Computer assignment 2b – Econometrics I – due Sep 12, 2017, 8.45am

THIS IS A GROUP EXERCISE. PLEASE SUBMIT PDF THROUGH BLACKBOARD TIMELY. PLEASE INCLUDE YOUR NAMES.

Goal of the assignment

In the previous computer assignment you learned how to estimate a treatment effect using the instrumental variable approach. Now you are on your own. Your task is to apply the approach and to interpret the results.

Setting

The data are from Angrist and Evans (1998).[1] Both the article and the data are downloadable from Blackboard. The authors address the question how having an additional child affects labor supply. They focus on women's labor supply. Obviously, a randomized trial is out of the question, but there are ways of isolating exogenous variation in the number of children in observational data. The data file contains data from the 1980 U.S. Census on 254,654 women between the age of 21 and 35.

Some relevant variables are:

- `weeksm1`= mom's weeks worked
- `morekids` = 1 if mom had more than two children and 0 otherwise
- `samesex` = 1 if first two children have the same sex (boy-boy or girl-girl) and 0 otherwise
- `boy1st` = 1 if first child was a boy and 0 otherwise
- `boy2nd` = 1 if second child was a boy and 0 otherwise
- `agem1` = age of mom at Census
- `black` = 1 if mom is black and 0 otherwise
- `hispan` = 1 if mom is Hispanic and 0 otherwise
- `othrace` = 1 if mom is not black, Hispanic or white and 0 otherwise

What to submit

What you should submit is one PDF file containing the points we ask for below and a copy and paste of your STATA do file (do not attach a separate .do file)

(1) First explore your data. This is important to get a feeling for the data. Look at a histogram of number of weeks worked, which varies from 0 to 52. Make a bar graph denoting weeks worked by the variable indicating whether a mom has two or more children. If you are able to install additional

[1] Joshua Angrist and William Evans, 1998, Children and Their Parents' Labor Supply. Evidence from Exogenous Variation in Family Size, *American Economic Review*, 88 (3), 450-477. This assignment is inspired by an exercise in Stock and Watson's textbook 'Introduction to Econometrics'.

programs into STATA, then a plot of the cumulative distribution function using the STATA command `distplot` would be instructive as well[i]:

```
ssc install distplot
distplot scatter weeksm1, by(morekids)
```

(2) Run a 'naïve' regression that relates the number of weeks worked to the variable indicating whether a mom has two or more children. Describe what you find. Interpret the estimated coefficient in terms that everybody can understand.

(3) Explain why the OLS regression estimated above is inappropriate for estimating the causal effect of fertility on women's labor supply. Provide a 'story' that revolves around an omitted variable.

(4) Are couples whose first two children are of the same sex more likely to have a third child? Run a regression to find this out. Remember that the data set contains a variable that indicates whether the first two children are of the same sex. Is the effect large? Is it statistically significant?

(5) Explain why this same-sex-indicator is a valid instrument for the instrumental variable regression of weeks worked on having two or more children. Discuss both the independence assumption and the exclusion restriction.

(6) Estimate the regression of weeks worked on having two or more children using the same-sex-indicator as an instrumental variable. How large is the fertility effect on women's labor supply?

(7) Do the estimation results change when you include the variables age of mother and the indicators of ethnic origin in the (instrumented) labor supply regression? Explain why or why not.

---

[i] A message from the TA on not being able to install programs on university computers: This is because Stata wants to write into directories where it does not have permissions. You can bypass it as follows:

```
sysdir set PLUS "U:\stata\plus"
sysdir set PERSONAL "U:\stata\personal"
```

where `U:\stata\plus` and `U:\stata\personal` need to be replaced by the path of your own home directory on the university network drive. The path will certainly include your user id (u123456 or something similar). You can find the path by browsing through your home directory using Windows explorer (or whatever the name for file explorer is in the current versions of Windows).

You can provide any directory you want, it doesn't need to be `stata\plus` or `stata\personal`.