Instrumental Variable Seminar Exercise

October 12, 2018

Introduction

In this seminar exercise, you will be exploring replication data for a notable IV paper, called "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." In this paper, the authors (Josh Angrist and Bill Evans) are interested in exploring how the number of children effect the labor supply of parents.

1. Why do you think that Angrist and Evans need an IV strategy? Why can't they get causal estimates by just running OLS regression of number of children on parental labor supply? Could they control for any ommitted variable bias by adding controls?

2. First, run OLS regressions to explore the relationship between number of children and labor supply.

- Perform three regressions, corresponding to the following dependent variables:
  - Whether the mother is employed (*workedind_moth)*
  - The numbers of hours worked by the mother (*hourswked_moth*)
  - The total income of the mother (*totalinc_moth*)
- Use the following explanatory variables:
  - whether or not the family had 3 or more kids (*morekids*)
  - If the first and second children were boys (*boy1st* and *boy2nd*)
  - Indicators for race/ethnicity of the mother: *black_mother*, *hisp_moth*, *othrace_moth*
- For both OLS and IV regression, run the regression only for the subset of individual for which the "main sample" indicator (*Main*) is equal to 1.
  - This corresponds to mothers aged 21-35, with at least two kids older than 4, and who were at least 15 when giving birth to their first child.

3. To avoid the endogeneity issue of total number of children, Angrist and Evans instrument for whether families have three or more children using whether or not the first two children were of the same sex (*samesex*).

- Why might the first two children being of the same sex serve as a valid instrument for having more than than 2 children? Consider both relevance and exogeneity.
- Repeat the regressions from step 2 now using *samesex* as an instrument.
  - How do the estimates change between OLS and IV? And the standard errors?
  - Evaluate the instrument relevance and weakness using the **ivreg2** output.
  - Can the test of overidentifying restrictions tell us anything about the instrument exogeneity in this case?

4. Repeat the analysis from (3), instead using as instruments both whether or not the first two children were boys (*boys2*) and whether or not the first two chidren were girls (*girls2*).

- What can you conclude from the test of overidentifying restrictions?