Applied Economics

Angrist and Krueger, *The Quarterly Journal of Economics*, 1991

Does compulsory school attendance affect schooling and earnings?

Economics Department
Universidad Carlos III de Madrid

See also Angrist and Pischke (ch. 4)

Returns to education

- A typical example of endogeneity appears when we want to estimate the return to education regressing income on education (and some other factors). An omitted variable in that regression may be the capacity or ability of the individuals.
- If more able individuals are on average more educated, and at the same time they earn higher wages, education will be correlated with the error term and OLS will yield inconsistent estimators.
- Angrist and Krueger in this paper suggest to solve the endogeneity problem using an instrumental variables estimation.
- The instrument is date of birth measured by quarter of birth.

Idea from US laws

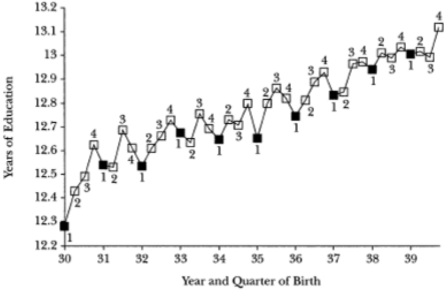
- The idea exploits the variation induced by compulsory schooling laws in the US: most states require students to enter school in the calendar year they turn 6. In addition, these laws require students to remain in school at least until their 16th birthday.
- For instance, a student born in January starts school at 6 (and 8 months), and at her 16th birthday she will have 9 years of completed schooling. A student born in December starts school at 5 (and 8 months), and when she turns 16 she will have 10 years of schooling.
- Then, depending on the date of birth students will be in different grades, or through a given grade to a different degree, when they reach the legal dropout age.

Instrument validity

- Exogeneity: quarter of birth should not affect income directly: not correlated with ability, motivation, family connections, etc.
- Relevance: correlated with educational attainment.
- Angrist and Krueger use data from the 1980 census in the US for men born between 1930 and 1959. We use data from the Joshua Angrist Dataverse, an extract for men born in 1930-1939.
- They show the following figure to argue that men born earlier in the calendar year tended to have lower average schooling levels.

Quarter of birth and education

Figure 1
Mean Years of Completed Education, by Quarter of Birth

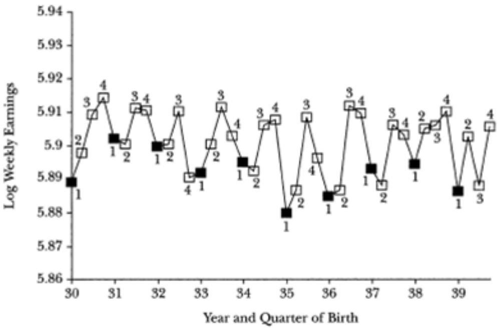


Source: Authors' calculations from the 1980 Census.

- This figure reflects the first stage (conditional on year of birth).
- The figure shows the increasing trend in education and also a pattern related to the quarter of birth.

Quarter of birth and earnings

Figure 2
Mean Log Weekly Earnings, by Quarter of Birth



Source: Authors' calculations from the 1980 Census.

- The figure shows the "reduced-form" relationship between the instrument and the dependent variable.

Intuition

- On average, men born in early quarters tend to earn less than those born later in the year.
- This reduced-form relation parallels the previous pattern.
- In general, the analysis of the first stage and the "reduced-form" is useful to understand the potential causality in the relation, and to motivate the use of the instruments.
- In this case, the graphs are used to motivate that income differences by quarter of birth are due to the schooling differences by quarter of birth.

More formally

To simplify the discussion we use as instrument a dummy variable taking the value one if the individual was born in the first quarter (Z_i).

Then, a mathematical representation of the story in the figures is:

- First figure (first stage):

  $E_i = \pi_0 + \pi_1 Z_i + \pi_2 Y_{1930} + \ldots \pi_{10} Y_{1938} + v_i,$

  where Y_j is a dummy variable taking the value one for being born in year j. The parameter π_1 captures the effect of the quarter of birth on education, conditional on year of birth.

- Second figure (reduced-form relation):

  $log(w_i) = \gamma_0 + \gamma_1 Z_i + \gamma_2 Y_{1930} + \ldots \gamma_{10} Y_{1938} + \varepsilon_i,$

  γ_1 captures the direct effect of the quarter of birth on earnings, conditional on year of birth.

First Estimations

- The estimations in the paper of the previous equations show that men born in the first quarter have around 1/10 less education and earn 1% less than men born in other quarters.
- We replicate the estimations in the paper.
- Note: it is necessary to generate dummy variables for the first quarter and for the controls (we present the results with no controls, to replicate Table III, panel B).

Replication

Estimations in `gretl`:

$$\widehat{\text{EDUC}} = \underset{(0.0066)}{12.797} - \underset{(0.0133)}{0.109}\ QB_1$$

$$\widehat{\text{LWKLYWGE}} = \underset{(0.0014)}{5.903} - \underset{(0.0027)}{0.011}\ QB_1$$

Table III in the paper:

	(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)

Main results

- The baseline equation estimated in the paper is:
  $log(w_i) = \beta_0 + \beta_1 X_i + \sum_c \xi_c Y_{ic} + \rho E_i + \mu_i,$
- with a first stage given by:
  $E_i = \pi_0 + \pi_1 X_i + \sum_c \delta_c Y_{ic} + \sum_c \theta_{cj} Y_{ic} Q_{ij} + \varepsilon_i$
- The instruments are the interactions between quarter and year of birth. Note: the IV variables are also called excluded instruments as opposed to included instruments, that are the exogenous controls in the original equation (Y_{ic} in this case).
- The authors add some additional regressors: dummies for the region of residence, age, age squared, race, marital status. Results are similar in all specifications.

OLS vs IV

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	0.0891	0.0711	0.0760	0.0632	0.0806	0.0632	0.0600
	(0.0003)	(0.0161)	(0.0003)	(0.0290)	(0.0003)	(0.0164)	(0.0003)	(0.0299)
Race (1 = black)	—	—	—	—	−0.2575	−0.2302	−0.2575	−0.2626
					(0.0040)	(0.0261)	(0.0040)	(0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763	0.1581	0.1763	0.1797
					(0.0029)	(0.0174)	(0.0029)	(0.0305)
Married (1 = married)	—	—	—	—	0.2479	0.2440	0.2479	0.2486
					(0.0032)	(0.0049)	(0.0032)	(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	−0.0801	—	—	−0.0760	−0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared	—	—	0.0008	0.0008	—	—	0.0008	0.0007
			(0.0007)	(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.

OLS vs IV

Authors' comments:

- In all the specifications in the table the models are overidentified. They perform Sargan tests and they don't reject in none of the cases the overidentifying restrictions.
- 2SLS estimations are in general similar than the OLS estimations.
- The small differences between them suggest that OLS may be underestimating the return to education.
- Overall, "this evidence casts doubt on the importance of omitted variables bias in OLS estimates" in this case.

In Applied Economics

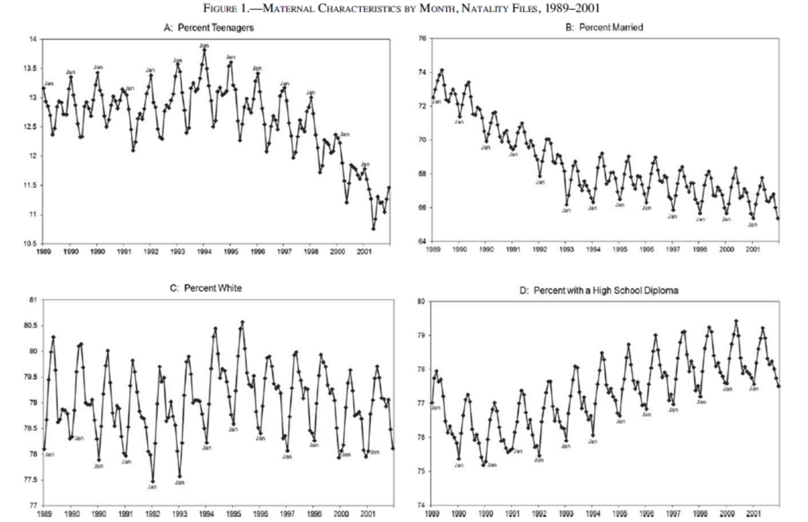
¿What would you have done after this course? Let's look at results in columns (3) and (4).

- Instruments exogeneity is tested with Sargan (in the table): $\chi^2 = 23.1$, $p - value = 0.679$.
- Are the instruments relevant? Look at first stage $F = 1.5909$ (recall that an F statistic below 10 is cause for concern).
- The consequence of instruments with little explanatory power is to increase bias in the IV estimators. If the explanatory power is "weak" asymptotic properties fail.
- "This evidence casts doubt on the importance of omitted variables bias" Should we use OLS or IV? Hausman test: p-value $= 0.864908$.

Critiques: Exogeneity

- Is quarter of birth an exogenous variable?
- Buckles and Hungerman in The Review of Economics and Statistics in 2013 argue that there is some seasonality in mother's characteristics, and this seasonality may be behind the seasonality in earnings.
- They find that family background controls explain nearly half of season-of-birth's relation to adult outcomes.
- The authors find that women giving birth in the winter look different from other women: they are younger, less educated, and less likely to be married.

Seasonality in maternal characteristics



The sample for each figure includes all births in the Natality Detail Files from 1989 to 2001, for 52,041,054 observations.

Critiques: Weak instrument

- Bound, Jaeger and Baker, in the Journal of the American Statistical Association in 1995 argue that this is a case of a weak instrument: quarter of birth is not highly correlated with education once we control for other variables.
- BJB used an irrelevant instrument (suggested by Krueger): they assign randomly a quarter of birth to each individual and use that quarter as an instrument. They argue that the 2SLS results were similar to those obtained with the real quarter of birth (see Table 3).
- They also conclude that the problem should have been detected looking at the first stage.

BJB results

Table 3. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Using Simulated Quarter of Birth (500 replications)

Table (column)	1 (4)	1 (6)	2 (2)	2 (4)
Estimated Coefficient				
Mean	.062	.061	.060	.060
Standard deviation of mean	.038	.039	.015	.015
5th percentile	−.001	−.002	.034	.035
Median	.061	.061	.060	.060
95th percentile	.119	.127	.083	.082
Estimated Standard Error				
Mean	.037	.039	.015	.015

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509.

- The first two columns are comparable to columns (6) and (8) in Table V in AK. The other two include additional instruments.