Note: IV estimation

The Inconsistency of the LS Estimator

- When *Cov(x,ε)* is nonzero, the LS estimate of β_2 would be inconsistent and the bias is approximately given in a simple linear model

$$\mathrm{Cov}(x, \varepsilon) / \mathrm{Var}(x) = \mathrm{Corr}(x, \varepsilon)\, \sigma_\varepsilon / \sigma_x$$

The Inconsistency of the LS Estimator

- $y = \beta_1 + \beta_2 x + e. \Rightarrow y - E(y) = \beta_2[x - E(x)] + e \quad (\leftarrow E(y) = \beta_1 + \beta_2 E(x))$

  $\Rightarrow [x - E(x)][y - E(y)] = \beta_2 [x - E(x)]^2 + [x - E(x)]e$

  $\Rightarrow E[x - E(x)][y - E(y)] = \beta_2 E[x - E(x)]^2 + E\{[x - E(x)]e\}$

  $\Rightarrow \mathrm{cov}(x,y) = \beta_2 \mathrm{var}(x) + \mathrm{cov}(x,e) \Rightarrow \beta_2 = \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} - \dfrac{\mathrm{cov}(x,e)}{\mathrm{var}(x)}$

- $b_2 = \dfrac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} = \dfrac{\sum(x_t - \bar{x})(y_t - \bar{y})/(T-1)}{\sum(x_t - \bar{x})^2/(T-1)} = \dfrac{\hat{\mathrm{cov}}(x,y)}{\hat{\mathrm{var}}(x)}$

- If $\mathrm{cov}(x,e) = 0$ then $b_2 = \dfrac{\hat{\mathrm{cov}}(x,y)}{\hat{\mathrm{var}}(x)} \to \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} = \beta_2$

- If $\mathrm{cov}(x,e) \neq 0$, then $\beta_2 = \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} - \dfrac{\mathrm{cov}(x,e)}{\mathrm{var}(x)} \Rightarrow b_2 \to \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} = \beta_2 + \dfrac{\mathrm{cov}(x,e)}{\mathrm{var}(x)} \neq \beta_2$

The Problem of Weak Instrument

- The Problem of Weak Instrument
  - *Cov(z,ε)=0* is never guaranteed. If *Cov(z, ε)* is nonzero, then the IV estimate of β_2 would be inconsistent and the bias is approximately

$$\mathrm{Cov}(z, \varepsilon) / \mathrm{Cov}(z,x) = \mathrm{Corr}(z, \varepsilon)\sigma_\varepsilon / [\mathrm{Corr}(z,x)\sigma_x]$$

  - If z is a weak instrument (i.e. Corr(z,x) is small), the 2SLS estimate of β_2 could be even more biased than the OLS estimate.

Variance of IV Estimates

- Using IV results in less "precise" estimate than using OLS (i.e. it is harder to reject null hypothesis, or less likely to draw conclusion from data).

$$\mathrm{Var}(\beta_2^{iv})/\mathrm{Var}(\beta_2^{ols}) = 1/r^2_{x,z} > 1$$

How much is the return to education?

- Return to education
  - What is the value of Schooling?
  - Do people really learn useful things in schools?
  - Do people learn more in college than in high school?
  - Should government spend more money on schools? What level of schools should get more funding?

How hard is it to measure the return to education? 7

- A naïve regression

  $\text{Income}_i = \alpha \cdot \text{Education}_i + \varepsilon_i$
- Problems
  - Measurement Errors (Income, Education)?
  - Underspecification?
  - Functional form misspecification?
  - Endogeneity?

Key Issue 8

- Endogeneity is the key issue
  - Due to the correlation between education and omitted variables that affect income (talents, family environment, …)
  - Due to measurement errors in education
- The problem may not be solved by adding more control variables or using panel data.

The Impact of Years of Education on Earnings (Angrist and Krueger) 9

- A naive model

  $Ln(Earnings) = \beta_0 + \beta_1 (years\ of\ edu) + u$
- IV: Quarter of birth
  - Children born in the fourth quarter (e.g. December 31) enter school at age 5.75. In contrast, those born in the first quarter (e.g. January 1) enter school at age 6.75.
  - Compulsory schooling laws requires students to remain in school until their 16th birthdays.
  - Hence, people born in different quarters have different education length on average. Moreover, the difference is possibly exogenously determined.

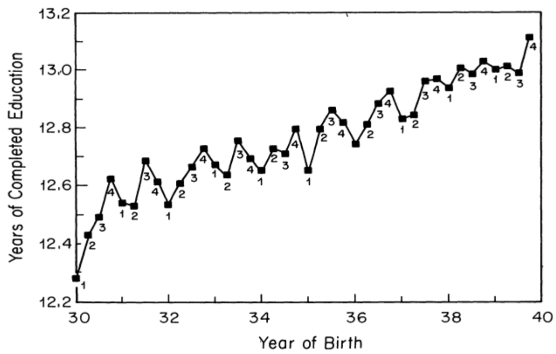
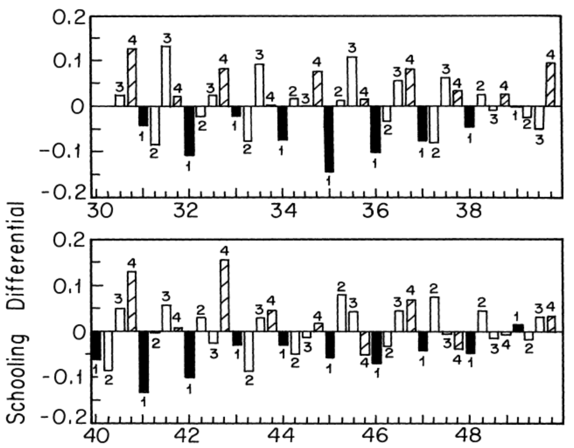


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.



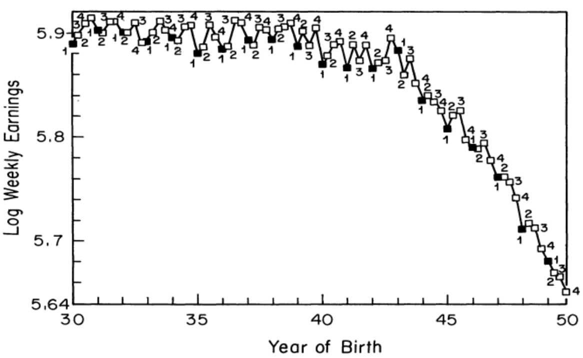


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

Findings

13

- Men born in the first quarter have about one-tenth of a year less schooling than men born in later quarters.
- Men born in the first quarter earn about 0.1 percent less than men born in later quarters

Other examples

14

- Acemoglu et al (2001)
  - A naïve model :
    - Ln(GDP per capita) = a + b(institution) + u
    - Instrument : European mortality rates
- Frankel and Romer (1999)
  - A naïve model "
    - Ln(GDP per capital) = a + b(Trade Volume) + u
    - Instrument : Geographic characteristics

Natural (Quasi-) Experiment

15

- In an ideal experiment we have a *treatment group* and a *control group*. The two groups are the same except that only the treatment group is affected by some factor of interest while the control group is not affected by the factor (both groups can still be affected by other confounding factors).
- A natural experiment is like an experiment. The difference is that in the latter case we have good control on which group can be affected by which factors. In a natural experiment, we can not control the factors, but we try to find some natural settings in which the nature has generated two similar groups, one of which is affected by the factors of interest and the other group is not.

Natural Experiment

16

- NE required the control and treatment groups to be matched not only by observed factors, but also unobserved factors.
  - Whether unobserved factors are "matched" is inferred from theory and other observed information.

Solution with Identical twins

17

- Arguments for using twins data
  - Identical twins should be equally smart; they should look similar; and they should have the same family background.
    - Therefore, if we could find twins with different education, we can compare their income. The difference should reflect only the effect of education but not of omitted variables.
    - Strictly speaking, we assume there is no unobserved inter-twin difference that is correlated with education.
- An Empirical Study(Bonjour et al., 2003)
  - 3,300 same-sex twins in 1999 in U.K.