

Economic Policy Analysis (EC406)

Instrumental Variables Methods

John Van Reenen
Centre for Economic Performance
London School of Economics

* These lecture notes are incomplete. Reading the lecture notes without attending lectures can be very misleading.

Motivation

What is the rate of return to education?

$$\ln W_i = X_i \beta + \rho E_i + \sum_c d_{ic} \xi_c + \mu_i$$

- cross-section data on individuals i born in years c (cohort c)
- W is weekly wage
- E is years of education
- X is vector of covariates (e.g. gender, race,...)
- d_{ic} is $\{0,1\}$ dummy for whether individual i is born in year c

Econometric problem

- The error term μ_i includes individual's unobserved ability
 - Determines weekly wage
 - Correlated with years of education

Economic Questions

- “Bell Curve” – ability/IQ drives (Almost) everything so subsidising schooling is inefficient
- Issue of “Sheepskin”/signalling effects. Could still matter even if we control for ability
- Strategies
 - Measure ability (e.g. Willis and Rosen, 1979)
 - Twin studies (Ashenfelter and Krueger)
 - IV

Instrumental Variables Approach

Use of instrumental variables to overcome omitted variables problems in estimates of causal relationships

Instruments derived from nature, institutions and government policy

- Instruments derived from “Natural Experiments”

Interpretation in terms of program evaluation

- Education is the treatment
- Endogenous participation decision (sample selection)

Search for an instrumental variable

- Correlated with an individual’s participation (education) decision
- Has no independent effect on the outcome of interest (weekly wage) other than through education

Empirical Application : Compulsory Schooling Laws

Does Compulsory School Attendance Affect Schooling and Earnings? (Angrist and Krueger, QJE, 1991)

- Most US school districts do not admit students to first grade unless they will attain age 6 by January 1 of the academic year in which they enter school
 - Students born early in the year are older when they start school than students born later in the year. Therefore (under any compulsory school leaving age) they can leave earlier and get less education
- Compulsory schooling laws
 - Generally require students to remain in school until their 16th or 17th birthday (cross state-time variation)

Empirical Application : Compulsory Schooling Laws

Does Compulsory School Attendance Affect Schooling and Earnings? (Angrist and Krueger, *Quarterly Journal of Economics*, 1991)

- Interaction of school entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months
- Season of birth generates exogenous variation in education that can be used to estimate
 - Impact of compulsory schooling on education (first-stage)
 - Effect of education on earnings (second-stage)
 - Impact of compulsory schooling on earnings (reduced form)

Other Empirical IV Applications

How much of a firm's (monopoly) profits are captured by workers?
(Van Reenen, *Quarterly Journal of Economics*, 1996)

- Under imperfect competition in labour market wages depend on company's profits as well as individual profitability (e.g. oligopoly and union bargaining)
- But a regression of firm average wages on firms average profitability will be biased downwards because a wage shock will lead to lower profits
- Use observed technological innovations as an instrument for profits. Should increase profits, but no direct effect on wage
- Find that existing measures of rent sharing based on OLS ten times too small over a quarter of profits captured by workers in sample.

Other Empirical Applications : IV and Program Evaluation

Effect of war veteran status on earnings (Angrist, AER, 1990)

- Instrument: US draft lotteries during Vietnam War
 - A Random Sequence Number (RSN) from 1-365 was assigned to birth dates in the cohort being drafted
 - Men called for induction by RSN up to a ceiling determined by the Defence Department
 - Only men with lottery numbers below the ceiling “draft eligible”

Effects of police on city crime rates (Levitt, AER, 1997)

- Instrument: timing of mayoral and gubernatorial elections
 - Increases in the size of the police force in large cities’ disproportionately concentrated in election years
 - Identifying assumption: timing of elections affects the size of the police force but does not directly affect crime rates

Structure of Presentation

(1) Instrumental variables estimation

- Discrete instrumental variables and the Wald estimator

(2) Does compulsory schooling affect years of education?

(3) OLS and Wald estimates of the return to education

(4) OLS and 2SLS estimates of the return to education

Instrumental Variables Estimation

Consider the following economic model:

$$y = \beta_0 + \beta_1 x + u$$

$$\text{Cov}(x, u) \neq 0$$

$$x = \gamma_0 + \gamma_1 z + \varepsilon$$

Identifying assumptions:

$$\text{Cov}(z, x) \neq 0$$

$$\text{Cov}(z, u) = 0$$

Instrumental variables estimator:

$$\hat{\beta}_1^{IV} = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}$$

Instrumental Variables Estimation

Suppose that z is a discrete variable that takes the value 1 for N_1 of the N obs in a cross-section and 0 for the remaining N_0 obs

The instrumental variables estimator β_1^{IV} is equivalent to the following Wald estimator

$$\hat{\beta}_1^{Wald} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

- \bar{y}_1 is mean of y across the N_1 obs where $z = 1$
- \bar{y}_0 is the mean of y across the N_0 obs where $z = 0$
- and analogously for x

Proof

Multiplying out the terms in brackets and simplifying

$$\hat{\beta}_1^{IV} = \frac{\sum_{i=1}^n y_i z_i - \bar{y} z_i}{\sum_{i=1}^n x_i z_i - \bar{x} z_i}$$

$$\hat{\beta}_1^{IV} = \frac{\sum_{i \in N_1} y_i z_i - \bar{y} z_i + \sum_{i \in N_0} y_i z_i - \bar{y} z_i}{\sum_{i \in N_1} x_i z_i - \bar{x} z_i + \sum_{i \in N_0} x_i z_i - \bar{x} z_i}$$

Noting that $z_i = 1$ for $i \in N_1$ and $z_i = 0$ for $i \in N_0$, and multiplying the numerator and denominator by $1/N_1$

Proof

$$\hat{\beta}_1^{IV} = \frac{\left(1/N_1 \sum_{i \in N_1} y_i\right) - \bar{y}}{\left(1/N_1 \sum_{i \in N_1} x_i\right) - \bar{x}} = \frac{\bar{y}_1 - \bar{y}}{\bar{x}_1 - \bar{x}}$$

Noting that $\bar{y} = (N_1/N)\bar{y}_1 + (N_0/N)\bar{y}_0$ and analogously for x , we obtain

$$\hat{\beta}_1^{IV} = \frac{(N_0/N)(\bar{y}_1 - \bar{y}_0)}{(N_0/N)(\bar{x}_1 - \bar{x}_0)} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0} = \hat{\beta}_1^{Wald}$$

Does Compulsory Schooling Affect Years of Education?

In virtually all birth cohorts, children born in first quarter of year have lower average education than children born later in the year

- If fraction of students who want to dropout prior to legal dropout age is independent of season of birth
 - Consistent with compulsory schooling constraining some students born later in the year to stay in school longer

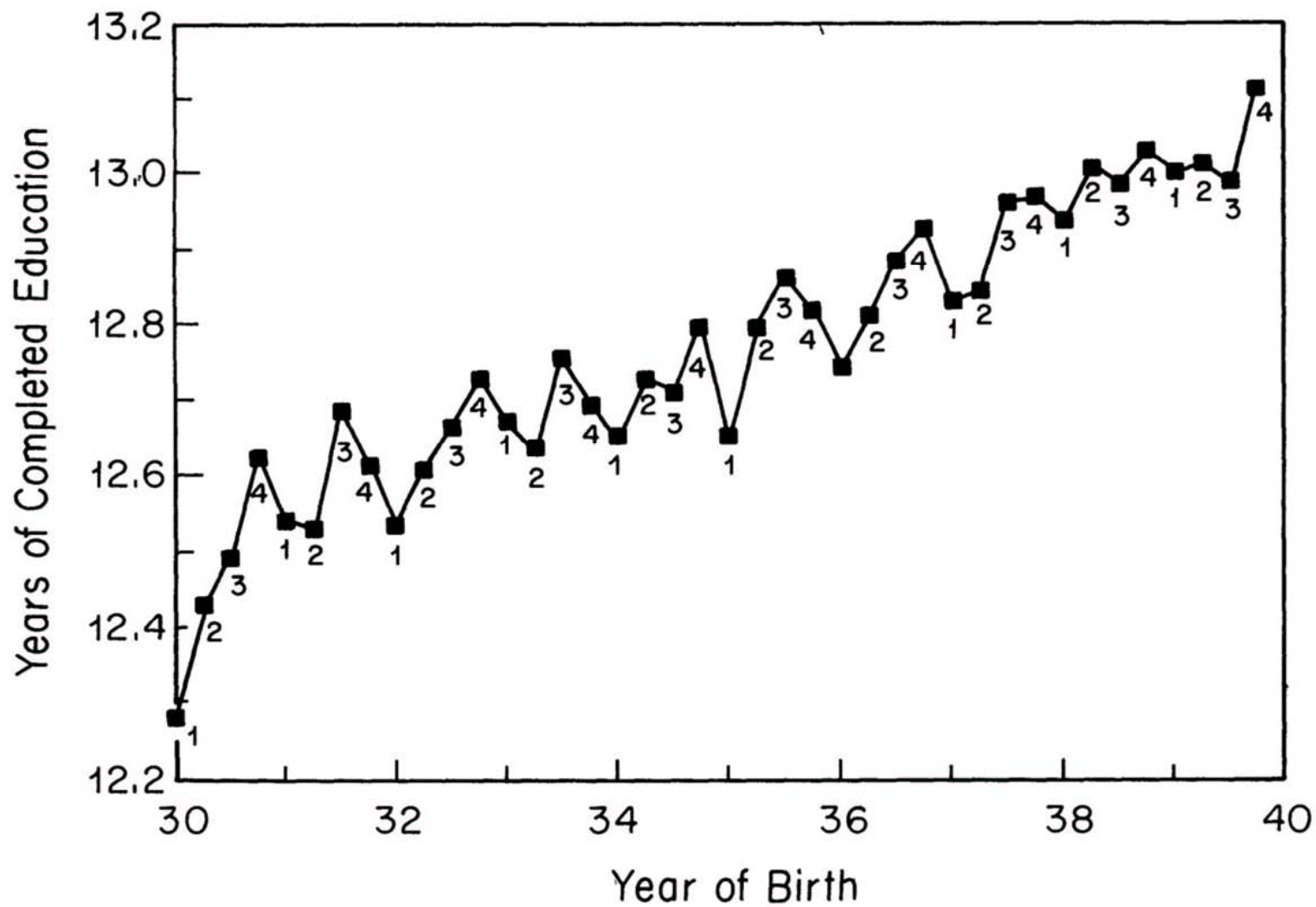


FIGURE I
 Years of Education and Season of Birth
 1980 Census
Note. Quarter of birth is listed below each observation.

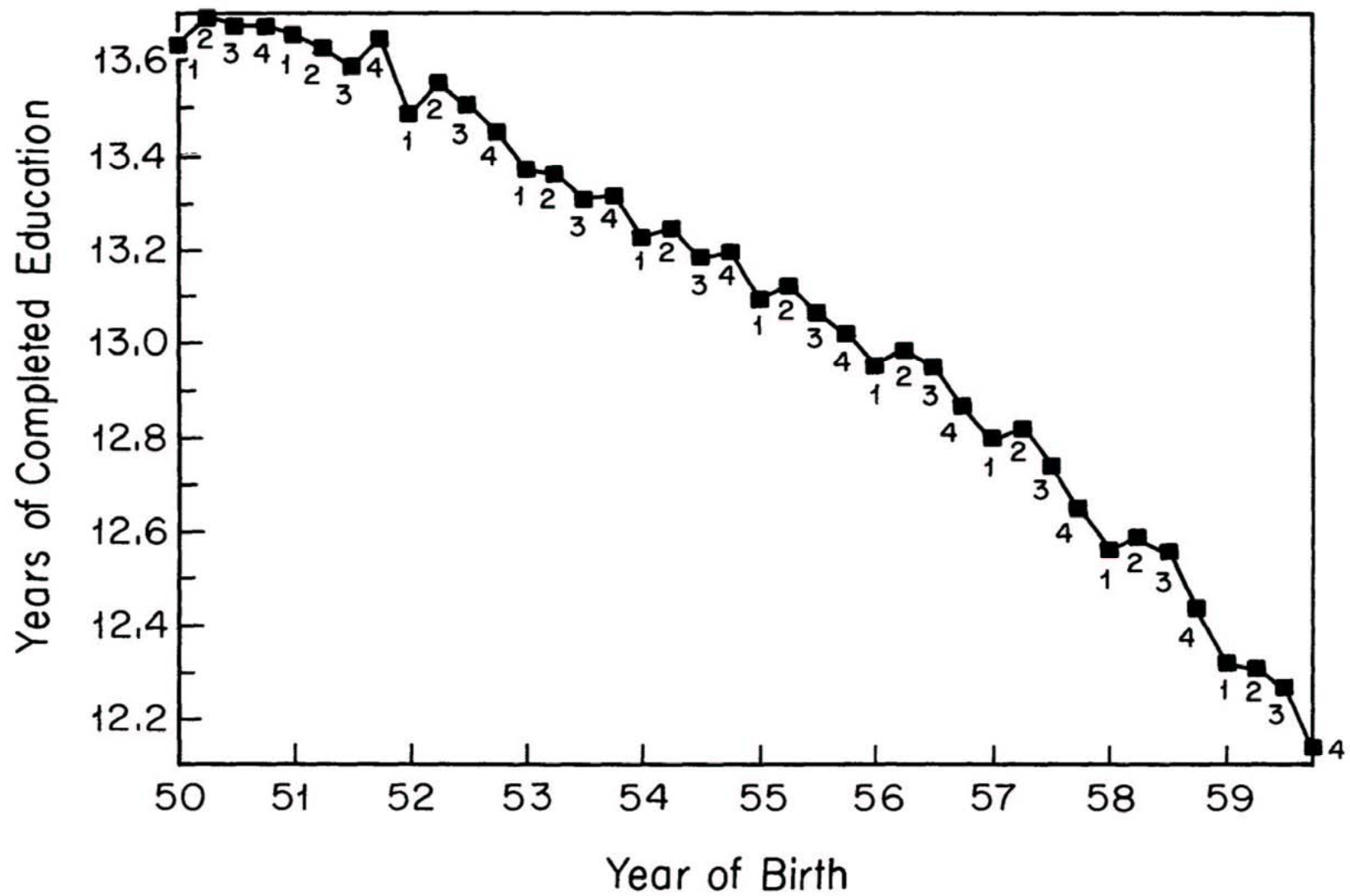


FIGURE III
 Years of Education and Season of Birth
 1980 Census
Note. Quarter of birth is listed below each observation.

A More Formal Analysis

Abstract from the trend in years of education across cohorts

- Subtract a moving average of the average education of cohorts born in the two preceding and two succeeding quarters

$$MA_{cj} = (E_{-2} + E_{-1} + E_{+1} + E_{+2}) / 4 \quad \text{year } c, \text{ quarter } j$$

Regress de-trended years of education on quarterly dummies

$$(E_{icj} - MA_{cj}) = \alpha + \sum_j^3 \beta_j q_{icj} + \omega_{icj} \quad \text{individual } i$$

1980 census

- For men born in the 1930s and 1940s, average number of completed years of schooling is about 1/10th of a year lower for those born in the first relative to the last quarter

Big effect

No effect

TABLE I
THE EFFECT OF QUARTER OF BIRTH ON VARIOUS EDUCATIONAL
OUTCOME VARIABLES

| Outcome variable | Birth cohort | Mean | Quarter-of-birth effect ^a | | | <i>F</i> -test ^b [<i>P</i> -value] |
|--|--------------|-------|--------------------------------------|-------------------|-------------------|---|
| | | | I | II | III | |
| Total years of education | 1930–1939 | 12.79 | -0.124 (0.017) | -0.086 (0.017) | -0.015 (0.016) | 24.9 [0.0001] |
| | 1940–1949 | 13.56 | -0.085 (0.012) | -0.035 (0.012) | -0.017 (0.011) | 18.6 [0.0001] |
| High school graduate | 1930–1939 | 0.77 | -0.019 (0.002) | -0.020 (0.002) | -0.004 (0.002) | 46.4 [0.0001] |
| | 1940–1949 | 0.86 | -0.015 (0.001) | -0.012 (0.001) | -0.002 (0.001) | 54.4 [0.0001] |
| Years of educ. for high school graduates | 1930–1939 | 13.99 | -0.004 (0.014) | 0.051 (0.014) | 0.012 (0.014) | 5.9 [0.0006] |
| | 1940–1949 | 14.28 | 0.005 (0.011) | 0.043 (0.011) | -0.003 (0.010) | 7.8 [0.0017] |
| College graduate | 1930–1939 | 0.24 | -0.005 (0.002) | 0.003 (0.002) | 0.002 (0.002) | 5.0 [0.0021] |
| | 1940–1949 | 0.30 | -0.003 (0.002) | 0.004 (0.002) | 0.000 (0.002) | 5.0 [0.0018] |
| Completed master's degree | 1930–1939 | 0.09 | -0.001 (0.001) | 0.002 (0.001) | -0.001 (0.001) | 1.7 [0.1599] |
| | 1940–1949 | 0.11 | 0.000 (0.001) | 0.004 (0.001) | 0.001 (0.001) | 3.9 [0.0091] |
| Completed doctoral degree | 1930–1939 | 0.03 | 0.002 (0.001) | 0.003 (0.001) | 0.000 (0.001) | 2.9 [0.0332] |
| | 1940–1949 | 0.04 | -0.002 (0.001) | 0.001 (0.001) | -0.001 (0.001) | 4.3 [0.0050] |

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. *F*-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

Additional Evidence

Additional evidence linking the seasonal pattern in education to compulsory schooling

- Seasonal pattern in years of education quite different for the subsample of individuals who have at least high school education (falsification exercise)
- Seasonal differences in years of education smaller for men born in the 1940s than for men born in the 1930s
 - Consistent with compulsory schooling laws being less of a binding constraint for more recent cohorts
- Find greater decline in enrollment of 16 year olds in states that permit 16 year olds to leave school than in states that have a 17 or 18 year old compulsory schooling requirement

Differences in Differences Analysis

Consider 1960 census and compare enrollment rates of students who had already turned 16 when the Census was taken (April 1) with those who turn 16 later in the year

- In States with an age 16 compulsory schooling requirement, the enrollment rate of those who have already turned 16 is 4.5% points lower
- In States with an age 17 or 18 compulsory schooling requirement, the enrollment rate of those who have already turned 16 is a statistically insignificant 0.6% points lower

Taking the difference between these two numbers (difference in differences, between both age groups and states)

- Compulsory schooling laws increased the enrollment rate by around 4% points in states with an age 17 or 18 requirement

TABLE II
PERCENTAGE OF AGE GROUP ENROLLED IN SCHOOL BY BIRTHDAY AND LEGAL
DROPOUT AGE^a

| Date of birth | Type of state law ^b | | Column (1) - (2) |
|--|----------------------------------|--|---------------------|
| | School-leaving age: 16 (1) | School-leaving age: 17 or 18 (2) | |
| Percent enrolled April 1, 1960 | | | |
| 1. Jan 1-Mar 31, 1944 (age 16) | 87.6 (0.6) | 91.0 (0.9) | -3.4 (1.1) |
| 2. Apr 1-Dec 31, 1944 (age 15) | 92.1 (0.3) | 91.6 (0.5) | 0.5 (0.6) |
| 3. Within-state diff. (row 1 - row 2) | -4.5 (0.7) | -0.6 (1.0) | -4.0 (1.2) |
| Percent enrolled April 1, 1970 | | | |
| 4. Jan 1-Mar 31, 1954 (age 16) | 94.2 (0.3) | 95.8 (0.5) | -1.6 (0.6) |
| 5. Apr 1-Dec 31, 1954 (age 15) | 96.1 (0.1) | 95.7 (0.3) | 0.4 (0.3) |
| 6. Within-state diff. (row 1 - row 2) | -1.9 (0.3) | 0.1 (0.6) | -2.0 (0.6) |
| Percent enrolled April 1, 1980 | | | |
| 7. Jan 1-Mar 31, 1964 (age 16) | 95.0 (0.1) | 96.2 (0.2) | -1.2 (0.2) |
| 8. Apr 1-Dec 31, 1964 (age 15) | 97.0 (0.1) | 97.7 (0.1) | -0.7 (0.1) |
| 9. Within-state diff. (row 1 - row 2) | -2.0 (0.1) | -1.5 (0.2) | 0.5 (0.3) |

a. Standard errors are in parentheses.

b. Data set used to compute rows 1-3 is the 1960 Census, 1 percent Public Use Sample; data set used to compute rows 4-6 is 1970 Census, 1 percent State Public Use Sample (15 percent form); data set used to compute rows 7-9 is the 1980 Census, 5 percent Public Use Sample. Each sample contains both boys and girls. Sample sizes are 4,153 for row 1; 12,512 for row 2; 7,758 for row 4; 24,636 for row 5; 42,740 for row 7; and 131,020 for row 8.

Why Do Compulsory Schooling Laws Work?

What is the underlying mechanism?

- Child labor laws restrict or prohibit children of compulsory schooling age from participating in the workforce
- Direct enforcement and policing of school attendance
 - Truant officers
 - Criminal penalties for parents of offenders

Weekly Earnings and Compulsory Schooling

Men born in the first quarter of the year also tend to earn less per week than men born later in the year

Secular age-earning profile

- Positively sloped for men aged between 30 and 39
- Flat for men aged between 40 and 49

Therefore, focus largely on men aged 40-49 in order to concentrate on the relationship between quarter of birth and weekly earnings after abstracting from age effects

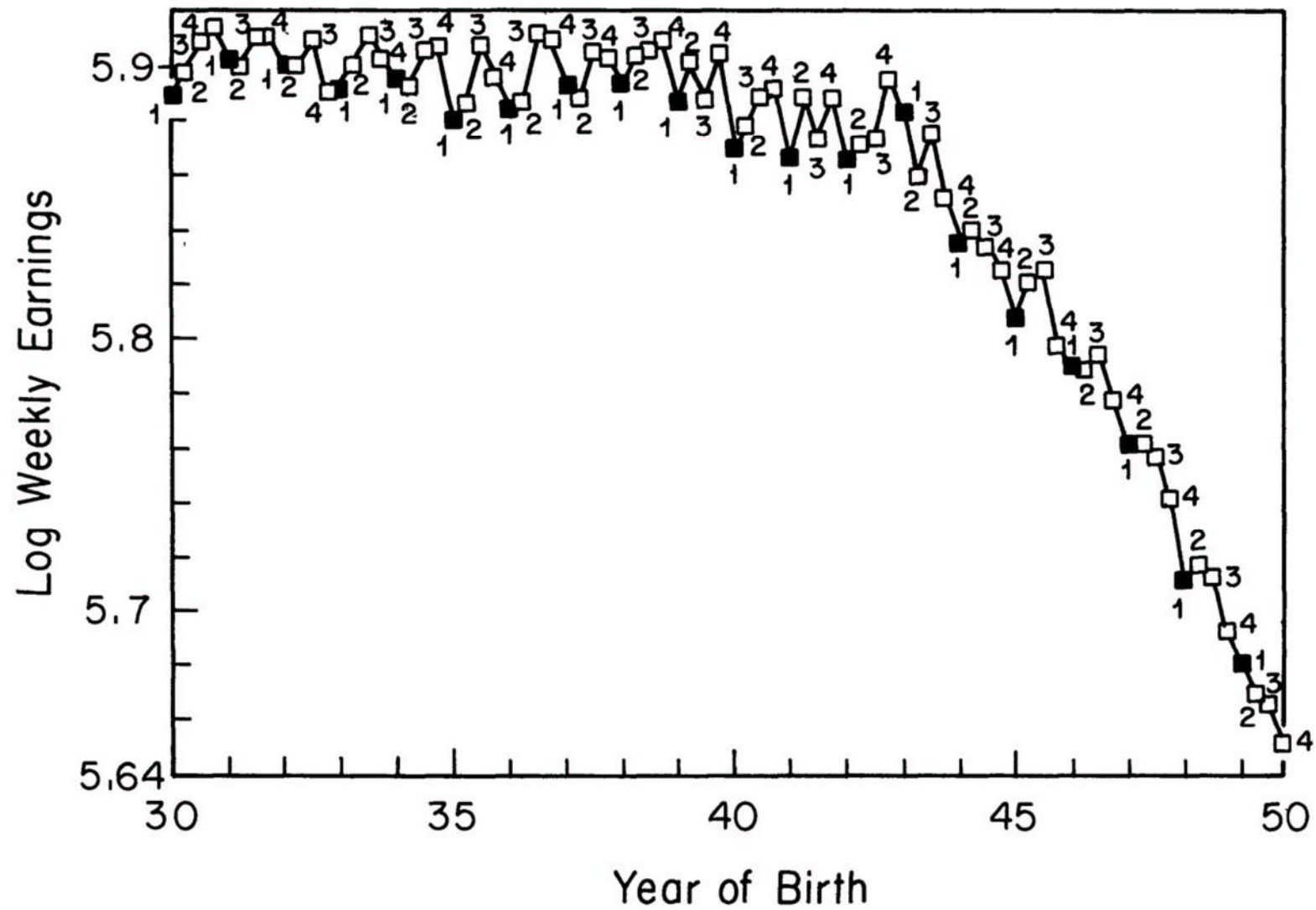


FIGURE V
 Mean Log Weekly Wage, by Quarter of Birth
 All Men Born 1930–1949; 1980 Census

Estimating the Rate of Return to Education

$$w_i = \beta_0 + \beta_1 E_i + u_i$$

$$E_i = \gamma_0 + \gamma_1 z_i + \varepsilon_i$$

- $w = \ln W$ is log weekly wage
- E is years of education
- z is a dummy which equals 1 if born in the first quarter of the year and 0 otherwise

Estimate β_1 using

- OLS
- Wald estimator

$$\hat{\beta}_1^{Wald} = \frac{\bar{w}_1 - \bar{w}_0}{\bar{E}_1 - \bar{E}_0}$$

TABLE III
 PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 1920–1929^a

| | (1) Born in 1st quarter of year | (2) Born in 2nd, 3rd, or 4th quarter of year | (3) Difference (std. error) (1) – (2) |
|--------------------------------------|--|---|--|
| ln (wkly. wage) | 5.1484 | 5.1574 | –0.00898 (0.00301) |
| Education | 11.3996 | 11.5252 | –0.1256 (0.0155) |
| Wald est. of return to education | | | 0.0715 (0.0219) |
| OLS return to education ^b | | | 0.0801 (0.0004) |

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

| | (1) Born in 1st quarter of year | (2) Born in 2nd, 3rd, or 4th quarter of year | (3) Difference (std. error) (1) – (2) |
|----------------------------------|--|---|--|
| ln (wkly. wage) | 5.8916 | 5.9027 | –0.01110 (0.00274) |
| Education | 12.6881 | 12.7969 | –0.1088 (0.0132) |
| Wald est. of return to education | | | 0.1020 (0.0239) |
| OLS return to education | | | 0.0709 (0.0003) |

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

Two Stage Least Squares Estimation

Consider the following economic model:

$$\ln W_i = X_i \beta + \rho E_i + \sum_c d_{ic} \xi_c + \mu_i \quad c=1, \dots, 10, \quad j=1, \dots, 3$$

$$E_i = X_i \pi + \sum_c d_{ic} \delta_c + \sum_c \sum_j d_{ic} q_{ij} \theta_{jc} + \varepsilon_i$$

- d_{ic} is $\{0,1\}$ dummy for whether individual i is born in year c
- q_{ij} is $\{0,1\}$ dummy for whether individual i is born in quarter j

Excluded instruments from the wage equation

- 3 quarter of birth dummies interacted with 9 year of birth dummies

Inclusion of birth year dummies in wage equation means ρ is identified from education variation across birth quarters within each birth year

- Identifying assumption: quarter of birth uncorrelated with μ ²⁷

Baseline 2SLS Empirical Results

1980 census, 40 - 49 year old men

- OLS estimate, Column (1), Table V, 0.0711 (*0.003*)
- 2SLS estimate, Column (2), Table V, 0.0891 (*0.0161*)
- *Very similar (see Card, 1999, HLE, for discussion)*

Test the model's overidentifying restrictions

- Unable to reject the null hypothesis that the excluded exogenous variables are orthogonal to the wage equation residuals (5% critical value for $\chi^2(29)$ is 42.6)

Consider augmented specifications that include

- Age and Age squared measured in quarters
- Dummies for race, residence in SMSA, marital status, and region of residence

Table 5 : Baseline 2SLS Empirical Results

TABLE V
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 | 0.0711 (0.0003) | 0.0891 (0.0161) | 0.0711 (0.0003) | 0.0760 (0.0290) | 0.0632 (0.0003) | 0.0806 (0.0164) | 0.0632 (0.0003) | 0.0600 (0.0299) |
| Race (1 = black) | — | — | — | — | -0.2575 (0.0040) | -0.2302 (0.0261) | -0.2575 (0.0040) | -0.2626 (0.0458) |
| SMSA (1 = center city) | — | — | — | — | 0.1763 (0.0029) | 0.1581 (0.0174) | 0.1763 (0.0029) | 0.1797 (0.0305) |
| Married (1 = married) | — | — | — | — | 0.2479 (0.0032) | 0.2440 (0.0049) | 0.2479 (0.0032) | 0.2486 (0.0073) |
| 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 |
| Age | — | — | -0.0772 (0.0621) | -0.0801 (0.0645) | — | — | -0.0760 (0.0604) | -0.0741 (0.0626) |
| Age-squared | — | — | 0.0008 (0.0007) | 0.0008 (0.0007) | — | — | 0.0008 (0.0007) | 0.0007 (0.0007) |
| χ^2 [dof] | — | 25.4 [29] | — | 23.1 [27] | — | 22.5 [29] | — | 19.6 [27] |

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Additional 2SLS Empirical Results

Allow for cross-state variability in the seasonal pattern of education

- Allows for variation in compulsory schooling laws and school start age policy across states

$$\ln W_i = X_i \beta + \rho E_i + \sum_c d_{ic} \xi_c + \sum_s d_{is} \lambda_s + \mu_i$$

$$E_i = X_i \pi + \sum_c d_{ic} \delta_c + \sum_s d_{is} \phi_s + \sum_c \sum_j d_{ic} q_{ij} \theta_{jc} + \sum_s \sum_j d_{is} q_{ij} \eta_{js} + \varepsilon_i$$

- d_{is} is $\{0,1\}$ dummy for whether individual i is born in state s

Inclusion of birth year and birth state dummies in wage equation means that ρ is identified from education variation across quarters of birth by birth year and birth state

Table 7 : Additional 2SLS Empirical Results

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 (0.0003) | 0.0928 (0.0093) | 0.0673 (0.0003) | 0.0907 (0.0107) | 0.0628 (0.0003) | 0.0831 (0.0095) | 0.0628 (0.0003) | 0.0811 (0.0109) |
| Race (1 = black) | — | — | — | — | -0.2547 (0.0043) | -0.2333 (0.0109) | -0.2547 (0.0043) | -0.2354 (0.0122) |
| SMSA (1 = center city) | — | — | — | — | 0.1705 (0.0029) | 0.1511 (0.0095) | 0.1705 (0.0029) | 0.1531 (0.0107) |
| Married (1 = married) | — | — | — | — | 0.2487 (0.0032) | 0.2435 (0.0040) | 0.2487 (0.0032) | 0.2441 (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 (0.0617) | -0.0880 (0.0624) | — | — | -0.0778 (0.0603) | -0.0876 (0.0609) |
| Age-squared | — | — | 0.0008 (0.0007) | 0.0009 (0.0007) | — | — | 0.0008 (0.0007) | 0.0009 (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.

Authors' Interpretation of Their Results

Exploiting exogenous variation due to the interaction of school start policy and compulsory schooling laws, 2SLS estimates of the rate of return to education are close to those estimated using OLS

- Typically not statistically significantly different

To the extent differences between OLS and 2SLS exist

- Suggest omitted variables or measurement error in education bias the OLS estimate of rate of return to education downwards

Compulsory schooling laws effective in compelling some students to attend school and those compelled to remain in school earn higher wages as a result of their extra schooling

- Need for a fuller cost-benefit analysis of the effects of compulsory schooling laws

Potential Causes for Concern (Bound, Jaeger and Baker , JASA, 1995)

Problem when instruments are only weakly correlated with the endogenous explanatory variables

- Results in large 2SLS standard errors
- More importantly
 - Even a weak correlation between the instruments and the error in the wage equation can lead to a large inconsistency in the IV estimates
 - In finite samples, IV estimates are biased in the same direction as the OLS estimates, with the magnitude of the bias approaching that of OLS as the R^2 between the instruments and the endogenous explanatory variables approaches 0

Potential Causes for Concern (Bound, Jaeger and Baker, JASA, 1995)

Cannot learn about the potential sizes of these biases from the second-stage regression (wage equation)

Important to examine the power of the instruments in the first-stage regression (education equation)

- F-test for the joint statistical significance of the excluded exogenous variables in the first-stage regression

Bound et al. (1995) present evidence that weak instruments are a potential concern in Angrist and Krueger (1991)

- Despite large samples sizes, the F-statistics for a test of the joint statistical significance of the excluded exogenous variables in the first-stage regression are close to 1
- Adding interaction terms to provide additional excluded exogenous variables may exacerbate the weak instruments problem

Partial R-squared

$$y = \beta_0 + \beta_1 x + \beta_2 z + u$$

Partial R^2 between y and z

- Regress y on x
- Regress z on x
- Regress y after conditioning on x against z after conditioning on x
- Partial R^2 is the R^2 of this last regression

$$R_{yz|x}^2 = \frac{R^2 - R_{yx}^2}{1 - R_{yx}^2}$$

Conclusions

Exploited exogenous variation in years of education due to compulsory schooling to estimate rate of return to education

- Interaction of school entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months
 - Provided evidence of a systematic relationship between quarter of birth and years of education
 - Used this relationship to provide Wald and 2SLS estimates of the rate of return to education

Example of instruments derived from nature, institutions and government policy (“natural experiments”)

- Identifying assumptions
- Power of the instruments