

Week 5: Weak Instruments

Compulsory schooling · PP5001 · Martinmas 2026

[Class slides](#) · [Slides PDF](#) · [Notes PDF](#) · [Quarto source](#) · [Problem Set 2](#)

Learning Objectives

By the end of this topic, you should be able to:

- Distinguish unbiasedness from consistency using the sampling distributions of estimators.
- Explain how a weak first stage can magnify violations of instrument validity and produce substantial finite-sample bias towards OLS.
- Interpret the first-stage F statistic and partial R^2 , and explain why the second-stage results alone cannot establish instrument strength.
- Assess the quarter-of-birth identification strategy and compare Wald and IV estimates under different sets of controls and instruments.

Before class

Read these notes and **Angrist and Krueger (1991)** and **Bound, Jaeger, and Baker (1995)**. Both papers are on Moodle. We will examine the compulsory-schooling argument, the first-stage evidence, and the interpretation of the IV estimates. We also use **Figure 1 in Bound and Jaeger (2000)** to discuss the exclusion restriction.

Textbook references

- **The Effect:** [Chapter 19, Instrumental Variables](#), particularly instrument relevance and weak instruments.
- **Causal Inference: The Remix:** [Chapter 7, Instrumental Variables](#), particularly the discussion of weak instruments.
- **Mastering 'Metrics:** Chapter 3, *Instrumental Variables*. [Companion resources](#). [Mastering Econometrics videos at MRU](#).
- **Mostly Harmless Econometrics — advanced reading:** Section 4.6 on IV with weak instruments and finite-sample behaviour.

1. Unbiasedness and consistency

Repeated samples

An estimate changes when the sample changes. Imagine repeatedly drawing samples of the same size from the same population and estimating the same relationship. The distribution of those estimates is the **sampling distribution**. Its centre and spread tell us different things about an estimator.

An estimator is **unbiased** if its expectation equals the true parameter:

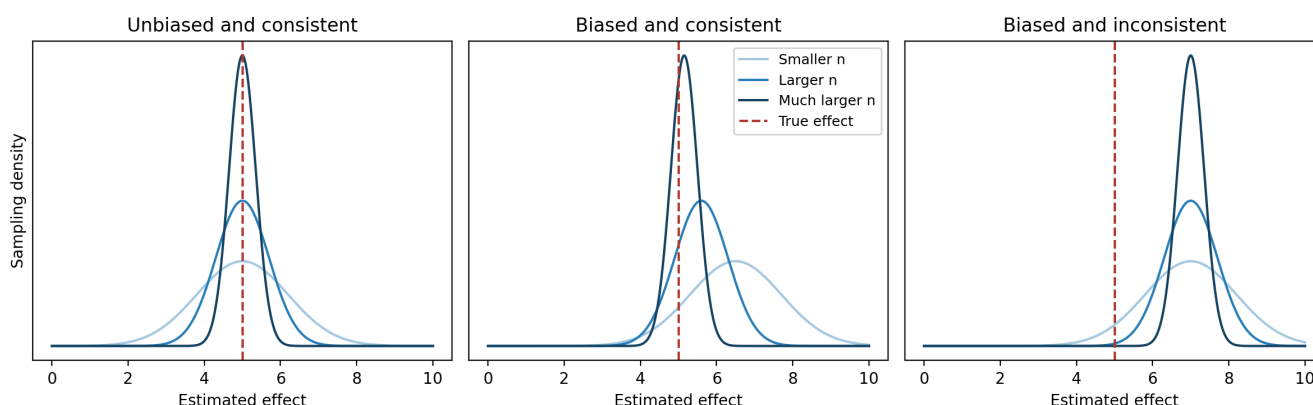
$$E[\hat{\beta}] = \beta.$$

This is a statement about the average over repeated samples of a particular size. It does not mean that any individual estimate equals the truth. An unbiased estimator can have a wide sampling distribution.

An estimator is **consistent** if, as the sample grows, its sampling distribution concentrates around the true parameter. For any fixed margin of error, the probability of falling outside that margin eventually becomes arbitrarily small. We can understand this graphically without deriving an asymptotic distribution.

Watch the narrated animation on unbiasedness and consistency (5½ minutes).

These distinctions matter directly for instrumental variables. OLS with an omitted variable behaves like the sample mean plus 20: more data gives a more precise estimate of the wrong quantity. IV with a valid, strong instrument behaves more like the sample mean plus $20/n$: biased in finite samples, but consistent. With a weak instrument, that finite-sample bias can remain large even in very big samples, and it pulls IV towards OLS. That is the problem we turn to next.



Schematic sampling distributions. The dashed line marks the true effect. All panels use the same scales.

The probability limit is the value that the estimate approaches as the sample size grows: the probability that the estimate differs from that value by more than any fixed positive amount goes to zero.

We write $\text{plim } \hat{\beta} = b$. An estimator is **consistent** if its probability limit equals the true parameter, $b = \beta$.

The curves above are schematic normal densities chosen to illustrate the properties. The left panel is centred on the truth at every sample size and becomes more concentrated as the sample grows. In the middle panel the estimator is biased at each of the displayed sample sizes, but the centre moves towards the truth while the spread shrinks. In the right panel the distribution becomes increasingly concentrated around the wrong value. More observations make that estimator more precise about the wrong target.

i Your turn

Which panels illustrate unbiasedness? Which illustrate consistency? Why is a smaller standard error insufficient to establish that an estimator is estimating the causal effect?

OLS and IV

When an omitted variable is correlated with the explanatory variable, OLS can converge to a slope different from the causal effect. As the sample size increases, the OLS estimate becomes more precise, but omitted-variable bias remains. We can therefore obtain a very precise estimate of the wrong population parameter. Under the IV assumptions, a relevant instrument allows IV to converge to the causal effect. Finite-sample IV estimates can nevertheless be biased and highly variable. With weak instruments, this distinction matters even in datasets containing hundreds of thousands of observations.

The concentration around a point describes the distribution of the **unscaled estimator**. A centred and standardised estimator is a different object. Our graphs here concern the coefficient estimates themselves.

2. What goes wrong when relevance is weak?

We use the notation in Bound, Jaeger, and Baker. Let x_i denote schooling, y_i log earnings, z_i an excluded instrument, and ϵ_i the outcome disturbance. Initially omit additional controls:

$$x_i = \pi_0 + \pi_1 z_i + \nu_i,$$

$$y_i = \beta_0 + \beta x_i + \epsilon_i.$$

Schooling x_i can be endogenous because it is related to unobserved determinants of earnings in ϵ_i . An instrument must be related to schooling and orthogonal to the outcome disturbance. This week asks what happens when the relationship between the instrument and schooling is very small.

For a binary instrument the Wald estimator is

$$\hat{\beta}^{IV} = \frac{\bar{y}_{z=1} - \bar{y}_{z=0}}{\bar{x}_{z=1} - \bar{x}_{z=0}}.$$

A small denominator makes the estimated effect sensitive to small changes in the estimated group means. Sampling variation can dominate the schooling difference. If the population schooling difference is zero, the instrument fails relevance entirely.

Deriving the inconsistency of OLS and IV

Start with $y_i = \beta_0 + \beta x_i + \epsilon_i$. Substituting into the OLS formula gives

$$\hat{\beta}_{OLS} = \beta + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) \epsilon_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2}.$$

The probability limit, written plim , is the value around which an estimator concentrates as the sample grows. Under the usual sampling conditions, the sample covariance and variance converge to their population counterparts:

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{\sigma_{x,\epsilon}}{\sigma_x^2}.$$

Here $\sigma_{x,\epsilon} = \text{Cov}(x_i, \epsilon_i)$ and $\sigma_x^2 = \text{Var}(x_i)$.

Let $\hat{x}_i$ be the first-stage fitted value. With an intercept, $\bar{\hat{x}} = \bar{x}$ and the first-stage residual is orthogonal to $\hat{x}_i$.

$$\hat{\beta}_{IV} = \frac{\sum_i (\hat{x}_i - \bar{x}) y_i}{\sum_i (\hat{x}_i - \bar{x}) x_i} = \beta + \frac{\frac{1}{n} \sum_i (\hat{x}_i - \bar{x}) \epsilon_i}{\frac{1}{n} \sum_i (\hat{x}_i - \bar{x})^2}.$$

Therefore,

$$\text{plim } \hat{\beta}_{IV} = \beta + \frac{\sigma_{\hat{x},\epsilon}}{\sigma_{\hat{x}}^2}.$$

The population moments refer to the population first-stage fitted component.

What Does R^2 Measure?

R^2 is the proportion of the sample variation in the dependent variable explained by the regression.

In the first stage, the dependent variable is x_i . For OLS with an intercept,

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{\sum_i (\hat{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} = 1 - \frac{\sum_i \hat{\nu}_i^2}{\sum_i (x_i - \bar{x})^2}.$$

Here $\hat{\nu}_i = x_i - \hat{x}_i$. An R^2 of **0.01** means the instruments explain **1%** of the variation in x_i .

The population counterpart is $R_{x,z}^2 = \sigma_{\hat{x}}^2 / \sigma_x^2$, the share of population variance explained by the instruments.

The total variation is measured by the sum of squared deviations of x_i from its sample mean. The explained variation is the corresponding sum for the fitted values. The residual sum of squares measures the variation left unexplained. With an intercept, OLS splits total variation into these two parts:

$$\sum_i (x_i - \bar{x})^2 = \sum_i (\hat{x}_i - \bar{x})^2 + \sum_i \hat{\nu}_i^2.$$

Thus $R^2 = 0$ means the regression explains none of the sample variation, while $R^2 = 1$ means it fits every observation exactly. In the population expression, $\sigma_{\hat{x}}^2$ refers to the variance of the population first-stage fitted component. This is the variance ratio that enters the next derivation. Later, when we include controls, partial R^2 will measure how much of the variation remaining after those controls the excluded instruments explain.

Relative Inconsistency of IV and OLS

Divide the two discrepancies from the true effect:

$$\frac{\text{plim } \hat{\beta}_{IV} - \beta}{\text{plim } \hat{\beta}_{OLS} - \beta} = \frac{\sigma_{\hat{x},\epsilon} / \sigma_{\hat{x}}^2}{\sigma_{x,\epsilon} / \sigma_x^2} = \frac{\sigma_{\hat{x},\epsilon} / \sigma_{x,\epsilon}}{\sigma_{\hat{x}}^2 / \sigma_x^2}.$$

The denominator is the population first-stage R^2 :

$$\boxed{\frac{\text{plim } \hat{\beta}_{IV} - \beta}{\text{plim } \hat{\beta}_{OLS} - \beta} = \frac{\sigma_{\hat{x},\epsilon} / \sigma_{x,\epsilon}}{R_{x,z}^2}}.$$

A small R^2 magnifies the consequences of imperfect instrument validity.

One Instrument: Expressing the Result in Correlations

With one instrument,

$$\text{plim } \hat{\beta}_{IV} - \beta = \frac{\sigma_{z,\epsilon}}{\sigma_{z,x}} = \frac{\rho_{z,\epsilon}}{\rho_{z,x}} \frac{\sigma_{\epsilon}}{\sigma_x},$$

while

$$\text{plim } \hat{\beta}_{OLS} - \beta = \rho_{x,\epsilon} \frac{\sigma_{\epsilon}}{\sigma_x}.$$

Dividing cancels $\sigma_{\epsilon} / \sigma_x$:

$$\frac{\text{plim } \hat{\beta}_{IV} - \beta}{\text{plim } \hat{\beta}_{OLS} - \beta} = \frac{\rho_{z,\epsilon} / \rho_{x,\epsilon}}{\rho_{x,z}}.$$

When Can IV Be Worse Than OLS?

IV has a larger absolute inconsistency when

$$\left| \frac{\rho_{z,\epsilon}}{\rho_{x,\epsilon}} \right| > |\rho_{x,z}|.$$

An instrument can be much less correlated with the disturbance than x_i is, yet still produce a larger inconsistency if it barely predicts x_i .

We can measure the instrument's predictive power. We must assess its validity using the institutional setting and supporting evidence.

These expressions assume a fixed, nonzero population first stage and finite population moments. They describe what happens with imperfect validity even as the sample becomes arbitrarily large. The relative-inconsistency ratios also assume that OLS inconsistency is nonzero. If the instrument is valid, $\sigma_{\hat{x},\epsilon} = 0$ and IV is consistent.

With several instruments, the population first-stage fitted component is the linear combination of instruments that predicts x_i . Its variance divided by the variance of x_i gives $R_{x,z}^2$. In the sample, the identity in the IV derivation follows by writing $x_i = \hat{x}_i + \hat{\nu}_i$ and using $\sum_i (\hat{x}_i - \bar{x}) \hat{\nu}_i = 0$. With additional controls, the same argument applies after removing their linear contribution from the variables. This motivates the partial R^2 below.

Finite-sample bias with valid instruments

Now restore instrument validity. The population first-stage fitted component is uncorrelated with ϵ_i . With a fixed, nonzero first stage, IV is consistent. To study finite-sample bias, however, we must consider the estimated first stage in repeated samples of a given size:

$$\hat{\beta}_{IV} - \beta = \frac{\sum_i (\hat{x}_i - \bar{x}) \epsilon_i}{\sum_i (\hat{x}_i - \bar{x})^2}.$$

Both parts of this ratio are random. Population orthogonality does not imply that its expectation is zero. In particular, the first-stage fitted values depend on the same observations used to estimate the outcome equation. The expectation of the ratio generally differs from the ratio of expectations.

Why the bias is related to OLS

Write the first stage with K excluded instruments as

$$x_i = \pi_0 + \sum_{j=1}^K \pi_j z_{ji} + \nu_i.$$

The instruments are exogenous, so the correlation between x_i and the outcome disturbance comes from ν_i :

$$\text{Cov}(x_i, \epsilon_i) = \text{Cov}(\nu_i, \epsilon_i) = \sigma_{\epsilon,\nu}.$$

In a particular sample, the instruments will also predict some of the variation in ν_i by chance. The fitted first stage consequently combines the instruments' true predictive contribution with some fitted noise. Because ν_i is related to ϵ_i , this fitted noise can introduce the same endogeneity that affects OLS into the second stage. When the instruments are weak, the fitted noise can be a large share of the variation in $\hat{x}_i$.

This explains the pull towards OLS in overidentified 2SLS. Adding instruments gives the first-stage regression more opportunities to fit noise. If they add little true predictive power, the finite-sample bias can increase.

A large sample helps when the population first stage is fixed and nonzero: the true predictive contribution accumulates as observations are added. If the population first stage is zero, more observations cannot create identification.

3. First-stage evidence

Partial R^2

Suppose we include L controls w_{ki} and three excluded instruments:

$$x_i = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i} + \pi_3 z_{3i} + \sum_{k=1}^L \delta_k w_{ki} + \nu_i.$$

The overall first-stage R^2 includes the predictive contribution of the controls. Our question is how much the **excluded instruments** add after accounting for those controls. Estimate a restricted regression containing the controls alone, and an unrestricted regression that adds the excluded instruments. Then

$$\text{partial } R^2 = \frac{R_U^2 - R_R^2}{1 - R_R^2} = \frac{SSR_R - SSR_U}{SSR_R},$$

where SSR is the sum of squared residuals. This measures the proportion of the schooling variation left after the controls that is explained by adding the excluded instruments. A value of 0.00012 means 0.012%, not 1.2%.

The first-stage F statistic

Test the joint hypothesis

$$H_0 : \pi_1 = \pi_2 = \pi_3 = 0.$$

The test concerns the excluded instruments. The overall regression test also tests the coefficients on the controls and answers a different question. With one excluded instrument, the corresponding first-stage F is the square of its t statistic.

Report the first-stage coefficients, their standard errors, the excluded-instrument F , and partial R^2 alongside the IV estimates. Inspect these for each control set because adding controls changes the variation available to identify the effect.

The $F < 10$ rule of thumb flags a possible weak-instrument problem. An F above 10 does not guarantee reliable conventional IV inference. The interpretation of critical values depends on the model and the assumptions about the disturbances. A heteroskedasticity-robust joint test and the conventional homoskedastic test need not give identical statistics.

Relative bias and the first-stage F

Deriving the finite-sample bias of IV is complicated. Interested readers can find the argument in Bound, Jaeger, and Baker (1995).

Under certain assumptions, the bias of IV relative to the bias of OLS is approximately

$$\frac{\text{Bias}(\hat{\beta}_{IV})}{\text{Bias}(\hat{\beta}_{OLS})} \approx \frac{1}{F}.$$

For example, $F = 10$ suggests an IV bias roughly one-tenth of the OLS bias.

This is why we report the first-stage F alongside the IV estimate: it helps us assess how serious finite-sample bias may be.

The ratio describes the share of OLS bias remaining in IV. For example, if OLS has a bias of 2, a relative bias of 10% corresponds to an IV bias of about 0.2.

This approximation provides guidance for overidentified IV under the assumptions used in the bias calculation. Its accuracy depends on the model and instrument strength. The first-stage F is a diagnostic, and the approximation does not guarantee a particular amount of bias in an application.

Inference

When instruments are weak, the usual IV coefficient divided by its reported standard error need not have the distribution used to calculate conventional p-values. Heteroskedasticity-consistent standard errors address heteroskedasticity, but do not solve weak identification. A seemingly precise second-stage result can be misleading.

One approach is to test a proposed effect directly. If $\beta = b$ and the instrument is valid, then $y_i - bx_i$ should have no relationship with the excluded instruments after accounting for the controls. An **Anderson-Rubin test** checks this implication. Using a version appropriate to the disturbance structure allows inference that remains valid with weak instruments. A confidence set contains the values of b that are not rejected. It may be very wide or unbounded when the data provide little information about the effect.

4. The simulation illustrations

How the data are generated

We retain the population relationships from the original lecture illustration:

$$x_i = 1 + \pi_1 z_i + 2w_i + \nu_i,$$

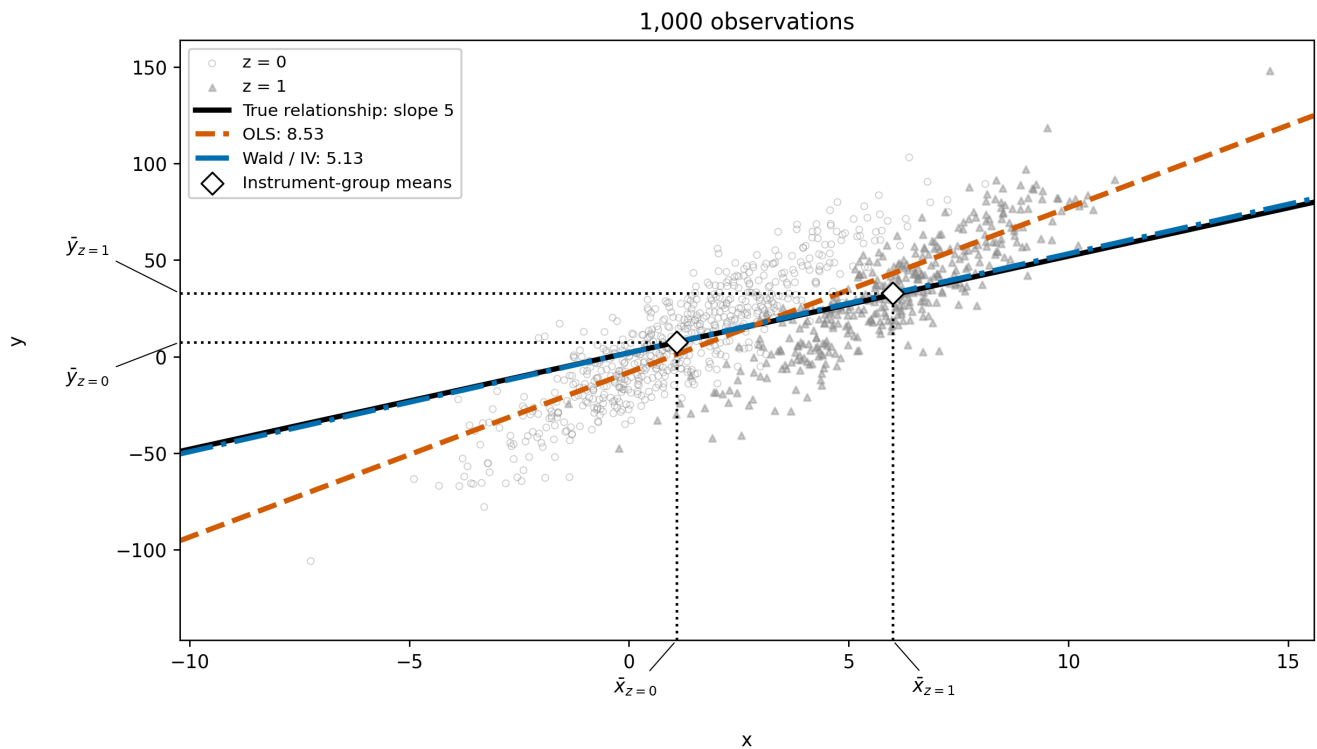
$$y_i = 2 + 5x_i + 20w_i + e_i.$$

The true causal slope is **5**. The omitted variable w_i raises both x_i and y_i . Draw w_i and ν_i independently from standard normal distributions and e_i independently from a normal distribution with standard deviation 10. The binary instrument has probability 0.4 of equalling one and is independent of these draws.

Set $\pi_1 = 5$ for the strong-instrument example. For the other two examples set $\pi_1 = 0$, as in the original slides. These illustrate the limiting case of weak relevance: **no population first stage**. We use 1,000 observations in the first two examples and 329,509 in the third.

Each figure shows the true relationship, the sample OLS line and the sample Wald/IV line, with the two instrument-group means marked. Every observation is plotted. The three figures use the same axis scales. Dotted lines project each group mean onto both axes, showing how the first-stage difference collapses when relevance fails.

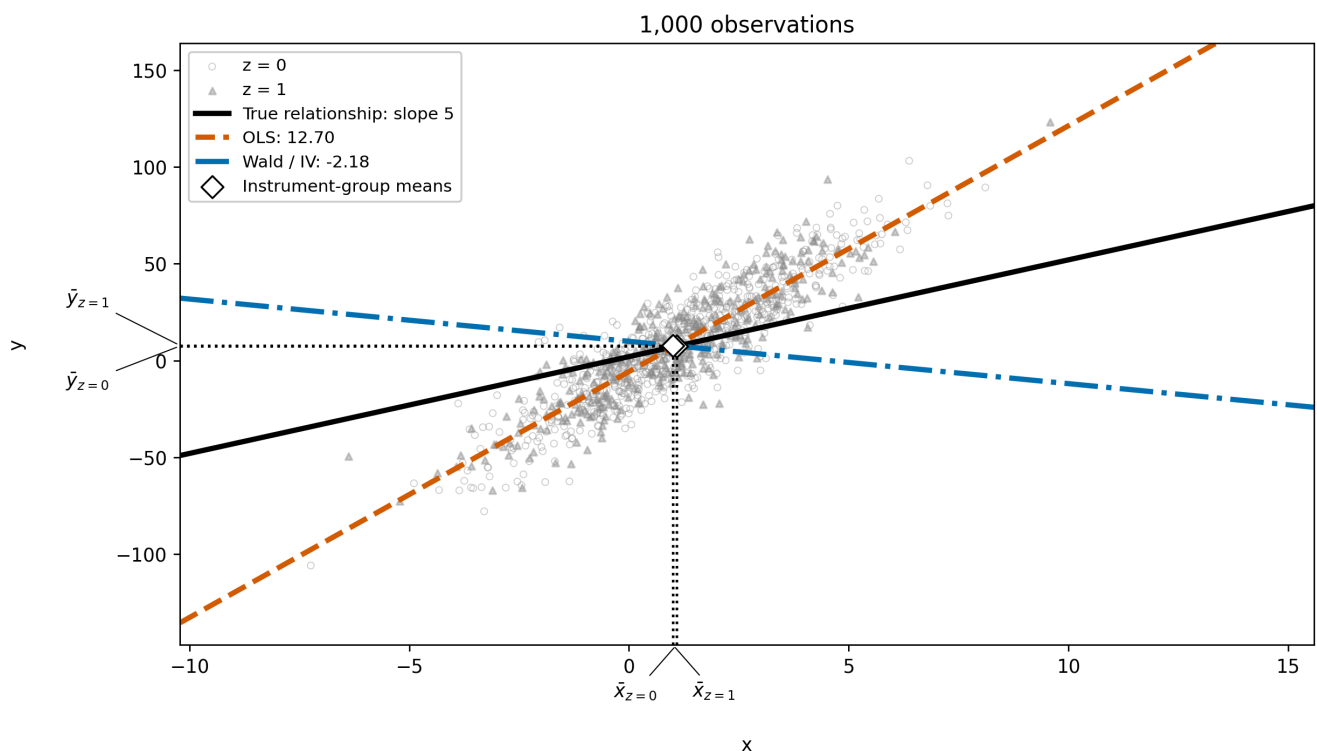
Strong instrument: 1,000 observations



Strong instrument.

The instrument creates a substantial difference in mean x_i . The IV line has slope about 5.13, close to the true slope of 5. OLS has slope about 8.53 because ability is omitted. The first-stage coefficient is about 4.91 and its robust F is about 1,262.

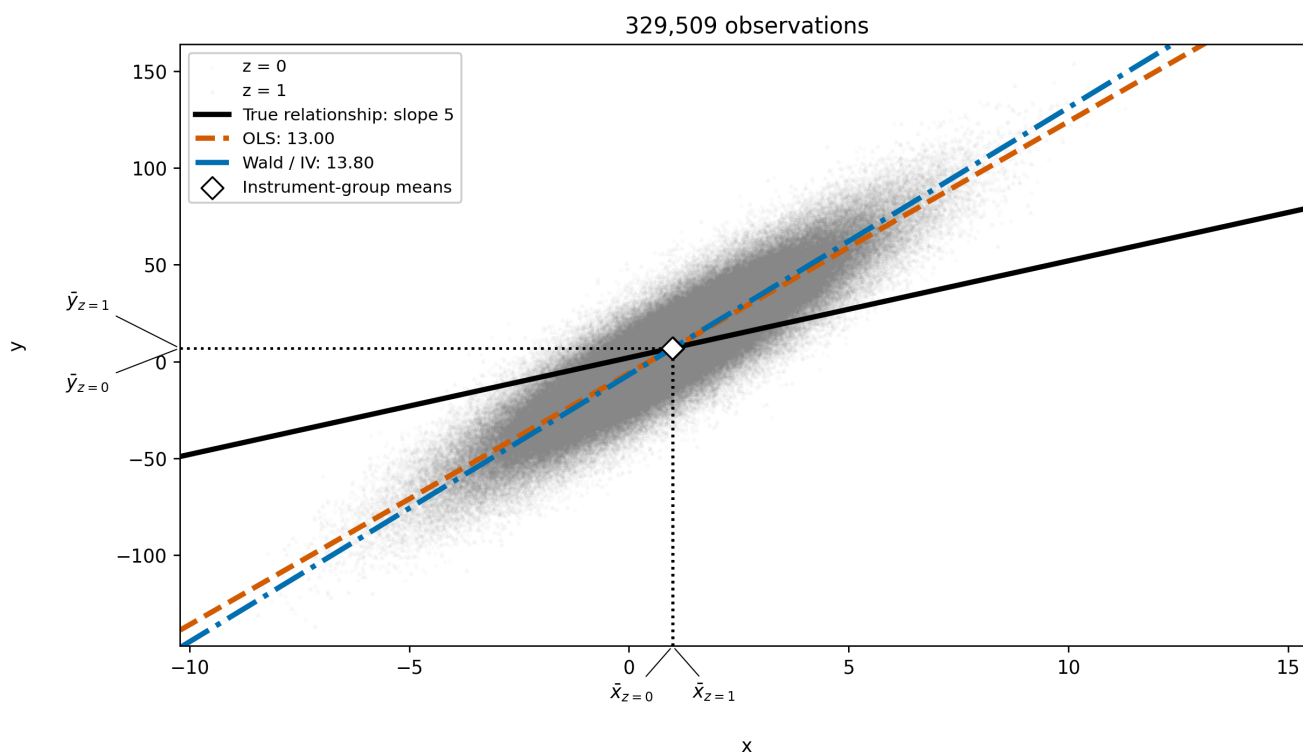
No population first stage: 1,000 observations



No population first stage, smaller sample.

The instrument-group means are close horizontally. The estimated IV slope is about -2.18 , with a reported standard error of about 25.78 . The first-stage F is about 0.39 . The sample provides no convincing evidence of relevance.

No population first stage: 329,509 observations



No population first stage, large sample.

In this draw, OLS is about 13.00 and IV about 13.80 , even though the true effect is 5 . The reported IV standard error is about 4.27 . Looking only at the second stage, we might regard the positive IV estimate and its proximity to OLS as reassuring. The first-stage F is only about 2.02 . The instrument has no population relevance, and the conventional second-stage inference is unreliable.

These are **illustrative draws**, using a fixed seed chosen to show that IV can resemble OLS when relevance fails. Repeating the simulation with different seeds can yield very different IV estimates. Agreement in this figure is not a claim that exactly identified IV converges to OLS when the instrument is irrelevant.

Your turn

Would the second-stage coefficient and reported standard error alone alert you to the problem? What changes when you inspect the first-stage results? Why do 329,509 observations fail to resolve the problem in this example?

Python code

[Download the Python simulation code](#). The code generates the three datasets, estimates OLS, first-stage and IV regressions, and saves the figures. Change the seed to examine other samples. The first graph uses schematic densities to illustrate unbiasedness and consistency.

Simulation code



5. Angrist and Krueger's natural experiment

School-entry rules and compulsory-schooling laws together can create a relationship between quarter of birth and completed schooling. Children entering school at different ages can reach the legal school-leaving age after completing different amounts of education. Angrist and Krueger use quarter of birth as an instrument to estimate the effect of schooling on log weekly earnings.

The 1980 Census sample used here contains **329,509 men born in 1930–1939**. In our notation, x_i is years of completed schooling and y_i is log weekly earnings. With a binary instrument, $z_i = 1$ denotes a first-quarter birth and $z_i = 0$ a birth in the other three quarters.

i Your turn: The policy and the instrument

Explain how school-entry and school-leaving rules generate a first stage. Whose schooling might change because of these rules? What assumptions are needed to interpret the quarter-of-birth comparison as an effect of schooling on earnings?

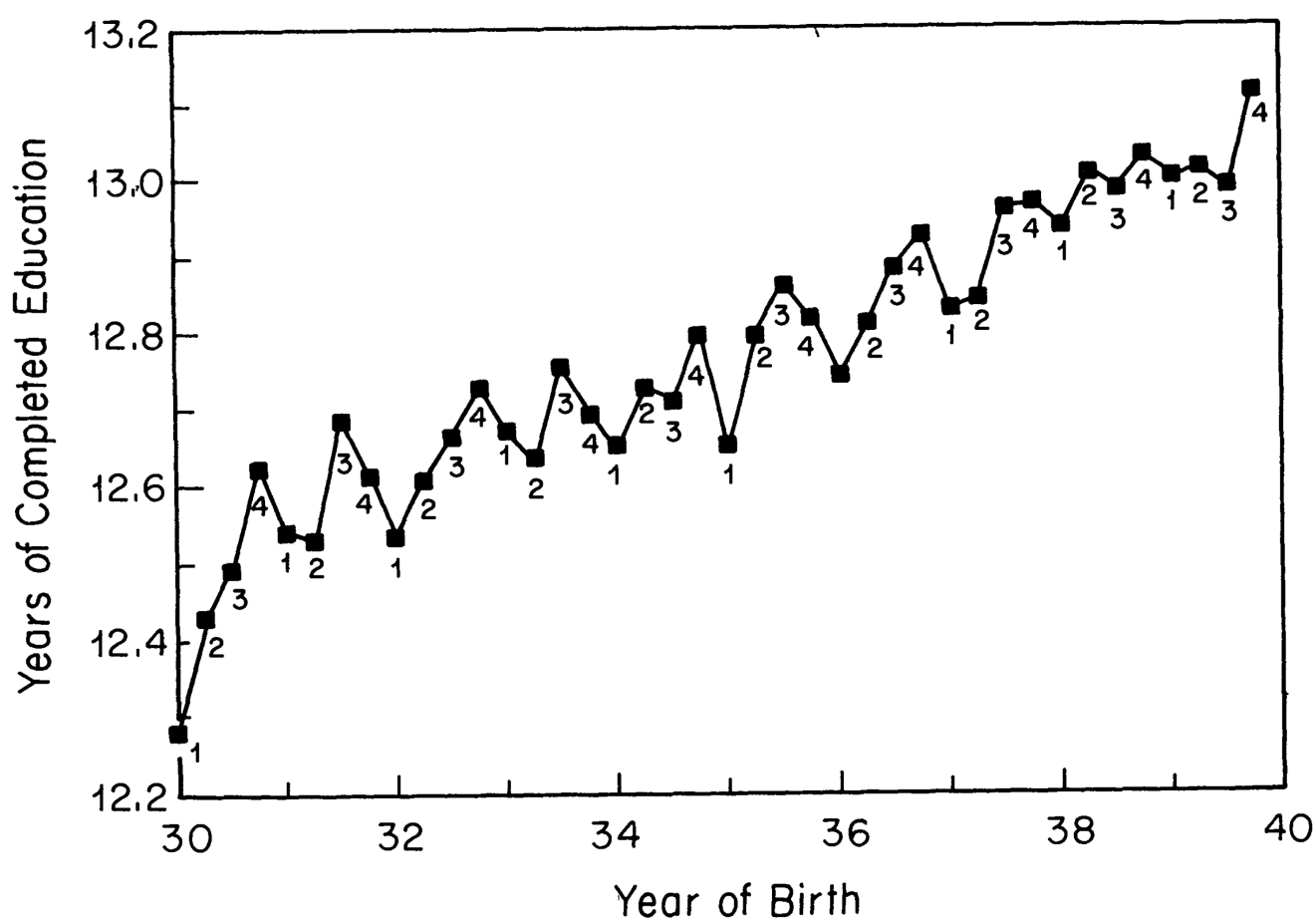


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

Angrist and Krueger (1991), Figure I.

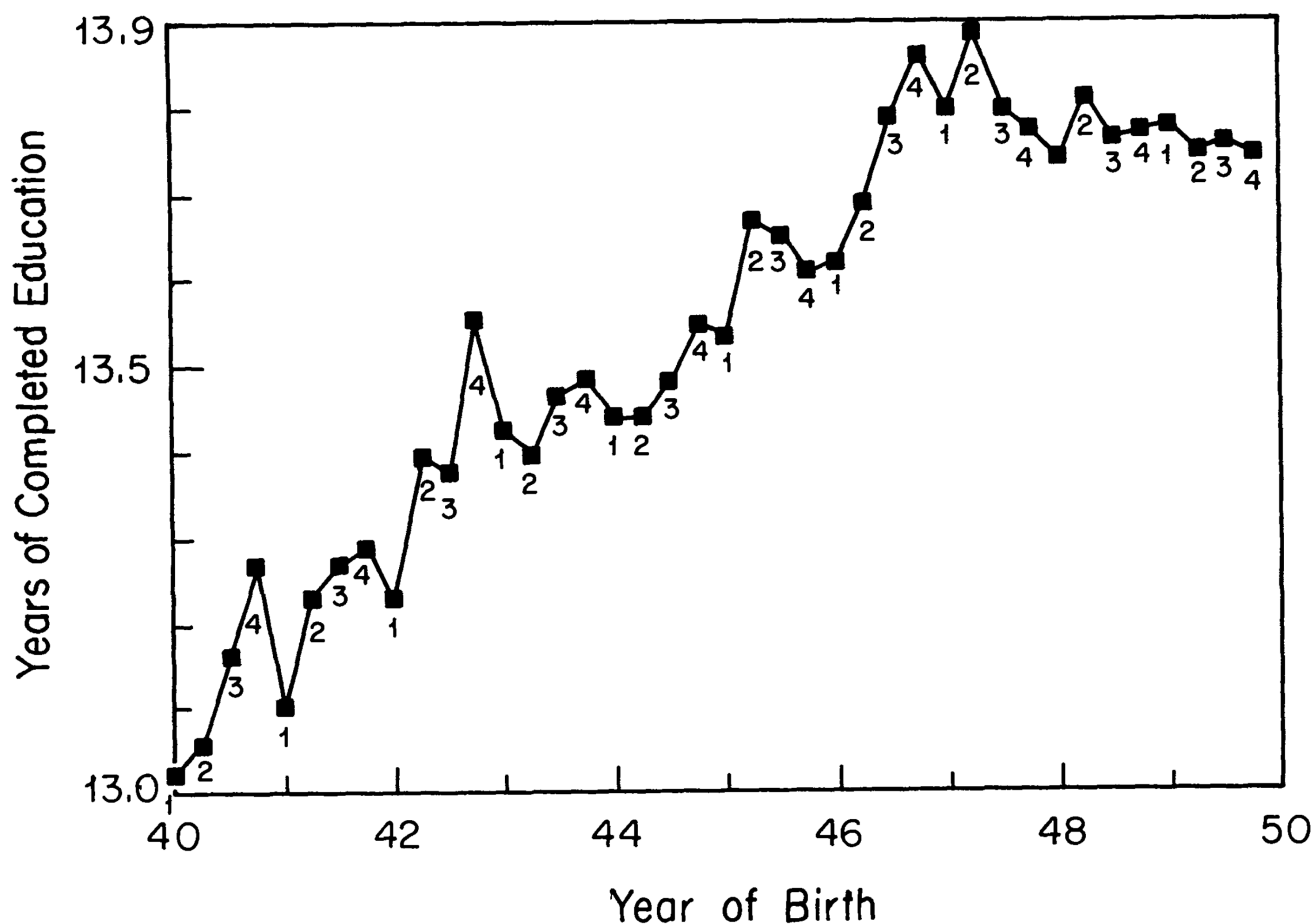


FIGURE II
 Years of Education and Season of Birth
 1980 Census
Note. Quarter of birth is listed below each observation.

Angrist and Krueger (1991), Figure II.

i Your turn: Figures I and II

Identify the outcome in each figure. Explain the pattern across quarters within birth years. What do the figures tell us about the proposed mechanism and the comparison underlying IV?

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) – (2)
ln (wkly. wage)	5.8916	5.9027	–0.01110 (0.00274)
Education	12.6881	12.7969	–0.1088 (0.0132)
Wald est. of return to education			0.1020 (0.0239)
OLS return to education			0.0709 (0.0003)

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

Angrist and Krueger (1991), Table III, Panel B.

The first-quarter group has mean schooling of about 12.6881 years and mean log weekly earnings of about 5.8916. For the other quarters the corresponding means are about 12.7969 and 5.9027. Dividing the earnings difference by the schooling difference gives a Wald estimate of approximately 0.1020. Use the unrounded sample means in the exercise. In a log-earnings equation this is approximately a 10.2% earnings increase for another year of schooling. The exact percentage transformation is $100 \left(\exp \left(\hat{\beta} \right) - 1 \right)$.

Controls and additional instruments

The empirical analysis also uses quarter-of-birth indicators and their interactions with year-of-birth indicators. Write the three quarter indicators as Q_{1i} , Q_{2i} and Q_{3i} , leaving the fourth quarter as the reference category. With controls w_{ki} , the three-instrument specification is

$$x_i = \pi_0 + \sum_{q=1}^3 \pi_q Q_{qi} + \sum_{k=1}^L \delta_k w_{ki} + \nu_i,$$

$$y_i = \beta_0 + \beta x_i + \sum_{k=1}^L \theta_k w_{ki} + \epsilon_i.$$

The same included controls belong in both equations. Additional interactions provide more excluded instruments in the paper's larger specifications.

TABLE V
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	0.0711 (0.0003)	0.0891 (0.0161)	0.0711 (0.0003)	0.0760 (0.0290)	0.0632 (0.0003)	0.0806 (0.0164)	0.0632 (0.0003)	0.0600 (0.0299)
Race (1 = black)	—	—	—	—	−0.2575 (0.0040)	−0.2302 (0.0261)	−0.2575 (0.0040)	−0.2626 (0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763 (0.0029)	0.1581 (0.0174)	0.1763 (0.0029)	0.1797 (0.0305)
Married (1 = married)	—	—	—	—	0.2479 (0.0032)	0.2440 (0.0049)	0.2479 (0.0032)	0.2486 (0.0073)
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
Age	—	—	−0.0772 (0.0621)	−0.0801 (0.0645)	—	—	−0.0760 (0.0604)	−0.0741 (0.0626)
Age-squared	—	—	0.0008 (0.0007)	0.0008 (0.0007)	—	—	0.0008 (0.0007)	0.0007 (0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Angrist and Krueger (1991), Table V.

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 (0.0003)	0.0928 (0.0093)	0.0673 (0.0003)	0.0907 (0.0107)	0.0628 (0.0003)	0.0831 (0.0095)	0.0628 (0.0003)	0.0811 (0.0109)
Race (1 = black)	—	—	—	—	−0.2547 (0.0043)	−0.2333 (0.0109)	−0.2547 (0.0043)	−0.2354 (0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705 (0.0029)	0.1511 (0.0095)	0.1705 (0.0029)	0.1531 (0.0107)
Married (1 = married)	—	—	—	—	0.2487 (0.0032)	0.2435 (0.0040)	0.2487 (0.0032)	0.2441 (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 (0.0617)	−0.0880 (0.0624)	—	—	−0.0778 (0.0603)	−0.0876 (0.0609)
Age-squared	—	—	0.0008 (0.0007)	0.0009 (0.0007)	—	—	0.0008 (0.0007)	0.0009 (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.

Angrist and Krueger (1991), Table VII.

Your turn: Tables V and VII

Explain what changes across the specifications, including the controls and excluded instruments. Compare OLS and IV estimates and their reported uncertainty. Which first-stage information would you want before judging the IV results?

6. Bound, Jaeger, and Baker

Bound, Jaeger, and Baker revisit the quarter-of-birth application to illustrate two concerns. Small violations of instrument validity can matter greatly when relevance is weak. Separately, even exogenous instruments can lead to poor finite-sample behaviour when they contain little information about the endogenous variable.

Table 1. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings
(standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
<i>F</i> (excluded instruments)		13.486		4.747		1.613
Partial R^2 (excluded instruments, $\times 100$)		.012		.043		.014
<i>F</i> (overidentification)		.932		.775		.725
<i>Age Control Variables</i>						
Age, Age ²	x	x			x	x
9 Year of birth dummies			x	x	x	x
<i>Excluded Instruments</i>						
Quarter of birth		x		x		x
Quarter of birth $\times$ year of birth				x		x
Number of excluded instruments		3		30		28

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. *F* (first stage) and partial R^2 are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basman (1960).

Bound, Jaeger, and Baker (1995), Table 1.

Your turn: Table 1

Compare the three IV specifications. Identify the excluded instruments and the controls. How does your assessment change when you read the first-stage *F* and partial R^2 alongside the second-stage estimates? Pay attention to the scaling of partial R^2 in the table.

Table 2. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Controlling for State of Birth
(standard errors of coefficients in parentheses)

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
<i>F</i> (excluded instruments)		2.428		1.869
Partial <i>R</i> ² (excluded instruments, ×100)		.133		.101
<i>F</i> (overidentification)		.919		.917
<i>Age Control Variables</i>				
Age, Age ²			x	x
9 Year of birth dummies	x	x	x	x
<i>Excluded Instruments</i>				
Quarter of birth		x		x
Quarter of birth × year of birth		x		x
Quarter of birth × state of birth		x		x
Number of excluded instruments		180		178

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables. *F* (first stage) and partial *R*² are for the instruments in the first stage of IV estimation. *F* (overidentification) is that suggested by Basmann (1960).

Bound, Jaeger, and Baker (1995), Table 2.

Table 3. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Using Simulated Quarter of Birth (500 replications)

Table (column)	1 (4)	1 (6)	2 (2)	2 (4)
<i>Estimated Coefficient</i>				
Mean	.062	.061	.060	.060
Standard deviation of mean	.038	.039	.015	.015
5th percentile	−.001	−.002	.034	.035
Median	.061	.061	.060	.060
95th percentile	.119	.127	.083	.082
<i>Estimated Standard Error</i>				
Mean	.037	.039	.015	.015

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509.

Bound, Jaeger, and Baker (1995), Table 3.

Your turn: Tables 2 and 3

For Table 2, explain what changes when state-of-birth controls and interactions are introduced. For Table 3, explain how the authors simulate quarter of birth and what they hold fixed. What would we expect if the instruments contained no information? Explain what the results reveal about second-stage estimates and the number of instruments.

In the exercise we keep the three quarter-of-birth instruments and vary the controls. Adding age changes the comparison because earnings and schooling can vary with age, while quarter of birth also determines age within a birth year. In BJB's Table 1, columns 1–2 include age and age squared, together with the demographic and region controls. The first-stage F in this specification is about 13.486, and partial R^2 is about 0.000123. The table reports the latter multiplied by 100.

7. Returning to exclusion

Bound and Jaeger examine whether compulsory-schooling laws alone account for the relationship between earnings and quarter of birth. Their Figure 1 compares third-quarter minus first-quarter differences across birth decades.

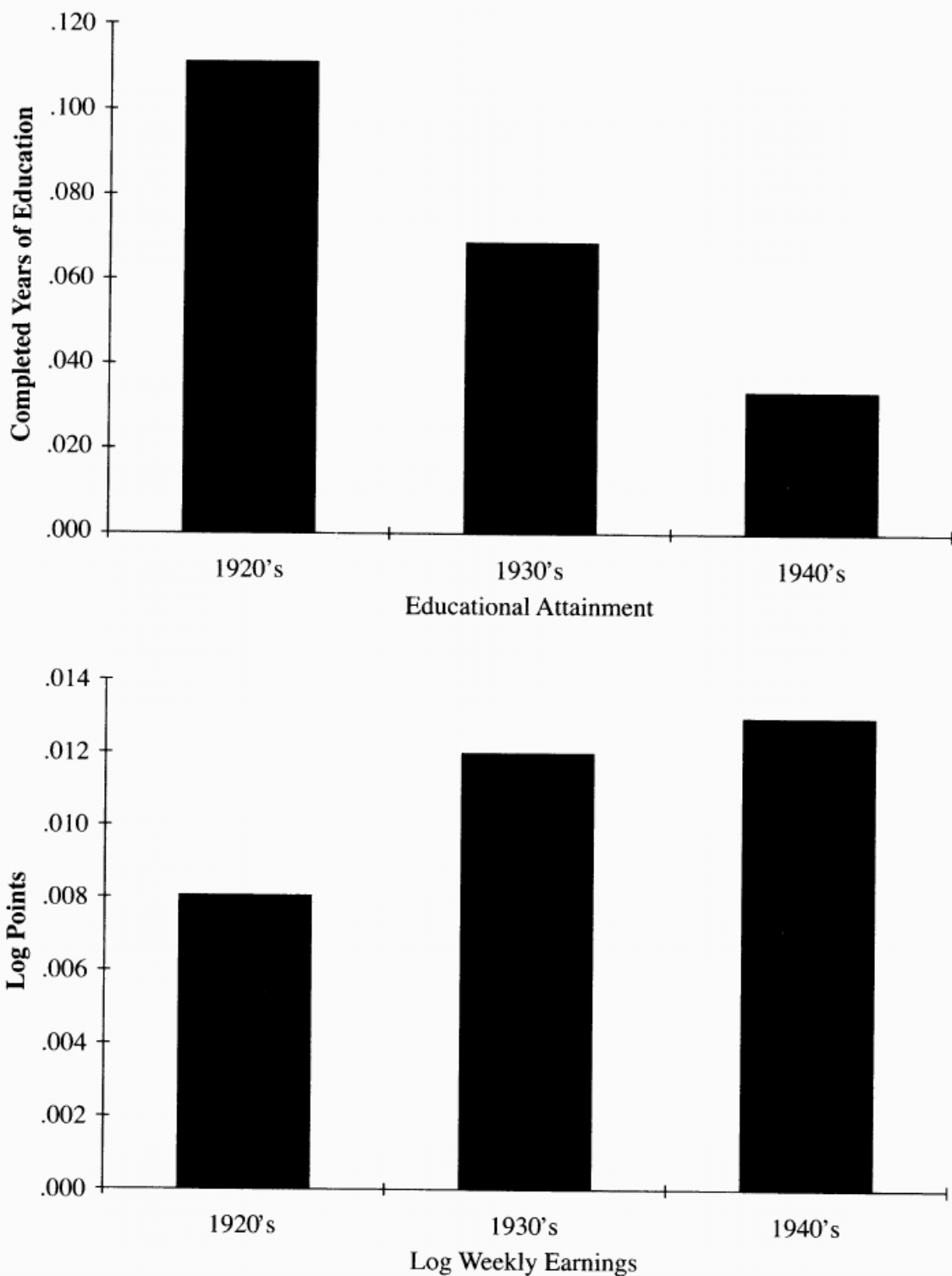


Fig. 1. Third Quarter-First Quarter Reduced Form Effect of Quarter of Birth on Educational Attainment and Log Weekly Earnings by Birth Decade. Source: Table 2.

Bound and Jaeger (2000), Figure 1.

Your turn: Exclusion

Compare the quarter-of-birth patterns in schooling and earnings across the three birth decades. If schooling were the only channel and its effect were stable, how would you expect the two patterns to relate? What further explanations or evidence would you consider before accepting the exclusion restriction?

This brings us back to the argument behind an instrument. The first-stage statistics assess relevance. Exclusion requires evidence and reasoning about the channels through which the instrument can affect the outcome. Weak relevance makes a small failure of that argument especially consequential for IV.

After class

Complete [Problem Set 2](#), due **Friday 30 October 2026 at 12 noon (UK time), in Week 7**. Submit the completed Quarto source and rendered PDF through Moodle. Calculate the Wald estimate from sample means, reproduce it with IV, and estimate the three-instrument specifications with different age controls. Present first-stage results alongside the estimated returns to schooling.

References

- Angrist, J. D., and A. B. Krueger (1991). “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics* 106(4): 979–1014. [Paper](#).
- Bound, J., D. A. Jaeger, and R. M. Baker (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(430): 443–450. [Paper](#).
- Bound, J., and D. A. Jaeger (2000). “Do Compulsory School Attendance Laws Alone Explain the Association between Earnings and Quarter of Birth?” *Research in Labor Economics* 19: 83–108. [Paper](#).
- Staiger, D., and J. H. Stock (1997). “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65(3): 557–586. [Paper](#).