approx 0.27$: the IV standard error should be about $1/0.27 \approx 3.7$ times the OLS one.
- ▶ If z barely moves x ($\rho_{xz} \approx 0$), the variance explodes: a first sign of **weak instruments**.

5. Practice (2/2): IV by hand, precision, sample size

Practice (in class, 2/2): IV by hand, precision, sample size

Part 4 of the notebook, second half (≈ 20 min)

- ▶ 4.3 IV by hand: $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$. Why does it give the same number as the Wald ratio?
- ▶ 4.4 standard errors: $\hat{\sigma}^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}$, compared with OLS
- ▶ 4.5 IV at different sample sizes, for `ovb` and `measerr`
- ▶ Back to the question left aside in 4.1: what does the first-stage F statistic tell you?

Predict first

- ▶ Will IV be more or less precise than OLS? By roughly how much?
- ▶ What will happen with $n = 100$?

Practice debrief (2/2): IV by hand, precision, sample size

Case ovb, $n = 10\,000$: $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ gives the Wald ratio (one binary instrument and a constant).

Estimator	Estimate	Standard error	95% CI
OLS	0.109	0.0017	[0.105, 0.112] excludes 0.08
IV	0.077	0.0064	[0.064, 0.0892]

The IV standard error is about 3.8 times the OLS one, as predicted.

Small samples

- ▶ ovb, $n = 100$: $F \approx 0.6$, $IV = 0.167$, standard error 0.224: useless.
- ▶ For $n \geq 10\,000$, IV is close to 0.08 in both cases, and the CI shrinks around the truth. With $n = 10\,000$, $F \approx 757$.
- ⇒ IV trades **precision** for **consistency**: it only uses the part of the variation in schooling driven by the offer.
- ⇒ Relevance can be tested (first stage). Exogeneity and exclusion cannot: here we know they hold because we simulated the data. In real applications, **they must be argued**.

Weak Instruments: Checking the First Stage

When **Z** barely moves **X**, IV is biased **towards OLS**, imprecise, and “ ± 1.96 SE” is not a valid CI. To check this, in the first stage, with the controls: a F test.

$$\mathbf{x} = \mathbf{Z}_1\pi_1 + \mathbf{X}_2\pi_2 + \mathbf{v}, \quad H_0 : \pi_1 = \mathbf{0} \text{ (the excluded instruments do not move } \mathbf{x}\text{)}.$$

With one instrument, $F = t^2$; with several, a joint test. Relevance is the **only** condition we can check.

Same coin toss, different n (practice, ovb):

n	F	IV (SE)
100	0.6	0.167 (0.224)
1 000	88	0.092 (0.018)
10 000	757	0.077 (0.006)
100 000	7 488	0.080 (0.002)

- ▶ F measures **strength**, not significance: it grows with n and with the share of x explained by z .
- ▶ Rule of thumb: $F \gtrsim 10$ (Stock and Yogo, 2005), much stricter than the 5% critical value. Below, the instrument is *weak*.
- ▶ Beware! An F test tells nothing about exogeneity or exclusion: a strong instrument can be invalid.

6. IV: 2SLS, what IV estimates, what can go wrong

6.1. Several instruments: two-stage least squares

Over-Identification ($L > K$) and Two-Stage Least Squares

- ▶ With $L > K$, $\mathbf{Z}'\mathbf{X}$ is $L \times K$: not square, no inverse. We have L moment conditions for only K parameters.
- ▶ The idea is to use the $\mathbf{Z}$ -projection of the residuals and make it as small as possible:

$$\hat{\beta}_{2SLS} = \arg \min_{\mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{P}_Z (\mathbf{y} - \mathbf{X}\mathbf{b}), \quad \mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

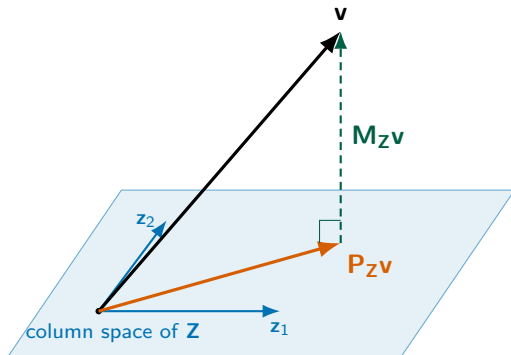
(the $\mathbf{P}_Z$ matrix projects any vector onto the column space of $\mathbf{Z}$, see next slide)

- ▶ The first-order condition gives

$$\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}$$

- ▶ $\mathbf{P}_Z$ is the projection matrix on the columns of $\mathbf{Z}$: symmetric ($\mathbf{P}_Z' = \mathbf{P}_Z$) and idempotent ($\mathbf{P}_Z\mathbf{P}_Z = \mathbf{P}_Z$).
- ▶ If $L = K$ (and $\mathbf{Z}'\mathbf{X}$ invertible), the formula collapses to $(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ (recall that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$): the just-identified IV estimator.

What Does $\mathbf{P_Z}$ Do? A Projection



- ▶ Suppose we have two instruments, $\mathbf{z}_1$ and $\mathbf{z}_2$
- ▶ The columns of $\mathbf{Z}$ span a “plane”: all the vectors we can build as $\mathbf{Zc} = c_1\mathbf{Z}_1 + c_2\mathbf{Z}_2$.
- ▶ $\mathbf{P_Z v} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}$ is the point of that plane **closest** to $\mathbf{v}$: the fitted values of the regression of $\mathbf{v}$ on $\mathbf{Z}$ (among all points $\mathbf{Zc}$ of the plane, look for the one nearest to $\mathbf{v}$).
- ▶ The residuals: $\mathbf{M_Z v} = \mathbf{v} - \mathbf{P_Z v} = (\mathbf{I} - \mathbf{P_Z})\mathbf{v}$, **orthogonal** to the plane: $\mathbf{Z}'\mathbf{M_Z v} = \mathbf{0}$.
- ▶ Projecting twice changes nothing: $\mathbf{P_Z P_Z} = \mathbf{P_Z}$.

- ▶ With $\mathbf{v} = \mathbf{x}$ (schooling): $\mathbf{P_Z x} = \hat{\mathbf{x}}$ is the part of schooling the instruments can explain, the **clean part**; $\mathbf{M_Z x}$ is the rest, where ability lives.

Why “Two Stages”?

First stage: regress each column of $\mathbf{X}$ on $\mathbf{Z}$

$$\hat{\mathbf{\Pi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}, \quad \hat{\mathbf{X}} = \mathbf{Z}\hat{\mathbf{\Pi}} = \mathbf{P}_Z\mathbf{X}$$

Second stage: regress $\mathbf{y}$ on $\hat{\mathbf{X}}$

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

Since $\mathbf{P}_Z$ is symmetric ($\mathbf{P}_Z' = \mathbf{P}_Z$) and idempotent ($\mathbf{P}_Z\mathbf{P}_Z = \mathbf{P}_Z$):

$$\hat{\mathbf{X}}'\hat{\mathbf{X}} = (\mathbf{P}_Z\mathbf{X})'\mathbf{P}_Z\mathbf{X} = \mathbf{X}'\mathbf{P}_Z\mathbf{X} \quad \text{and} \quad \hat{\mathbf{X}}'\mathbf{y} = \mathbf{X}'\mathbf{P}_Z\mathbf{y}$$

hence

$$(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y} = \hat{\beta}_{2SLS}.$$

- ▶ $\hat{\mathbf{X}}$ is the part of $\mathbf{X}$ explained by the instruments: **exogenous by construction**.
- ▶ Exogenous regressors $\mathbf{X}_2$ are columns of $\mathbf{Z}$: $\mathbf{P}_Z\mathbf{X}_2 = \mathbf{X}_2$. They enter the second stage **unchanged** (and must be in the first stage).

Beyond a Binary Instrument

In general (one x , one z , no controls), the IV estimator is a **ratio of two slopes**:

$$\hat{\beta}_{IV} = \frac{\widehat{\text{Cov}}(z, y)}{\widehat{\text{Cov}}(z, x)} = \frac{\widehat{\text{Cov}}(z, y) / \widehat{\text{Var}}(z)}{\widehat{\text{Cov}}(z, x) / \widehat{\text{Var}}(z)} = \frac{\text{slope of } y \text{ on } z}{\text{slope of } x \text{ on } z}$$

- ▶ **Binary z** : with $p = \Pr(z = 1)$, $\text{Cov}(z, y) = p(1 - p)(E[y | z = 1] - E[y | z = 0])$, and the same for x . The $p(1 - p)$ cancels: we get back the **Wald estimator**.
- ▶ **Continuous z** (rainfall, distance, land gradient): same logic, with slopes instead of differences in means. “By how much does y move per unit of z , relative to how much x moves per unit of z ?”
- ▶ **Multi-valued z (discrete)** (quarter of birth, number of siblings): IV is a **weighted average of Wald estimators** comparing adjacent values of z , with more weight where z moves x the most.
- ▶ **Several instruments** (three quarter-of-birth dummies): 2SLS combines them into one, the first-stage fitted value $\hat{x}$. Graphically, the 2SLS estimate is the slope of a line through the points $(\bar{x}_g, \bar{y}_g)$, one point per value of the instruments.

6.2. What does IV estimate?

What Does IV Estimate? (1/3) Heterogeneous Effects and Types

So far, the return to schooling β was the **same for everyone**. It may differ across people. Consider a binary instrument z_i (scholarship offer) and binary treatment D_i (goes to college).

- **Potential outcomes:** $Y_i(1)$ = wage of i with college, $Y_i(0)$ without. **Individual effect:** $\tau_i = Y_i(1) - Y_i(0)$.
- **Potential treatments:** $D_i(1)$ = would i go to college *if offered*? $D_i(0)$ = *if not offered*?
- We observe only one of each: $D_i = D_i(z_i)$ and $Y_i = Y_i(D_i)$. So we never know a person's type.

Four types, depending on how D_i reacts to the offer:

Type	$D_i(0)$ (<i>not offered</i>)	$D_i(1)$ (<i>offered</i>)	Who?
Compliers	0	1	go to college only because of the offer
Always-takers	1	1	go anyway (e.g., wealthy families)
Never-takers	0	0	do not go, even with the offer
Defiers	1	0	go only if not offered (implausible)

What Does IV Estimate? (2/3) The Wald Ratio Under Heterogeneity

Assumptions: **independence** (z random: same mix of types in both groups), **exclusion** (z affects Y only through D), **monotonicity** (no defiers). Let π_C, π_A, π_N be the shares of compliers, always-takers and never-takers ($\pi_D = 0$ under monotonicity assumption).

First stage: always-takers have $D = 1$ in both groups, never-takers $D = 0$: only compliers change.

$$E[D \mid z = 1] - E[D \mid z = 0] = (\pi_A + \pi_C) - \pi_A = \pi_C.$$

Reduced form: by exclusion, the wages of always- and never-takers do not depend on z ; they cancel.

$$E[Y \mid z = 1] = \pi_A E[Y(1) \mid A] + \pi_N E[Y(0) \mid N] + \pi_C E[Y(1) \mid C]$$

$$E[Y \mid z = 0] = \pi_A E[Y(1) \mid A] + \pi_N E[Y(0) \mid N] + \pi_C E[Y(0) \mid C]$$

$$\Rightarrow \frac{E[Y \mid z = 1] - E[Y \mid z = 0]}{E[D \mid z = 1] - E[D \mid z = 0]} = \frac{\pi_C E[\tau_i \mid C]}{\pi_C} = E[\tau_i \mid \text{complier}] = \text{LATE}$$

The **local average treatment effect**: the average effect for compliers (Imbens and Angrist, 1994).

What Does IV Estimate? (3/3) What “Local” Means

- ▶ **Only the compliers are informative**: the instrument does not move always-takers or never-takers, so it tells us nothing about **their** effect. $LATE \neq$ the average effect in the population (ATE), unless effects are homogeneous.
- ▶ **The first stage measures the share of compliers** (binary treatment): in [Angrist and Krueger \(1991\)](#), very few men change their schooling because of their birth date, so the LATE concerns a small group.
- ▶ **Different instruments, different compliers, different estimates**, and all of them can be valid:
 - quarter of birth: men who would have left school at the legal age (low-education margin);
 - a college scholarship: students at the margin of going to college.
- ▶ **Why monotonicity matters**: defiers would enter the Wald ratio with a **negative** weight, and it would no longer be the average effect for anyone.
- ▶ **You can ask yourself: who are the compliers, and is their effect the one the policy question needs?**

6.3. What can go wrong

Credibility of an IV Design

- ▶ Exogeneity, $E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}$, involves the unobserved ε : it cannot be tested directly. What we can do to have a convincing argument?
 - 1 **Tell the institutional story**: why is z as-if random? Why would it affect y only through x ?
 - 2 **Balance checks**: is z unrelated to observable pre-determined characteristics?
 - 3 **Placebo tests**: look for an effect of z on y in a group where x cannot respond to z . If there is one, exclusion fails.
 - 4 **Over-identification test** when $L > K$ (the J test, see the appendix): it checks that the instruments *agree with each other*. Necessary, not sufficient: if all instruments share the same flaw, J does not detect it.

In Practice

- ▶ **Put the controls in both stages** : every exogenous regressor of the second stage must be in $\mathbf{Z}$ (hence in the first stage).
- ▶ **Always report** :
 - the **first stage** (the coefficients on the excluded instruments $\mathbf{Z}_1$, and the F -statistic testing that they are all zero),
 - the **reduced form** (effect of the instruments directly on $\mathbf{y}$),
 - the OLS estimate, for comparison.

7. An application: Angrist and Krueger (1991)

Angrist and Krueger (1991)

Concept	In the paper
Outcome y	log (weekly) earnings
Endogenous regressor $\mathbf{X}_1$ ($K_1 = 1$)	years of schooling
Exogenous regressors $\mathbf{X}_2$	constant, year-of-birth dummies (and other controls)
Omitted variable $\mathbf{u}$	ability, motivation, family background
Excluded instruments $\mathbf{Z}_1$	quarter-of-birth dummies (and interactions in some specifications)
Coefficient of interest	return to one extra year of schooling

Back to the answers you gave at the beginning of the session:

- 1 Why is OLS a problem here?
- 2 Why should birth date be an instrument?
- 3 Is it relevant? Is it exogenous?
- 4 What do they find, and for whom?

The problem: OLS on the Mincer equation

The Mincer earnings function (Mincer 1974, NBER):

$$\ln w_i = \alpha + \rho S_i + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

- ▶ Same model as before, in the paper's notation: S_i = years of schooling (our educ_i), ρ = the return to schooling (our β), X_i = potential experience.
- ▶ As before, $\varepsilon_i = \gamma u_i + v_i$ contains ability u_i (and motivation, family background. . .).
- ▶ Ability plausibly raises both schooling and wages: $\gamma > 0$ and $\text{Cov}(S_i, u_i) > 0$.
- ▶ By the omitted variable bias formula, OLS overstates ρ (*ability bias*).

What is needed: an Instrument

*A source of variation in schooling that has **nothing to do with ability**.*

The natural experiment: two rules

- ▶ **Rule 1, school entry**: school districts typically require a child to have turned six by January 1 of the year in which he or she enters school.
- ▶ **Rule 2, compulsory schooling**: students may leave school once they reach the legal age (16–17, depending on the state).

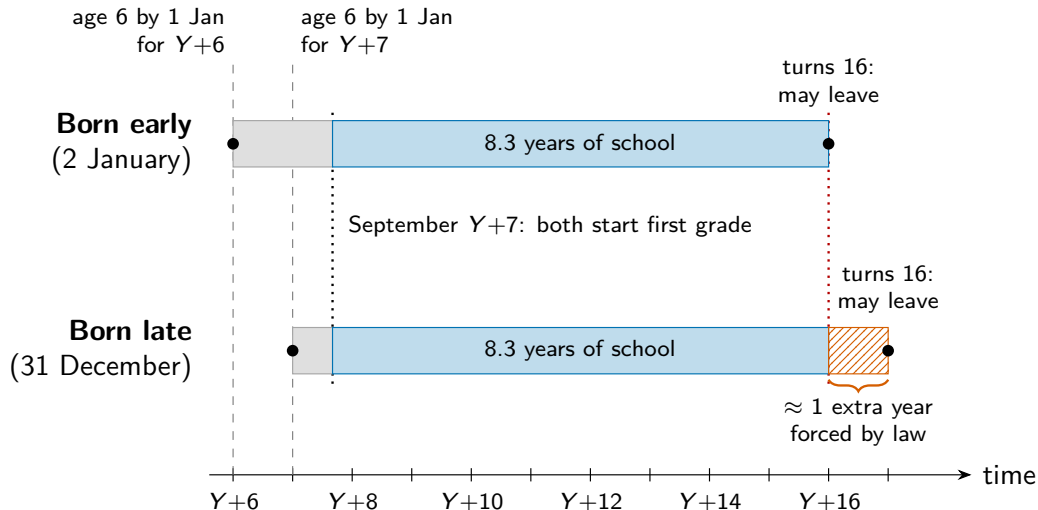
Hence:

- ⇒ Children **born early** in the year enter school **older** and reach the legal dropout age **earlier in their schooling**.
- ⇒ Children **born late** in the year enter school **younger**, and the law forces them to stay **longer**.

The instrument

Z = quarter of birth. It changes schooling through the rules, and is (arguably) unrelated to ability.

Two trajectories, same school entry, different legal exit



Relevance: does birth date move schooling?

First stage (relevance of the instrument): is $\pi_1 \neq 0$ in

$$S_i = \pi_0 + \pi_1 Z_i + \eta_i, \quad Z_i = 1 \text{ if born late in the year, } 0 \text{ if born early}$$

- ▶ Angrist and Krueger (1991): men born late in the year have slightly more schooling on average than men born early ($\pi_1 > 0$). The effect is **nonzero but small**.
- ▶ Robustness to convince that the *law* is the channel, and not seasonality in general:
 - 1 **College and graduate school**: no seasonal pattern. The law compels nobody beyond high school, so the birth-date gap should vanish there. It does.
 - 2 **Enrollment of 15- and 16-year-olds**: the birth-date gap in enrollment at 16 is larger in states that let 16-year-olds leave than in states that require attendance until 17.
- ▶ Hence, relevance holds: birth date shifts schooling, and the compulsory schooling law explains why.

Reduced form, first stage, and the Wald ratio

Compare men born early ($Z = 0$) with men born late in the year ($Z = 1$):

$$\hat{\rho}_{IV} = \frac{E[\ln w \mid Z = 1] - E[\ln w \mid Z = 0]}{E[S \mid Z = 1] - E[S \mid Z = 0]} = \frac{\text{gap in earnings}}{\text{gap in schooling}}$$

- ▶ **Numerator** (reduced form): late-born men also earn slightly more.
- ▶ **Denominator** (first stage): late-born men have slightly more schooling.
- ▶ Because both gaps are **small**, the ratio is delicate (the **weak instrument** problems arise).
- ▶ In the paper, the actual estimation is by **2SLS**: year-of-birth controls in both stages, several quarter-of-birth instruments, so $L_1 > K_1$ (*over-identified*).

Exogeneity: independence and exclusion

Season of birth might affect earnings through channels other than schooling (relative age in class, family background, health...).

The authors' placebo test

- ▶ College graduates are not constrained by compulsory schooling: for them, the law channel is switched off, while any other channel would still operate.
- ▶ If season of birth affected earnings for other reasons, there would be a gap among college graduates.
- ▶ They find **no relationship** between earnings and season of birth for this group.

Caveat

Indirect evidence: college graduates are a selected group, and a null result may also reflect low statistical power.

Result and interpretation

- ▶ Men born 1930–39 (1980 Census), several quarter-of-birth instruments: the 2SLS return to one extra year of schooling is between **6% and 9%** depending on the controls, against 6.3% to 7.1% for OLS.
- ▶ IV is **close to OLS**, and often slightly above. Recall that OLS was expected to be biased upward by ability: this suggests that ability bias may be small, or offset by another bias (e.g., measurement error in schooling).
- ▶ **Whose return is it?** The estimate applies to the **compliers** (LATE):
 - men who would have left school at the legal age, but were held in school by the law,
 - not the always-takers (e.g., those who would go to college anyway), whose schooling does not respond to birth date.
- ▶ **Monotonicity**: being born later in the year should never *reduce* schooling.

What happened next: the instrument under scrutiny

Relevance: weak instruments

- ▶ [Bound et al. \(1995, JASA\)](#): season of birth explains very little of schooling, especially with many quarter-of-birth $\times$ year-of-birth interactions as instruments.
- ▶ Replacing it with **randomly generated** data still gives estimates that look precise.
- ▶ Weak instruments $\Rightarrow$ bias towards OLS, misleading confidence intervals.
- ▶ Now standard practice: report the first-stage F statistic.

Exogeneity: is season of birth random?

- ▶ [Buckles and Hungerman \(2013, REStat\)](#): winter births are more frequent among teenage, unmarried or less-educated mothers.
- ▶ Family background explains up to half of the link between season of birth and adult outcomes.
- ▶ Season of birth may affect outcomes through **other channels**: exclusion and independence are at risk.

Lesson for your reading

- ▶ Every empirical paper rests on an **identifying assumption** (an assumption that allows you to estimate a causal effect from observed data).
 - Find it, and ask what could violate it.
- ▶ Look for what the literature said **after** the paper: comments, replies, replications.
- ▶ From now on, each reading assignment includes one **critical question**.

Take-aways

- ▶ **IV** solves endogeneity by using an **external source of variation**: the instrument.
- ▶ A valid instrument is **relevant** (testable) and **exogenous** (independence + exclusion: an argument, not a test).
- ▶ $\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ when $L = K$; $\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}$ in general; with one instrument, reduced form / first stage.
- ▶ Weak instruments amplify small violations of exogeneity.
- ▶ IV estimates a **local** effect (LATE) for the compliers.
- ▶ Angrist and Krueger (1991, QJE): birth date + entry rule + compulsory schooling $\Rightarrow$ variation in schooling unrelated to ability, whose credibility relies on the story and on the checks.

8. Your turn: IV detective

Your turn: IV detective

Four groups, four research questions. Each group gets one case.

1 Without the paper (12 min): read the problem.

- Why would OLS be biased? In which direction?
- Which variable could serve as an **instrument**? Is it relevant? Can you defend the exclusion restriction?

2 With the paper (15 min): read the introduction.

- Which instrument did the authors use? How do they argue relevance and exclusion?
- Was your instrument better or worse? Why?

3 Report (4 min per group): the problem, your instrument, theirs, one threat.

Do not look the paper up before step 2: the point is to find an instrument yourself.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: the instruments

Case	Instrument Z	Why Z moves X	What IV finds
1. Children	first two children of the same sex	parents prefer a mixed sibling-sex composition	effect smaller than OLS; none for husbands
2. Electricity	land gradient	steeper land raises the cost of a connection	female employment +9 to 9.5 points
3. Dams	river gradient	irrigation dams need a moderate, neither flat nor steep, gradient	production up downstream; poverty up in the dam's district
4. Pollution	daily changes in wind direction	wind brings pollution produced elsewhere	effect larger than OLS

- ▶ OLS can be biased **upward** (case 1) or **toward zero** (case 4; possible reasons: noisy measures of exposure, avoidance behaviour).
- ▶ Note the controls: the instrument has to be as good as random **conditional on** fixed effects, weather, geography.

IV detective: what do the four instruments have in common?

Case	Instrument Z	A threat to exclusion
1. Children	same-sex siblings	same-sex siblings share clothes and rooms: a cheaper third child
2. Electricity	land gradient	flat areas may have better roads, markets, farming
3. Dams	river gradient	gradient affects farming directly (soil, slopes)
4. Pollution	wind direction	wind also brings temperature, rain, other pollutants

- ▶ The instruments come from **nature, geography, or chance**, not from the people studied.
- ▶ Relevance is **shown** (first stage); exclusion is **argued**, and it is where researchers push.

IV detective: what do the four instruments have in common?

Case	Instrument Z	A threat to exclusion
1. Children	same-sex siblings	same-sex siblings share clothes and rooms: a cheaper third child
2. Electricity	land gradient	flat areas may have better roads, markets, farming
3. Dams	river gradient	gradient affects farming directly (soil, slopes)
4. Pollution	wind direction	wind also brings temperature, rain, other pollutants

- ▶ The instruments come from **nature, geography, or chance**, not from the people studied.
- ▶ Relevance is **shown** (first stage); exclusion is **argued**, and it is where researchers push.

9. Next session

Reading assignment for Session 4 (October 28, 2026)

Topic: **The credibility revolution: why research design matters**

Read:

- ▶ [Angrist and Pischke \(2010, JEP\)](#): The credibility revolution in empirical economics: how better research design is taking the con out of econometrics.

This is not an empirical paper: the authors build an **argument** about how economists should produce evidence. Read it as such:

- ▶ what is claimed,
- ▶ which examples are used to support each claim,
- ▶ which research designs are praised, and on which assumptions they rely.

If you summarised [Ashenfelter and Card \(1985\)](#) for Session 2, bring it back: we use it for the difference-in-differences part.

Reading assignment for Session 4: what to prepare

Part 1. Map the argument

- ▶ List the main claims of the paper, one short sentence per claim.
- ▶ For each claim, say in one sentence what supports it (reasoning, evidence, example).

Part 2. Summary (approximately 1 page)

- 1 What is the paper about?
- 2 Why is this an important question?
- 3 What is the argument? What evidence and examples do the authors use?
- 4 What are the main conclusions?
- 5 Why is this subject of interest to an economist?

Part 3. Critical question

- ▶ Pick one research design praised by the authors (randomized trial, natural experiment, instrumental variables, difference-in-differences, regression discontinuity).
- ▶ Which assumption does it rely on? Give one situation in which this assumption would fail.

10. Appendix

References I

- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106: 979–1014, doi: 10.2307/2937954.
- Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24: 3–30, doi: 10.1257/jep.24.2.3.
- Ashenfelter, O. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics* 67: 648, doi: 10.2307/1924810.
- Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443–450, doi: 10.2307/2291055.
- Buckles, K. S. and Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *The Review of Economics and Statistics* 95: 711–724, doi: 10.1162/REST_a_00314.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475, doi: 10.2307/2951620.
- Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings*. NBER, 41–63.
- Stock, J. H. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*. Cambridge University Press. 80–108, doi: 10.1017/cbo9780511614491.006.

12. Backup: more on instrumental variables

Back to the wage equation

- ▶ The model we want to estimate:

$$\ln w_i = \alpha + \beta \text{educ}_i + \varepsilon_i, \quad \varepsilon_i = \gamma u_i + v_i,$$

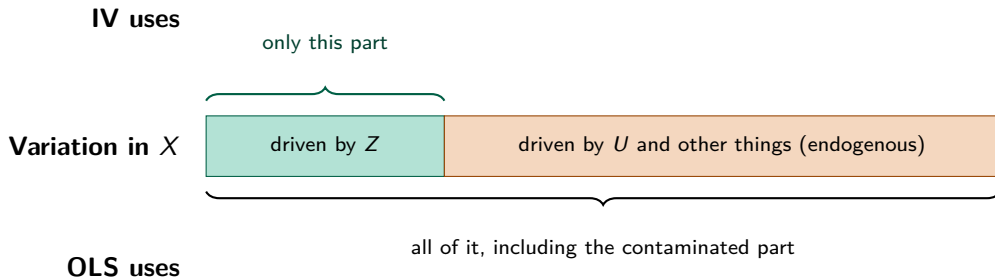
where u_i is **ability** (unobserved) and v_i is a wage shock.

- ▶ OLS converges to

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{\text{Cov}(\text{educ}_i, \varepsilon_i)}{\text{Var}(\text{educ}_i)}.$$

- ▶ If more able people study longer, $\text{Cov}(\text{educ}_i, u_i) > 0$, so $\text{Cov}(\text{educ}_i, \varepsilon_i) \neq 0$: OLS is **inconsistent**, and more data does not help.
- ▶ The same happens with measurement error: the noise in educ_i is also in ε_i .
- ▶ The problem: **part** of the variation in schooling comes from ability. OLS cannot separate the effect of schooling from the effect of ability.

Use Only the “Clean” Part of the Variation in X



- ▶ An **instrument** Z is a variable that shifts X but has no other route to y .
- ▶ The variation in X induced by Z is **as good as random** with respect to U : we follow it, and only it, to see what happens to y .

OLS as a Special Case of IV ($\mathbf{Z} = \mathbf{X}$)

Take the IV recipe and use the regressors as their own instruments, $\mathbf{z}_i = \mathbf{x}_i$:

	IV (general $\mathbf{Z}$)	OLS ($\mathbf{Z} = \mathbf{X}$)
Moment condition	$E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}$	$E[\mathbf{x}_i \varepsilon_i] = \mathbf{0}$
Estimator	$(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$	$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
Rank condition	$\text{rank } E[\mathbf{z}_i \mathbf{x}_i'] = K$	$\text{rank } E[\mathbf{x}_i \mathbf{x}_i'] = K$

- ▶ The rank condition is automatic for OLS (a regressor is perfectly correlated with itself): all the difficulty lies in exogeneity.
- ▶ If a regressor is endogenous, it cannot be its own instrument: we need another variable $\mathbf{z}$ that moves it, and that is exogenous.
- ▶ Exogenous regressors can (and do) instrument themselves: $\mathbf{Z} = [\mathbf{Z}_1 \ \mathbf{X}_2]$.

2SLS: Consistency and Variance

Consistency. Plugging $\mathbf{y} = \mathbf{X}\beta + \varepsilon$:

$$\hat{\beta}_{2SLS} = \beta + (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\varepsilon \xrightarrow{P} \beta \quad \text{because} \quad \frac{1}{n}\mathbf{Z}'\varepsilon \xrightarrow{P} \mathbf{0}.$$

Variance (homoskedastic case): $\widehat{\text{Var}}(\hat{\beta}_{2SLS}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}$, with

$$\hat{\sigma}^2 = \frac{1}{n}\hat{\varepsilon}'\hat{\varepsilon}, \quad \hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS} \quad (\text{computed with } \mathbf{X}, \text{ not } \hat{\mathbf{X}}).$$

- ▶ Running the second stage by hand uses $\mathbf{y} - \hat{\mathbf{X}}\hat{\beta}$: **wrong standard errors**. Use a dedicated command. Heteroskedasticity-robust versions exist (sandwich form).
- ▶ Since $\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{P}_Z\mathbf{X} + \mathbf{X}'\mathbf{M}_Z\mathbf{X}$ with $\mathbf{M}_Z = \mathbf{I} - \mathbf{P}_Z$ positive semi-definite, $(\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1} \succeq (\mathbf{X}'\mathbf{X})^{-1}$: IV is **always less precise than OLS**.

Over-Identification Test (Sargan–Hansen J)

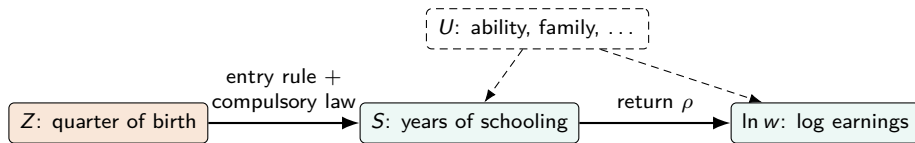
- ▶ With $L > K$, there are more moment conditions than parameters: we can check whether they are **compatible**.
- ▶ With $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}_{2SLS}$ (homoskedastic case):

$$J = \frac{\hat{\varepsilon}' \mathbf{P}_Z \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon} / n} \stackrel{a}{\sim} \chi^2(L - K) \text{ under } H_0 : E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}.$$

- ▶ Intuition: if all instruments are valid, the 2SLS residuals should be (nearly) unexplained by $\mathbf{Z}$.
- ▶ It checks that the instruments *agree with each other*: necessary, not sufficient (if all instruments share the same flaw, J does not detect it). With heterogeneous effects, valid instruments can also disagree, because they move different compliers.

13. Backup: alternative slides on Angrist and Krueger (1991)

The IV diagram of the paper

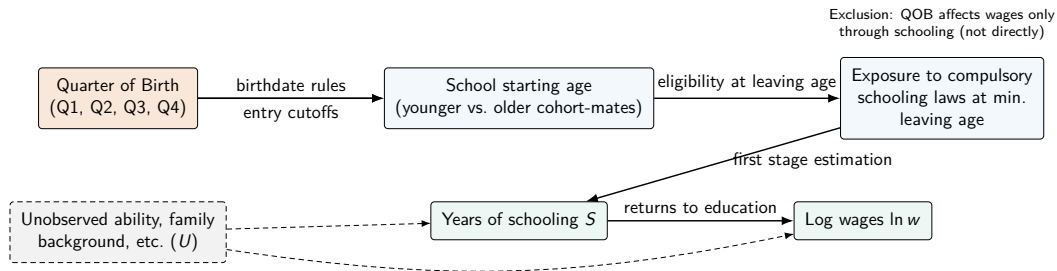


Independence: Z as-if random w.r.t. U

Exclusion: $Z \not\rightarrow \ln w$ except via S

- The **story** told in the paper is what makes Z a candidate instrument.

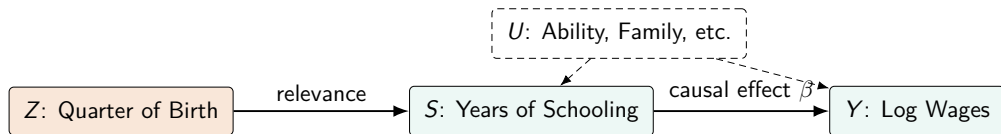
Angrist and Krueger (1991): Quarter of birth as an instrument



Idea: $\Rightarrow$

- ▶ Quarter of birth shifts school start/exit timing.
- ▶ This induces variation in year of schooling that is unrelated to U (unobserved characteristics).

Angrist and Krueger (1991): IV DAG and 2SLS



Exclusion: $Z \not\rightarrow Y$ except via S

The estimation is done using two-stage least squares:

► **First stage:**

$$S_i = \pi_0 + \pi_1 Z_i + \pi' \mathbf{X}_i + u_i$$

► **Second stage:**

$$Y_i = \alpha + \beta \hat{S}_i + \gamma' \mathbf{X}_i + \varepsilon_i$$

IV assumptions:

- **Relevance:** $\pi_1 \neq 0$ (the instrument must be correlated with the endogenous regressor).
- **Exclusion:** $Z \perp \varepsilon_i \mid \mathbf{X}_i$ (the instrument must affect the outcome only through the regressor)
- **Independence:** Z as-if random wrt unobservables

Angrist and Krueger (1991) (1/2): weak instruments

- ▶ Bound et al. (1995, JASA) re-examine Angrist and Krueger (1991).
- ▶ Quarter of birth explains **very little** of the variation in schooling (the first stage is **weak**), especially when many interactions are used as instruments.
- ▶ Their striking test: they **replace quarter of birth with randomly generated data**, unrelated to schooling or wages,
 - and still obtain estimates that **look precise**.
- ▶ With weak instruments, the IV estimate is biased towards the OLS estimate, and standard confidence intervals are misleading.
- ▶ Their recommendation, now standard practice: **always report the first-stage F statistic** (and partial R^2).

⇒ **Which IV assumption is under pressure here?**

Angrist and Krueger (1991) (2/2): is season of birth random?

- ▶ Buckles and Hungerman (2013, REStat) use U.S. birth certificate and census data.
- ▶ Children born in **winter** are disproportionately born to mothers who are **teenagers**, **unmarried**, or **without a high school degree**.
- ▶ Family background controls explain **up to half** of the relationship between season of birth and adult outcomes.
- ▶ Implication: season-of-birth instruments **may produce inconsistent estimates**.

⇒ **Which IV assumption is under pressure here? Look back at the DAG: which arrow is missing?**