

In-Class Activity: Instrumenting Education

A labor economist wants to estimate the **causal effect of years of education on log hourly wages**. She uses data on 3,010 U.S. men born between 1930 and 1939 from the 1980 Census. After hanging out with the ECON3500 crowd for too long, she is appropriately worried that OLS estimates are biased because **ability** is unobserved: more able individuals get more education *and* earn higher wages, even holding education constant.

Her proposed instrument: **quarter of birth**. U.S. compulsory schooling laws require students to stay in school until age 16. But students' age when they first enter school is staggered because of age cut-offs. Hence, students born earlier in the year reach age 16 earlier in the school year, so they can legally drop out with less total education. The idea is that quarter of birth thus affects years of education completed but should not directly affect wages.

(Stylized example in the spirit of Angrist and Krueger, 1991 – numbers constructed to prove my specific pedagogical points.)

REGRESSION OUTPUT

Dep. variable: Method:	(1) log(wage) OLS	(2) educ First stage	(3) log(wage) Reduced form	(4) log(wage) 2SLS
Years of education	0.0700*** (0.0035)			0.142** (0.061)
Born Q1 (Jan–Mar)		−0.152** (0.067)	−0.0216** (0.0093)	
Black	−0.236*** (0.018)	−1.47*** (0.12)	−0.342*** (0.022)	−0.133* (0.075)
Married	0.121*** (0.016)	0.38*** (0.11)	0.129*** (0.020)	0.075* (0.043)
Region (South)	−0.098*** (0.015)	−0.54*** (0.10)	−0.137*** (0.018)	−0.060 (0.038)
Constant	5.02*** (0.052)	12.84*** (0.078)	5.93*** (0.062)	4.11*** (0.78)
N	3,010	3,010	3,010	3,010
R^2	0.132	0.087	0.041	—
First-stage F	—	5.15	—	—

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column (3) is OLS of log(wage) on Born Q1 and controls. Column (4) instruments for years of education using Born Q1.

QUESTIONS

1. Interpret the OLS coefficient on years of education. Be precise about units and magnitude.
2. Interpret the coefficient on “Born Q1” in the first-stage regression. What does it tell us about the relationship between quarter of birth and years of education?
3. The first-stage F -statistic on the excluded instrument is $F = 5.15$. What does this tell us? Should we be concerned? Why or why not?
4. Interpret the 2SLS coefficient on years of education. How does it compare to the OLS estimate?
5. With a single instrument and the same controls in all regressions, the 2SLS coefficient equals the ratio of the **reduced-form** coefficient on Born Q1 (column 3) to the **first-stage** coefficient on Born Q1 (column 2):

$$\hat{\beta}_1^{2SLS} \approx \frac{\hat{\gamma}_{Q1}^{RF}}{\hat{\pi}_{Q1}^{FS}}.$$

Verify this numerically using the table. Then explain why, when the first-stage coefficient (the denominator) is small, the IV estimate gets amplified — and why a weak first stage is so dangerous.

FOR DISCUSSION/OPTIONAL

6. Given that the OLS estimate (0.070) is *smaller* than the 2SLS estimate (0.142), what does this imply about the direction of bias in OLS? Is this surprising? Why or why not?

Hint: Think about what you would expect if ability bias drives OLS upward. What else could be going on?

7. The IV estimate is a **Local Average Treatment Effect (LATE)**. In this context, who are the *compliers* – the group whose behavior is actually affected by the instrument?