

Assignment 2: Instrumental variables – Due: March 10th, 2020 in class

Directions: Please turn in your answers **stapled**, typed, and beautifully written with beautiful tables automated using -estout-, -outreg2-, or R equivalent packages. If there are figures, please create **beautiful figures**.

Replicate parts of Angrist and Evans (1998) “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size”

Background and set up

1. Write a brief (1-3 paragraph) summary of Angrist and Evans’ project. What is their question and the theory behind it? What data do they use and what is their research design? What do they find?
2. Create a new directory (separate from the old Hansen RDD one from Assignment 1), labeled appropriately, with the following subdirectories: /Article, /Do, /Data, /Misc, /Figures, /Tables, /Writing. Place your data /data.
3. Create a new do file for this project and store it in the /Do subdirectory. Your do file should have a commented out header stating the name of the file, a description of the file, the file’s authorship, and the date of last time you worked on the do file.
4. Include in the do file the following lines:
 - a. clear
 - b. and “cd ...” which will change working directory to /do.
5. Download pums80.dta from canvas and store in the /data subdirectory.

Estimation

1. Replicate the 1980 PUMS column of Table 1 for the variables listed. Remember to create a beautiful table appropriately labeled so that I know exactly how to interpret the table without reading their paper or anything else you’ve written.
2. Replicate the 1980 PUMS column of Table 2 for the variables listed.
3. Create a single table like Cunningham and Finlay’s methamphetamine and foster care Table 3 showing the following analysis.
 - a. Note that your treatment variable is morekids.¹ There are four outcomes: worked,² weeksm1,³ hourswm,⁴ incomem.⁵ Your instrument is samesex.⁶
 - b. Present OLS estimates in the first column, covariate adjusted IV regression models in the second column, manual 2SLS in the third column, 2SLS based on - ivregress 2sls- or some R equivalent in the fourth column. There should be a total of two tables. The first table conducts the following analysis for worked and weeksm1 as outcomes. The second table conducts the following analysis for hourswm and incomem as the outcomes. Label all of this cleanly and beautifully so that a reader can easily interpret your table.

¹ Dummy equaling one if the mother had more than two kids.

² Did mom work for pay?

³ Moms weeks worked in 1979

⁴ Hours of work per week in 1979

⁵ Labor income per week in 1979

⁶ First two kids are same sex

Assignment 2: Instrumental variables – Due: March 10th, 2020 in class

- i. Column 1: OLS estimates of outcomes onto treatment with heteroskedastic robust standard errors in parentheses.
 - ii. Column 2: Covariate adjusted IV regression (CIV). This is the ratio of the coefficient on the instrument in the reduced form regression to the first stage regression. Do not calculate a standard error – present only the point estimate.
 - iii. Column 3: Manually construct the two stage least squares estimate (M2SLS). In the first stage, regress the endogenous treatment variable onto the instrument and report it at the bottom like Cunningham and Finlay Table 3. Then using the fitted values from that regression, estimate a second stage regression of the outcome onto the fitted treatment variable from the first stage. Use heteroskedastic robust standard errors for both stages
 - iv. Column 4: Automated 2SLS using `-ivregress 2sls-` in Stata or some R equivalent (2SLS). The standard errors are “correct” here for heteroskedastic robust adjustments.
 - v. In your writeup, interpret all of these results. Explain what you found in easy to understand English.
4. Write a discussion section that explains your results. Your discussion should, using beautiful scientific rhetoric and exposition, include answers to the following questions.
 - a. What is your main finding per labor market outcome under 2SLS?
 - b. Why was OLS biased and inconsistent?
 - c. How does IV solve this inconsistency problem? What assumptions are required for this?
 - d. How does 2SLS compare with OLS in terms of sign and magnitude?
 - e. Compare your standard errors and point estimate with 2SLS in column 4 to manual 2SLS you did in column 3. Compare point estimates of 2SLS with covariate adjusted IV (column 2).