Computer assignment 3b – Econometrics I – due September 19, 8.45 am, 2017

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 regression discontinuity approach. Now you are on your own – although we will provide some guidance. Your task is to apply the approach and to interpret the results.

Setting

The data for this assignment are from Lalive and Zweimüller 2009, downloadable from Blackboard.[1] They look into the causal relationship between parental leave policies and fertility and labor supply of women in Austria. The paper review should give you a good idea of the paper.

The data contains outcomes for women with a first child born around the treatment cutoff date of July 1, 1990. All variables are relative to the birthday of the first child (`bd`). The dummy variable `july` is 1 if that birthday is July 1 or later and zero otherwise. Please note that the cutoff point is at the value of `11139`. In STATA, 11139 stands for July 1, 1990. This is the number of days since January 1, 1960 (with January 1, 1960 being day 0).

The outcomes include whether the women had a second child within three years (`uncb3`), within ten years (`uncb10`), and whether the women returned to work within three years (`uncj3`) and within ten years (`uncj10`).

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) Explore your data. This is important for getting a feeling for the data.

(a) How long is the time window around July 1 in the data?

(b) How many women are included in the dataset?

(c) What is the average number of women included in the dataset per birthday?

To help you with the last point, do the following: generate a variable with the number of observations per birthday:

```
bys bd: gen number=_N
```

[1] R. Lalive and J. Zweimüller, 2009, How Does Parental Leave Affect Fertility and Return-to-Work? Evidence from Two Natural Experiments, *Quarterly Journal of Economics*, 124 (3), 1363-1402.

You already know the STATA command to get the mean of this variable.

To get an idea of the variation in the number of observations per birthday:

```
graph twoway (scatter number bd)
```

(d) Of all women, what percentage had an additional birth within three years?

(e) Of all women, what percentage returned to work within three years?

(f) Of all women, what percentage is between 30 and 34 years of age?

(2) Produce graphical evidence of a treatment effect.

(a) Produce a graph that shows whether take up of parental leave in the second year (`ikar4`) is affected by the policy of extending parental leave from 12 to 24 months. Use the STATA command `cmogram` (or follow some other way to produce similar output). Use the option `cutright` to include border observations on the right side of the split graph. Compare your output with Table II on page 1379.

(b) Do the same thing for months of employment within the three years after delivery of the first child: `pbexp3`. Do the treatment and control group behave very differently? Compare your output with Table V on page 1389.

(3) Conduct a visual check on the smoothness of all other variables.

Check whether the following characteristics of women follow a smooth path around the July 1 cutoff: `indWholesale`, `unEmpl`, `laborEarnings`.

(4) Estimate the treatment effect.

(a) Estimate the treatment effect on, first fertility: having an additional child within three years (`uncb3`), and second, on short run labor market performance: returning to work within three years (`uncj3`). First, write out the regression equations in mathematical terms, including the subscripts. Then, run two simple regressions for these two outcomes. To run these regressions, you should replace the variable birthday (`bd`) that you used so far by a treatment dummy (`july`).

What is the size of the estimated effect? Is the estimated coefficient statistically significantly different from zero? Compare your regression output with Table IV on page 1384 and with Table V on page 1389.

(b) How did the policy affect the long run labor market performance of these women? Check this by using the variables `pbexp10` (months of employment later on) and `pbinc_tot10` (monthly earnings later on).

(5) Further check on the validity of the approach.

One of the important assumptions underlying the regression discontinuity approach is that subjects cannot manipulate treatment assignment. In this context: the timing of birth of the first child is independent from the policy change. Given the ongoing political discussions about the policy up until three months before July 1, 1990, the date of conception of the child is an unlikely source of manipulation (see page 1375). But delivery of the baby can be fine-tuned to July 1 or later by planned caesarean sections, and, less likely, postponing induced births. Check this assumption visually. First create a variable `bandwidth` as follows:

```
gen day=day(bd)
gen month=month(bd)
gen bandwidth=day-1 if month==7
replace bandwidth=day-31 if month==6
```

Use the command `cmogram` to check for a drop in births before July 1 and a spike after that. As usual, birthday (`bd`) is on the horizontal axis, now the variable `number` that you created before should be on the vertical axis. Right before the comma that starts the options of the command (so after `bd`) add the following condition: `if bandwidth>-10&bandwidth<10`. What does the graph suggest? Be very specific in your answer. What could this imply for the type of women that end up in the control group and in the treatment group? And could this bias the estimated treatment effect? Next, run a regression, just as you did earlier, to see whether the difference is statistically significant for this variable `number`. Make sure you add the condition `if bandwidth>-10&bandwidth<10` at the end of the `regres` command.