

# Returns to Schooling

Department of Economics  
University of Wisconsin-Madison

March 14, 2008

# Returns to Schooling

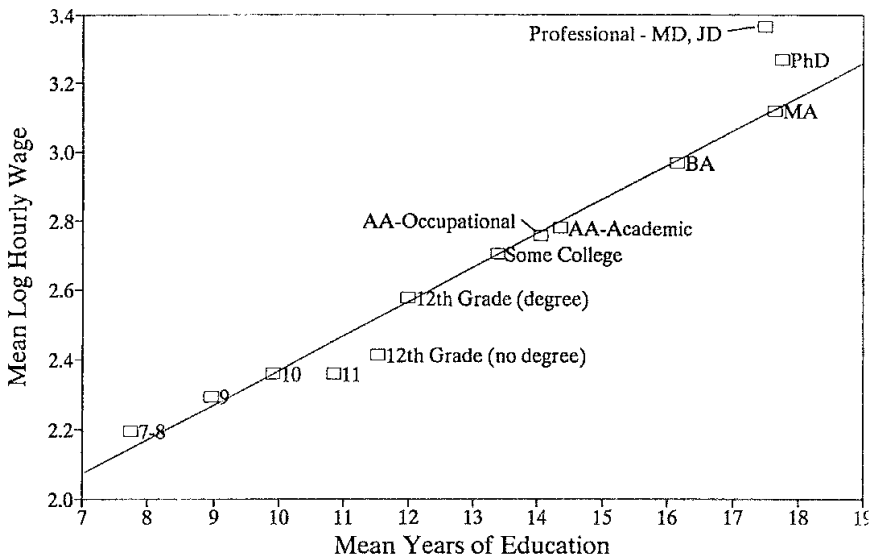
This comes from the Card Handbook Chapter

Lets assume that

$$\log(W_i) = a + bS_i + g(X_i) + \varepsilon_i$$

where  $W_i$  is wages,  $S_i$  is schooling, and  $X_i$  is experience.

Is schooling Really linear?



The figure shows that without controlling for ability bias, it seems to be pretty close.

Park, 1994 finds somewhat different results. He finds drop at 15 and jump at 16 but otherwise approximately linear.

We are worried about ability bias we want to use instrumental variables

A good instrument should have two qualities:

- It should be correlated with schooling
- It should be uncorrelated with ability (and other unobservables)

Many different things have been tried. Lets go through some of them

# Family Background

If my parents earn quite a bit of money it should be easier for me to borrow for college

Also they might put more value on education

This should make me more likely to go

This has no direct effect on my income-Wisconsin did not ask how much education my Father had when they made my offer

But is family background likely to be uncorrelated with unobserved ability?

# Closeness of College

If I have a college in my town it should be much easier to attend college

- I can live at home
- If I live on campus
  - I can travel to college easily
  - I can come home for meals and to get my clothes washed
- I can hang out with my friends from High school

But is this uncorrelated with unobserved ability?

# Quarter of Birth

This is the most creative

Consider the following two aspects of the U.S. education system (this actually varies from state to state and across time but ignore that for now),

- People begin Kindergarten in the calendar year in which they turn 5
- You must stay in school until you are 16

Now consider kids who:

- Can't stand school and will leave as soon as possible
- Obey truancy law and school age starting law
- Are born on either December 31, 1972 or January 1, 1973

## Those born on December 31 will

- turn 5 in the calendar year 1977 and will start school then (at age 4)
- will stop school on their 16th birthday which will be on Dec. 31, 1988
- thus they will stop school during the winter break of 11th grade

## Those born on January 1 will

- turn 5 in the calendar year 1978 and will start school then (at age 5)
- will stop school on their 16th birthday which will be on Jan. 1, 1989
- thus they will stop school during the winter break of 10th grade

The instrument is a dummy variable for whether you are born on Dec. 31 or Jan 1

This is pretty cool:

- For reasons above it will be correlated with education
- No reason at all to believe that it is correlated with unobserved ability

The Fact that not everyone obeys perfectly is not problematic:

An instrument just needs to be correlated with schooling, it does not have to be perfectly correlated

In practice we can't just use the day as an instrument, use "quarter of birth" instead

# Policy Changes

Another possibility is to use institutional features that affect schooling

Here often institutional features affect one group or one cohort rather than others

TABLE II  
OLS AND IV ESTIMATES OF THE RETURN TO EDUCATION WITH INSTRUMENTS BASED ON FEATURES OF THE SCHOOL SYSTEM

Author	Sample and Instrument	Schooling Coefficients		
		OLS	IV	
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and indicators for race, marital status, urban residence.	1920–29 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930–39 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940–49 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth. LIML estimates.	1930–39 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940–49 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience. Note: Schooling measured in units of college credit equivalents.	Models without test score or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) Instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with parental education. Controls include race, experience (treated as endogenous), region, and parental education	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)
		Models that use college proximity $\times$ family background as instrument	—	0.097 (0.048)

5. Conneely and Uusitalo (1997)	Finnish men who served in the army in 1982, and were working full time in civilian jobs in 1994. Administrative earnings and education data. Instrument is living in university town in 1980. Controls include quadratic in experience and parental education and earnings.	Models that exclude parental education and earnings	0.085 (0.001)	0.110 (0.024)
		Models that include parental education and earnings	0.083 (0.001)	0.098 (0.035)
6. Harmon and Walker (1995)	British Family Expenditure Survey 1978–86 (men). Instruments are indicators for changes in the minimum school leaving age in 1947 and 1973. Controls include quadratic in age, survey year, and region.		0.061 (0.001)	0.153 (0.015)
7. Ichino and Winter-Ebmer (1998)	Austria: 1983 Census, men born before 1946. Germany: 1986 GSOEP for adult men. Instrument is indicator for 1930–35 cohort. (Second German IV also uses dummy for father's veteran status). Controls include age, unemployment rate at age 14, and father's education (Germany only). Education measure is dummy for high school or more.	Austrian Men	0.518 (0.015)	0.947 (0.343)
		German Men	0.289 (0.031)	0.590/0.708 (0.844) (0.279)
8. Lemieux and Card (1998)	Canadian Census, 1971 and 1981: French-speaking men in Quebec and English-speaking in Ontario. Instrument is dummy for Ontario men age 19–22 in 1946. Controls include full set of experience dummies and Quebec-specific cubic experience profile.	1971 Census:	0.070 (0.002)	0.164 (0.053)
		1981 Census:	0.062 (0.001)	0.076 (0.022)
9. Meghir and Palme (1999)	Swedish Level of Living Survey (SLLS) data for men born 1945–55, with earnings in 1991, and Individual Statistics (IS) sample of men born in 1948 and 1953, with earnings in 1993. Instrument is dummy for attending “reformed” school system at age 13. Other controls include cohort, father's education, and county dummies. Models for IS data also include test scores at age 13.	SLLS Data (Years of education)	0.028 (0.007)	0.036 (0.021)
		IS Data (Dummy for 1–2 years of college relative to minimum schooling)	0.222 (0.020)	0.245 (0.082)

TABLE II—Continued

Author	Sample and Instrument		Schooling Coefficients	
			OLS	IV
10. Maluccio (1997)	Bicol Multipurpose Survey (rural Philippines): male and female wage earners age 20–44 in 1994, whose families were interviewed in 1978. Instruments are distance to nearest high school and indicator for local private high school. Controls include quadratic in age and indicators for gender and residence in a rural community.	Models that do not control for selection of employment status or location	0.073 (0.011)	0.145 (0.041)
		Models with selection correction for location and employment status	0.063 (0.006)	0.113 (0.033)
11. Duflo (1999)	1995 Intercensal Survey of Indonesia: men born 1950–72. Instruments are interactions of birth year and targeted level of school building activity in region of birth. Other controls are dummies for year and region of birth and interactions of year of birth and child population in region of birth. Second IV adds controls for year of birth interacted with regional enrollment rate and presence of water and sanitation programs in region.	Model for hourly wage	0.078 (0.001)	0.064/0.091 (0.025) (0.023)
		Model for monthly wage with imputation for self-employed.	0.057 (0.003)	0.064/0.049 (0.017) (0.013)

Notes: See text for sources and more information on individual studies.

Consistently IV estimates are higher than OLS

Why?

- Bad Instruments
- Ability Bias
- Measurement Error
- Publication Bias
- Discount Rate Bias

# Discount Rate Bias

This is a simplified version of it (and my version of it)

Lang and Card are more complicated and explain it somewhat differently

Suppose 2 levels of schooling and 2 values of instrument

$$S_i = \begin{cases} 0 & \text{High School} \\ 1 & \text{College} \end{cases}$$

$$Z_i = \begin{cases} z_h & \text{with probability } p_h \\ z_l & \text{with probability } 1 - p_h \end{cases}$$

$$\log(W_i) = \alpha_i + \beta_i S_i + u_i$$

$E(u_i) = 0$  and is uncorrelated with  $S_i$

$S_i$  is potentially correlated with  $(\alpha_i, \beta_i)$

Suppose that we have an instrument  $Z_i$  which is correlated with  $S_i$  but not with  $(\alpha_i, \beta_i, u_i)$

$$E(W_i | Z_i) = E(a_i | Z_i) + E(\beta_i S_i | Z_i)$$

If  $\beta_i = \beta_0$  so it is constant for everyone

$$E(W_i | Z_i) = E(a_i | Z_i) + \beta_0 E(S_i | Z_i)$$

so IV works

However if  $\beta_i$  varies across persons then in general

$$E(\beta_i S_i | Z_i) \neq E(\beta_i) E(S_i | Z_i)$$

To see what it converges to I draw on Imbens and Angrist (EMA, 1994)

Assume that

$$\Pr(S_i = 1 | Z_i = z_h) > \Pr(S_i = 1 | Z_i = z_\ell)$$

There are 4 different types of people those who attend college when:

- ①  $Z_i = z_h, Z_i = z_\ell$
- ② never
- ③  $Z_i = z_h$  only
- ④  $Z_i = z_\ell$  only

Imbens and Angrist's monotonicity rules out 4 as a possibility

This seems reasonable when we look at the instruments

Let  $P_1, P_2, P_3$  be the fraction of the population in each of the relevant 3 groups that are listed above

Let  $q_i$  represent the group to whom you belong (unobserved to econometrician)

$$\begin{aligned}\widehat{\beta}_{IV} &\xrightarrow{p} \frac{\text{cov}(\log(W_i), Z_i)}{\text{cov}(S_i, Z_i)} \\ &= \frac{E(\beta_i S_i Z_i) - E(\beta_i S_i) E(Z_i)}{E(S_i Z_i) - E(\beta_i S_i) E(Z_i)}\end{aligned}$$

Lets look at the pieces, first the numerator

$$\begin{aligned} & E(\beta_i S_i Z_i) - E(\beta_i S_i) E(Z_i) \\ = & \sum_{j=1}^3 E(\beta_i S_i Z_i | q_i = j) - \left( \sum_{j=1}^3 E(\beta_i S_i | q_i = j) P_j \right) E(Z_i) \\ = & P_1 E(Z_i) E(\beta_i | q_i = 1) + P_3 P_h z_h E(\beta_i | q_i = 3) \\ & - (P_1 E(\beta_i | q_i = 1) + P_3 P_h E(\beta_i | q_i = 3)) E(Z_i) \\ = & P_3 P_h (z_h - E(Z_i)) E(\beta_i | q_i = 3) \end{aligned}$$

then the denominator

$$\begin{aligned} & E(S_i Z_i) - E(S_i) E(Z_i) \\ = & \sum_{j=1}^3 E(S_i Z_i | q_i = j) - \left( \sum_{j=1}^3 E(S_i | q_i = j) P_j \right) P_h \\ = & P_1 E(Z_i) + P_3 P_h z_h - (P_1 + P_3 P_h) E(Z_i) \\ = & P_3 P_h (z_h - E(Z_i)) \end{aligned}$$

Thus

$$\begin{aligned}\widehat{\beta}_{IV} &\xrightarrow{p} \frac{\text{cov}(\log(W_i), Z_i)}{\text{cov}(S_i, Z_i)} \\ &= \frac{E(\beta_i S_i Z_i) - E(\beta_i S_i) E(Z_i)}{E(S_i Z_i) - E(\beta_i S_i) E(Z_i)} \\ &= \frac{P_3 P_h (z_h - E(Z_i)) E(\beta_i | q_i = 3)}{P_3 P_h (z_h - E(Z_i))} \\ &= E(\beta_i | q_i = 3)\end{aligned}$$

Thus  $\widehat{\beta}_{IV}$  may be high because  $E(\beta_i | q_i = 3)$  may be high

There are a number of reasons why this might be the case:

- Borrowing Constraints
- Nonlinearities in schooling

# Twins

$$\log(w_{if}) = \alpha_f + \beta S_{if} + u_{if}$$

The problem is that  $\alpha_f$  is correlated with  $S_{if}$

We can solve by differencing

$$E(\log(w_{if}) - \log(w_{jf})) = \beta E(S_{if} - S_{jf})$$

Use this to get consistent estimates of  $\beta$

Table 6

Cross-sectional and within-family differenced estimates of the return to education for twins<sup>a</sup>

Author	Sample and specification		Cross-sectional OLS	Differenced	
				OLS	IV
1. Ashenfelter and Rouse (1998)	1991–1993 Princeton Twins Survey. Identical male and female twins. Controls include quadratic in age, gender and race. Added controls include tenure, marital status and union status.	Basic	0.110 (0.010)	0.070 (0.019)	0.088 (0.025)
		Basic + added controls	0.113 (0.010)	0.078 (0.018)	0.100 (0.023)
2. Rouse (1997)	1991–1995 Princeton Twins Survey. Identical male and female twins. Basic controls as above.		0.105 (0.008)	0.075 (0.017)	0.110 (0.023)
3. Miller et al. (1995)	Australian Twins Register. Identical and fraternal twins. Controls include quadratic in age, gender, marital status. Incomes imputed from occupation	Identical twins	0.064 (0.002)	0.025 (0.005)	0.048 (0.010)
		Fraternal twins	0.066 (0.002)	0.045 (0.005)	0.074 (0.008)
4. Behrman et al. (1994)	NAS-NRC white male twins born 1917–1927, plus male twins born 1936–1955 from Minnesota Twins Registry. Controls include quadratic in age <sup>b</sup>	Identical twins	0.071 (0.002)	0.035 (0.005)	0.056 –
		Fraternal twins	0.073 (0.003)	0.057 (0.005)	0.071 –
5. Isacson (1997)	Swedish same-sex twins with both administrative and survey measures of schooling. Controls include sex, marital status, quadratic in age, and residence in a large city <sup>c</sup>	Identical twins	0.049 (0.002)	0.023 (0.004)	0.024 (0.008)
		Fraternal twins	0.051 (0.002)	0.040 (0.003)	0.054 (0.006)

Problems:

- Twins aren't a random sample of population (and often not a random sample of twins)
- Need to have variation in  $S_{if} - S_{jf}$
- Is  $\alpha_{if}$  really the same for identical twins?

In Willis and Rosen two things affect schooling choices:

- $r_i$
- ability differences (Roy model style)

These should be the same for both

Suppose that  $\alpha_{if} \neq \alpha_{jf}$

We expect that:

- $\text{corr}(S_{if}, \alpha_{if}) > 0$
- $\text{corr}(S_{if} - S_{jf}, \alpha_{if} - \alpha_{jf}) > 0$

While most of the variation in  $\alpha_{if}$  may be explained by family effects, it may also be that most of the variation in  $S_{if}$  is explained by family effects as well

Since

$$\beta_{OLS} = \beta + \frac{\text{cov}(S_{if}, \alpha_{if})}{\text{var}(S_{if})}$$
$$\beta_{FE} = \beta + \frac{\text{cov}(S_{if} - S_{jf}, \alpha_{if} - \alpha_{jf})}{\text{var}(S_{if} - S_{jf})}$$

If  $\text{var}(S_{if} - S_{jf})$  is small the bias could be large

It is not clear which has bigger bias