Econ 191
Replication Exercise

Due November 10th

Replicating a paper is a great way to learn how empirical work is practiced. In this exercise, we ask you to read a paper published in economics, and then we will guide you in replicating the main results. You may use whatever software you like, although programming hints will be provided for Stata only. Many of the commands that we ask you to run in Stata are also available in Stata's dropdown menu. You can use it, but make sure to write each command in a do-file as we have shown in class.

Please hand in, via bCourses, three files: (i) your log file, (ii) your do-file, (iii) and a concise write up of your answers to the questions below.

You are welcome to work in groups of up to 3 but you must each upload your own submission to bcourses and write the names of people you worked with at the top of your write up.

1 Overview of Paper

Let's begin by motivating and reviewing the research design employed in the paper. Around 2-4 sentences is sufficient for each of the questions in this section. Precision and brevity will be rewarded!

1. Imagine you observe cross-sectional data across a city for a given month. In particular, for each district in the city, you observe the average number of police officers on duty each day (police_officers), the total number of car thefts (car_thefts) and a set of district-level characteristics $(X_1, X_2, ..., X_K)$. You run the following OLS regression of car thefts on the number of police officers, controlling for district characteristics.

$$\text{car\_thefts}_d = \alpha + \beta \text{police\_officers}_d + \sum_{k=1}^{K} \gamma_k X_{k,d} + \varepsilon_d$$

Does the estimated $\widehat{\beta}$ capture the causal effect of police presence on car thefts? Why or why not? If $\widehat{\beta}$ is a biased estimate of the true causal effect, in which direction do you think it is biased?

2. Summarize the research design/identification strategy employed in Di Tella & Schargrodsky (2004). What variation in police presence does their estimator use and how does this uncover the causal effect of police presence on car thefts?

3. What is the key identifying assumption in their research design[1]? Describe what this identifying assumption means in this particular setting and evaluate its plausibility.

2 Setup

1. Download the dataset MonthlyPanel2 from Bcourses.

2. Open Stata, set up the directory, and create a new do file to keep track of your steps.

3. Open a log file called replication.txt and load the data MonthlyPanel2.dta.

4. Describe the data to understand the different variables.

3 Data Creation

1. Create a categorical variable named *category* that takes the value of 1 if there is a Jewish institution on the block (variable *distanci* equals 0); the value of 2 if location is one block from the nearest institution (variable *distanci* equals 1); the value of 3 if the location is 2 blocks away; and the value of 4 if it is more than two blocks away from the nearest Jewish institution.

2. Drop the observations for which *mes* takes the value of 72. [2]

---

[1]Hint: The key identifying assumption for all difference-in-differences designs.

[2]The authors use mes 72 and 73 as different periods from the second half of July, but for simplicity we aggregated those into mes 73 for the variable *totrob2*

3. Generate a variable *month* that takes the same values as mes, but takes the value of mes+1 if mes is above 7.

4. Replace the variable month with the value of 8 if the value of *month* is 74. (Hint: Use the tab command to make sure your month variable goes from 4 to 13).

5. Use the label define and label values commands to label the months (4 should correspond to April, .., 7 to July(1-17), 8 to July(18-31), 9 to August, ... 13 to December). (Hint: Use the *label define* and *label values* commands.) [3]

6. Save the database as DataClean.dta.

4 Descriptives[4]

1. Construct a graph that shows the evolution of the average number of car thefts over the months by the distance category you constructed.

   - To do that, it is useful to first collapse the data using the mean of the car thefts (*totrob2*) by month and category.

   - Then, use the *twoway* and *connected* commands to make plot the evolution. Label the categories according to the distance, name the y-axis "Average number of car thefts", the x-axis "Months" and title the graph "Evolution of Average car thefts by category".[5]

   - Export the graph and include it in your write up.

2. Replicate columns (A)-(D) of Table 2 of the paper.

   - Use the database DataClean.dta.

   - Use the estpost and tabstat commands to store the mean and standard deviation of the total number of car thefts by month (*totrob2*), for each one of the categories depending on the distance to the nearest Jewish institution (The variable you constructed). [6]

---

[3]Refer to $https://stats.idre.ucla.edu/stata/modules/labeling-data/$ for help

[4]After each item, there is a suggested way of constructing tables/figures. If If you use a software different from Stata, or prefer using different commands, no need to follow the steps, just replicate the graph and table and answer Q4.3

[5]Hint: We did a very similar think in the Rep. Exercise Tutorial.

[6]i.e: estpost tabstat totrob2 if category==x , statistics(mean sd) by(month) nototal

   est store ax

- After storing, use the esttab command to export the file using a rtf format, naming the columns "No Policy in place". Depending on your version of stata, your code should look like:

  ```
  1) esttab a1 a2 a3 a4 using table2.rtf, cells(Mean(fmt(5)) ///
  SD(par fmt(3))) collabels("No Policy in place") replace label

  2) esttab a1 a2 a3 a4 using table2.rtf, main(mean) aux(sd) ///
  collabels("No Policy in place") nostar replace label
  ```

- Compute the number of blocks and paste it on the table after opening it with Word. (Hint: you can use the summarize command of the block id variable by category restricting to one particular month, or the unique command for the block id variable restricting to a particular category and month).[7]

- Change the name of the columns, and make a title and a footnote to make it more similar to the table on the paper. Include the table in your write up.

3. Looking at the months before the terrorist attack, what can you conclude about the car thefts on the blocks closer/further from the Jewish institutions? Are they comparable? What can you say about the difference after the attack? Explain.

5 Diff-in-Diff

1. Replicate columns (A)-(C) of Table 3 of the paper.[8]

   - Drop observations from July(18-31) (i.e. those for which month == 8)

   - Create a post-treatment time indicator. Namely, generate a variable called *post* that equals 1 if and only if month > 8.

   - Label the variables you just constructed according to the names on Table 3 of the paper.

   - Consider the specification in the first column of Table 3 [9].

---

[7]i.e. unique observ if month==4 & category==y

[8]Again, we present a suggested way to replicate the table, but feel free to use your own software/code.

[9]See p191 of paper.

$$\text{Car Theft }_{it} = \alpha_0 \text{ Same Block Police }_{it} + M_t + F_i + \varepsilon_{it}$$

The code for running and outputting this regression looks like this:

```
gen same_block = (distanci==0)
gen same_block_police = same_block*post

areg totrob same_block_police i.month, absorb(observ) robust
outreg2 using table3.doc, keep(*_block_police) replace lab nocons
```

Use this code block and write the rest of the code to replicate columns (A)-(C) of Table 3, which you can export to a .doc. Include the table in your write up.

2. In column (A) of Table 3 we see that the coefficient on Same-Block Police is -0.07752. What is the interpretation for this?

3. Describe what we learn from Table 4. Why do you think the authors include this table in their paper?