Causal Inference: Instrumetal Variables

POLI 803 Research Methods in PS

Howard Liu

2025

Roadmap

- Instrumental variables
  - Background
  - Intuition
- Estimators
  - Two Step
  - Weak instruments

The basic idea

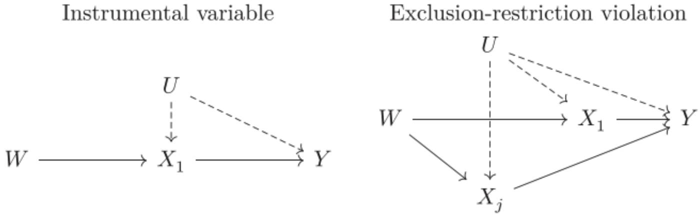


Figure: DAGs illustrating instrumental variables (IVs) and endogeneity

A popular example: rainfall

- Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression (Ritter and Conrad, 2016)
- "Although **rainfall** can dissuade dissenters from challenges, we argue that rainfall is exogenous to repression. Repression is carried out under authoritative orders to contain threat, regardless of the weather, and **is only related to rainfall through rain's effect on dissent.**"

A popular example: rainfall

- Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression (Ritter and Conrad, 2016)
- “Although **rainfall** can dissuade dissenters from challenges, we argue that rainfall is exogenous to repression. Repression is carried out under authoritative orders to contain threat, regardless of the weather, and **is only related to rainfall through rain’s effect on dissent.**”
- Do you know any popular IV in your field?

Rainfall: violation of exclusion restriction

- "Rain, rain, go away: 194 potential exclusion-restriction violations for studies using weather as an instrumental variable' (Mellon, APSR 2025)
  $\rightarrow$ weather can have all kinds of indirect effects on the outcome (e.g., civil wars, state repression)
- "Rainfall shocks and state repression: How rainfall shocks incentivize governments to commit human rights abuses" (Appel and Smith, CPMS 2024) $\rightarrow$ reduced food production

Rainfall: violation of exclusion restriction

- IV can be a powerful design but it also needs careful justification and a whole lot of rebustness checks

IV from my research

- The effect of infrastructural damages on repression using **earthquake shocks** as an instrument (Liu and Sullivan, JCR 2021)
- The effect of protest on police repression using **weekends** as an instrument (Liu and Radford, ISQ 2025)

Uses of instrumental variables

- When you have unobserved confounders, IV may work (most common)
- When you have reverse causality, IV could help (most common)
- When you have the treatment and outcome variables being determined simultaneously (like in markets with supply and demand), then IV can step in (most common)
- When you run an experiment (RCT, AB tests, etc.) but not everyone obeys their treatment assignment, IV will help you
- When your treatment variable is measured with error, IV can help

Key Learning Goal Today

1. Focus on **intuition** – what is an instrument and what do you do with it?
2. Show how the estimation is done
3. Emphasize key identifying assumptions
4. Teach you how to **justify your identification strategy**

Instruments aren't labeled

- They don't just fall from the sky
- No data codebook saying "instrumental variable". So how do you find it?

Instruments aren't labeled

- They don't just fall from the sky
- No data codebook saying "instrumental variable". So how do you find it?
- **Need a strong substantive knowledge about the context (e.g., earthquakes in Guatemala, women conscription lottery in Denmark)**

Some questions you need to be asking

1. Is our instrument highly correlated with the treatment? With the outcome? Can we see evidence for this?
2. Are there random reasons why our treatment changes? Why do you think that?
3. Is the instrument independent of confounders? Why do you think that?
4. Could the instrument affect outcomes directly? Why do you think that?

Roadmap

Instrumental variables

- Background
- Intuition

Estimators

- Two Step
- Weak instruments

Two step vs Minimum Distance

- The two-stage least squares (2SLS) estimator was developed by Theil (1953) and Basman (1957) independently
- Kolesaŕ has a helpful distinction: two step (Wald, 2 Sample IV, JIVE, UJIVE, 2SLS) vs minimum distance estimators (LIML)
- Too much to review as IV is a *huge* area, so I will focus on a few things, starting with two stage least squares (2SLS)
- **2SLS is basically the workhorse IV model**, though it can have some issues because of its finite sample bias with weak instruments

Two-stage least squares concepts

- Causal model.: Your main research question:

$$Y_i = \alpha + \delta S_i + \eta_i$$

- **First-stage regression**: Gets the name because of two-stage least squares:

$$S_i = \gamma + \rho Z_i + \zeta_i$$

- **Second-stage regression**: Notice the fitted values, $\widehat{S}$:

$$Y_i = \beta + \delta \widehat{S}_i + \nu_i$$

- **Reduced form regression**: Y regressed onto the instrument:

$$Y_i = \psi + \pi Z_i + \varepsilon_i$$

Intuition of 2SLS

- Intuition is that 2SLS replaces S with the fitted values $\widehat{S}$ from the first stage regression of S onto Z and all other covariates
- By using the fitted values of the endogenous regressor from the first stage regression, our regression now uses *only* the quasi-random variation in the treatment due to the instrumental variable itself (only the random parts of schooling remain)
- Our regression now uses *only* the exogenous variation in the regressor due to the instrumental variable itself

Weak instruments

"In instrumental variables regression, the instruments are called weak if their correlation with the endogenous regressors, conditional on any controls, is close to zero." – Andrews, Stock and Sun (2018)

Weak instruments

- Weak instruments can happen if the two variables are independent or the sample is small
- If you have a weak instrument, then the bias of 2SLS is centered on the bias of OLS and the cure ends up being worse than the disease

Compulsory schooling example

- In the US, you could drop out of school once you turned 16
- "School districts typically require a student to have turned age six by January 1 of the year in which he or she enters school" (Angrist and Krueger 1991, p. 980)
- Children have different ages when they start school, though, and this creates different lengths of schooling at the time they turn 16 (potential drop out age):

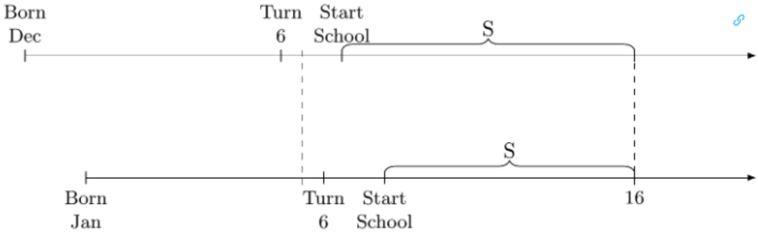


Figure 7.4: Compulsory schooling start dates by birthdates.

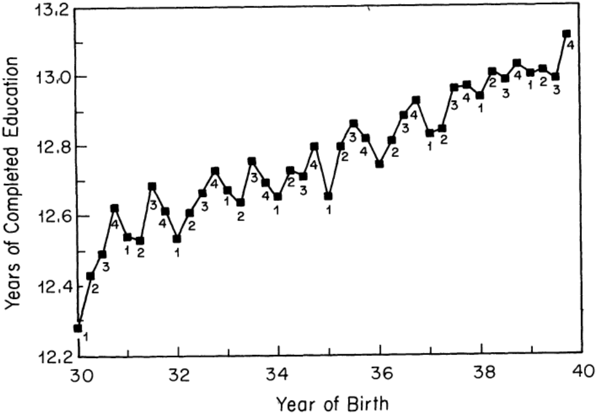
If you’re born in the fourth quarter, you hit 16 with more schooling than those born in the first quarter (because you need to wait for almost another year to enter school)

Visuals

- You need good data visualization for IV partly because of the scrutiny around the design
- The two pieces you should be ready to build pictures for are the first stage and the reduced form
- Angrist and Krueger (1991) provide simple, classic and compelling pictures of both

First Stage

Children born earlier in the year, like Jan 1st, have lower schooling. This indicates that there is a first stage. Notice all the 3s and 4s at the top. But then notice how it attenuates over time …



Those 3s and 4s on average have more years of education than those

Two Stage Least Squares model

- The causal model is

$$Y_i = X\pi + \delta S_i + \varepsilon$$

- The first stage regression is:

$$S_i = X\pi_{10} + \pi_{11} Z_i + \eta_{1i}$$

- The reduced form regression is:

$$Y_i = X\pi_{20} + \pi_{21} Z_i + \eta_{2i}$$

- The sample analog of the Wald estimator that adjusts for covariates: $\frac{\pi_{21}}{\pi_{11}}$

Two Stage Least Squares

- Angrist and Krueger instrument for schooling using three quarter of birth dummies: a dummies for 1st, 2nd and 3rd QoB
- Their initial first-stage regression is:

$$S_i = X\pi_{10} + Z_{1i}\pi_{11} + Z_{2i}\pi_{12} + Z_{3i}\pi_{13} + \eta_1$$

- The second stage is the same as before (including all controls X), but the fitted values are from the new first stage

$$Y_i = X\pi + \delta\widehat{S}_i + \epsilon$$

First stage regression results

Quarter of birth is a strong predictor of total years of education

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect[a] I	II	III	F-test[b] [P-value]
Total years of education	1930–1939	12.79	−0.124	−0.086	−0.015	24.9
			(0.017)	(0.017)	(0.016)	[0.0001]
	1940–1949	13.56	−0.085	−0.035	−0.017	18.6
			(0.012)	(0.012)	(0.011)	[0.0001]
High school graduate	1930–1939	0.77	−0.019	−0.020	−0.004	46.4
			(0.002)	(0.002)	(0.002)	[0.0001]
	1940–1949	0.86	−0.015	−0.012	−0.002	54.4
			(0.001)	(0.001)	(0.001)	[0.0001]
Years of educ. for high school graduates	1930–1939	13.99	−0.004	0.051	0.012	5.9
			(0.014)	(0.014)	(0.014)	[0.0006]
	1940–1949	14.28	0.005	0.043	−0.003	7.8
			(0.011)	(0.011)	(0.010)	[0.0017]
College graduate	1930–1939	0.24	−0.005	0.003	0.002	5.0
			(0.002)	(0.002)	(0.002)	[0.0021]
	1940–1949	0.30	−0.003	0.004	0.000	5.0
			(0.002)	(0.002)	(0.002)	[0.0018]

IV Estimates Birth Cohorts 20-29, 1980 Census

Independent variable	(1) OLS	(2) TSLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)
Race (1 = black)	—	—
SMSA (1 = center city)	—	—
Married (1 = married)	—	—
9 Year-of-birth dummies	Yes	Yes
8 Region-of-residence dummies	No	No
Age	—	—
Age-squared	—	—
χ^2 [dof]	—	25.4 [29]

QoB $\rightarrow$ years of education $\rightarrow$ earnings

180 instruments

- To improve ”precision” in their two stage least squares model, they include more instruments (**causes 40 percent reduction in standard errors in 2SLS**)
- More instruments can increase variation in the predicted schooling variable, lowering standard errors and tightening confidence intervals
- Three QoB dummies interacted with 50 state-of-birth dummies (3 x 50) plus 3 QoB dummies interacted with 9 year-of-birth dummies (3 x 9)
- Includes 50 state-of-birth dummies so variability in education in 2SLS is solely due to differences in seasons of birth and this is allowed to vary by state and birth year for the first time

More instruments

TABLE VII
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673	0.0928	0.0673	0.0907	0.0628	0.0831	0.0628	0.0811
	(0.0003)	(0.0093)	(0.0003)	(0.0107)	(0.0003)	(0.0095)	(0.0003)	(0.0109)
Race (1 = black)	—	—	—	—	−0.2547	−0.2333	−0.2547	−0.2354
					(0.0043)	(0.0109)	(0.0043)	(0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705	0.1511	0.1705	0.1531
					(0.0029)	(0.0095)	(0.0029)	(0.0107)
Married (1 = married)	—	—	—	—	0.2487	0.2435	0.2487	0.2441
					(0.0032)	(0.0040)	(0.0032)	(0.0042)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	−0.0757	−0.0880	—	—	−0.0778	−0.0876
			(0.0617)	(0.0624)			(0.0603)	(0.0609)
Age-squared	—	—	0.0008	0.0009	—	—	0.0008	0.0009
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	163 [179]	—	161 [177]	—	164 [179]	—	162 [177]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 150 quarter-of-birth times state-of-birth interactions. Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is the same as in Table VI. Sample size is 329,509.

Weak Instruments

- Important paper suggesting OLS and 2SLS were pretty similar, plus introduces modern notion of seeking "plausibly exogenous instruments"
- **But** in the early 1990s, a number of papers showed that IV can be *severely* biased with weak instruments and many instruments for one endogenous variable
- In the worst case, if the instruments are so weak that **there is no first stage**, then the 2SLS sampling distribution is centered on the probability limit of OLS

Weak Instruments - Adding More Instruments

- Adding more weak instruments will increase the bias of 2SLS
  - $\rightarrow$ By adding further instruments without predictive power, the first stage F**-statistic** goes toward zero and the bias increases
  - $\rightarrow$ F**-statistic** needs to be larger than 10 (if with one IV)
- If the model is "just identified" – mean the same number of instrumental variables as there are endogenous covariates – weak instrument bias is less of a problem
- Exclusion restriction: as long as you can justify that all your IVs are exogenous

Bound, Jaeger and Baker (1995)

Remember, AK present findings from expanding their instruments to include many interactions (i.e., saturated model)

1. Quarter of birth dummies $\rightarrow$ 3 instruments
2. Quarter of birth dummies $+$ (quarter of birth) $\times$ (year of birth) $+$ (quarter of birth) $\times$ (state of birth) $\rightarrow$ 180 instruments

So if **any** of these are weak, then the approximate bias of 2SLS gets worse

Adding instruments in Angrist and Krueger

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063	.142	.063	.081
	(.000)	(.033)	(.000)	(.016)
F (excluded instruments)		13.486		4.747
Partial R^2 (excluded instruments, ×100)		.012		.043
F (overidentification)		.932		.775
	Age Control Variables			
Age, Age^2	x	x		
9 Year of birth dummies			x	x
	Excluded Instruments			
Quarter of birth		x		x
Quarter of birth × year of birth				x
Number of excluded instruments		3		30

Adding more weak instruments reduced the first stage F-statistic and increases the bias of 2SLS. Notice its also moved closer to OLS.

Adding instruments in Angrist and Krueger

	(1) OLS	(2) IV
Coefficient	.063	.083
	(.000)	(.009)
F (excluded instruments)		2.428
Partial R^2 (excluded instruments, ×100)		.133
F (overidentification)		.919
Age Control Variables		
Age, Age^2		
9 Year of birth dummies	x	x
Excluded Instruments		
Quarter of birth		x
Quarter of birth × year of birth		x
Quarter of birth × state of birth		x
Number of excluded instruments		180

More instruments increase precision, but drive down F, therefore we know the problem has gotten worse

IV advice: Weak instruments

"In the leading case with a single endogenous regressor, we recommend that researchers judge instrument strength based on the effective F-statistic of Montiel Olea and Pflueger (2013). If there is a single instrument, we recommend reporting identification robust Anderson-Rubin confidence intervals. These are effective regardless of the strength of the instruments, and so should be reported regardless of the value of the first stage F. Finally, if there are multiple instruments, the literature has not yet converged on a single procedure, but we recommend choosing from among the several available robust procedures that are efficient when the instruments are strong." – Andrews, Stock and Sun (2018)