Computer assignment 4a – Econometrics I – due September 26, 8.45 am, 2017

THIS IS A GROUP EXERCISE. PLEASE SUBMIT PDF THROUGH BLACKBOARD TIMELY. PART II AND III OF THE ASSIGNMENT ONLY. PDF SHOULD SHOW COMMANDS + OUTPUT + (IF ASKED) EXPLANATION, NO DO-FILES. PLEASE INCLUDE YOUR NAMES.

Goal of the assignment

In this computer assignment you are asked to test the effect of a treatment using the difference-in-differences approach. It is based on a study by Card and Krueger, a link to the article is available on Blackboard.[1, 2] The authors study the effects of an increase in the minimum wage on employment. More specifically, they look at the labor market effects of a raise in the minimum wage in 1992 in New Jersey (NJ) from $4.25 to $5.05 per hour, while keeping eastern Pennsylvania (PA) with a constant minimum wage of $4.25 as a control. The study concentrates on fast-food restaurants, which were sampled one month before and eight months after the policy change. Information was collected at each restaurant about variables such as the number of employees, product prices, and store hours.

I. Preparing the data file for analysis: creating your do file

(a) Open STATA and then open the STATA do file editor. Give a short description in the first line of the do file, for instance:

```
* Computer Assignment 4 Econometrics, Sep 2017
```

then save the do file under a name or your choosing.

The next line in the do file should tell STATA to open the data from the folder where you saved your data, for instance:

```
use "C:\Users\Nick\Data\minwage_280915.dta", clear
```

Run this first command of your do file (using the appropriate path) as follows: select the line of code and then click on:  (Execute selection). This loads the data into memory.

II. Descriptive statistics

Before you run any regressions, you should first conduct exploratory data analysis. This gives you the opportunity to check your data for errors, and also to get a feeling for the data (every paper includes a section 'descriptive statistics').

[1] D. Card and A.B. Krueger, 1994, Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review*, 84 (4), 772-793.

[2] This computer assignment is inspired by Steve Pischke's "Problems to go with Mastering Metrics".

The data are based on two interview waves, represented by the variable `time` {0,1}. Outcome measures are: (1) `fte` refers to full time equivalent employment; (2) `wage_st` reflects the starting wage.

The dummy variable `sample` is 1 if both wage and employment data are available in the first and second interview wave, and 0 otherwise. You should conduct your analysis for those data for which the variable `sample` is equal to 1. In that way, you compare the same set of restaurants before and after the policy change.

(a) What is the number of restaurants within the relevant sample and what is the share of the restaurants located in New Jersey? (note that the variable `state` refers to the state (1 if NJ; 0 if PA))

(b) What is the minimum and maximum number of full time equivalents in restaurants within the relevant sample?

(c) What is the minimum and maximum starting wage in restaurants within the relevant sample?

III. Test of treatment effect

Now you are asked to conduct a test for the presence of a treatment effect.

(a) First, show graphically what happened in the two states.

1. Change in wages. Compute the means by state (`mean_wage_st`) as follows:

```
bys state time: egen mean_wage_st=mean(wage_st) if sample==1
```

Then create a line graph showing the change in wages in both states over time:

```
graph twoway (line mean_wage_st time if state==1) (line
mean_wage_st time if state==0), yscale(range(0)) ylabel(0(1)6)
```

The first option scales the y axis down to 0; the second sets the range of the labels of the y axis from 0 to 6.

2. Do the same thing for the change in employment (`fte`), by creating the variable `mean_fte`, etc. Do not forget the condition relating to the sample. Obviously, you will have to adjust the label of the y axis since you are now dealing with a variable with different values.

(b) Compute the difference-in-differences estimate by hand. Take the following steps:

1. Calculate the average starting wage (`wage_st`) separately for restaurants in NJ and in PA, both for each interview wave, as follows:

```
table state time if sample==1, contents(mean wage_st)
```

2. Calculate the difference in the average wages between the second and first interviews.

3. Calculate the difference between NJ and PA of the time differences just obtained.

4. What is the interpretation of such a difference-in-differences estimate of the wage effect? Under what conditions does this provide a valid estimate of the minimum wage increase on wages in the fast food industry?

5. Interpret your finding.

(c) Repeat the same exercise as in (b) for full time equivalent employment. What is the impact of the minimum wage increase on relative employment in NJ restaurants?

(d) Difference-in-difference estimates can also be calculated from the regression:

$$Y_{i,s,t} = \alpha + \delta\,(TREAT_{i,s} * POST_t) + \beta\,TREAT_{i,s} + \gamma\,POST_t + \varepsilon_{i,s,t}$$

where $Y_{i,s,t}$ is employment in restaurant *i* in state *s* and period *t*, TREAT is an indicator for the treatment area (NJ), POST is an indicator for the treatment period, and TREAT * POST is the interaction of these two dummies. δ is the parameter of interest. In the dataset, this interaction term is the variable `state_time`. Again: on the final exam you will be asked to write out an equation like the one above, including all the correct subscripts, the error term, etc. If you fail to include any element, then you receive no credit.

1. What are the regression DD estimates on wages and employment using this regression? How do they compare to the results you found in (b) and (c)?

2. The regression allows you to control for other factors. Repeat the regressions, including, first, a dummy variable for whether the restaurant is company owned as compared to franchised (`co_owned`), and, second, also three dummy variables for three of the four chains in the dataset (Burger King, KFC, Roy Rogers, and Wendy's). For the chain dummies either you construct them from the variable `chain` or you simply include `i.chain` in your regression equation, then STATA will do this for you and automatically drop one of the four (recall why you should always drop one).

3. Do your results change when you include restaurant-specific covariates compared to the results under (d)1? Be specific. Would you have expected the results to change? Explain why or why not.