**Mixtape IV Workshop: Coding Lab 1**

This lab will get your IV hands dirty with data from the Angrist and Krueger (1991) quarter-of-birth study. To start, load the *angrist\_krueger.dta* data file into Stata or R. This is a subset of the original study data, for one cohort (men born in 1940, with earnings measured in 1980).

1. Estimate the bivariate statistical relationship between log wages (*lwage*) and completed years of schooling (*educ*) using OLS. Report your coefficient and robust standard error. Visualize this relationship with a simple graph of your choice.

|  |  |
| --- | --- |
| *OLS Estimate:* | *Standard Error:* |

*Graph:*

1. Estimate the returns to schooling using an indicator for individuals being born in the first quarter of the year as an instrument for completed years of schooling (and no other controls). Report your coefficient and robust standard error. What interesting things do you notice vs. the answer in 1?

|  |  |
| --- | --- |
| *Simple 2SLS Estimate:* | *Standard Error:* |

*Interesting Things:*

1. Estimate the average log wages and completed years of schooling for individuals who are and are not born in the first quarter. Check that you can get the 2SLS estimate in 2 manually from these numbers, using the Wald IV formula:

|  |  |  |  |
| --- | --- | --- | --- |
| *2SLS Estimate =* | *E[Y|Z=1]:* | *…minus…* | *E[Y|Z=0]:* |
| *Divided by:* | | |
| *E[D|Z=1]:* | *…minus…* | *E[D|Z=0]:* |

1. Add indicators for being born in the second and third quarter of the year as instruments to your specification in 2. Report your coefficient and robust standard error. What interesting things do you notice this vs. the answer in 1?

|  |  |
| --- | --- |
| *Overidentified 2SLS Estimate:* | *Standard Error:* |

*Interesting Things:*

1. Collapse your data into means of log wages and completed years of schooling by quarter of birth. Plot average log wages against average years of schooling. What is the slope of this relationship? Are you surprised? Explain what we’ve shown here

*Plot:*

*Explanation:*

1. Let’s put the 2S in 2SLS. First add to your overidentified specification in 4 indicators for an individual’s year-of-birth as controls. Report your coefficient and robust standard error. Now obtain exactly the same coefficient estimate in two steps, where the second step involves a regression on OLS fitted values. Comment on the difference in the standard errors and any other 2SLS diagnostics.

|  |  |
| --- | --- |
| *Overidentified+Controlled 2SLS Estimate:* | *Standard Error:* |

*Interesting Things:*

1. Ok, now let’s get crazy. Add to the previous 2SLS specification interactions of the three quarter-of-birth indicators with all of the year-of-birth indicators as instruments (keeping the year-of-birth “main effects” as controls). Report your coefficient and standard error. How do these compare with the coefficients and standard errors in part 1 and 2? Comment on any other 2SLS diagnostics and how they affect how you feel about this estimate of the returns to schooling.

|  |  |
| --- | --- |
| *Overidentified+Controlled 2SLS Estimate:* | *Standard Error:* |

*Interesting Things:*