Economics 423 Summer 2020
Intermediate Econometrics

Computer Project #3

Instructions:
-You must submit this to soriano@umd.edu by 11:59 PM by the due date specified on the most recent version of the syllabus.
-A proper submission includes a log file (or whatever the equivalent is in R) in a text or PDF format. I should not have to open Stata, R, or Python to see your code and output.
-For each line of code, include a brief comment that explains what you are intending to do with that line. (For example: sum y x z *Summarize variables I am using). This is your best way to receive partial credit if you make a mistake. This is especially true for people using R or Python, since I do not use R or Python myself.
-Comment your answers to the questions at the beginning or the end of the log. Keep them all in one place so I do not need to search the output for each answer.
-Please see my solution Computer Project #1 for guidance on formatting.
-At the beginning of the file, specify the one person that you worked with (if anyone).
-Please note: I reserve the right to ask anyone to stay in office hours to explain their answers if I have reason to believe their work in not their own.

Background and Purpose

The purpose of this project is to estimate the impact of children on female labor supply. In the process, you will learn how to estimate econometric models using instrumental variables/two-stage least squares.

Women with children work less outside the home than women without children. In models where labor supply (either weeks worked or a dummy indicating whether the woman is in the labor force) is regressed on number of children in the household, the coefficient on children is usually negative and statistically significant. This does not mean, however, that additional children actually cause the drop in labor supply. It is more likely that the decision to work is jointly determined with the decision to have children. Thus, unmeasured factors that affect labor supply will be correlated with number of children. (Number of children is endogenous.)

To obtain a consistent estimate of the impact of children on labor supply, some authors have suggested using whether a mother had twins on her first birth as an instrument for the number of children in the household. Twins are in many respects random, and by definition increase the number of children in the household. The rarity of twin births, however, requires constructing a dataset in which mothers with twins are over-represented relative to their occurrence in the population.

The data for this problem were constructed from the 1980 Census Public Use Micro sample by Bill Evans (former UMD prof). The dataset consists of over 12,000 mothers aged 21-40 with at least one child. 6,000 of these mothers have twins; the rest do not. The data are located on ELMS. There are 11 variables in the dataset:

agem "mom’s age at time of census"
agefst "mom’s age when she first gave birth"
race "1=white, 2=black, 3=other race"
educm "mom’s years of education"
married "=1 if mom currently married"
kids "# of kids ever born to mom"
boy1st "=1 if 1st kid a boy"
twin1st "=1 if first birth is a twin"
weeks "weeks worked in 1979"
worked "=1 if worked at all in 1979"
lincome "mom’s labor income, 1979"

Tasks

Points for each item are for correct commands and output and written answers to questions.

1. Open the dataset and look at it. (In STATA, use the "describe" command and "browse." In other software, display a list of variables in the data with information like their type of variable [e.g. string, int, double, etc])

2. (10) Construct an indicator that equals 1 if a woman has a second child. Call this SECOND. What fraction of women have a second child?

3. (10) What fraction of women work? What is the average weeks worked among women who work? What is the median labor earnings for women who work?

4. (10) Run an OLS regression in which weeks worked (WEEKS) is regressed on SECOND. What is the coefficient on SECOND in this regression?

5. (10) Now use twins on the first birth (Z) as an instrument for SECOND. To check that it is an appropriate instrument, regress SECOND on Z. What do you conclude? What is R^2_{xz}? Why is the coefficient on TWIN1ST < 1?

6. (10) When you use Z as an instrument for SECOND, what happens to the coefficient of SECOND in the labor supply equation? How does its standard error compare to what you obtained in 4.? Explain why this occurred.

7. (10) Test whether SECOND is really endogenous. Report and interpret the test result.

8. (10) To conduct the test of over-identifying restrictions add BOY1ST as a second instrument. Report and interpret the results of the test.

9. (10) Now estimate a model via OLS in which more variables are used to explain WEEKS. Add EDUCM, AGEFST, MARRIED and dummies for Black and Other races, in addition to SECOND. What is the impact of SECOND in this equation? How does it compare with 4.?

10. (10) Contrast the result in 9. with estimating the equation via 2SLS with TWIN1ST used as an instrument for SECOND. How do the results compare? Check that TWIN1ST is a valid instrument by running the first stage regression as well.

11. (10) If a colleague asked you what is the impact of a second child on the number of weeks that a woman works, what would you tell him or her? Why?