Computer assignment 4b – Econometrics I – due September 26, 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 difference-in-differences approach. In this assignment, we take the step from two groups and two periods towards two groups and multiple periods.

Setting

The data are taken from a field experiment conducted in the city of Tilburg during December 2014-July 2015. The data are downloadable from Blackboard (`bat_did_2017.dta`).

Treatment

Goal of the treatment was to stimulate households to separate their waste. To that end, households received a warning label on their garbage container if incorrect separation of waste was detected by enforcement officers. In case of repeated wrongdoing, households could get fined. The enforcement campaign ran for four weeks. It was announced in a letter than was sent out three weeks before the start in the campaign. For simplicity, we use one treatment period which is 1 as soon as the first treatment starts and 0 otherwise. Main outcome variable is the weight of residual waste collected.

Data sample

The dataset is a subsample of the overall data generated in the field experiment. The dataset contains averaged outcomes for 10 routes that received the treatment during the weeks included in this dataset and for another group of 10 routes that did not receive the treatment.

The following variables are included:

- `residual_weight`: the weight (in tons) of residual waste collected per week and route, averaged over the treatment group of 10 routes and the control group of 10 routes
- `treatment_group`: indicator variable for the group of routes that were treated
- `treatment_period`: the weeks during which the treatment group was treated
- `calendar_week`: the unit of time

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.

First tell STATA the structure of your data:

```
xtset treatment_group calendar_week
```

(a) What is the average value of the outcome variable for both groups together before the treatment had started in any group (the so-called 'baseline period')? Given that each route contains some 1,050 households, and that each group contains 10 routes, what does this average imply for how much residual waste an average household produces per week?

(b) Is the outcome variable normally distributed? Just look at the shape of the distribution, do not worry about a statistical test for normality.

(2) Graphical evidence of a treatment effect

(a) Is the common trend assumption satisfied? Plot the trend in the outcome variable during the baseline period as follows:

```
graph twoway (line residual_weight calendar_week if treatment_group==0) (line residual_weight calendar_week if treatment_group==1) if treatment_period~=1
```

(b) Compute the difference-in-difference estimate by hand. Actually, STATA can provide most of the means that you need to compute by using the command that you learned in Computer Assignment 4a, part III (b) 1. And what is the treatment effect in percentage terms?

(c) Create a plot of how the outcome variable varies before and after the treatment was administered in both groups:

```
graph twoway (line residual_weight calendar_week if treatment_group==0) (line residual_weight calendar_week if treatment_group==1), xline(23) yscale(range(0)) ylabel(0(2)12)
```

The option `xline(23)` creates the vertical line at the start of the treatment period. Is the plot in line with what you computed under (b)?

(3) Statistical test of a treatment effect

(a) Write out the d-i-d regression equation in mathematical terms, including subscripts etc. Remember that you are dealing with two groups and multiple periods.

(b) Run a regression that produces the difference-in-differences estimator. Generate a treatment dummy as follows:

```
gen treatment_dummy= treatment_group*treatment_period
```

You can include time fixed effects as follows: `i.calendar_week`

Is the estimated effect statistically significant? Let STATA tell you what the estimated effect in percentage terms is.