Instrumental Variables

RMDA II — Spring 2026

Sebastian Tello-Trillo

University of Virginia

Today's Roadmap

Part I: Intuition

1. The endogeneity problem
2. The idea: exogenous variation
3. Twins example
4. The IV "picture"

Part II: Application

5. Reduced form way
6. First stage + ITT
7. Two-stage least squares
8. Stata implementation

Part III: Assumptions & LATE

9. Relevance
10. Exclusion restriction
11. LATE & complier types
12. Monotonicity
13. Empirical evidence

Part IV: 16 and Pregnant

Part V: Other Considerations + Group Exercise

Key Insight

IV is an empirical strategy that uses an external source of variation to isolate the **causal effect** of an endogenous variable—when adding controls is not enough.

Part I: Developing Intuition

The Problem: Endogeneity (A Quick Review)

- We want to estimate the causal effect of D on Y:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

The Problem: Endogeneity (A Quick Review)

- We want to estimate the causal effect of D on Y:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

- But if $\text{Cov}(D, \epsilon) \neq 0$, then OLS is **biased**.

The Problem: Endogeneity (A Quick Review)

- We want to estimate the causal effect of D on Y:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

- But if $\text{Cov}(D, \epsilon) \neq 0$, then OLS is **biased**.
- We've been dealing with this by adding controls:

$$Y_i = \beta_0 + \beta_1 D_i + \gamma \vec{X}_i + \epsilon_i$$

The Problem: Endogeneity (A Quick Review)

- We want to estimate the causal effect of D on Y:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

- But if $\text{Cov}(D, \epsilon) \neq 0$, then OLS is **biased**.
- We've been dealing with this by adding controls:

$$Y_i = \beta_0 + \beta_1 D_i + \gamma \vec{X}_i + \epsilon_i$$

- But what if we can **never** control for everything?
  - Unobserved ability, motivation, preferences…
  - The list could be endless.

A Motivating Example

Does **family size** affect **educational investment** in children?

A Motivating Example

Does **family size** affect **educational investment** in children?

$$\text{Years of Schooling}_i = \alpha_0 + \delta \cdot \text{FamilySize}_i + \epsilon_i$$

A Motivating Example

Does **family size** affect **educational investment** in children?

$$\text{Years of Schooling}_i = \alpha_0 + \delta \cdot \text{FamilySize}_i + \epsilon_i$$

Question

What is a measure of family size? What is a measure of educational investment? Why might family size be endogenous?

Why Is Family Size Endogenous?

- Family size is correlated with things in ϵ:

Why Is Family Size Endogenous?

- Family size is correlated with things in ϵ:
  - Socioeconomic status (wealthier families may choose fewer kids and invest more)
  - Parental education (more educated parents may have fewer children)
  - Cultural preferences

Why Is Family Size Endogenous?

- Family size is correlated with things in ϵ:
  - Socioeconomic status (wealthier families may choose fewer kids and invest more)
  - Parental education (more educated parents may have fewer children)
  - Cultural preferences
- We can add controls, but this could be **endless**:

$$\text{Schooling}_i = \alpha_0 + \delta \cdot \text{FamSize}_i + \gamma \vec{X}_i + \epsilon_i$$

Why Is Family Size Endogenous?

- Family size is correlated with things in ϵ:
  - Socioeconomic status (wealthier families may choose fewer kids and invest more)
  - Parental education (more educated parents may have fewer children)
  - Cultural preferences
- We can add controls, but this could be **endless**:

$$\text{Schooling}_i = \alpha_0 + \delta \cdot \text{FamSize}_i + \gamma \vec{X}_i + \epsilon_i$$

- There will always be something unobserved. So what else can we do?

The Key Idea: Find Exogenous Variation

- The problem: $\text{Cov}(\text{FamSize}, \epsilon) \neq 0$

The Key Idea: Find Exogenous Variation

- The problem: $\text{Cov}(\text{FamSize}, \epsilon) \neq 0$
- What if we could "extract" only the **clean variation** in FamilySize—the part that is not correlated with the error?

The Key Idea: Find Exogenous Variation

- The problem: $\text{Cov}(\text{FamSize}, \epsilon) \neq 0$
- What if we could "extract" only the **clean variation** in FamilySize—the part that is not correlated with the error?
- We need a variable Z (an **instrument**) that:
    1. **Moves** D (family size)
    2. Does **not** affect Y (schooling) through any other channel

The Key Idea: Find Exogenous Variation

- The problem: $\text{Cov}(\text{FamSize}, \epsilon) \neq 0$
- What if we could "extract" only the **clean variation** in FamilySize—the part that is not correlated with the error?
- We need a variable Z (an **instrument**) that:
  1. **Moves** D (family size)
  2. Does **not** affect Y (schooling) through any other channel

Key Insight

An instrument is a source of exogenous variation in the endogenous variable. It lets us isolate the part of D that is "clean" and use only that part to estimate the effect on Y.

Thinking about the instrument

- Imagine a family that chooses to have a second child…

Thinking about the instrument

- Imagine a family that chooses to have a second child…
- …and then surprise: they get **two**!

Thinking about the instrument

- Imagine a family that chooses to have a second child…
- …and then surprise: they get **two**!
- Twinning is an "exogenous shock" to family size.

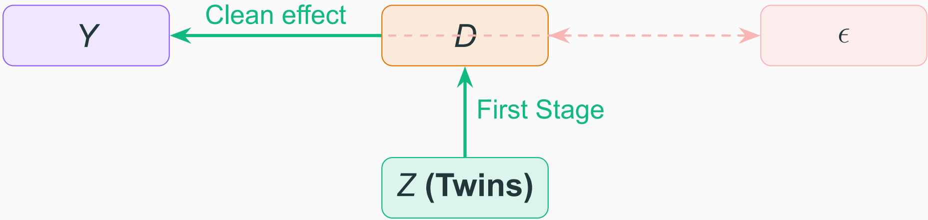
Thinking about the instrument

- Imagine a family that chooses to have a second child…
- …and then surprise: they get **two**!
- Twinning is an "exogenous shock" to family size.
- It seems unrelated to future investment in children otherwise.

Thinking about the instrument

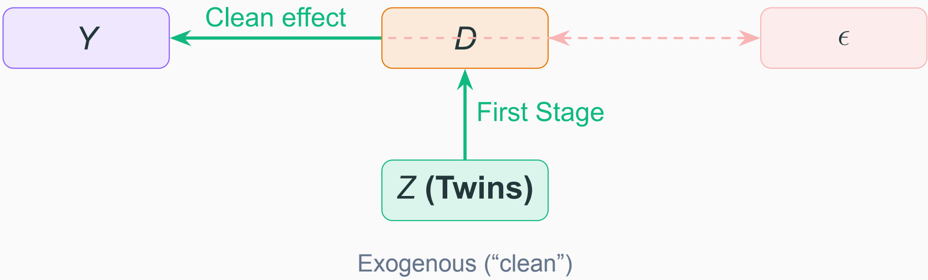
- Imagine a family that chooses to have a second child…
- …and then surprise: they get **two**!
- Twinning is an “exogenous shock” to family size.
- It seems unrelated to future investment in children otherwise.
- This gives us **clean variation** in family size that we can use!

The Solution: Introduce Z



Exogenous ("clean")

The Solution: Introduce Z

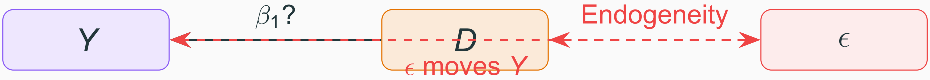


The **green arrows** show the clean path: $Z \to D \to Y$. The instrument only affects Y through D—bypassing the endogeneity problem entirely.

More IV Examples →

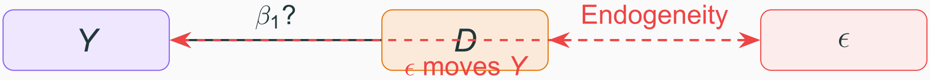
The Problem: A Picture

Think of the regression $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$:



The Problem: A Picture

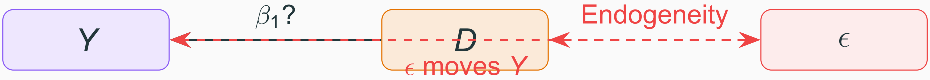
Think of the regression $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$:



- The **red two-way arrows** are the problem: D and ϵ move together, and ϵ moves Y.

The Problem: A Picture

Think of the regression $Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$:



- The **red two-way arrows** are the problem: D and ϵ move together, and ϵ moves Y.
- So we can't tell if Y changed because of D or because of ϵ.

Setting Up the Language

- Before we proceed, let's fix our notation:

Setting Up the Language

- Before we proceed, let's fix our notation:
  - Y = the **outcome** (what we care about)
  - D = the **endogenous variable** (the "treatment")
  - Z = the **instrument** (the source of clean variation)

Setting Up the Language

- Before we proceed, let's fix our notation:
  - Y = the **outcome** (what we care about)
  - D = the **endogenous variable** (the "treatment")
  - Z = the **instrument** (the source of clean variation)

Question

In our twins example, what is Y, D, and Z?

Answer: The Language of IV

- Y = Years of Schooling (outcome)

Answer: The Language of IV

- Y = Years of Schooling (outcome)
- D = Family Size (endogenous treatment)

Answer: The Language of IV

- Y = Years of Schooling (outcome)
- D = Family Size (endogenous treatment)
- Z = Having twins at second birth (instrument)

Key Insight

Once you identify Y, D, and Z, you can implement IV. There are **two ways**: the reduced form way and two-stage least squares (2SLS).

Check Your Understanding: Intuition

Question

Q1: A researcher wants to estimate the effect of exercise on mental health. A colleague suggests using "distance to nearest gym" as an instrument.

1. What are Y, D, and Z?
2. Draw the IV diagram. Does this seem like a reasonable instrument?
3. What might go wrong with this instrument?

Answer: Exercise and Mental Health

- Y = mental health, D = exercise, Z = distance to gym

Answer: Exercise and Mental Health

- Y = mental health, D = exercise, Z = distance to gym
- **Relevance**: People closer to a gym probably exercise more. Plausible.

Answer: Exercise and Mental Health

- Y = mental health, D = exercise, Z = distance to gym
- **Relevance**: People closer to a gym probably exercise more. Plausible.
- **Concern**: Distance to gym is correlated with neighborhood characteristics (income, urbanity, green space) that also affect mental health.

Answer: Exercise and Mental Health

- Y = mental health, D = exercise, Z = distance to gym
- **Relevance**: People closer to a gym probably exercise more. Plausible.
- **Concern**: Distance to gym is correlated with neighborhood characteristics (income, urbanity, green space) that also affect mental health.
- This violates the exclusion restriction unless we control for those factors.

Warning

Just because Z "sounds random" doesn't mean it is. Always think about what else Z might be correlated with.

Check Your Understanding: Propose an Instrument

Question

Q2: For each research question, propose an instrument and evaluate whether it is likely valid:

1. Does **class size** affect **student test scores**?
2. Does **incarceration** affect **future employment**?
3. Does **smoking** affect **birth weight**?

Answer: Proposed Instruments

1. **Class size** $\rightarrow$ **test scores**: Maimonides' Rule—schools must split classes at 40 students. Discontinuous jump creates exogenous variation in class size.

Answer: Proposed Instruments

1. **Class size** $\rightarrow$ **test scores**: Maimonides' Rule—schools must split classes at 40 students. Discontinuous jump creates exogenous variation in class size.
2. **Incarceration** $\rightarrow$ **employment**: Random assignment of judges (some judges are harsher). Judge leniency as instrument.

Answer: Proposed Instruments

1. **Class size** $\rightarrow$ **test scores**: Maimonides' Rule—schools must split classes at 40 students. Discontinuous jump creates exogenous variation in class size.
2. **Incarceration** $\rightarrow$ **employment**: Random assignment of judges (some judges are harsher). Judge leniency as instrument.
3. **Smoking** $\rightarrow$ **birth weight**: State cigarette taxes. Higher taxes $\rightarrow$ less smoking, unlikely to affect birth weight otherwise (conditional on income, etc.).

Part II: How to Apply IV

Two Ways to Implement IV

- Once we have Y, D, and Z, there are two approaches:

Two Ways to Implement IV

- Once we have Y, D, and Z, there are two approaches:
  1. **Reduced Form Way** (the intuitive approach)
  2. **Two-Stage Least Squares (2SLS)** (the practical tool)

Two Ways to Implement IV

- Once we have Y, D, and Z, there are two approaches:
  1. **Reduced Form Way** (the intuitive approach)
  2. **Two-Stage Least Squares (2SLS)** (the practical tool)
- Let's start with the reduced form way.

Step 1: The First Stage

Estimate the effect of the instrument (Z) on the treatment (D):

$$\text{FamilySize}_i = \phi_0 + \phi_1 \cdot \text{Twins}_i + \nu_i$$

Step 1: The First Stage

Estimate the effect of the instrument (Z) on the treatment (D):

$$\text{FamilySize}_i = \phi_0 + \phi_1 \cdot \text{Twins}_i + \nu_i$$

- In Stata: `regress D Z`

Step 1: The First Stage

Estimate the effect of the instrument (Z) on the treatment (D):

$$\text{FamilySize}_i = \phi_0 + \phi_1 \cdot \text{Twins}_i + \nu_i$$

- In Stata: `regress D Z`
- This is the **First Stage**.
- It tells us: how much does the instrument move D?

Step 1: The First Stage

Estimate the effect of the instrument (Z) on the treatment (D):

$$\text{FamilySize}_i = \phi_0 + \phi_1 \cdot \text{Twins}_i + \nu_i$$

- In Stata: `regress D Z`
- This is the **First Stage**.
- It tells us: how much does the instrument move D?
- If $\hat{\phi}_1 = 0$, the instrument doesn't work at all.

Step 2: The Reduced Form (ITT)

Estimate the effect of the instrument (Z) on the outcome (Y):

$$\text{YearsSchooling}_i = \rho_0 + \rho_1 \cdot \text{Twins}_i + \mu_i$$

Step 2: The Reduced Form (ITT)

Estimate the effect of the instrument (Z) on the outcome (Y):

$$\text{YearsSchooling}_i = \rho_0 + \rho_1 \cdot \text{Twins}_i + \mu_i$$

- In Stata: `regress Y Z`

Step 2: The Reduced Form (ITT)

Estimate the effect of the instrument (Z) on the outcome (Y):

$$\text{YearsSchooling}_i = \rho_0 + \rho_1 \cdot \text{Twins}_i + \mu_i$$

- In Stata: `regress Y Z`
- This is the **Reduced Form** or **Intent-to-Treat (ITT)**.
- It captures the overall effect of Z on Y (through D).

Step 2: The Reduced Form (ITT)

Estimate the effect of the instrument (Z) on the outcome (Y):

$$\text{YearsSchooling}_i = \rho_0 + \rho_1 \cdot \text{Twins}_i + \mu_i$$

- In Stata: `regress Y Z`
- This is the **Reduced Form** or **Intent-to-Treat (ITT)**.
- It captures the overall effect of Z on Y (through D).
- If $\rho_1 = 0$, then either Z doesn't affect D, or D doesn't affect Y.

Step 3: The Wald Estimator

Divide the reduced form by the first stage:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}}$$

Step 3: The Wald Estimator

Divide the reduced form by the first stage:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}}$$

Key Insight

Think of it as the "chain rule": Effect of Z on Y = (Effect of Z on D) $\times$ (Effect of D on Y).

So: What we want $= \frac{\text{Reduced Form}}{\text{First Stage}}$

Let's Step Back: What Did Each Piece Do?

First Stage ($\hat{\phi}_1$)

- How much does Z move D?

Let's Step Back: What Did Each Piece Do?

First Stage ($\hat{\phi}_1$)

- How much does Z move D?
- Measures "compliance"—how many people respond to the instrument
- The "first arrow" in the DAG

Reduced Form / ITT ($\hat{\rho}_1$)

- What is the total effect of Z on Y?

Let's Step Back: What Did Each Piece Do?

First Stage ($\hat{\phi}_1$)

- How much does Z move D?
- Measures "compliance"—how many people respond to the instrument
- The "first arrow" in the DAG

Reduced Form / ITT ($\hat{\rho}_1$)

- What is the total effect of Z on Y?
- If the exclusion restriction holds, this effect comes only through D
- The "full chain" from Z to Y

Exercise: Compute the IV Estimate

A researcher studies the effect of **family size** on **children's years of schooling** using twin births as an instrument ($N = 180{,}000$ families):

	First Stage *Dep. var: FamilySize*	**Reduced Form** *Dep. var: YearsSchooling*
`Twins`	0.700^{***}	-0.140^{***}
	(0.018)	(0.028)
N	180,000	180,000

SE in parentheses. $^{***}p < 0.01$.

Question

1. Compute the Wald (IV) estimate: $\hat{\lambda}_{\text{IV}} = \hat{\rho}_1 / \hat{\phi}_1$.
2. Interpret the result in one sentence.

Answer: Computing the Wald Estimate

Step 1 — First stage ($\hat{\phi}_1$): twins increase family size by **0.700 children**.

Answer: Computing the Wald Estimate

Step 1 — First stage ($\hat{\phi}_1$): twins increase family size by **0.700 children**.

Step 2 — Reduced form ($\hat{\rho}_1$): twins reduce children's schooling by **0.140 years**.

Answer: Computing the Wald Estimate

Step 1 — First stage ($\hat{\phi}_1$): twins increase family size by **0.700 children**.

Step 2 — Reduced form ($\hat{\rho}_1$): twins reduce children's schooling by **0.140 years**.

Step 3 — Wald estimate:

$$\hat{\lambda}_{\text{IV}} = \frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{-0.140}{0.700} = -0.200$$

Answer: Computing the Wald Estimate

Step 1 — First stage ($\hat{\phi}_1$): twins increase family size by **0.700 children**.

Step 2 — Reduced form ($\hat{\rho}_1$): twins reduce children's schooling by **0.140 years**.

Step 3 — Wald estimate:

$$\hat{\lambda}_{\text{IV}} = \frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{-0.140}{0.700} = -0.200$$

Key Insight

Interpretation: One additional child causes a reduction of approximately **0.2 years** in each child's schooling—consistent with the quantity–quality tradeoff.

Why Do We Need 2SLS?

- The reduced form / Wald estimator is intuitive, but:

Why Do We Need 2SLS?

- The reduced form / Wald estimator is intuitive, but:
  - It doesn't give us **standard errors** easily
  - It's hard to add **controls**
  - Most people use a single command in Stata

Why Do We Need 2SLS?

- The reduced form / Wald estimator is intuitive, but:
  - It doesn't give us **standard errors** easily
  - It's hard to add **controls**
  - Most people use a single command in Stata
- Two-Stage Least Squares (2SLS) does the same thing but better.
- In fact, with one instrument and no controls, 2SLS = Wald estimator.

2SLS: The First Stage

- First stage (same as before):

$$D_i = \alpha_0 + \phi Z_i + \epsilon_i$$

2SLS: The First Stage

- First stage (same as before):

$$D_i = \alpha_0 + \phi Z_i + \epsilon_i$$

- Obtain $\hat{\alpha}_0$ and $\hat{\phi}$, then compute the **fitted values**:

$$\hat{D}_i = \hat{\alpha}_0 + \hat{\phi} Z_i$$

2SLS: The First Stage

- First stage (same as before):

$$D_i = \alpha_0 + \phi Z_i + \epsilon_i$$

- Obtain $\hat{\alpha}_0$ and $\hat{\phi}$, then compute the **fitted values**:

$$\hat{D}_i = \hat{\alpha}_0 + \hat{\phi} Z_i$$

- $\hat{D}_i$ is the part of D that is predicted by the instrument only.

2SLS: The Second Stage

- Use $\hat{D}_i$ in the "second stage":

$$Y_i = \gamma + \lambda_{2SLS} \hat{D}_i + \nu_i$$

2SLS: The Second Stage

- Use $\hat{D}_i$ in the "second stage":

$$Y_i = \gamma + \lambda_{\text{2SLS}} \hat{D}_i + \nu_i$$

- Since $\hat{D}_i$ only contains variation from Z, and Z is exogenous…

2SLS: The Second Stage

- Use $\hat{D}_i$ in the "second stage":

$$Y_i = \gamma + \lambda_{\text{2SLS}} \hat{D}_i + \nu_i$$

- Since $\hat{D}_i$ only contains variation from Z, and Z is exogenous…
- …the endogeneity problem is solved!

2SLS: The Second Stage

- Use $\hat{D}_i$ in the "second stage":

$$Y_i = \gamma + \lambda_{2SLS}\hat{D}_i + \nu_i$$

- Since $\hat{D}_i$ only contains variation from Z, and Z is exogenous…
- …the endogeneity problem is solved!

Key Insight

$\hat{D}_i$ is D "purified" of endogeneity. We stripped out all the problematic variation and kept only the clean part driven by Z.

Adding Controls to 2SLS

- First Stage:

$$D_i = \alpha_0 + \phi Z_i + \gamma_1 A_i + \epsilon_i$$

Adding Controls to 2SLS

- First Stage:

$$D_i = \alpha_0 + \phi Z_i + \gamma_1 A_i + \epsilon_i$$

- Second Stage:

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma_2 A_i + \nu_i$$

Adding Controls to 2SLS

- First Stage:

$$D_i = \alpha_0 + \phi Z_i + \gamma_1 A_i + \epsilon_i$$

- Second Stage:

$$Y_i = \alpha_2 + \lambda_{2SLS}\hat{D}_i + \gamma_2 A_i + \nu_i$$

Warning

Controls must appear in **both stages**. If you include a control in one stage but not the other, the estimates are wrong.

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

- Multiple instruments:

```
ivregress 2sls Y controls (D = Z1 Z2), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

- Multiple instruments:

```
ivregress 2sls Y controls (D = Z1 Z2), first
```

- Our example:

```
ivregress 2sls yrsschooling age race (famsize = twins), first
```

The `first` option displays the first-stage results. Always check them!

Sample Output: First Stage

```
.  ivregress 2sls yrsschooling (famsize = twins), first

First-stage regressions
-----------------------

                                        Number of obs     =         50
                                        F(   1,      48)  =      40.69
                                        Prob > F          =     0.0000
                                        R-squared         =     0.4588
                                        Adj R-squared     =     0.4475
                                        Root MSE          = 11722.1199

------------------------------------------------------------------------------
     famsize |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       twins |   4.081283   .6398354     6.38   0.000     2.794808    5.367758
       _cons |  -31100.69   12586.39    -2.47   0.017    -56407.32    -5794.06
------------------------------------------------------------------------------
```

Look at the coefficient on the instrument and the F-statistic. Is the instrument strong?

Sample Output: Second Stage (2SLS)

```
Instrumental variables (2SLS) regression          Number of obs   =        50
                                                  Wald chi2(1)    =     63.36
                                                  Prob > chi2     =    0.0000
                                                  R-squared       =    0.4359
                                                  Root MSE        =    26.287

------------------------------------------------------------------------------
yrsschooling |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     famsize |   .0027983   .0003516     7.96   0.000     .0021093    .0034874
       _cons |   99.08714   17.44581     5.68   0.000     64.89399    133.2803
------------------------------------------------------------------------------
Instrumented:  famsize
Instruments:   twins
```

The coefficient on FamilySize is now the IV estimate—using only the “clean” variation from twins.

Check Your Understanding: Standard Errors

Question

Q4: Your colleague manually runs two separate OLS regressions (first stage, then second stage using predicted values). They report: "The IV estimate is 0.15 with a standard error of 0.03." What is wrong with this approach?

Answer: The SE Problem

- The standard errors from the manual two-step procedure are **wrong**.

Answer: The SE Problem

- The standard errors from the manual two-step procedure are **wrong**.
- The second-stage regression doesn't know that $\hat{D}$ is estimated—it treats it as if it were measured without error.

Answer: The SE Problem

- The standard errors from the manual two-step procedure are **wrong**.
- The second-stage regression doesn't know that $\hat{D}$ is estimated—it treats it as if it were measured without error.
- This makes the SEs **too small** (overstates precision).

Answer: The SE Problem

- The standard errors from the manual two-step procedure are **wrong**.
- The second-stage regression doesn't know that $\hat{D}$ is estimated—it treats it as if it were measured without error.
- This makes the SEs **too small** (overstates precision).
- `ivregress 2sls` automatically corrects the standard errors.

Warning

Never run 2SLS "by hand" with two separate `reg` commands. Always use `ivregress 2sls`, which computes the correct standard errors.

Part III: Assumptions & LATE

Assumption 1: Relevance

$$\text{Cov}(Z, D) \neq 0$$

- The instrument must **cause variation** in *D*.

Assumption 1: Relevance

$$\text{Cov}(Z, D) \neq 0$$

- The instrument must **cause variation** in D.
- In our example: twins must actually increase family size.

Assumption 1: Relevance

$$\text{Cov}(Z, D) \neq 0$$

- The instrument must **cause variation** in D.
- In our example: twins must actually increase family size.
- This is **testable**! Just run the first stage and check.

Assumption 1: Relevance

$$\text{Cov}(Z, D) \neq 0$$

- The instrument must **cause variation** in D.
- In our example: twins must actually increase family size.
- This is **testable**! Just run the first stage and check.
- Rule of thumb: first-stage F-statistic > 10.

Assumption 1: Relevance

$$\text{Cov}(Z, D) \neq 0$$

- The instrument must **cause variation** in D.
- In our example: twins must actually increase family size.
- This is **testable**! Just run the first stage and check.
- Rule of thumb: first-stage F-statistic > 10.

Question

What happens if the instrument is “weak” (F-stat is small)?

What is the F-statistic?

Answer: Weak Instruments

- If Z barely moves D, you're dividing by something close to zero:

$$\hat{\lambda}_{\text{IV}} = \frac{\hat{\rho}_1}{\hat{\phi}_1} \quad \xrightarrow{\hat{\phi}_1 \to 0} \quad \text{unstable, biased}$$

Answer: Weak Instruments

- If Z barely moves D, you're dividing by something close to zero:

$$\hat{\lambda}_{\text{IV}} = \frac{\hat{\rho}_1}{\hat{\phi}_1} \quad \xrightarrow{\hat{\phi}_1 \to 0} \quad \text{unstable, biased}$$

- Weak instruments bias IV estimates **toward OLS**.
- Standard errors become unreliable.
- You lose the whole point of doing IV!

Warning

A weak first stage is one of the most common problems in applied IV. Always report the first-stage F-statistic.

Assumption 2: Exclusion Restriction

$$\text{Cov}(Z, \epsilon) = 0$$

- The instrument affects Y **only through** D.

Assumption 2: Exclusion Restriction

$$\text{Cov}(Z, \epsilon) = 0$$

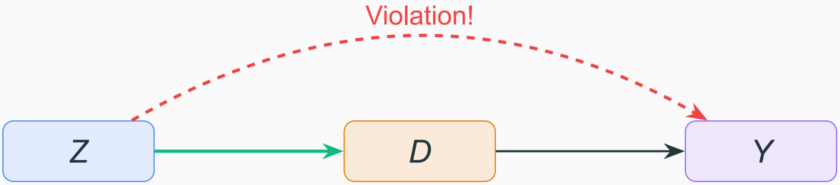
- The instrument affects Y **only through** D.
- Equivalently: conditional on controls, Z is as good as randomly assigned.

Assumption 2: Exclusion Restriction

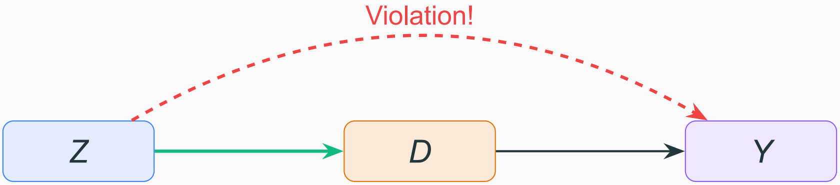
$$\text{Cov}(Z, \epsilon) = 0$$

- The instrument affects Y **only through** D.
- Equivalently: conditional on controls, Z is as good as randomly assigned.
- This is **NOT testable**—it requires a theoretical argument.

Exclusion Restriction: The Picture

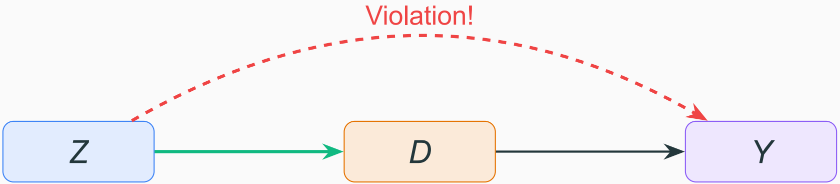


Exclusion Restriction: The Picture



- If Z affects Y directly (the red dashed arrow), the exclusion restriction fails.

Exclusion Restriction: The Picture



- If Z affects Y directly (the red dashed arrow), the exclusion restriction fails.
- Example: if twins cause stress that directly affects parenting quality (beyond family size), the exclusion restriction is violated.

Why Does the Exclusion Restriction Matter?

Taking the expected value of the IV estimator:

$$\mathbb{E}(\hat{\beta}_{\text{IV}}) = \beta_{\text{LATE}} + \frac{\text{Cov}(Z, \epsilon)}{\text{Cov}(Z, D)}$$

Why Does the Exclusion Restriction Matter?

Taking the expected value of the IV estimator:

$$\mathbb{E}(\hat{\beta}_{\text{IV}}) = \beta_{\text{LATE}} + \frac{\text{Cov}(Z, \epsilon)}{\text{Cov}(Z, D)}$$

- We need $\text{Cov}(Z, D) \neq 0$ (relevance—first assumption).
- To eliminate bias, we need $\text{Cov}(Z, \epsilon) = 0$ (exclusion restriction).

Why Does the Exclusion Restriction Matter?

Taking the expected value of the IV estimator:

$$\mathbb{E}(\hat{\beta}_{\text{IV}}) = \beta_{\text{LATE}} + \frac{\text{Cov}(Z, \epsilon)}{\text{Cov}(Z, D)}$$

- We need $\text{Cov}(Z, D) \neq 0$ (relevance—first assumption).
- To eliminate bias, we need $\text{Cov}(Z, \epsilon) = 0$ (exclusion restriction).
- If the exclusion restriction holds, the bias term vanishes and we get the causal effect.

▸ Derivation of this formula ▸ Potential Outcomes Framework

Making the Exclusion Restriction More Plausible

- In the twins example: twins are more common among older women and certain ethnicities.

Making the Exclusion Restriction More Plausible

- In the twins example: twins are more common among older women and certain ethnicities.
- These characteristics might also affect schooling investment.

Making the Exclusion Restriction More Plausible

- In the twins example: twins are more common among older women and certain ethnicities.
- These characteristics might also affect schooling investment.
- Solution: **add controls**! Now the condition becomes:

$$\text{Cov}(Z, \epsilon \mid \text{Age, Ethnicity}) = 0$$

Making the Exclusion Restriction More Plausible

- In the twins example: twins are more common among older women and certain ethnicities.
- These characteristics might also affect schooling investment.
- Solution: **add controls**! Now the condition becomes:

$$\text{Cov}(Z, \epsilon \mid \text{Age, Ethnicity}) = 0$$

- “Conditional on mother’s age and ethnicity, is twinning as good as randomly assigned?”

Making the Exclusion Restriction More Plausible

- In the twins example: twins are more common among older women and certain ethnicities.
- These characteristics might also affect schooling investment.
- Solution: **add controls**! Now the condition becomes:

$$\text{Cov}(Z, \epsilon \mid \text{Age, Ethnicity}) = 0$$

- “Conditional on mother’s age and ethnicity, is twinning as good as randomly assigned?”
- Much more believable than the unconditional version!

What Does the IV Estimate Actually Capture?

- Going back to the Wald estimator:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}} = \text{LATE}$$

What Does the IV Estimate Actually Capture?

- Going back to the Wald estimator:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}} = \text{LATE}$$

- LATE = **Local** Average Treatment Effect.

What Does the IV Estimate Actually Capture?

- Going back to the Wald estimator:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}} = \text{LATE}$$

- LATE = **Local** Average Treatment Effect.
- "Local" to whom? To the **compliers**—people whose behavior was actually changed by the instrument.

What Does the IV Estimate Actually Capture?

- Going back to the Wald estimator:

$$\frac{\hat{\rho}_1}{\hat{\phi}_1} = \frac{\text{ITT}}{\text{First Stage}} = \hat{\lambda}_{\text{IV}} = \text{LATE}$$

- LATE = **Local** Average Treatment Effect.
- "Local" to whom? To the **compliers**—people whose behavior was actually changed by the instrument.
- To understand LATE, we need to understand the different types of people.

Four Types of People

Let's use the **beer tax** example (Z = tax, D = drinking).

Complier

- Changes behavior when Z changes
- Higher tax $\rightarrow$ less drinking
- These are the people IV captures

Four Types of People

Let's use the **beer tax** example (Z = tax, D = drinking).

Complier

- Changes behavior when Z changes
- Higher tax $\rightarrow$ less drinking
- These are the people IV captures

Never-Taker

- Never drinks, regardless of tax
- Demand is always zero

Four Types of People

Let's use the **beer tax** example (Z = tax, D = drinking).

Complier

- Changes behavior when Z changes
- Higher tax $\rightarrow$ less drinking
- These are the people IV captures

Never-Taker

- Never drinks, regardless of tax
- Demand is always zero

Always-Taker

- Always drinks the same, regardless of tax
- Demand is constant

Four Types of People

Let's use the **beer tax** example (Z = tax, D = drinking).

Complier

- Changes behavior when Z changes
- Higher tax $\rightarrow$ less drinking
- These are the people IV captures

Never-Taker

- Never drinks, regardless of tax
- Demand is always zero

Always-Taker

- Always drinks the same, regardless of tax
- Demand is constant

Defier

- Does the opposite: higher tax $\rightarrow$ drinks more
- Rare in practice

LATE Is the Effect for Compliers Only

- LATE is **uninformative** about:

LATE Is the Effect for Compliers Only

- LATE is **uninformative** about:
  - Never-takers (the instrument doesn't move them)
  - Always-takers (the instrument doesn't move them either)
  - Defiers (assumed not to exist—more on this next)

LATE Is the Effect for Compliers Only

- LATE is **uninformative** about:
  - Never-takers (the instrument doesn't move them)
  - Always-takers (the instrument doesn't move them either)
  - Defiers (assumed not to exist—more on this next)

Question

In the twins example, who are the compliers? Who are the always-takers?

Answer: Compliers in the Twins Example

- **Compliers**: Families who have more children because of twins (they wouldn't have had as many otherwise).

Answer: Compliers in the Twins Example

- **Compliers**: Families who have more children because of twins (they wouldn't have had as many otherwise).
- **Always-takers**: Families who would have had many children regardless (twins or not).

Answer: Compliers in the Twins Example

- **Compliers**: Families who have more children because of twins (they wouldn't have had as many otherwise).
- **Always-takers**: Families who would have had many children regardless (twins or not).
- **Never-takers**: Families who stop after twins and don't have more.

Answer: Compliers in the Twins Example

- **Compliers**: Families who have more children because of twins (they wouldn't have had as many otherwise).
- **Always-takers**: Families who would have had many children regardless (twins or not).
- **Never-takers**: Families who stop after twins and don't have more.

Key Insight

When is LATE policy-relevant?

- When compliers are the only ones treated (no always-takers)
- When the instrument is the policy (e.g., a tax change)

Assumption 3: Monotonicity

[label=mono]

Monotonicity Assumption

There are **no defiers**.

Assumption 3: Monotonicity

[label=mono]

Monotonicity Assumption

There are **no defiers**.

- Now that you know what defiers are, this assumption says: no one does the opposite of what the instrument pushes.

Assumption 3: Monotonicity

[label=mono]

Monotonicity Assumption

There are **no defiers**.

- Now that you know what defiers are, this assumption says: no one does the opposite of what the instrument pushes.
- Beer tax example: no student drinks more when beer gets more expensive.

Assumption 3: Monotonicity

[label=mono]

Monotonicity Assumption

There are **no defiers**.

- Now that you know what defiers are, this assumption says: no one does the opposite of what the instrument pushes.
- Beer tax example: no student drinks more when beer gets more expensive.
- Twins example: no family has fewer total children because of twins.

Assumption 3: Monotonicity

[label=mono]

Monotonicity Assumption

There are **no defiers**.

- Now that you know what defiers are, this assumption says: no one does the opposite of what the instrument pushes.
- Beer tax example: no student drinks more when beer gets more expensive.
- Twins example: no family has fewer total children because of twins.
- Usually plausible—but think about whether defiers could exist in your setting.

Check Your Understanding: Application

Question

Q3: A researcher runs `ivregress 2sls lwage (educ = nearc4)` and gets $\hat{\beta}_{\text{IV}} = 0.125$. The OLS estimate was $\hat{\beta}_{\text{OLS}} = 0.074$.

1. Why might the IV estimate be larger than OLS?
2. What does each estimate measure?

Answer: Why IV $>$ OLS?

- **OLS** estimates the return for the average person—but is biased by ability.

Answer: Why IV $>$ OLS?

- **OLS** estimates the return for the average person—but is biased by ability.
- If ability bias is upward, shouldn't OLS be larger? Not necessarily—measurement error biases downward.

Answer: Why IV $>$ OLS?

- **OLS** estimates the return for the average person—but is biased by ability.
- If ability bias is upward, shouldn't OLS be larger? Not necessarily—measurement error biases downward.
- **IV** estimates the return for **compliers**—people on the margin whose education was changed by college proximity. For them, the return may be especially high.

Key Insight

IV estimates **LATE**—the effect for compliers, not the population ATE.

Part IV: Providing Evidence for Assumptions

Providing Evidence for Assumptions

1. **Does Z affect D?** (Relevance—testable)
    - Run the first stage, report the F-statistic (> 10)

Providing Evidence for Assumptions

1. **Does Z affect D?** (Relevance—testable)
   - Run the first stage, report the F-statistic (>10)
2. **Is Z as-good-as random?** (Supports exclusion)
   - Show a balance table: covariates should look similar across $Z=0$ vs $Z=1$

Providing Evidence for Assumptions

1. **Does Z affect D?** (Relevance—testable)
    - Run the first stage, report the F-statistic (> 10)
2. **Is Z as-good-as random?** (Supports exclusion)
    - Show a balance table: covariates should look similar across $Z = 0$ vs $Z = 1$
3. **Does Z affect Y through other channels?** (Exclusion)
    - Theoretical argument
    - Add controls for any observable confounders
    - No-first-stage subsample test

The No-First-Stage Subsample Test

- Find a subsample where Z does **not** affect D (no first stage).

The No-First-Stage Subsample Test

- Find a subsample where Z does **not** affect D (no first stage).
- How? Use theory: cut by demographics (age, gender, etc.).

The No-First-Stage Subsample Test

- Find a subsample where Z does **not** affect D (no first stage).
- How? Use theory: cut by demographics (age, gender, etc.).
- Then run the reduced form (reg Y Z) in that subsample.

The No-First-Stage Subsample Test

- Find a subsample where Z does **not** affect D (no first stage).
- How? Use theory: cut by demographics (age, gender, etc.).
- Then run the reduced form (reg Y Z) in that subsample.
- If there's no first stage and Z still affects Y, then Z must be affecting Y through another channel $\Rightarrow$ exclusion violation!

The No-First-Stage Subsample Test

- Find a subsample where Z does **not** affect D (no first stage).
- How? Use theory: cut by demographics (age, gender, etc.).
- Then run the reduced form (reg Y Z) in that subsample.
- If there's no first stage and Z still affects Y, then Z must be affecting Y through another channel $\Rightarrow$ exclusion violation!

Real-World Application

Example: Medicaid expansion as IV for health insurance. Test on elderly (who already have Medicare). If the Medicaid instrument still "affects" health in that group, something is wrong.

Check Your Understanding: Assumptions

Question

Q5: A researcher instruments for "having health insurance" using "winning a Medicaid lottery."

1. Which assumption is satisfied by design?
2. Which assumption still needs to be argued?
3. Who are the compliers in this setting?

Answer: Medicaid Lottery

1. **Exclusion restriction** is plausible by design: lottery is random, so $\text{Cov}(Z, \epsilon) = 0$.

Answer: Medicaid Lottery

1. **Exclusion restriction** is plausible by design: lottery is random, so $\text{Cov}(Z, \epsilon) = 0$.
2. **Relevance** needs to be shown: does winning the lottery actually increase insurance take-up? (Not everyone who wins enrolls.)

Answer: Medicaid Lottery

1. **Exclusion restriction** is plausible by design: lottery is random, so $\text{Cov}(Z, \epsilon) = 0$.
2. **Relevance** needs to be shown: does winning the lottery actually increase insurance take-up? (Not everyone who wins enrolls.)
3. **Compliers**: People who get insurance because they won the lottery. Not those who would have found insurance anyway (always-takers) or those who don't enroll even after winning (never-takers).

Check Your Understanding: Exclusion Restriction

Question

Q6: In the twins example, older women are more likely to have twins and may invest differently in children's education.

1. Does this violate **relevance** or **exclusion**?
2. How would you address this concern?

Answer: Twins and Mother's Age

- This threatens the **exclusion restriction**: age affects both twinning (Z) and schooling investment (Y), creating a back-door path.

Answer: Twins and Mother's Age

- This threatens the **exclusion restriction**: age affects both twinning (Z) and schooling investment (Y), creating a back-door path.
- It does not threaten relevance (twins still cause larger families).

Answer: Twins and Mother's Age

- This threatens the **exclusion restriction**: age affects both twinning (Z) and schooling investment (Y), creating a back-door path.
- It does not threaten relevance (twins still cause larger families).
- **Solution**: control for mother's age (and ethnicity). Now:

$$\text{Cov}(Z, \epsilon \mid \text{Age}) = 0$$

- "Among mothers of the same age, is twinning random?" Much more believable.

Part V: Other Considerations

Multiple Instruments

- You can have more than one instrument:

Multiple Instruments

- You can have more than one instrument:
- First Stage:

$$D_i = \alpha_0 + \phi_1 Z_{1i} + \phi_2 Z_{2i} + \gamma A_i + \epsilon_i$$

Multiple Instruments

- You can have more than one instrument:
- First Stage:

$$D_i = \alpha_0 + \phi_1 Z_{1i} + \phi_2 Z_{2i} + \gamma A_i + \epsilon_i$$

- Second Stage (unchanged—still just $\hat{D}_i$):

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma A_i + \nu_i$$

Multiple Instruments

- You can have more than one instrument:
- First Stage:

$$D_i = \alpha_0 + \phi_1 Z_{1i} + \phi_2 Z_{2i} + \gamma A_i + \epsilon_i$$

- Second Stage (unchanged—still just $\hat{D}_i$):

$$Y_i = \alpha_2 + \lambda_{2SLS} \hat{D}_i + \gamma A_i + \nu_i$$

- Example: instrument family size with both twins and same-sex siblings.

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

- Multiple instruments:

```
ivregress 2sls Y controls (D = Z1 Z2), first
```

How to Do IV in Stata

- Basic syntax:

```
ivregress 2sls Y (D = Z), first
```

- With controls:

```
ivregress 2sls Y controls (D = Z), first
```

- Multiple instruments:

```
ivregress 2sls Y controls (D = Z1 Z2), first
```

- Our example:

```
ivregress 2sls yrsschooling age race (famsize = twins), first
```

The `first` option displays the first-stage results. Always check them!

Can You Have Too Many Instruments?

- Yes! With multiple instruments, you should check whether they agree.

Can You Have Too Many Instruments?

- Yes! With multiple instruments, you should check whether they agree.
- **Overidentification test**: run IV separately with each instrument.
  - If the estimates are very different, one instrument may be invalid.
  - You won't know which one.

Can You Have Too Many Instruments?

- Yes! With multiple instruments, you should check whether they agree.
- **Overidentification test**: run IV separately with each instrument.
    - If the estimates are very different, one instrument may be invalid.
    - You won't know which one.
- With 2+ instruments, $\hat{\lambda}_{\text{2SLS}}$ is a **weighted average** of the individual IV estimates.

Can You Have Too Many Instruments?

- Yes! With multiple instruments, you should check whether they agree.
- **Overidentification test**: run IV separately with each instrument.
    - If the estimates are very different, one instrument may be invalid.
    - You won't know which one.
- With 2+ instruments, $\hat{\lambda}_{\text{2SLS}}$ is a **weighted average** of the individual IV estimates.
- If they have similar effect sizes, your estimate becomes more precise.

RCTs as the Perfect IV

- In an RCT, the instrument is **treatment assignment**:

RCTs as the Perfect IV

- In an RCT, the instrument is **treatment assignment**:
  - Z = assigned to treatment group
  - D = actually took the treatment
  - Relevance: assignment predicts take-up ✓
  - Exclusion: assignment is random by design ✓
  - Monotonicity: assignment only pushes toward treatment ✓

RCTs as the Perfect IV

- In an RCT, the instrument is **treatment assignment**:
  - Z = assigned to treatment group
  - D = actually took the treatment
  - Relevance: assignment predicts take-up ✓
  - Exclusion: assignment is random by design ✓
  - Monotonicity: assignment only pushes toward treatment ✓
- The lottery makes the perfect instrument!
- IV in RCTs adjusts for non-compliance (people who don't follow their assignment).

Summary: The IV Checklist

1. **Identify** Y, D, and Z

2. **Argue relevance**: Does Z move D? Show first stage, F-stat > 10

3. **Argue exclusion**: Does Z affect Y only through D? Is Z as good as randomly assigned? Theory + balance table

4. **Run 2SLS**: `ivregress 2sls Y controls (D = Z), first`

5. **Interpret as LATE**: Effect for compliers. Who are they?

6. **Report diagnostics**: First-stage F, balance table, robustness

Part VI: Example 16 and Pregnant

Can TV Influence Behavior?

- A broad policy question: how does media influence behaviors?

Can TV Influence Behavior?

- A broad policy question: how does media influence behaviors?
  - Does sexual content lead to riskier sexual behavior?
  - Does suicide coverage lead to more suicides?
  - Does food advertising lead to more consumption?

Can TV Influence Behavior?

- A broad policy question: how does media influence behaviors?
  - Does sexual content lead to riskier sexual behavior?
  - Does suicide coverage lead to more suicides?
  - Does food advertising lead to more consumption?

Real-World Application

Kearney & Levine study the impact of MTV's 16 and Pregnant on teen birth rates. Did the show increase or decrease teen pregnancies?

The Naive Approach

$$\text{Teen Births}_{ct} = \alpha_0 + \beta_1 \cdot \text{MTVShowRating}_{ct} + \epsilon$$

The Naive Approach

$$\text{Teen Births}_{ct} = \alpha_0 + \beta_1 \cdot \text{MTVShowRating}_{ct} + \epsilon$$

Question

What are the problems with this regression?

Answer: Endogeneity of TV Ratings

- MTV show ratings are **endogenous**:

Answer: Endogeneity of TV Ratings

- MTV show ratings are **endogenous**:
    - Areas with more teen pregnancies might watch more MTV
    - Urban vs. rural differences in both viewing and teen births
    - Unobserved cultural factors

Answer: Endogeneity of TV Ratings

- MTV show ratings are **endogenous**:
  - Areas with more teen pregnancies might watch more MTV
  - Urban vs. rural differences in both viewing and teen births
  - Unobserved cultural factors
- We need an **instrument** for MTV ratings.

The Instrument: Pre-Show Ratings

- Kearney & Levine use **spatial variation in ratings for the time slot** that 16 and Pregnant took over from.

The Instrument: Pre-Show Ratings

- Kearney & Levine use **spatial variation in ratings for the time slot** that 16 and Pregnant took over from.
- Some areas had high ratings for that time slot before the show aired.

The Instrument: Pre-Show Ratings

- Kearney & Levine use **spatial variation in ratings for the time slot** that 16 and Pregnant took over from.
- Some areas had high ratings for that time slot before the show aired.
- These areas were more likely to watch the new show too.

The Instrument: Pre-Show Ratings

- Kearney & Levine use **spatial variation in ratings for the time slot** that 16 and Pregnant took over from.
- Some areas had high ratings for that time slot before the show aired.
- These areas were more likely to watch the new show too.
- Key question: does pre-show time-slot popularity affect teen births through any channel other than watching 16 and Pregnant?

Variation in MTV Ratings

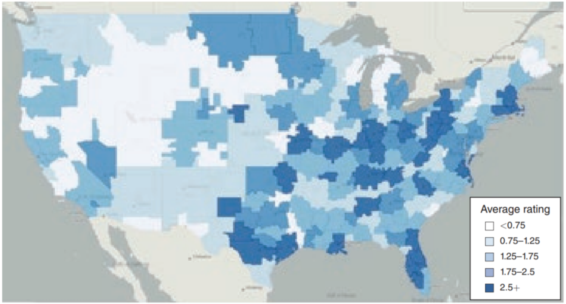


Figure 1. Nielsen Ratings for *16 and Pregnant*/*Teen Mom*/*Teen Mom 2* for those Aged 12 to 24, by Designated Market Area

Geographic variation in MTV show ratings across media markets.

First Stage Evidence

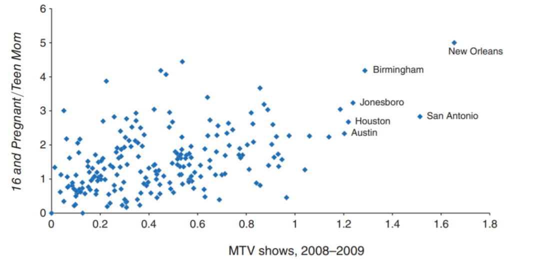


FIGURE 4. RELATIONSHIP BETWEEN MTV RATINGS IN 2008/2009 AND RATINGS FOR *16 AND PREGNANT/TEEN MOM* (*Verification of IV Monotonicity Assumption*)

Pre-show ratings predict 16 and Pregnant viewership. The first stage works.

The Naive OLS Results

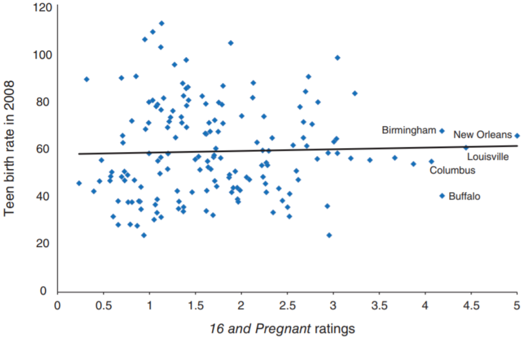


FIGURE 3. RELATIONSHIP BETWEEN *16 AND PREGNANT* RATINGS AND TEEN BIRTH RATES IN 2008

The IV Results

The IV equation takes the form of the OLS equation as represented in equation (1) above, but the variable of interest is replaced by predicted ratings:

$$(2)\quad \ln(B_{jt}) = \beta_0 + \beta_1 \widehat{Rate16P_j} \times post_t + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \boldsymbol{\theta}_t + \boldsymbol{\delta}_{js} + \varepsilon_{jt},$$

where $\widehat{Rate16P_j} \times post_t$ is generated by the first stage regression in the IV framework:

$$(3)\quad Rate16P_j \times post_t = \beta_0 + \beta_1 MTV0809_j \times post_t + \beta_2 U_{jy} + \mathbf{X}_{jy}\gamma + \boldsymbol{\theta}_t + \boldsymbol{\delta}_{js} + \varepsilon_{jt}.$$

The variable *MTV0809* represents the ratings among those between ages 12 and 24 for shows that aired between 9:00 PM to 10:00 PM on MTV between July of 2008 and May of 2009 in each media market. It is time invariant. Its interaction with *post* is consistent with that between *Rate16P* and the *post* indicator, creating an instrumental variable (labeled here as the interaction of two variables) that takes on the value of zero in the quarters before the show was introduced and the value of ratings for shows from 2008–2009 in the following quarters.[28]

Detailed Results

TABLE 1—ESTIMATES OF THE IMPACT OF *16 AND PREGNANT* RATINGS ON TEEN BIRTH RATES

	OLS (1)	First stage (2)	IV (3)	Reduced form (4)
Dependent variable:	ln(birth rate)	*16 and Pregnant* ratings	ln(birth rate)	ln(birth rate)
16 and Pregnant ratings	−1.020*		−2.368**	
	(0.552)		(0.942)	
MTV ratings 2008–2009		1.511***		−3.581**
		(0.204)		(1.512)
Unemployment rate	−1.440***	−0.001	−1.487***	−1.485***
	(0.401)	(0.026)	(0.375)	(0.432)
F-statistic on omitted instrument		48.1		

Notes: The birth data used for this analysis represents quarterly birth rates by DMA for conceptions leading to live births between 2005 and 2010. The sample size in each model is 4,919 (205 DMAs, 24 quarters, and one observation was dropped because there were no teen births). Coefficients and standard errors (reported in parentheses) in birth rate regressions are multiplied by 100. Each model also includes the percentage of a DMA's female teen that is Hispanic and non-Hispanic black along with quarter and DMA × season fixed effects. Regressions are weighted by the relevant sample sizes for each outcome. Reported standard errors are clustered at the DMA level.

Interpretation of Results

- An increase in ratings of 16 and Pregnant decreases the teen birth rate by **2.4 percent**.

Interpretation of Results

- An increase in ratings of 16 and Pregnant decreases the teen birth rate by **2.4 percent**.
- Average ratings: 1.8 points. Predicted effect: $2.37 \times 1.8 = 4.27\%$ reduction.

Interpretation of Results

- An increase in ratings of 16 and Pregnant decreases the teen birth rate by **2.4 percent**.
- Average ratings: 1.8 points. Predicted effect: $2.37 \times 1.8 = 4.27\%$ reduction.
- Teen birth rate fell 17.6% over this period.

Interpretation of Results

- An increase in ratings of 16 and Pregnant decreases the teen birth rate by **2.4 percent**.
- Average ratings: 1.8 points. Predicted effect: $2.37 \times 1.8 = 4.27\%$ reduction.
- Teen birth rate fell 17.6% over this period.
- The show explains $\frac{4.27}{17.6} =$ **24.3%** of the decline.

Interpretation of Results

- An increase in ratings of 16 and Pregnant decreases the teen birth rate by **2.4 percent**.
- Average ratings: 1.8 points. Predicted effect: $2.37 \times 1.8 = 4.27\%$ reduction.
- Teen birth rate fell 17.6% over this period.
- The show explains $\frac{4.27}{17.6}$ = **24.3%** of the decline.

Key Insight

A single TV show could account for roughly **one quarter** of the decline in teen pregnancies during this period. Media can be a powerful driver of social outcomes.

Mechanisms: Google Trends Evidence

TABLE 3—IMPACT OF EPISODE RELEASE ON NATIONAL GOOGLE SEARCHES AND TWEETS ABOUT SEXUAL ACTIVITY, CONTRACEPTION, AND ABORTION

	Google Trends: search index			Twitter: ln(tweet rate)	
	"How get birth control"	"How get birth control pill"	"How get abortion"	"Birth control"	"Abortion"
Week new *16 and Pregnant* episode released	0.825 (1.216)	2.227 (2.010)	−1.627 (1.730)	—	—
Day new *16 and Pregnant* episode released	—	—	—	0.120*** (0.047)	0.142*** (0.036)
Day new *16 and Pregnant* episode released—lagged	—	—	—	0.230*** (0.058)	0.212*** (0.046)
Number of weeks (searches)/days (tweets)	209	209	209	336	336

Google Trends and Twitter data suggest the show raised teens' interest in birth control and abortion.

Check Your Understanding: 16 and Pregnant

Question

Q7: In the Kearney & Levine study of 16 and Pregnant:

1. Who are the **compliers**? (Whose behavior does the instrument change?)
2. Is the **exclusion restriction** plausible? What might violate it?

Answer: 16 and Pregnant IV

- **Compliers**: Teens in areas that watched 16 and Pregnant because of pre-existing viewing habits for that time slot—not because of interest in teen pregnancy specifically.

Answer: 16 and Pregnant IV

- **Compliers**: Teens in areas that watched 16 and Pregnant because of pre-existing viewing habits for that time slot—not because of interest in teen pregnancy specifically.
- **Exclusion**: Pre-show time-slot ratings should only affect teen births through watching the show. Potential concern: areas with high MTV viewership may differ culturally in ways that independently affect teen birth trends.

Answer: 16 and Pregnant IV

- **Compliers**: Teens in areas that watched 16 and Pregnant because of pre-existing viewing habits for that time slot—not because of interest in teen pregnancy specifically.
- **Exclusion**: Pre-show time-slot ratings should only affect teen births through watching the show. Potential concern: areas with high MTV viewership may differ culturally in ways that independently affect teen birth trends.
- The authors address this by controlling for area-level demographics and using a difference-in-differences design within the IV framework.

Group Exercise

Group Exercise: Returns to Education

Real-World Application

Research question: What is the causal effect of education on wages?

Data: Card (1995) — uses proximity to a 4-year college as an instrument for years of education.

Variables: `lwage` (log wage), `educ` (years of education), `nearc4` (near a 4-year college), plus controls (`age`, `married`, `smsa`, `black`, `south76`).

Work in groups. Each step builds on the previous one. We'll discuss answers together.

Step 1: The Naive OLS

Question

Run the OLS regression:

```
reg lwage educ age married smsa black south76, robust
```

1. What is the estimated return to one additional year of education?
2. Why might this estimate be biased? In which direction?

Answer: Step 1

- OLS coefficient on `educ` ≈ 0.074: one more year of education is associated with about 7.4% higher wages.

Answer: Step 1

- OLS coefficient on `educ` ≈ 0.074: one more year of education is associated with about 7.4% higher wages.
- **Bias**: ability is omitted. More able people get more education and earn more.
- Sign of ability bias: $(+) \times (+) =$ positive $\Rightarrow$ OLS is biased **upward**.

Answer: Step 1

- OLS coefficient on `educ` ≈ 0.074: one more year of education is associated with about 7.4% higher wages.
- **Bias**: ability is omitted. More able people get more education and earn more.
- Sign of ability bias: $(+) \times (+) =$ positive $\Rightarrow$ OLS is biased **upward**.
- But: measurement error in education biases **downward** (attenuation bias).
- Net direction is ambiguous!

Step 2: The First Stage

Question

Run the first stage:

```
reg educ nearc4 age married smsa black south76, robust
```

1. What is the coefficient on `nearc4`?
2. Is the instrument strong? What is the F-statistic?
3. Interpret: what does it mean to live near a 4-year college?

Answer: Step 2

- Coefficient on `nearc4` ≈ 0.32: living near a 4-year college is associated with about 0.32 more years of education.

Answer: Step 2

- Coefficient on `nearc4` ≈ 0.32: living near a 4-year college is associated with about 0.32 more years of education.
- F-statistic ≈ 15: above the rule-of-thumb threshold of 10. Instrument is reasonably strong.

Answer: Step 2

- Coefficient on `nearc4` ≈ 0.32: living near a 4-year college is associated with about 0.32 more years of education.
- F-statistic ≈ 15: above the rule-of-thumb threshold of 10. Instrument is reasonably strong.
- Interpretation: geographic proximity to college lowers the cost of attendance, making some people attend who otherwise wouldn't.

Step 3: Reduced Form

Question

Run the reduced form:

```
reg lwage nearc4 age married smsa black south76, robust
```

1. What is the coefficient on `nearc4`?
2. Can you compute the Wald estimate manually?
   Hint: $\hat{\lambda}_{\text{IV}} = \hat{\rho}_1 / \hat{\phi}_1$

Answer: Step 3

- Reduced form coefficient on `nearc4` ≈ 0.04: living near a college is associated with about 4% higher wages.

Answer: Step 3

- Reduced form coefficient on `nearc4` ≈ 0.04: living near a college is associated with about 4% higher wages.
- Wald estimate: $\frac{0.04}{0.32} \approx 0.125$

Answer: Step 3

- Reduced form coefficient on `nearc4` ≈ 0.04: living near a college is associated with about 4% higher wages.
- Wald estimate: $\frac{0.04}{0.32} \approx 0.125$
- This means: one additional year of education (induced by college proximity) increases wages by about 12.5%.

Answer: Step 3

- Reduced form coefficient on `nearc4` ≈ 0.04: living near a college is associated with about 4% higher wages.
- Wald estimate: $\frac{0.04}{0.32} \approx 0.125$
- This means: one additional year of education (induced by college proximity) increases wages by about 12.5%.
- Compare to OLS (7.4%). The IV estimate is larger!

Step 4: Run 2SLS

Question

Run the full IV regression:

```
ivregress 2sls lwage age married smsa black south76 (educ =
nearc4), first robust
```

1. Does the 2SLS estimate match your Wald calculation?
2. Compare the IV standard error to the OLS standard error. Which is larger?

Answer: Step 4

- 2SLS estimate ≈ 0.125—same as the manual Wald estimate!

Answer: Step 4

- 2SLS estimate ≈ 0.125—same as the manual Wald estimate!
- IV standard errors are **much larger** than OLS standard errors.

Answer: Step 4

- 2SLS estimate ≈ 0.125—same as the manual Wald estimate!
- IV standard errors are **much larger** than OLS standard errors.
- This is normal: IV uses less variation (only the part from Z), so estimates are less precise.

Key Insight

IV trades **precision** for **consistency** (less bias). Less biased estimate, but wider confidence intervals—the fundamental IV trade-off.

Step 5: Discussion

Question

1. Why is the IV estimate larger than OLS? (Think about who the compliers are.)
2. Is the exclusion restriction plausible? What might violate it?
3. For a policymaker, which estimate would you emphasize? Why?

Answer: Discussion

- **Compliers**: People who get more education because a college was nearby—on the margin of attendance. Returns may be especially high for them.

Answer: Discussion

- **Compliers**: People who get more education because a college was nearby—on the margin of attendance. Returns may be especially high for them.
- **Exclusion**: Proximity could affect wages through neighborhood quality, peer effects, or local labor markets—not just education. Controls help.

Answer: Discussion

- **Compliers**: People who get more education because a college was nearby—on the margin of attendance. Returns may be especially high for them.
- **Exclusion**: Proximity could affect wages through neighborhood quality, peer effects, or local labor markets—not just education. Controls help.
- **For a policymaker**: If the policy is "build more colleges," the IV/LATE estimate is exactly right—it captures the effect for those who respond to proximity.

Thank You

Questions?

sebastian.tellotrillo@virginia.edu

Monotonicity in RCTs

- In RCTs, monotonicity is almost always satisfied:

Monotonicity in RCTs

- In RCTs, monotonicity is almost always satisfied:
  - Being assigned to treatment can only encourage taking the treatment.
  - Nobody takes the treatment because they were assigned to control.

Monotonicity in RCTs

- In RCTs, monotonicity is almost always satisfied:
  - Being assigned to treatment can only encourage taking the treatment.
  - Nobody takes the treatment because they were assigned to control.
- RCTs make the **perfect** IV setup:
  - Z = treatment assignment (random by design $\Rightarrow$ exclusion holds)
  - D = actual treatment received
  - Relevance: assignment is correlated with take-up
  - Monotonicity: assignment only pushes toward treatment

◂ Back to Slide 8

Appendix: Other Examples of Instruments

Beer tax $\rightarrow$ Drinking $\rightarrow$ Grades

- Z: State beer tax
- D: Alcohol consumption
- Y: Academic grades

College proximity $\rightarrow$ Educ. $\rightarrow$ Wages

- Z: Grew up near a 4-year college
- D: Years of education
- Y: Log wages

Rainfall $\rightarrow$ Civil war $\rightarrow$ Growth

- Z: Rainfall shocks
- D: Civil conflict
- Y: Economic growth

Draft lottery $\rightarrow$ Military $\rightarrow$ Earnings

- Z: Vietnam draft lottery number
- D: Military service
- Y: Lifetime earnings

◂ Back to Slide 8

Appendix: Derivation of the IV Bias Formula

◂ Back to Slide

Appendix: IV — Potential Outcomes Setup

Define: $Y_i(d)$ = potential outcome if treatment $= d$; $D_i(z)$ = treatment if instrument $= z$.

What we observe (the reduced form / ITT):

$$\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]$$

What we want (LATE — causal effect for compliers):

$$\mathbb{E}\big[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)\big]$$

Under the **exclusion restriction** $Y_i(z, d) = Y_i(d)$, the ITT rewrites as:

$$\mathbb{E}\big[Y_i(D_i(1)) \mid Z_i{=}1\big] - \mathbb{E}\big[Y_i(D_i(0)) \mid Z_i{=}0\big]$$

▸ Decomposition →

Let's Do Some Magic

$$\mathbb{E}\left[Y_i(D_i(1)) \mid Z_i{=}1\right] - \mathbb{E}\left[Y_i(D_i(0)) \mid Z_i{=}0\right]$$

Let’s Do Some Magic

$$\mathbb{E}\left[Y_i(D_i(1)) \mid Z_i{=}1\right] - \mathbb{E}\left[Y_i(D_i(0)) \mid Z_i{=}0\right]$$

$$\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] + \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]$$

Let's Do Some Magic

$$\mathbb{E}\left[Y_i(D_i(1)) \mid Z_i{=}1\right] - \mathbb{E}\left[Y_i(D_i(0)) \mid Z_i{=}0\right]$$

$$\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] + \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]$$

$$\underbrace{\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1]}_{\text{Causal chain}} + \underbrace{\mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]}_{\text{Bias}}$$

Let's Do Some Magic

$$\mathbb{E}\left[Y_i(D_i(1)) \mid Z_i{=}1\right] - \mathbb{E}\left[Y_i(D_i(0)) \mid Z_i{=}0\right]$$

$$\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] + \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]$$

$$\underbrace{\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1]}_{\text{Causal chain}} + \underbrace{\mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]}_{\text{Bias}}$$

Under independence $\left(Z_i \perp (Y_i(\cdot), D_i(\cdot))\right)$:

$$\mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] = \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0] \quad \Rightarrow \text{Bias} = 0$$

Let's Do Some Magic

$$\mathbb{E}\left[Y_i(D_i(1)) \mid Z_i{=}1\right] - \mathbb{E}\left[Y_i(D_i(0)) \mid Z_i{=}0\right]$$

$$\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] + \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]$$

$$\underbrace{\mathbb{E}[Y_i(D_i(1)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1]}_{\text{Causal chain}} + \underbrace{\mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] - \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0]}_{\text{Bias}}$$

Under independence $\left(Z_i \perp (Y_i(\cdot), D_i(\cdot))\right)$:

$\mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}1] = \mathbb{E}[Y_i(D_i(0)) \mid Z_i{=}0] \Rightarrow$ Bias $= 0$

Divide by first stage & apply monotonicity $\Rightarrow$ ITT / First Stage $=$ **LATE**

$$\frac{\mathbb{E}[Y_i \mid Z_i{=}1] - \mathbb{E}[Y_i \mid Z_i{=}0]}{\mathbb{E}[D_i \mid Z_i{=}1] - \mathbb{E}[D_i \mid Z_i{=}0]} = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)\right]$$

← Setup

Back to Slide

Appendix: What Is the First-Stage F-Statistic?

▸ Worked Example ⟶

◂ Back to Relevance

Appendix: F-Statistic — Worked Example

The Stata output from our twins first stage:

```
. ivregress 2sls yrsschooling (famsize = twins), first

First-stage regressions
-----------------------

                                                Number of obs     =         50
                                                F(   1,      48)  =      40.69
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4588
                                                Adj R-squared     =     0.4475
                                                Root MSE          = 11722.1199

------------------------------------------------------------------------------
     famsize |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       twins |   4.081283   .6398354     6.38   0.000     2.794808    5.367758
       _cons |  -31100.69   12586.39    -2.47   0.017    -56407.32    -5794.06
------------------------------------------------------------------------------
```

From the output: $n = 50$, $R^2 = 0.459$, $t_{\texttt{twins}} = 6.38$

Method 1 — via R^2

$$F = \frac{R^2/q}{(1 - R^2)/(n - k - 1)}$$

$$= \frac{0.4588/1}{0.5412/48} \approx \mathbf{40.69}$$

Method 2 — via t-stat (one instrument only)

$$F = t^2 = 6.38^2 = \mathbf{40.69} \checkmark$$

Strong instrument. *In plain English: the twins signal is $\sim$40$\times$ louder than noise.*

← F-stat definition

With covariates →

Back to Relevance

Appendix: F-Statistic with Covariates

ivregress 2sls yrsschooling pcturban (famsize = twins), first

```
.  ivregress 2sls yrsschooling pcturban (famsize = twins), first

First-stage regressions
-----------------------

                                                Number of obs     =         50
                                                F(    2,     47)  =      23.76
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5027
                                                Adj R-squared     =     0.4816
                                                Root MSE          = 11354.8710

------------------------------------------------------------------------------
     famsize |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pcturban |   281.4376   138.0671     2.04   0.047     3.682484    559.1927
       twins |   3.184166   .7601527     4.19   0.000     1.654936    4.713396
       _cons |  -32448.99   12209.99    -2.66   0.011    -57012.32   -7885.653
------------------------------------------------------------------------------
```

Watch out: `F(2,47)=23.76` tests all regressors jointly. **Not** the IV relevance F.

We need the partial F for the instrument alone, partialling out covariates.

One instrument: partial $F = t^2_{\texttt{twins}} = 4.19^2 \approx \mathbf{17.56}$

Stata command (after `ivregress`):

`estat firststage`

Reports partial F for excluded instruments only.

Still strong. *In plain English: even with controls, the twins signal is $\sim 17\times$ louder than noise.*

← No-covariate example

Back to Relevance