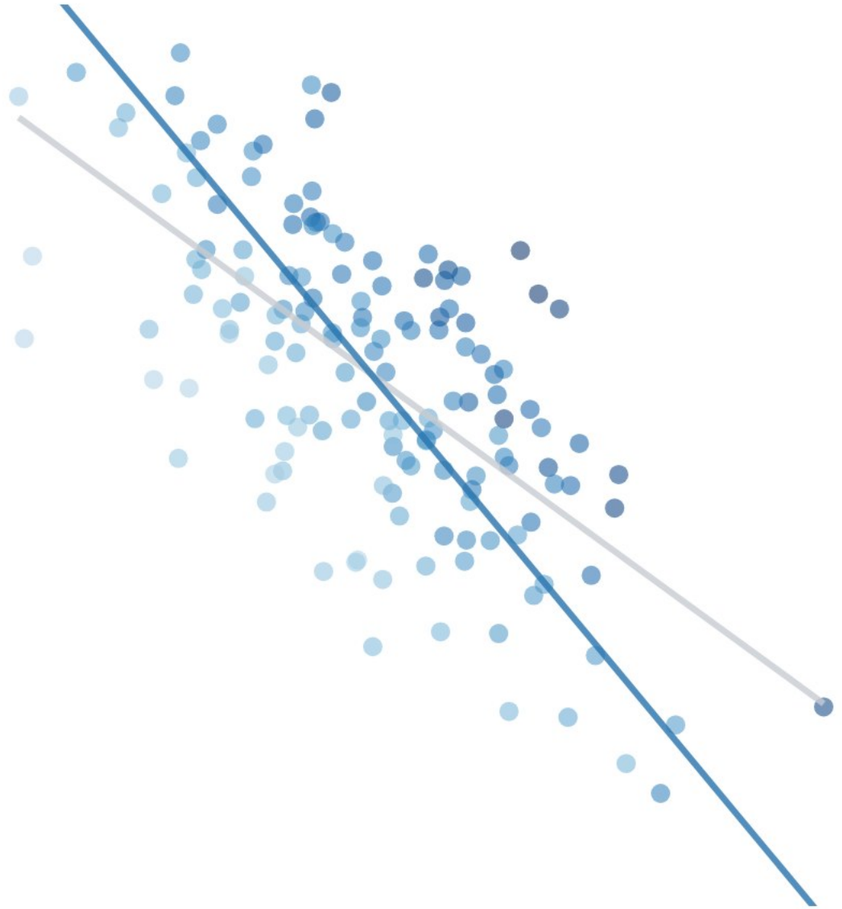
ECON 104 · DISCUSSION · FALL 2026

Endogeneity and Instrumental Variables

Jake Anderson
UCLA



Plan for Today

1

Endogeneity and 2SLS

Chinese imports and U.S. manufacturing jobs. Why do OLS and 2SLS give different answers?

Which way is OLS biased, and how does an instrument remove the bias?

Spring 2026 Final (Boswell), Question 5

2

Instrument strength

Do the instruments move the endogenous regressor enough?

Compute the first-stage F statistic from R output and compare it to 10.

Spring 2023 Midterm V1, Question 14

3

Surplus instruments

Two instruments for one endogenous regressor. Is the extra instrument valid?

The Sargan test, and how to read all three diagnostic rows in R.

Spring 2023 Midterm V1, Question 16

The Setup: The China Syndrome

This paper estimates the effect of rising Chinese import competition on local U.S. labor markets using commuting zones (CZs) as the unit of analysis. Because regions differ in their initial manufacturing industry mix, they face differential exposure to Chinese imports. The authors' main estimating equations are:

$$\Delta L_{it}^{m} = \gamma_t + \beta_1 \, \Delta IPW_{uit} + \text{controls}_{it} + \varepsilon_{it} \qquad \text{(Second stage)}$$

$$\Delta IPW_{uit} = \gamma_t + \delta_1 \, \Delta IPW_{oit} + \text{controls}_{it} + \nu_{it} \qquad \text{(First stage)}$$

- ΔL_{it}^{m}: change in manufacturing employment as a share of working-age population in CZ i, period t.
- ΔIPW_{uit}: change in U.S. import exposure from China in CZ i (endogenous).
- ΔIPW_{oit}: analogous Chinese import exposure in eight other high-income countries (instrument).
- controls_{it}: baseline controls (education, gender, race shares; share of routine jobs). γ_t: period fixed effect.

Reported results: First-stage F-statistic $= 11.1$; $\hat{\beta}_1^{\text{OLS}} = -0.365$ (SE 0.074); $\hat{\beta}_1^{\text{2SLS}} = -0.596$ (SE 0.137).

Autor, Dorn, and Hanson (2013), "The China Syndrome," *American Economic Review* 103(6).

Source: Spring 2026 Final (Boswell), setup for Questions 1 to 8

The Question: How Much Larger Is the 2SLS Estimate?

$\hat{\beta}_1^{\text{OLS}} = -0.365$ (SE 0.074); $\hat{\beta}_1^{\text{2SLS}} = -0.596$ (SE 0.137). β_1 is the effect of import exposure ΔIPW_{uit} on manufacturing employment ΔL_{it}^{m}.

Using the reported estimates, compute the approximate percentage by which the 2SLS estimate of β_1 exceeds the OLS estimate in absolute magnitude, and interpret the direction of OLS bias.

a 39% larger; OLS is downward biased, overstating the negative employment effect.

b 39% larger; OLS is upward biased, understating the negative employment effect.

c 63% larger; OLS is upward biased, understating the negative employment effect.

d 63% larger; OLS is downward biased, overstating the negative employment effect.

e OLS and 2SLS are not meaningfully different once standard errors are considered.

Source: Spring 2026 Final (Boswell), Question 5

A Commuting Zone Is a Local Labor Market

COUNTIES

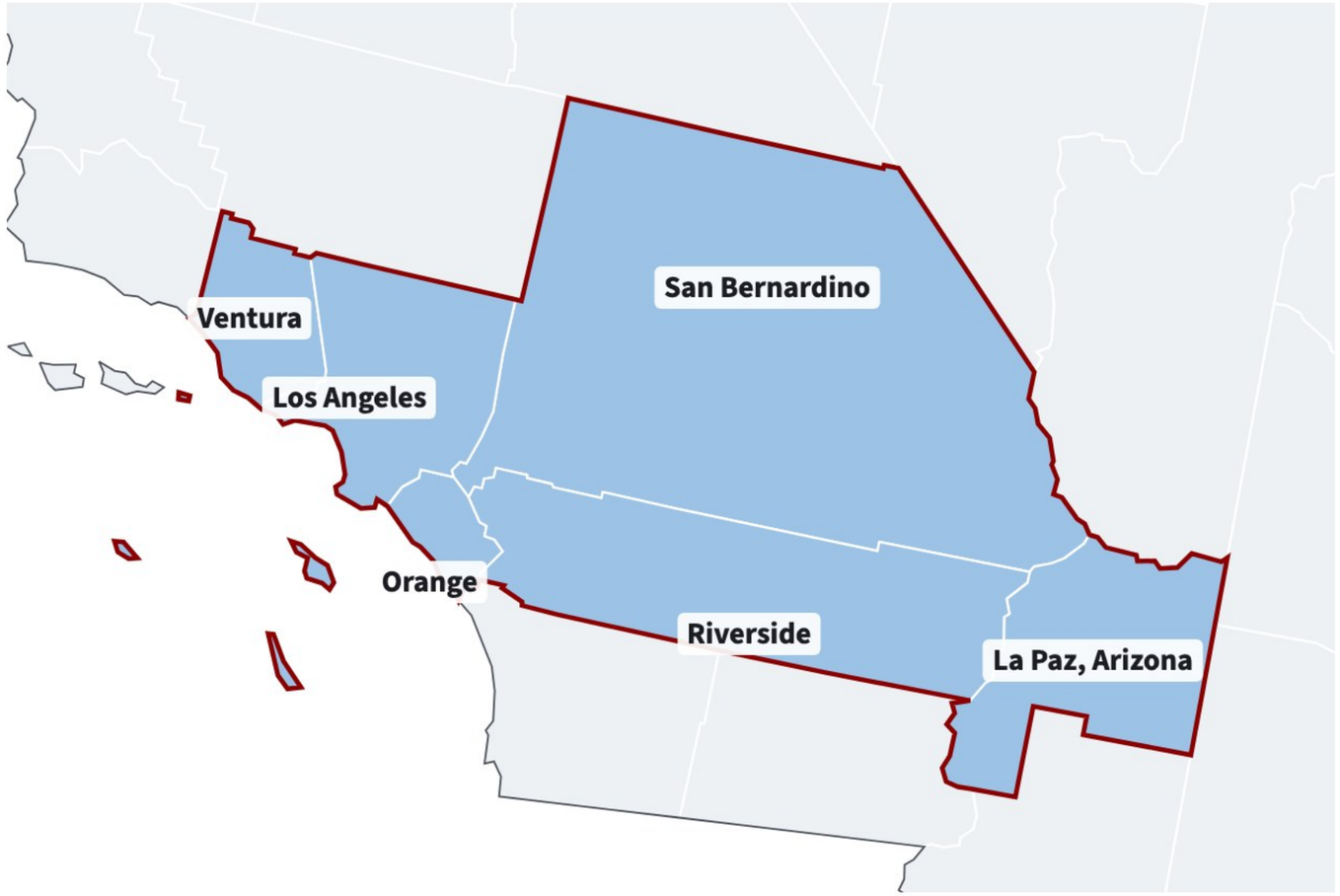
About 3,100 counties. Many people **live in one county and work in another**.

COMMUTING ZONES

Counties grouped by 1990 commuting data (Tolbert and Sizer 1996). Most residents work inside their zone $\Longrightarrow$ each zone is **one local labor market**.

EXAMPLE: LOS ANGELES

Six counties: Los Angeles, Orange, Riverside, San Bernardino, Ventura, and La Paz (Arizona).



3,108 counties (2017 boundaries) in 722 commuting zones, from David Dorn's crosswalk.

A Commuting Zone Is a Local Labor Market

COUNTIES

About 3,100 counties. Many people **live in one county and work in another**.

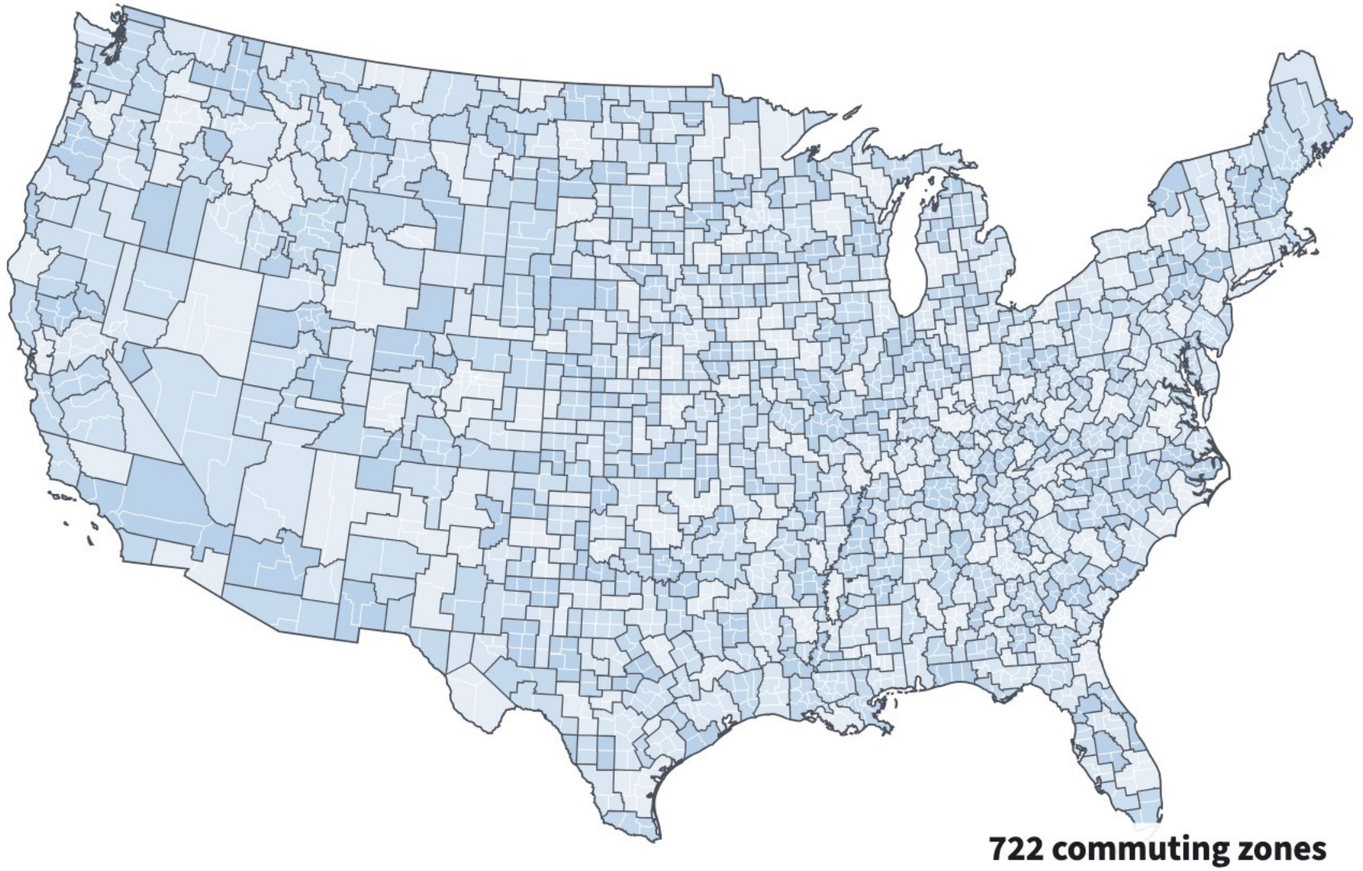
COMMUTING ZONES

Counties grouped by 1990 commuting data (Tolbert and Sizer 1996). Most residents work inside their zone $\Longrightarrow$ each zone is **one local labor market**.

EXAMPLE: LOS ANGELES

Six counties: Los Angeles, Orange, Riverside, San Bernardino, Ventura, and La Paz (Arizona).

722 zones $\times$ 2 periods (1990 to 2000, 2000 to 2007) $=$ **1,444 observations**.



3,108 counties (2017 boundaries) in 722 commuting zones, from David Dorn's crosswalk.

U.S. Imports from China Grew Fastest in a Few Industries

THE TOTAL

From **26 billion** dollars in 1991 to **330 billion** in 2007.

BY INDUSTRY

5 of 395 industries make up **32%** of the rise from 2000 to 2007.

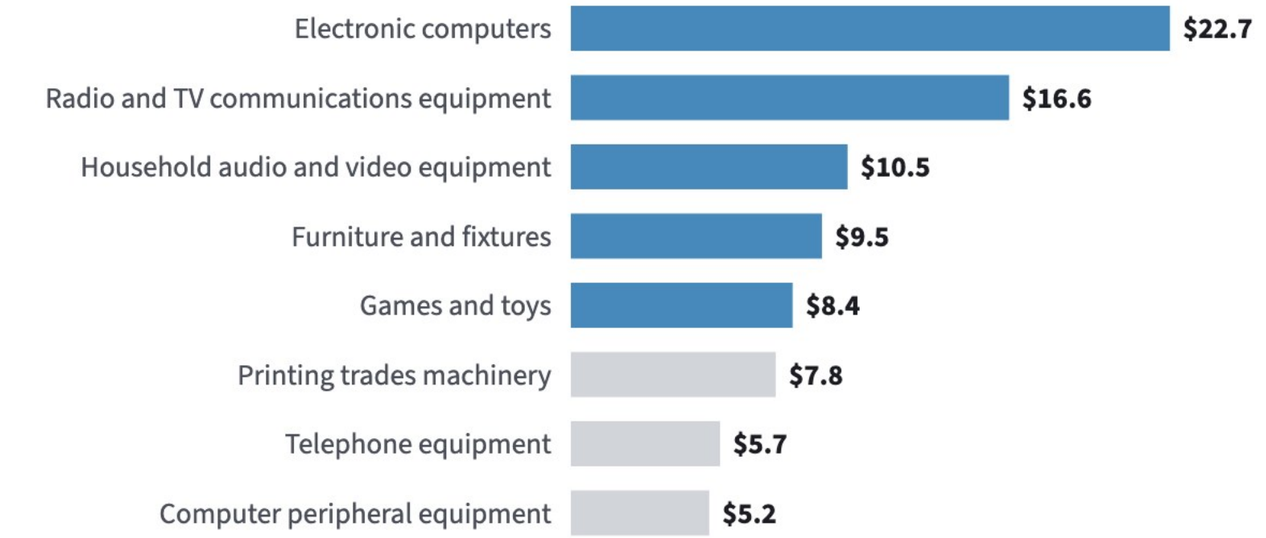
THE CAUSE

Market reforms, lower trade barriers, and WTO entry in 2001: **changes inside China** $\Longrightarrow$ more supply of Chinese goods.

U.S. imports from China (billions of 2007 dollars)



Increase from 2000 to 2007: the 8 largest of 395 industries (billions)



Autor, Dorn, and Hanson (2013), Table 1 and their industry trade data (UN Comtrade). 2007 dollars.

Exposure Depends on What Each Zone Makes

$$\Delta IPW_{ui} = \sum_j \frac{L_{ij}}{L_{uj}} \cdot \frac{\Delta M_{ucj}}{L_i}$$

L_{ij}/L_{uj}	zone i's share of all U.S. jobs in industry j
ΔM_{ucj}	rise in U.S. imports from China in industry j
L_i	workers in zone i

Unit: thousands of dollars of new Chinese imports per worker, per ten years. Map: 2000 to 2007.

SAN JOSE, CA: 7.3

45% from **computers**.

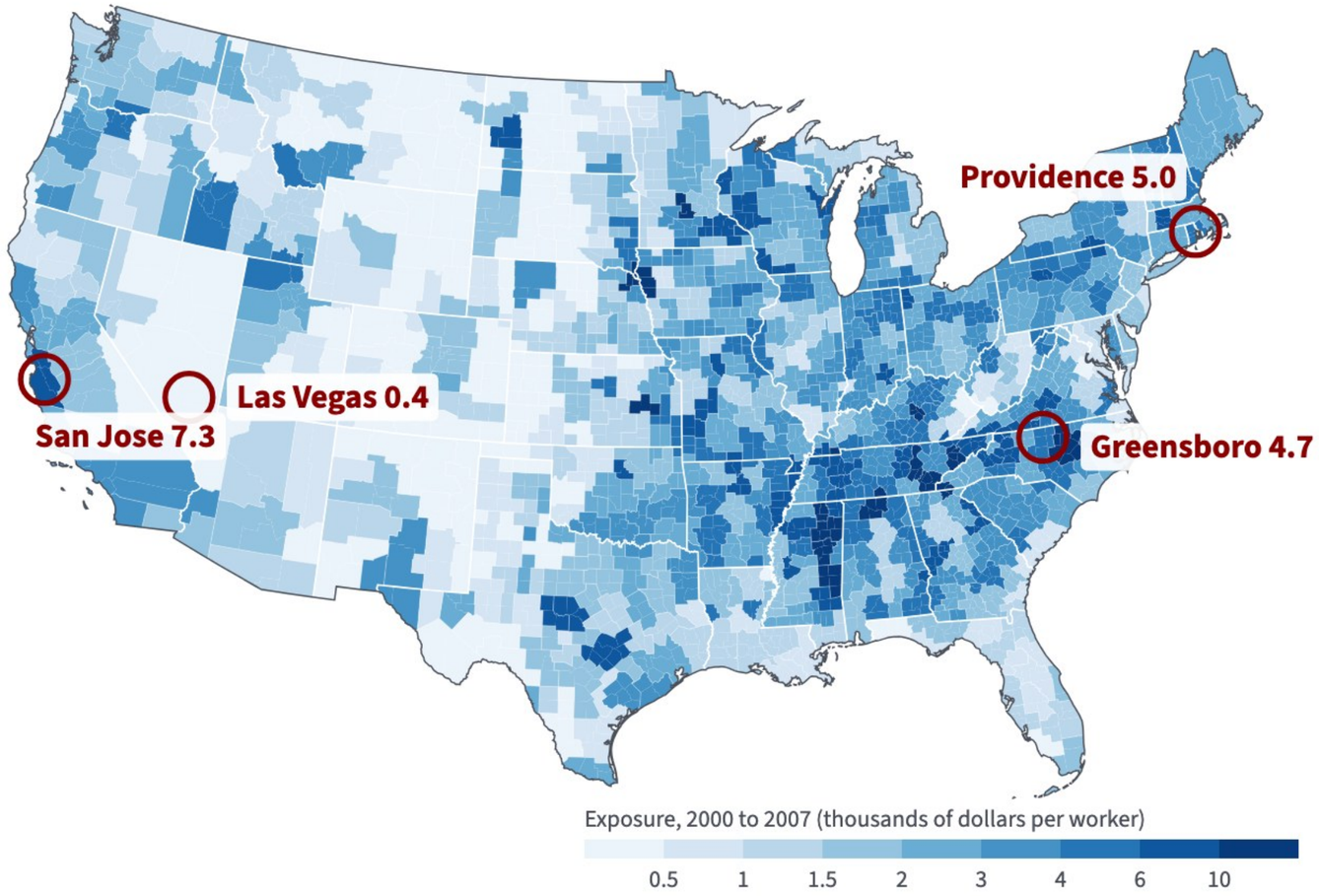
GREENSBORO, NC: 4.7 · PROVIDENCE, RI: 5.0

Greensboro: 28% from **furniture**. Providence: largest share from **costume jewelry** (18%).

LAS VEGAS, NV: 0.4

Only **3.8% of jobs** are in manufacturing.

Every zone faces the same import growth in each industry $\Longrightarrow$ exposure differs **only through the industry mix**.



Exposure: authors' replication files. Industry shares: computed from County Business Patterns, 1999.

The Instrument: Same Calculation, Other Countries' Imports

$$\Delta IPW_{oi} = \sum_j \frac{L_{ij,t-1}}{L_{uj,t-1}} \cdot \frac{\Delta M_{ocj}}{L_{i,t-1}}$$

Dark red: the two changes from ΔIPW_{ui}.

CHANGE 1: WHOSE IMPORTS

ΔM_{ocj}: imports from China into **8 other rich countries** (Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, Switzerland).

CHANGE 2: OLDER JOB COUNTS

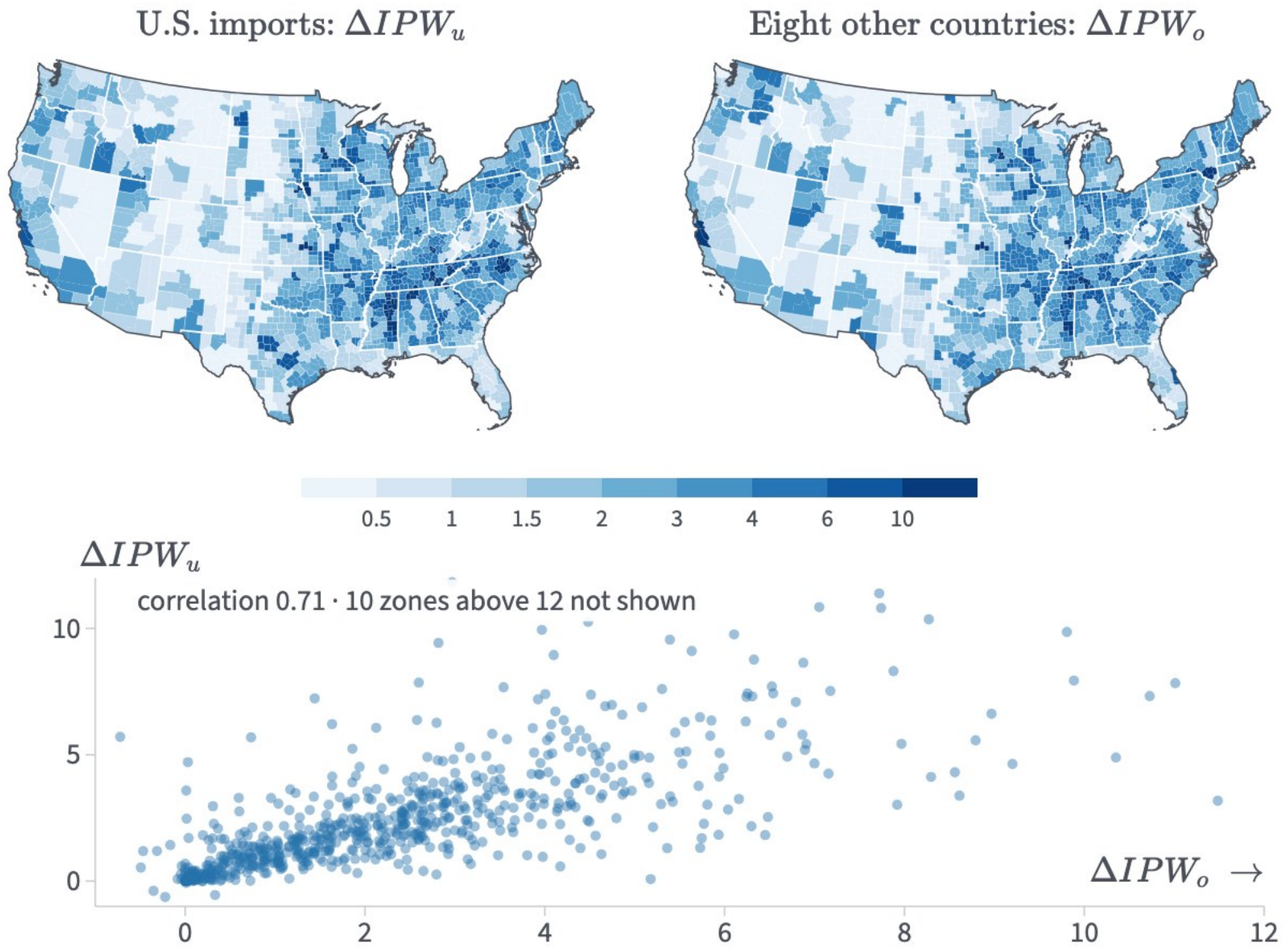
Employment from **ten years earlier**, so it reacts less to expected Chinese imports.

RELEVANCE

The two measures have correlation **0.71** across the 722 zones.

EXOGENEITY

A demand boom in Los Angeles raises U.S. imports, but **not imports into Germany or Japan**.



2000 to 2007, authors' replication files. Same color scale in both maps.

A Demand Shock Moves Both Imports and Employment

THE DATA

Each point is one commuting zone. OLS slope: -0.365.

SOMETHING OLS DOES NOT OBSERVE

Shading: the local demand shock, part of the error ε. Strong demand (dark) $\Longrightarrow$ **more imports and more factory jobs**.

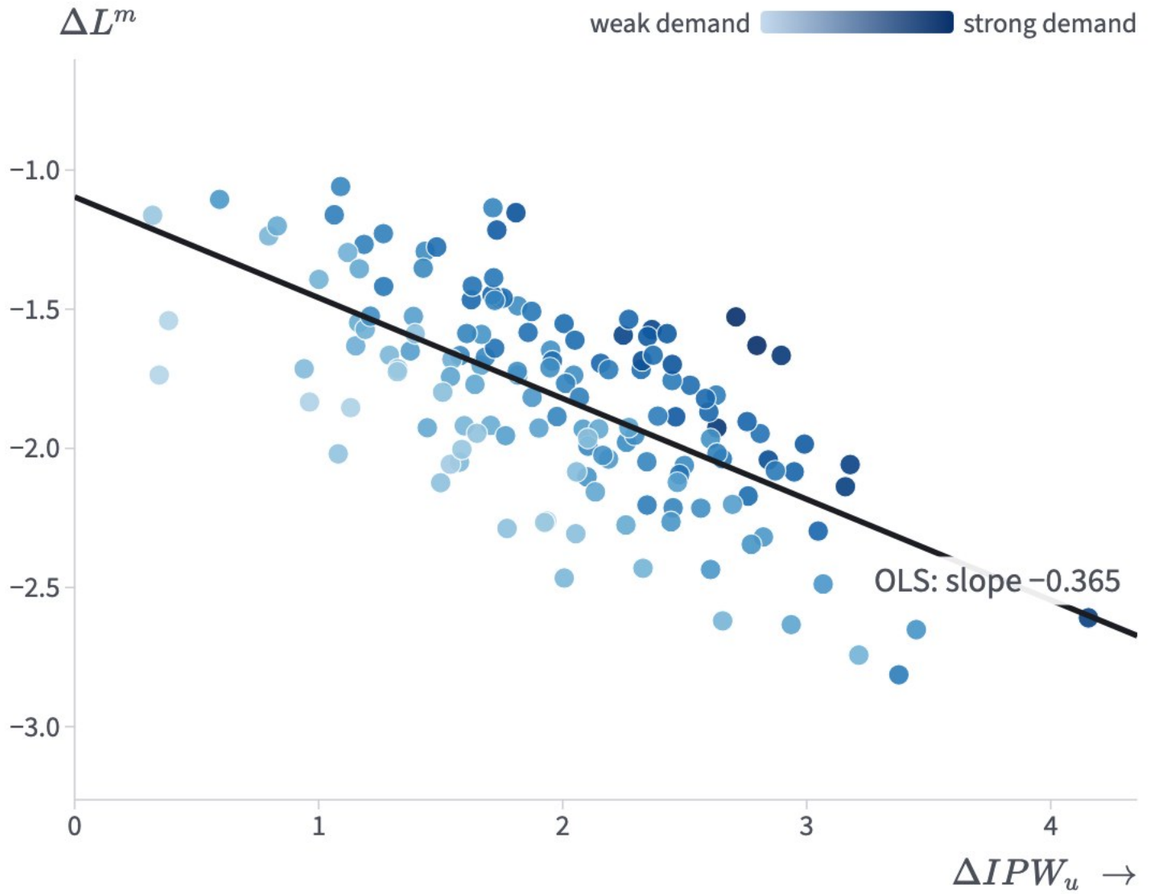


Illustration: 150 zones built so that OLS gives -0.365 and the true effect is -0.596.

A Demand Shock Moves Both Imports and Employment

THE DATA

Each point is one commuting zone. OLS slope: -0.365.

SOMETHING OLS DOES NOT OBSERVE

Shading: the local demand shock, part of the error ε. Strong demand (dark) $\Longrightarrow$ **more imports and more factory jobs**.

REMOVE THE SHOCK

Slope -0.596: the effect of imports alone. Real data do not show the shock, so we cannot do this.

OLS BIAS

$$\operatorname{plim} \hat{\beta}_1^{\text{OLS}} = \beta_1 + \frac{\operatorname{Cov}(\Delta IPW_u,\, \varepsilon)}{\operatorname{Var}(\Delta IPW_u)}$$

plim: the value OLS approaches in large samples. $\operatorname{Cov} > 0$ $\Longrightarrow$ OLS is biased **toward zero**.

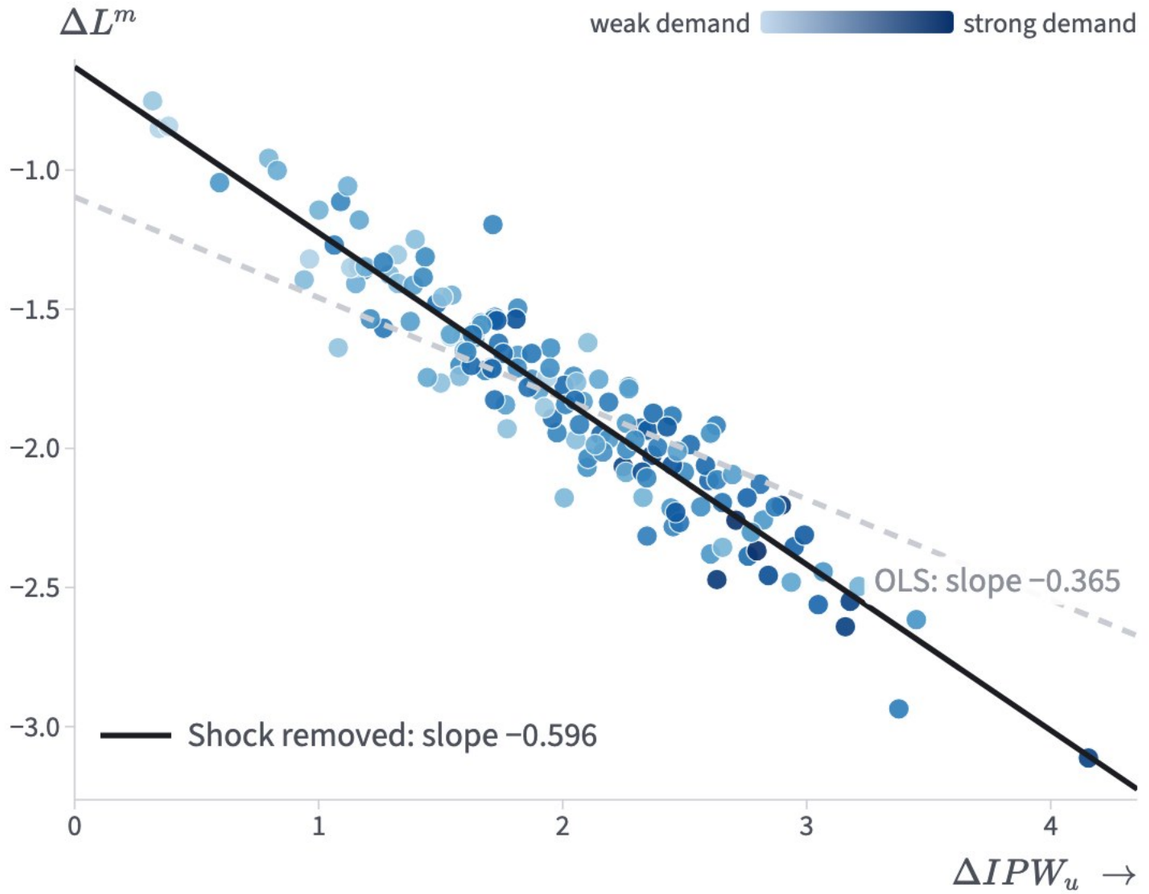


Illustration: 150 zones built so that OLS gives -0.365 and the true effect is -0.596.

A Valid Instrument Needs Two Conditions

THE PROBLEM

The shock is in ε and changes imports $\implies \text{Cov}(\Delta IPW_u, \varepsilon) \neq 0$: imports are **endogenous**.

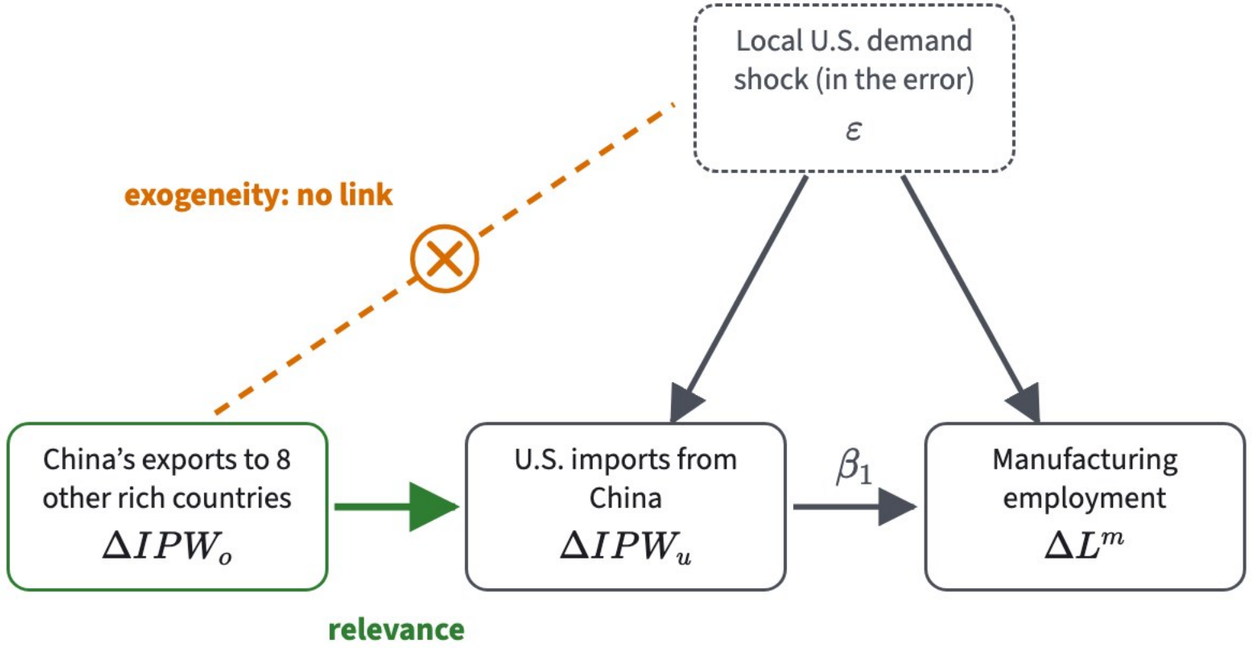
CONDITION 1: RELEVANCE

$\text{Cov}(\Delta IPW_o, \Delta IPW_u) \neq 0$. Check it with the **first-stage F** (Problem 2).

CONDITION 2: EXOGENEITY

$\text{Cov}(\Delta IPW_o, \varepsilon) = 0$. With one instrument, it **cannot be tested** (Problem 3).

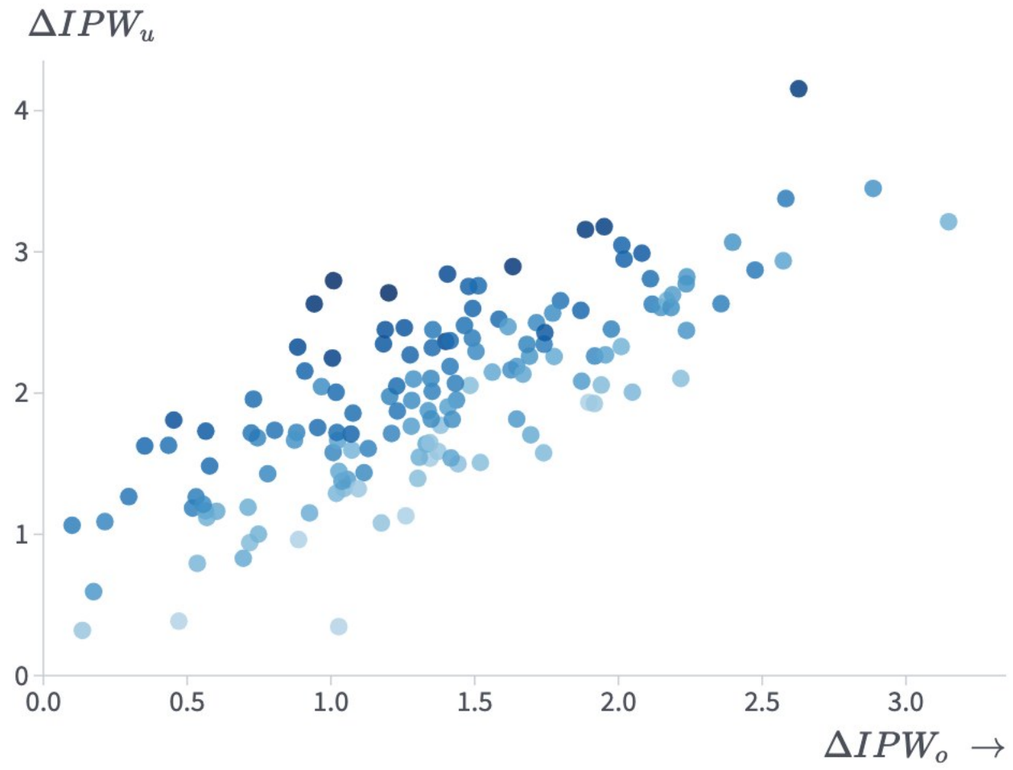
The instrument affects employment **only through** U.S. imports.



2SLS Uses Only the Part of Imports Driven by China

STAGE 1

Regress ΔIPW_u on ΔIPW_o and the controls.



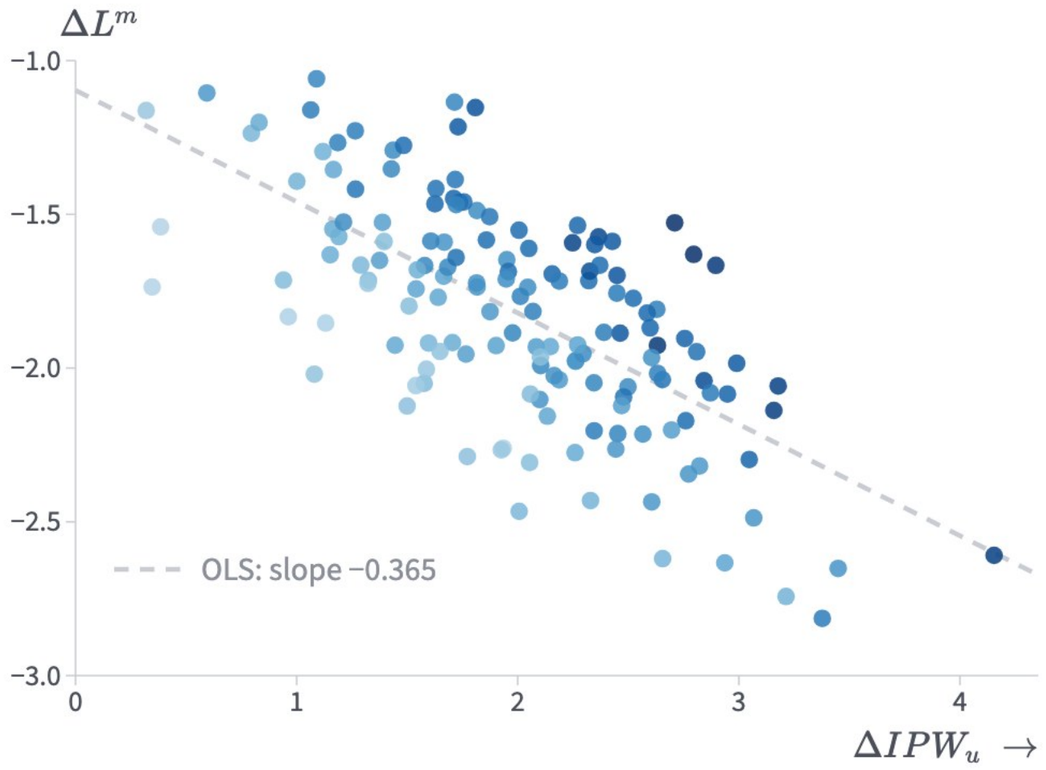


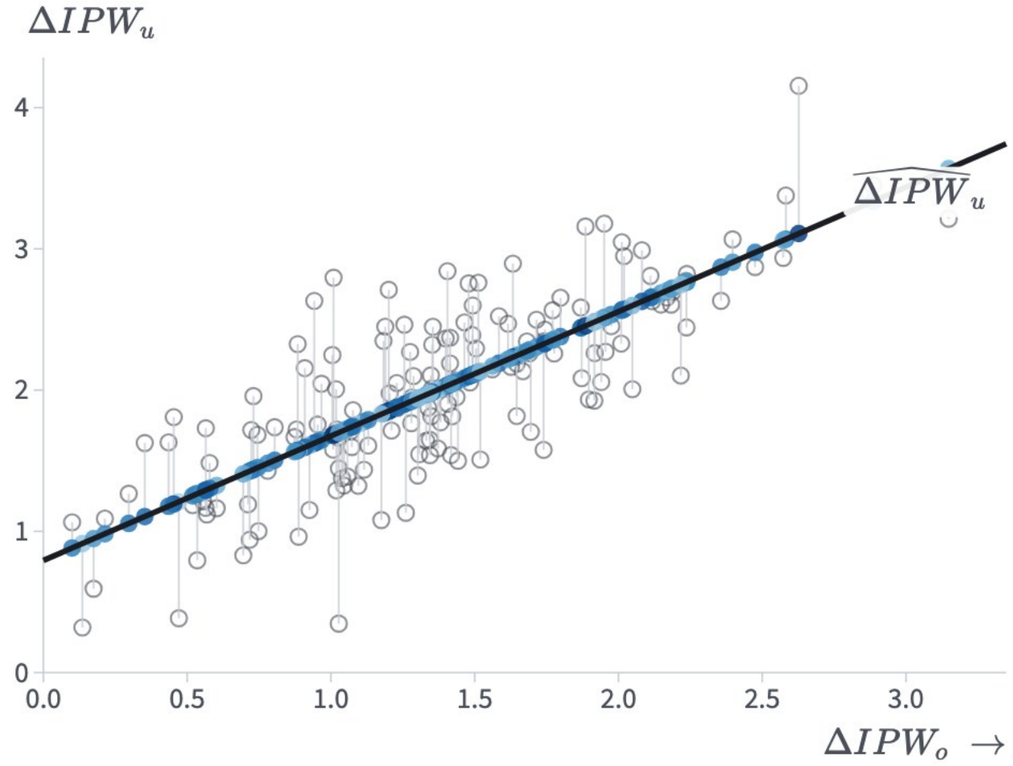
Illustration: the same 150 commuting zones. Shading: local demand shock (darker = stronger demand).

2SLS Uses Only the Part of Imports Driven by China

STAGE 1

Regress ΔIPW_u on ΔIPW_o and the controls.

Fitted value $\widehat{\Delta IPW}_u$: the part of imports **predicted by China's exports elsewhere**.



STAGE 2

Replace imports with $\widehat{\Delta IPW}_u$. The shading is now mixed: **unrelated to the shock**.

Regress ΔL^m on $\widehat{\Delta IPW}_u$: slope -0.596.

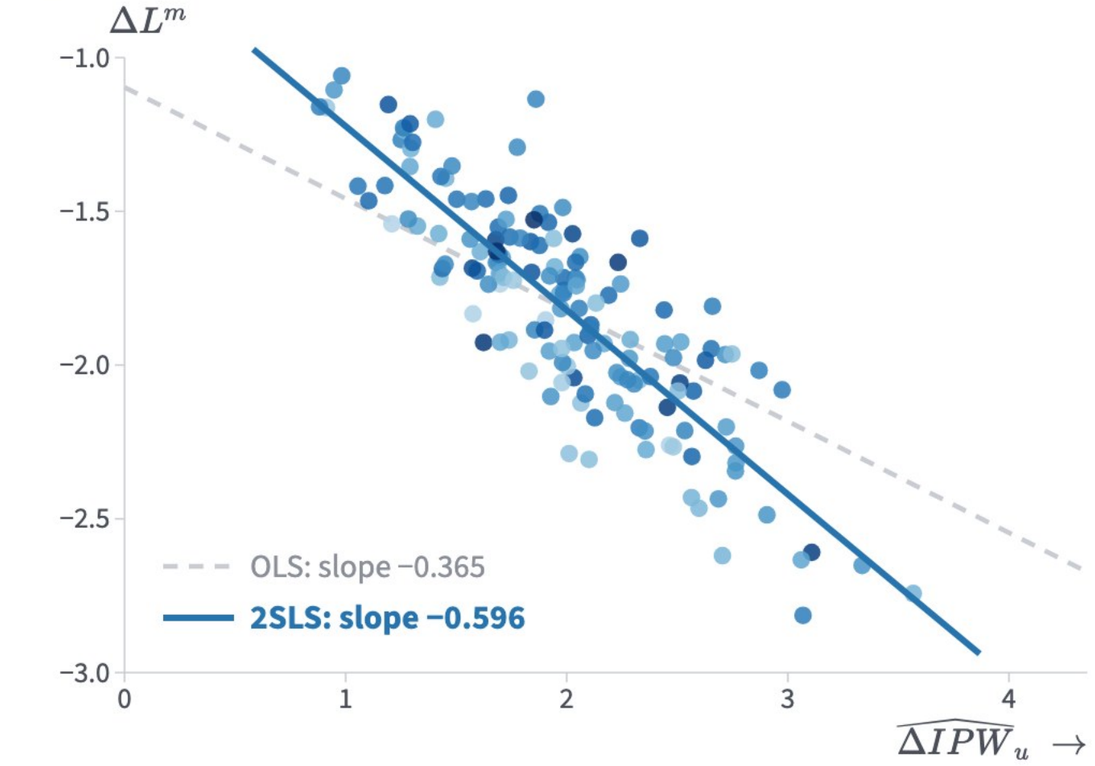


Illustration: the same 150 commuting zones. Shading: local demand shock (darker = stronger demand).

Compare the Two Estimates

DROP THE MINUS SIGNS

$|-0.596| = 0.596$ and $|-0.365| = 0.365$

GAP

$0.596 - 0.365 = 0.231$

PERCENT LARGER

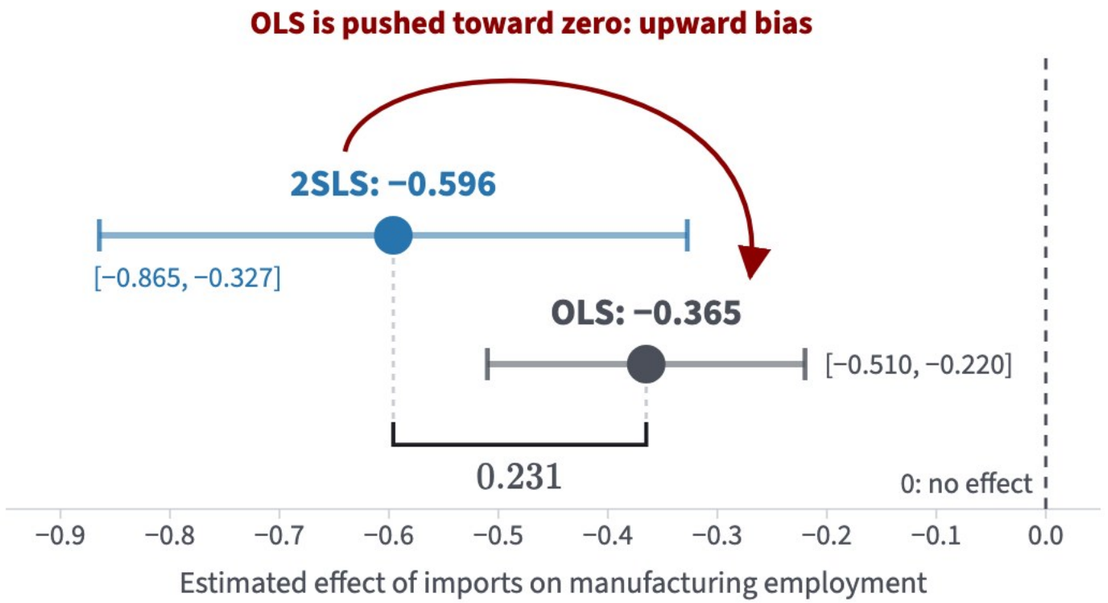
Compare the gap **to OLS**:

$$\frac{0.231}{0.365} = 0.633 \implies \text{about } 63\% \text{ larger}$$

DIRECTION OF OLS BIAS

2SLS (-0.596) is the target. OLS is **closer to zero** $\implies$ biased **upward**: it understates the job loss.

> $0.231/0.596 = 39\%$ answers a different question: how much smaller OLS is.



95% CONFIDENCE INTERVALS

estimate $\pm$ $1.96\times$ SE

OLS: $-0.365 \pm 1.96(0.074) = [-0.510,\ -0.220]$

2SLS: $-0.596 \pm 1.96(0.137) = [-0.865,\ -0.327]$

> Overlapping intervals do not test whether OLS and 2SLS differ. That needs a **Hausman test**.

Checking Each Answer Choice

a 39% larger; OLS is downward biased, overstating the negative employment effect.

Divides by the wrong estimate, and the direction is reversed.

b 39% larger; OLS is upward biased, understating the negative employment effect.

Right direction, but divides by 0.596 instead of 0.365.

d 63% larger; OLS is downward biased, overstating the negative employment effect.

Wrong direction: OLS is closer to zero, so it understates the effect.

e OLS and 2SLS are not meaningfully different once standard errors are considered.

Testing the gap needs a Hausman test, which is not reported.

c 63% larger; OLS is upward biased, understating the negative employment effect.

Answer: (c)

INTERPRETATION

Imports cut manufacturing employment by 0.596 per unit of exposure, not 0.365.

SOURCE OF THE BIAS

A demand shock raises imports **and** jobs $\Longrightarrow$ $\text{Cov}(\Delta IPW_u, \varepsilon) > 0 \Longrightarrow$ upward bias.

The Question: Are the Instruments Strong?

We are interested in the following model: $\log(Wage_i) = EXPER_i + EXPER_i^2 + EDUC_i + \epsilon_i$. However, we are worried that $EDUC$ is endogenous so we use $MOTHERCOLL$ and $FATHERCOLL$ as instruments. We run the following R code:

```
# Estimate first stage
first.stage2 <- lm(educ ~ exper + I(exper^2) + mothercoll + fathercoll, data = mroz_subset)
# Test joint significance of MOTHERCOLL and FATHERCOLL
linearHypothesis(first.stage2, c("mothercoll = 0", "fathercoll = 0"))
## Model 1: restricted model
## Model 2: educ ~ exper + I(exper^2) + mothercoll + fathercoll
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    425 2219.2
## 2    423 1748.3  2    470.88 56.963 < 2.2e-16 ***
```

What can we conclude about the instruments?

a The instruments are strong: The F statistic is significant at the 5% level.

b The instruments are strong: The F statistic is greater than 10.

c The instruments are weak: The F statistic is less than 10.

d The instruments are weak: The F statistic is less than 10.

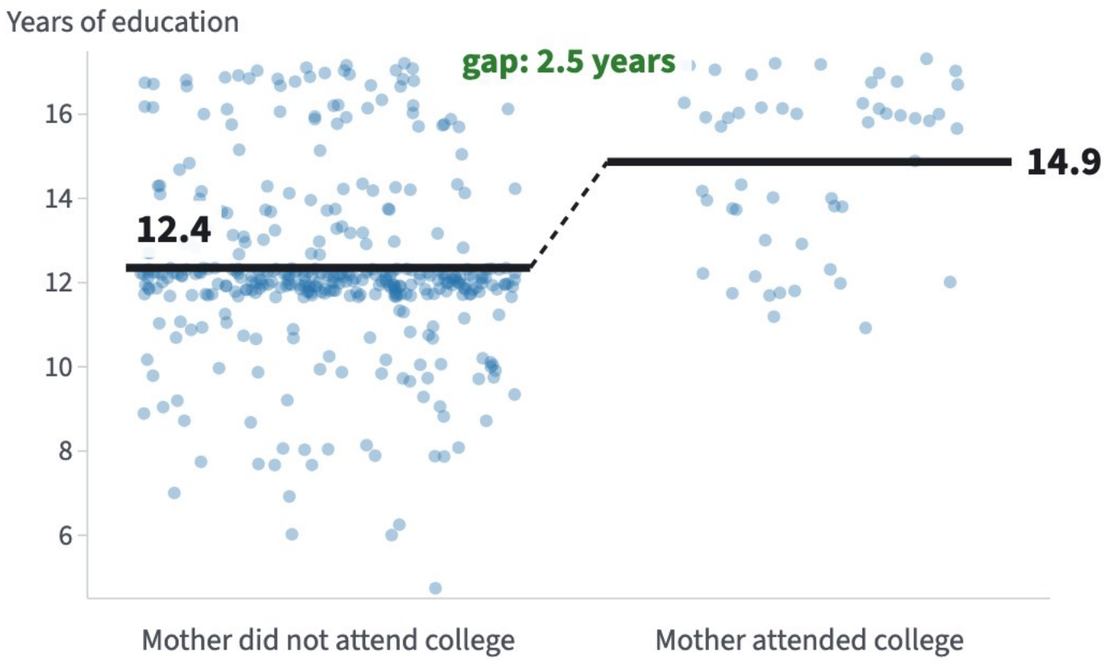
e The strength of the instruments is inconclusive: the second stage of the regression is needed.

Source: Spring 2023 Midterm V1, Question 14 (some echoed lines omitted)

A Relevant Instrument Moves Education

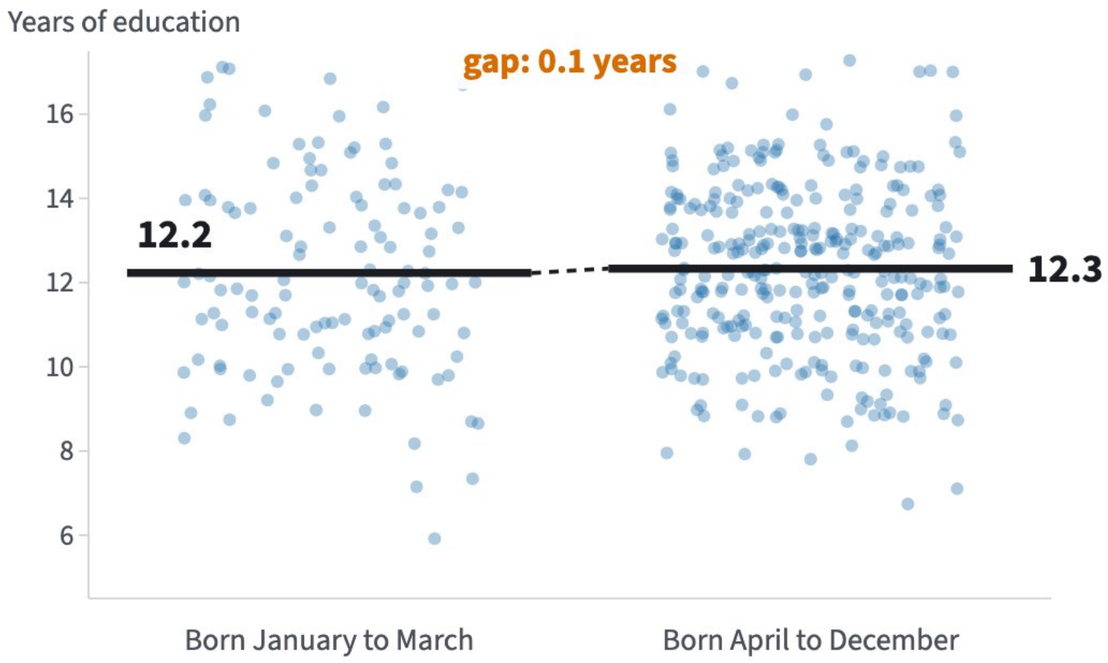
MOTHER WENT TO COLLEGE (REAL DATA)

428 working women from the exam's data. Dash: group average. Gap: **2.5 years**.



A WEAK INSTRUMENT: QUARTER OF BIRTH (ILLUSTRATION)

Born January to March $\Longrightarrow$ about **0.1 fewer years** of school (Angrist and Krueger 1991).



2SLS uses only the variation in education that the instrument creates: **a lot** on the left, **very little** on the right. With 329,509 men, quarter of birth gives a first-stage F of 13.5 with 3 dummies and **1.6** with 28 (Bound, Jaeger, and Baker 1995).

With a Weak Instrument, 2SLS Is Unreliable

REPEAT THE STUDY 200 TIMES

Each dot: 2SLS on 428 new women. True value: 0.06 (black line).

STRONG INSTRUMENT

First-stage F about 62: **171 of 200** estimates are within 0.05 of 0.06.

WEAK INSTRUMENT

First-stage F about 2: only **45 of 200** are within 0.05.

Rule of thumb: first-stage $F > 10$.

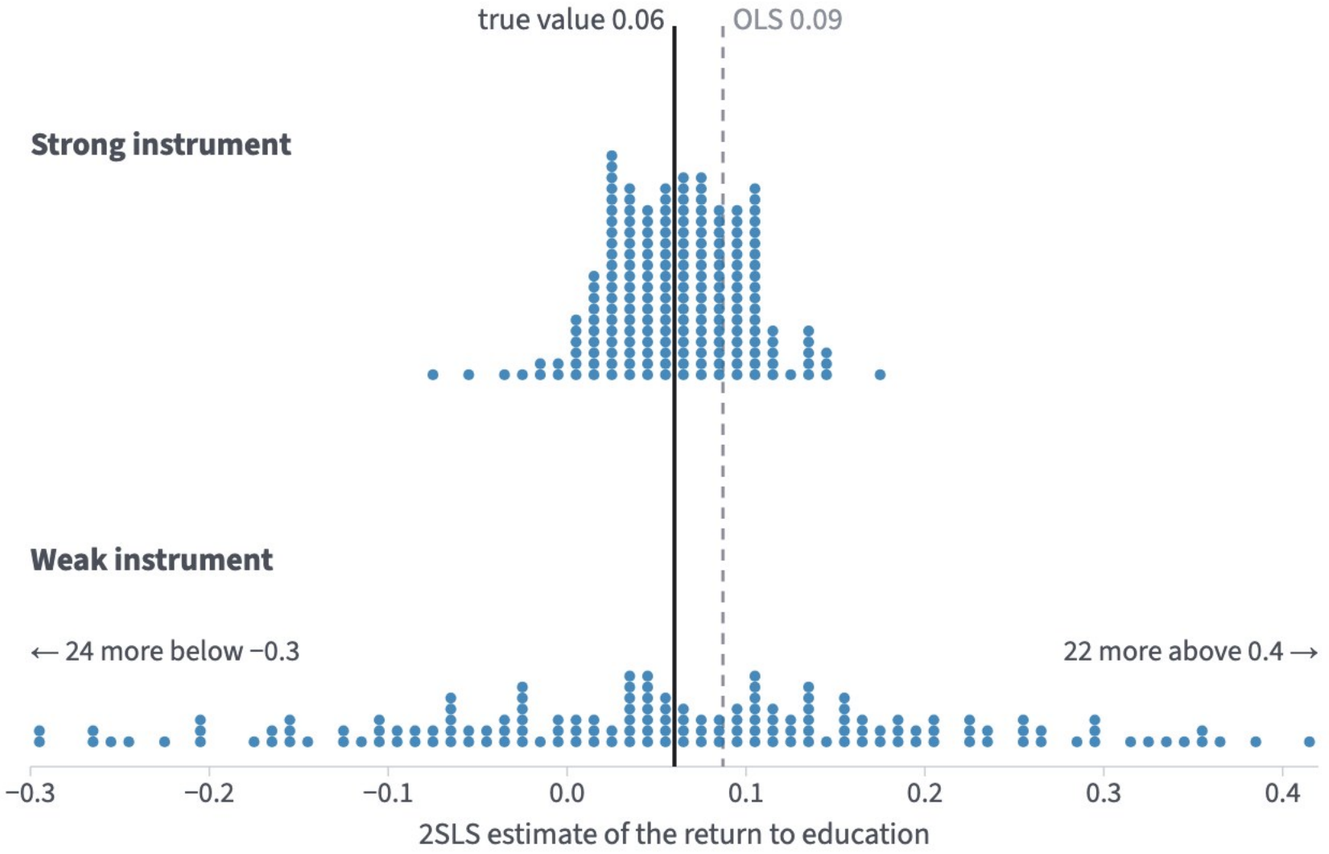


Illustration: unobserved ability raises education and wages, so OLS is biased upward (dashed gray, about 0.09).

Compute the First-Stage F Statistic

```
## Model 1: restricted model
## Model 2: educ ~ exper + I(exper^2) + mothercoll + fathercoll
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    425 2219.2
## 2    423 1748.3  2    470.88 56.963 < 2.2e-16 ***
```

`Model 2`	unrestricted first stage: educ on all exogenous variables, including both instruments
`Model 1`	restricted: the same regression with mothercoll and fathercoll dropped
`RSS`	residual sum of squares, $\sum \hat{v}_i^2$. The textbook calls it SSE.
`Res.Df`	$N - K$: observations minus coefficients ($428 - 5 = 423$)
`Df`	J: the number of restrictions tested (2 instruments)

FIRST STAGE

$$EDUC = \gamma_1 + \gamma_2 EXPER + \gamma_3 EXPER^2 + \theta_1 MOTHERCOLL + \theta_2 FATHERCOLL + v$$

HYPOTHESES

$H_0: \theta_1 = \theta_2 = 0 \iff$ the instruments do not move education

H_1: at least one $\theta \neq 0$

FORMULA

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

PLUG IN

$$F = \frac{(2219.2 - 1748.3)/2}{1748.3/423} = \frac{235.45}{4.133}$$

$$F \approx 56.97$$

> R prints 56.963 because it uses the unrounded sums of squares.

Compare F to Two Thresholds

UNDER H_0

$F \sim F_{(2,\,423)}$

SIGNIFICANCE

5% critical value: 3.02. Above it $\Longrightarrow$ **relevant**.

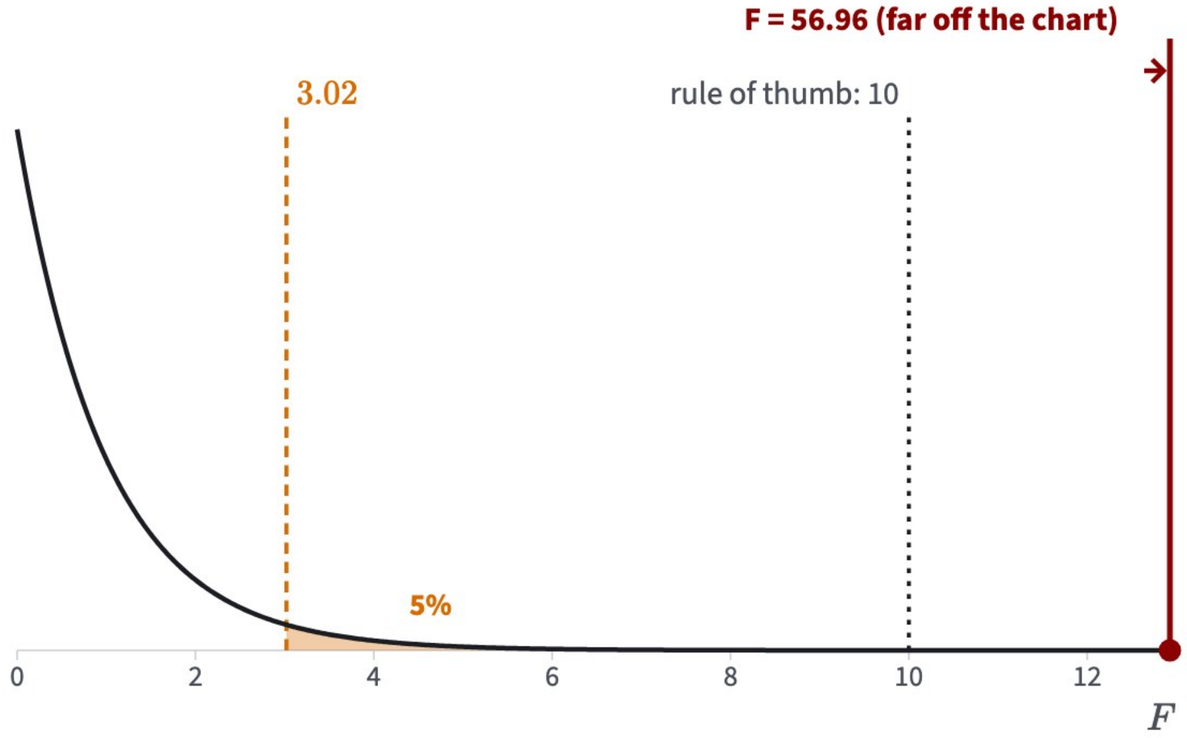
STRENGTH

Rule of thumb: $F > 10 \Longrightarrow$ **strong**. Stricter than significance.

OUR STATISTIC

$F = 56.96$: above both.

STRONG $56.96 > 10 \Longrightarrow$ mother's and father's college strongly predict education.



Checking Each Answer Choice

c The instruments are weak: The F statistic is less than 10.

56.96 is not less than 10.

d The instruments are weak: The F statistic is less than 10.

Same text as (c) on the exam.

e The strength of the instruments is inconclusive: the second stage of the regression is needed.

Strength is a first-stage property.

a The instruments are strong: The F statistic is significant at the 5% level.

Right conclusion, wrong reason: 5% significance only needs $F > 3.02$.

b The instruments are strong: The F statistic is greater than 10.

Answer: (b)

SIGNIFICANT IS NOT STRONG

$F = 5$ is significant ($5 > 3.02$) but **weak** ($5 < 10$).

COMPARE: THE CHINA SYNDROME

First-stage $F = 11.1$: **just above 10** (Spring 2026 Final, Question 3).

The Question: Is the Extra Instrument Valid?

Below we estimate a wage equation with the purpose of evaluating the overidentifying restrictions. In the model, we only need one instrument (e.g. *mothereduc*) for years education (*educ*), therefore, we want to evaluate if the extra instrument (e.g. *fathereduc*) is valid. Based on the results below, which conclusion is correct?

```
mroz1 <- mroz[mroz$lfp==1,] #restricts sample to lfp=1.
mroz1.iv1 <- ivreg(log(wage)~educ+exper+I(exper^2)|
           exper+I(exper^2)+mothereduc+fathereduc,
           data=mroz1)
summary(mroz1.iv1, diagnostics=TRUE)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0481003  0.4003281   0.120  0.90442
educ         0.0613966  0.0314367   1.953  0.05147 .
exper        0.0441704  0.0134325   3.288  0.00109 **
I(exper^2)  -0.0008990  0.0004017  -2.238  0.02574 *
Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   2 423    55.400  <2e-16 ***
Wu-Hausman         1 423     2.793  0.0954 .
Sargan             1  NA     0.378  0.5386
```

a Based on the Sargan test, we fail to reject the null, and conclude that the extra instrument is valid.

b Based on the Sargan test, we fail to reject the null, and conclude that the extra instrument is not valid.

c Based on the Wu-Hausman test, we fail to reject the null, and conclude that the extra instrument is valid.

d Based on the Wu-Hausman test, we reject the null, and conclude that the extra instrument is valid.

e We cannot come to a conclusion from the information given.

Source: Spring 2023 Midterm V1, Question 16 (residuals, fit statistics, and significance codes omitted)

Two Instruments Give Two Estimates

ONE INSTRUMENT AT A TIME

Each instrument gives its own 2SLS estimate of β_{educ}.

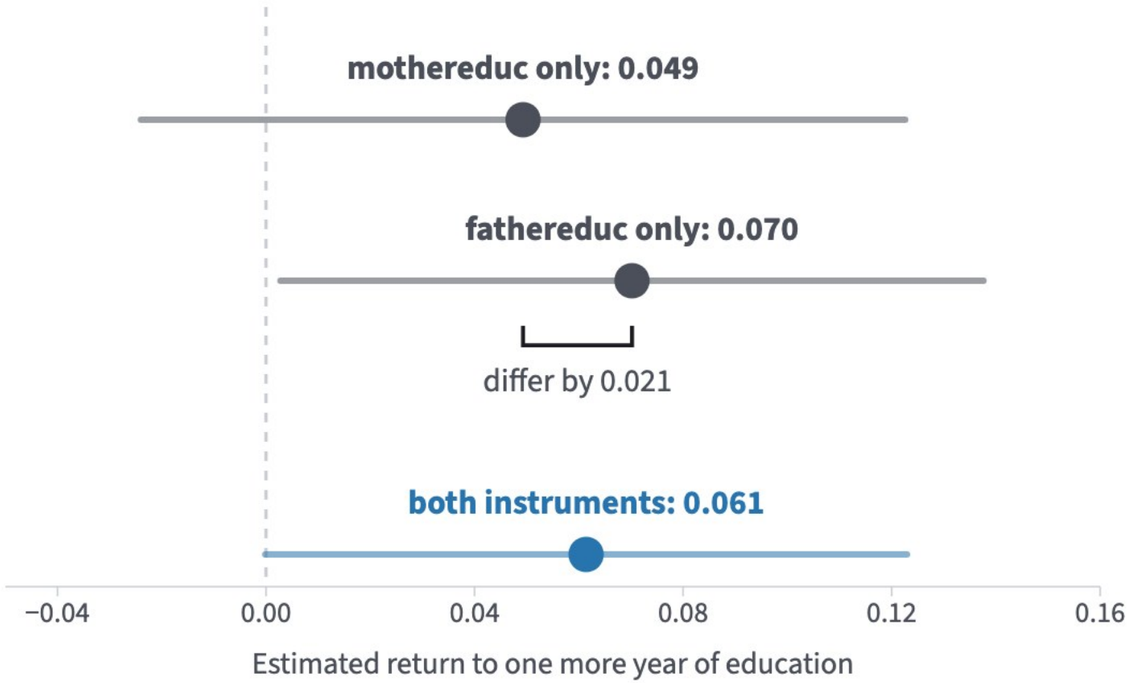
DO THEY AGREE?

Both valid $\implies$ estimates close. Gap: 0.021, small next to SEs of about 0.035.

IF ONE WERE INVALID

Its estimate would differ from β_{educ}, and the two would disagree.

Sargan asks: do the instruments **disagree by more than chance**? Both together: 0.061.



Exam data (mroz, 428 working women), exper and exper² as controls. Bars: 95% confidence intervals.

Count the Instruments and the Endogenous Regressors

```
ivreg(log(wage) ~ educ + exper + I(exper^2) |
                  exper + I(exper^2) + mothereduc + fathereduc)
```

`before |` the wage equation we want to estimate

`after |` every exogenous variable: the controls and the instruments

ENDOGENOUS REGRESSORS: $B = 1$

educ: in the model, **missing after** the `|` sign.

CONTROLS

exper, exper^2: on both sides. Do not count them.

INSTRUMENTS: $L = 2$

mothereduc, fathereduc: **only after** the `|` sign.

Count	Name	Sargan test?
$L < B$	not identified	no estimate at all
$L = B$	exactly identified	**no: nothing extra to test**
$L > B$	overidentified	**yes, with $L - B$ df**

$L - B = 1$ surplus instrument $\Longrightarrow$ Sargan has `df1 = 1`. China Syndrome: $L = B = 1$ $\Longrightarrow$ **no Sargan test**.

Compute the Sargan Statistic

HYPOTHESES

H_0: every instrument is uncorrelated with ε (the surplus instrument is valid)
H_1: at least one instrument is correlated with ε

1. ESTIMATE BY 2SLS

Both instruments. Save the residuals $\hat{\varepsilon}_i$.

2. AUXILIARY REGRESSION

Regress $\hat{\varepsilon}_i$ on all exogenous variables. Save R^2.

3. TEST STATISTIC

Valid instruments $\implies$ R^2 **near 0**.

$$\text{Sargan} = N \times R^2 \sim \chi^2_{(L-B)}$$

	df1	df2	statistic	p-value
Sargan	1	NA	0.378	0.5386

FROM THE OUTPUT

$N \times R^2 = 0.378$, with $L - B = 1$ df.

THE AUXILIARY R^2

$0.378/428 = 0.0009$: **almost none** of the residuals explained.

Compare Sargan to the $\chi^2_{(1)}$ Distribution

UNDER H_0

$\text{Sargan} \sim \chi^2_{(1)}$

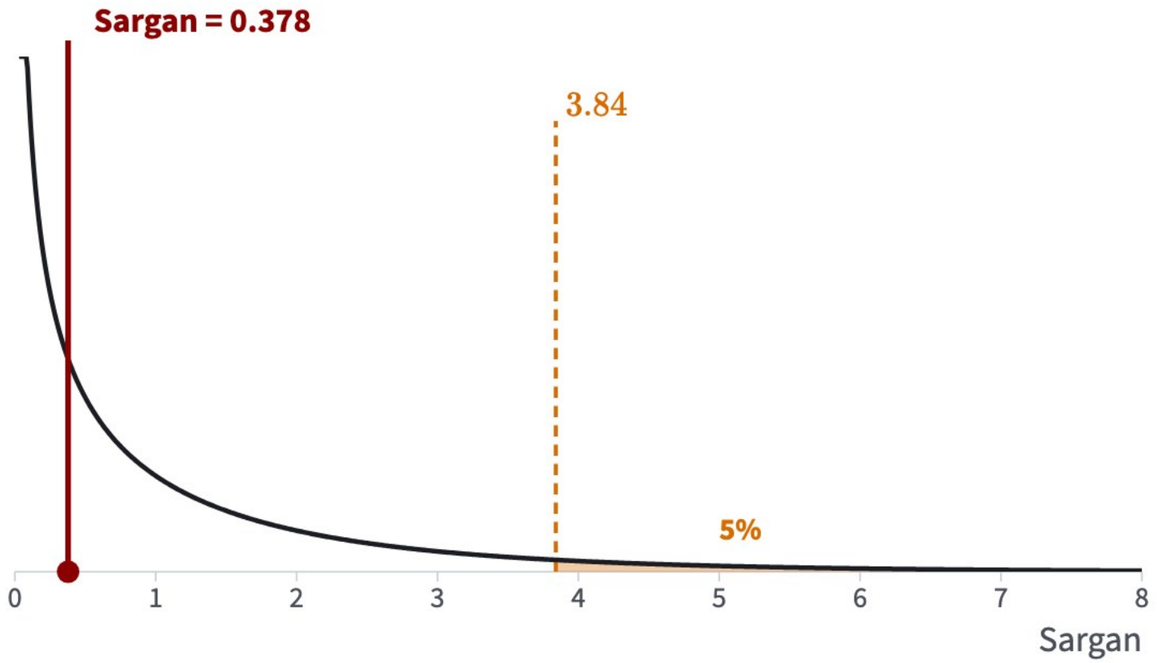
REJECTION REGION

5% critical value: 3.84

OUR STATISTIC

Sargan $= 0.378$ is far below 3.84 ($p = 0.54$).

FAIL TO REJECT H_0 $0.378 < 3.84 \implies$ no evidence against the extra instrument.



Each Diagnostic Row Answers a Different Question

Row	Question it answers	H_0	Statistic	p-value	Decision at 5%
`Weak instruments`	Do the instruments move educ?	first-stage instrument coefficients $= 0$	55.40	$< 2 \times 10^{-16}$	**reject: strong** ($55.4 > 10$)
`Wu-Hausman`	Is **educ** endogenous? Compares OLS (0.107) with 2SLS (0.061).	educ is exogenous	2.793	0.095	fail to reject (would reject at 10%)
`Sargan`	Are the **instruments** valid? Needs a surplus instrument.	all instruments are valid	0.378	0.54	fail to reject

INSTRUMENT VALIDITY

Only **Sargan** tests the instruments. Wu-Hausman tests the regressor educ.

ORDER TO READ THEM

Strength ($F > 10$), then **validity** (Sargan), then **was 2SLS needed** (Wu-Hausman).

Checking Each Answer Choice

c Based on the Wu-Hausman test, we fail to reject the null, and conclude that the extra instrument is valid.

Wu-Hausman tests whether educ is endogenous, not whether an instrument is valid.

d Based on the Wu-Hausman test, we reject the null, and conclude that the extra instrument is valid.

Wrong test, and $p = 0.095$ does not reject at 5%.

b Based on the Sargan test, we fail to reject the null, and conclude that the extra instrument is not valid.

Failing to reject gives no evidence that the instrument is invalid.

e We cannot come to a conclusion from the information given.

The Sargan row is printed: $0.378, p = 0.54$.

a Based on the Sargan test, we fail to reject the null, and conclude that the extra instrument is valid.

Answer: (a)

CAREFUL WITH "CONCLUDE VALID"

Fail to reject = **no evidence against** the instrument, not proof that it is valid.

ASSUMPTION

At least one instrument is valid. Sargan only **compares the instruments** with each other.

Thank you!

jakeanderson@g.ucla.edu