17.802 – Recitation 8

Instrumental Variables

Yiqing Xu

MIT

April 18, 2014

2SLS Lingo

- Y - dependent variables: the outcome variables
- S - variables of interest: the endogenous variables to be instrumented
- Z - instruments: exogenous
- X - covariates: exogenous

- In any IV analysis, variables are either dependent, (other) endogenous variables, instruments, or covariates

The IV Calculator

$$\begin{aligned} Y_i &= \alpha + \rho S_i + A_i'\gamma + v_i \\ &= \alpha + \rho S_i + \eta_i \end{aligned}$$

- Now we have an instrument Z_i for S_i, and $Cov(\eta_i, Z_i) = 0$
- Exclusion restriction is implicitly assumed (no Z_i in the formula)
- IV estimator

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(S_i, Z_i)}$$

Many Faces of 2SLS

$$Y_i = \alpha + \rho S_i + X_i'\pi + (A_i'\gamma + v_i)$$

- 2SLS estimator
  1. Regress S_i on Z_i and X_i; obtain $\hat{S}_i$
  2. Regress Y_i on $\hat{S}_i$ and X_i, obtain ρ
- Indirect Least Squares (ILS) estimator

$$\begin{aligned} Y_i &= X_i'\pi_{10} + \pi_{11}Z_i + \zeta_{1i} \\ S_i &= X_i'\pi_{20} + \pi_{21}Z_i + \zeta_{2i} \\ \hat{\rho} &= \hat{\pi}_{21}/\hat{\pi}_{21} \end{aligned}$$

- IV estimator ($\hat{S}_i^*$ as the instrument)

$$\rho = \frac{Cov(Y_i, \hat{S}_i^*)}{Cov(S_i, \hat{S}_i^*)} = \frac{Cov(Y_i, \tilde{Z}_i)}{Cov(S_i, \tilde{Z}_i)} = \pi_{21}/\pi_{21}$$

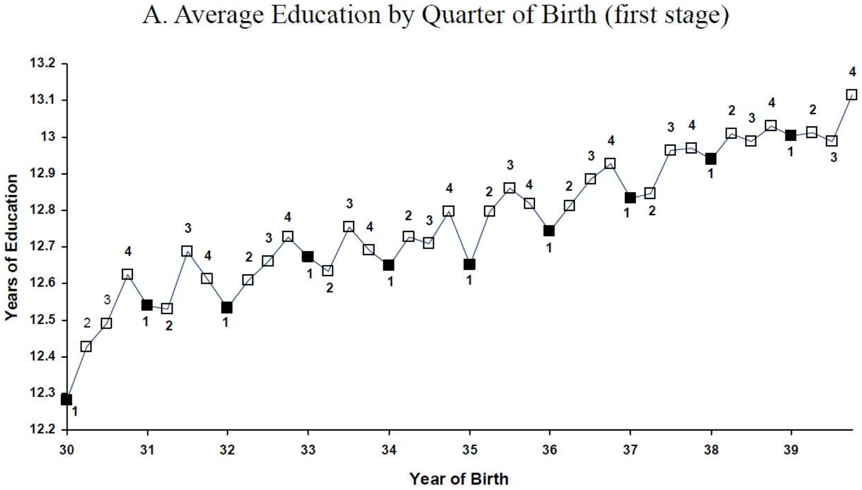
$\hat{S}_i^*$ residual from Reg: $\hat{S}_i$ on X_i; $\tilde{Z}_i$ residual from Reg: Z_i on X_i

Quarter of Birth, Schooling, and Earnings

Why one's birthday is connected with his/her schooling, and future earnings?

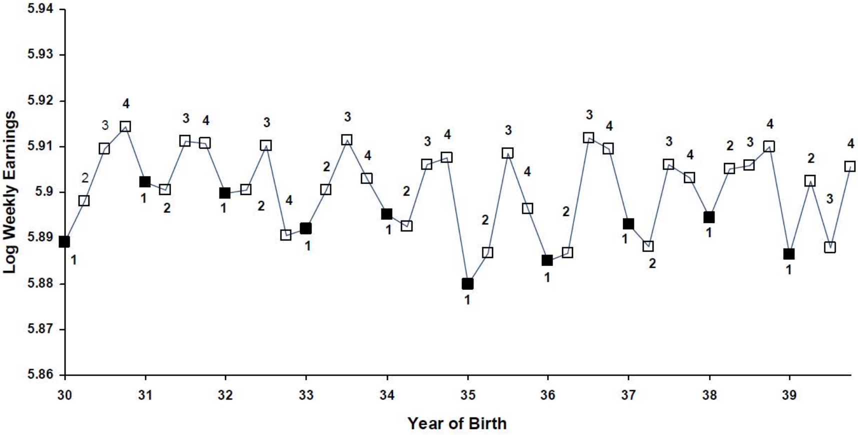
- Age at school entry: kids born early in a calender year are typically older when entering school
- Compulsory schooling laws

Angrist and Krueger (1991)



Angrist and Krueger (1991)

B. Average Weekly Wage by Quarter of Birth (reduced form)



Angrist and Krueger (1991)

Table 4.1.1
2SLS estimates of the economic returns to schooling

	OLS		2SLS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of education	.071	.067	.102	.13	.104	.108	.087	.057
	(.0004)	(.0004)	(.024)	(.020)	(.026)	(.020)	(.016)	(.029)
Exogenous Covariates								
Age (in quarters)								✓
Age (in quarters) squared								✓
9 year-of-birth dummies		✓			✓	✓	✓	✓
50 state-of-birth dummies		✓			✓	✓	✓	✓
Instruments								
dummy for QOB = 1			✓	✓	✓	✓	✓	✓
dummy for QOB = 2				✓		✓	✓	✓
dummy for QOB = 3				✓		✓	✓	✓
QOB dummies interacted with year-of-birth dummies (30 instruments total)							✓	✓

Notes: The table reports OLS and 2SLS estimates of the returns to schooling using the Angrist and Krueger (1991) 1980 census sample. This sample includes native-born men, born 1930–39, with positive earnings and nonallocated values for key variables. The sample size is 329,509. Robust standard errors are reported in parentheses. QOB denotes quarter of birth.

Other Stuff You Should Know

- Bias of 2SLS
  - 2SLS estimates are consistent for causal effects but biased towards OLS estimates
  - The bias comes from weak instruments. The bigger the first-stage F, the small the bias
  - You don't have much of the problem in the just-identified case
  - The reduced form is unbiased. Always check it first.
  - Limited maximum likelihood (LIML) is median-unbiased
- Forbidden regressions
  - Non-linear first stage

$$Y = X'\pi + \rho\hat{S} + \zeta + \rho(S - \hat{S})$$

  - Non-linear second stage

The LATE Framework

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \alpha_i + \rho_i D_i + \eta_i$$

- The Wald estimator

$$\rho = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

- Assumptions
  - Independence
    (Sufficient for identifying the causal effects of the instrument)
  - Exclusion
    (No direct effect; should be more controversial)
  - Monotonicity
    (Not always true)

The Story of Draft Lottery

Lottery number (RSN) $\rightarrow$ Veteran status $\rightarrow$ Earnings

- Angrist and Krueger (1990)
- Angrist and Krueger (1999)
- Angrist and Chen (2007)
- Angrist, Chen, and Song (2011)

Angrist and Krueger (1999)

IV Estimates of the Effects of Military Service on the Earnings of White Men born in 1950

	Earnings		Veteran Status		Wald Estimate of Veteran Effect
Earnings year	Mean	Eligibility Effect	Mean	Eligibility Effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8 (210.5)	.267	.159 (.040)	-2,741 (1,324)
1971	3,338	-325.9 (46.6)			-2050 (293)
1969	2,299	-2.0 (34.5)			

Note: Adapted from Table 5 in Angrist and Krueger (1999) and author tabulations. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.

Angrist and Krueger (1990)

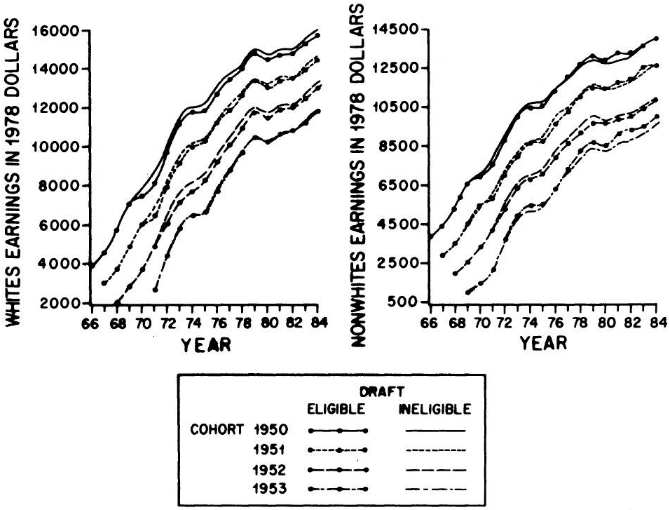


FIGURE 1. SOCIAL SECURITY EARNINGS PROFILES BY DRAFT-ELIGIBILITY STATUS

Angrist and Krueger (1990)

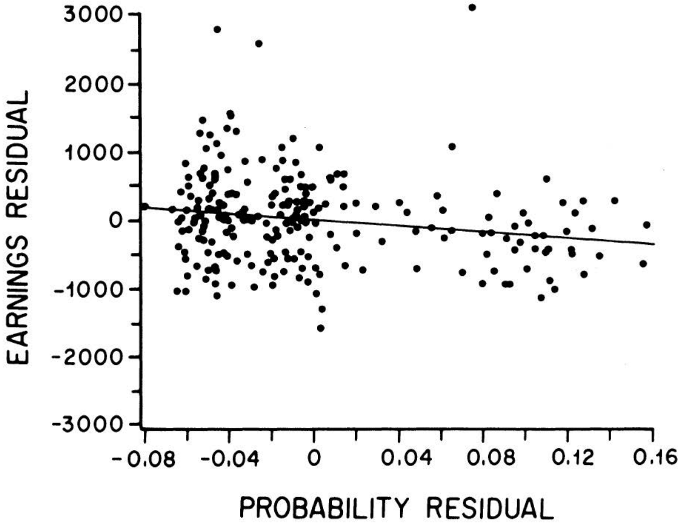
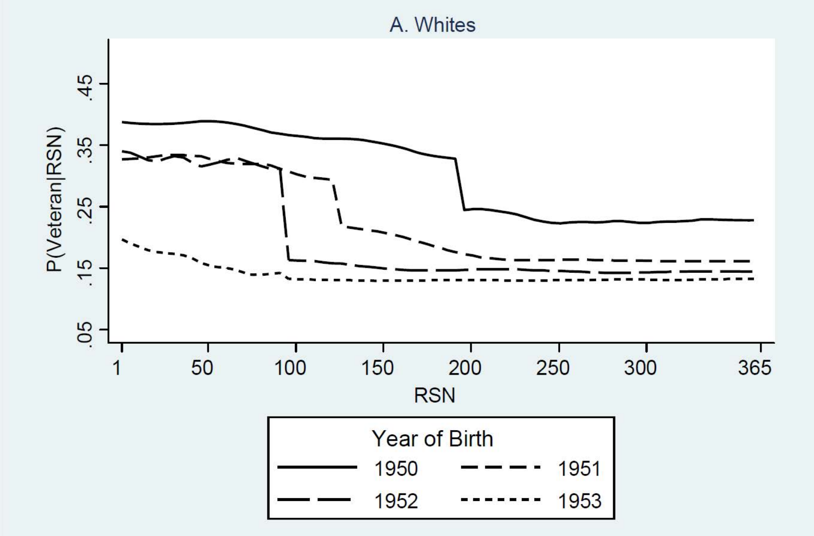


FIGURE 3. EARNINGS AND THE PROBABILITY OF VETERAN STATUS BY LOTTERY NUMBER

Angrist and Chen (2007)



Angrist, Chen, and Song (2011)

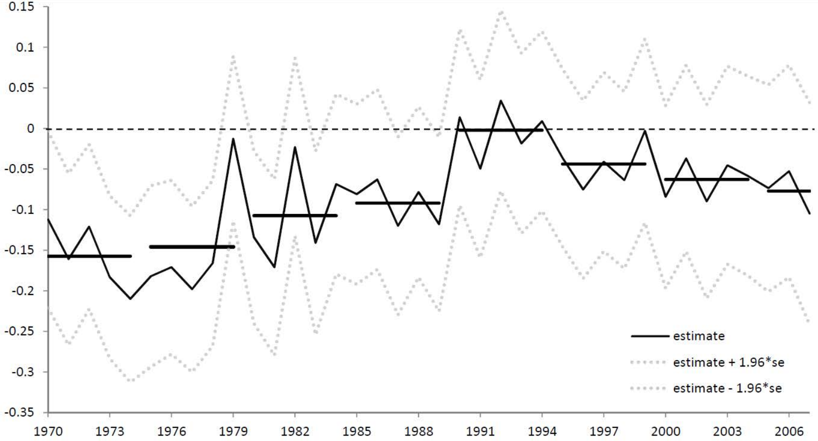


Figure 1. Draft-lottery Estimates of Vietnam-era Service Effects on ln(Earnings) for White Men Born 1950-52

Other Issues

- You cannot identify compliers, but you can count them (first stage) and characterize them (LARF)
- Two-sample IV is feasible, as long as the two samples are representative of the same population
- External validity comes from many experiments (instruments) or the same experiment in many occasions/places

In Summary

- IVs are powerful (w/o RCT, w/ non-complier)
- But they are to find
  - more and more from experiments
  - unanticipated shocks (e.g., natural disaster)
  - binding institutional rules
- The most demanding assumption is exclusion restriction
- Transparency is always the most important (show plots)
- The reduced form gives you a rough idea of whether it's gonna work (The S.E.s are pretty much the same)
- External validity is an issue, but let's worry about it later (counting and characterizing compliers help)