

ECO372H1F Assignment 04

Patrick Blanchenay

Due 2021-04-07, 11.59pm

1 Assignment presentation

This assignment tests your comprehension and application of the Difference-in-Differences estimation strategy by asking to investigate key results of the Agan & Starr (2018) paper on “Ban-the-box” policy, already discussed in GD3 and OQ3. You must submit the answers to the questions in Section 8. You will have to submit three files:

- A PDF document giving the answers to the exercises
- The unique Stata do-file that you used to generate the answers to all questions
- The unique log file produced by Stata when running the do-file

Each of these elements is detailed in Section 5.

2 Warning: plagiarism and academic offenses

This is an individual assignment: do not share you work with anyone, and do not use anyone else’s assignment work in yours. I count on your help to make sure the course is interesting and fair. I am saying this because there have been cases of Academic misconduct in the past; I hate reporting them, it takes some of my valuable time, but I have to in order to maintain fairness and integrity in the course. Some students have faced drastic consequences from their ill-considered decisions, that they regret almost immediately. Let’s avoid all of this. Do your part: play fair and square.

The files you submit will normally be checked in Turnitin for plagiarism at the time of upload. To limit possible infractions, you must cite any sources that you use. Any elements taken from the papers must be cited between quotes (“ ”). You should not need any external sources, but if you do use them, you must cite them adequately.

Any suspicious similarities with other submissions or any existing work will be carefully examined. Since the assignment is worth 10% of your final grade, I have to report any suspicion of academic offense to the Undergraduate Chair.

3 Before you start: Stata How-To

Make sure to read and walk through the following Stata How-To (available on Quercus):

- Stata How-To: OLS regressions (as a reminder), in particular on how to include fixed effects, and how to use interaction terms.
- Stata How-To: Conditions, subsetting data (in particular sections 2, 3, and 4)
- Stata How-To: Exporting Regression Results as formatted tables

4 Before you start: preparing the files

1. Decide, on your computer, which folder will be your working directory for this project. It is recommended that you use a folder with automatic backup such as Dropbox, OneDrive, Google Drive, iCloud Drive.
2. Download the Zip archive `EC0372_Assignment4_2021Apr.zip`, from Quercus in Assignments > Assignment 3, and save it in the working directory.
3. **Extract** the archive `EC0372_Assignment4_2021Apr.zip` in your working directory. This should extract the files and create the following folder structure:

```
[Working directory]
EC0372_Assignment4_2021Apr.pdf
EC0372_Assignment4_SURNAME_FirstName.do
|- datasets
   AganStarr2018.dta
```

4. Rename the do-file template `EC0372_Assignment4_SURNAME_FirstName.do` by replacing "SURNAME" and "FirstName" by your surname and first name, as they appear on ACORN. For instance, mine would be called `EC0372_Assignment4_BLANCHENAY_Patrick.do`. Note that there is a 10 point penalty for failing to name your do-file appropriately.
5. Open the newly-renamed do-file:

- On line 24, set the working directory to the folder where the do-file is on your computer.
- On lines 27 and 30, replace BLANCHENAY by your last name as it appears on ACORN, and Patrick by your first name as it appears on ACORN.
- On line 33, replace 12345678 by your student number.
- Save the do-file before doing any further changes.

5 Documents to upload

5.1 Results PDF document

Filename: EC0372_Assignment4_SURNAME_FirstName.pdf

The Results PDF should be a single document with your answers to all exercises. Most questions will require to perform analyses using Stata and provide suitable explanations and interpretations of the results. You can copy and paste results directly from Stata output window (beware of the keeping the font monospaced). You do not need to write anything in the PDF for questions that simply ask you to generate a new variable.

The answers you provide should only use results that are directly produced by your do-file. Conversely, you should not copy-paste ALL Stata output into your Results document. Only put the parts that are used to answer the questions.

Answers will be graded based on the quality of the explanations. It is not enough to paste Stata output. You have to explain how the output answers the specific question.

The PDF document must be uploaded to both the Quercus assignment by the deadline.

Format

- PDF only. No other file type will be accepted (in particular, no MS Word document).
- Letter-sized. Font should be at least 10 points, everything should be easily readable, including the Stata output.
- Top line of the document should contain : [SURNAME] [First name] - ECO372 Assignment 1
- Second line: Student Nb: [Student Nb]
- Answers should be clearly numbered; do not copy the text of the questions.
- Filename should be: EC0372_Assignment4_SURNAME_FirstName.pdf. For instance, mine would be called EC0372_Assignment4_BLANCHENAY_Patrick.pdf.

5.2 Stata Do-file

Filename: EC0372_Assignment4_SURNAME_FirstName.do

Use the provided do-file template as a starting base.

You can insert your commands in the space indicated in the provided template. Your code should produce all analyses and output necessary for all exercises and questions, from one single do-file.

Your do-file must be able to run in one go if placed on a computer with the same datasets available. The only thing I should need to change in your do-file, to reproduce exactly your results, is to change the working directory. In particular, this requires to keep the do-file in your working directory, and for the /datasets/ folder to be in your working directory. If you're not sure, try on a classmate's computer. If you get error when running your do-file (red lines in Stata output), correct the errors, then re-run the do-file again, until the whole do-file can execute in one pass.

Comment your code. You do not need to comment every instruction, but you should comment the big steps, or the big blocks of code. Explain why you are doing such or such instructions, and what you expect Stata to do. Indentation is also useful to make your code more readable.

Part of your grades depends on code formatting & commenting. You can use the provided do-file as an indication of a well-formatted, well-commented code.

Format

- Only ASCII characters should be used; no accented characters, no characters from extended alphabets or writing systems.
- Filename: EC0372_Assignment4_[SURNAME]_[FirstName].do, e.g. EC0372_Assignment4_BLANCHENAY_Patrick.do.

5.3 Stata log file

Filename: EC0372_Assignment4_SURNAME_FirstName.log

If you followed the steps in Section 4, your log file should be created automatically when you run your do-file. And it will be automatically named EC0372_Assignment4_SURNAME_FirstName.log, where SURNAME and Firstname have been appropriately replaced by your ACORN surname and your ACORN first name. For instance, mine would be called EC0372_Assignment4_BLANCHENAY_Patrick.log. Again, this should happen automatically if you are using the do-file template provided, and if you have configured it appropriately (see step 5 in Section 4).

Anything in your log file should come from your do-file, not from instructions typed in Stata command window. That is, if I re-run your do-file, I should obtain exactly the same log file (apart from the path to the working directory).

If you get error when running your do-file, correct the errors, then re-run the do-file in its entirety to generate an error-free log file.

Format

- Text file only, not in SMCL.
- Filename should be: EC0372_Assignment4_SURNAME_FirstName.log.

6 Submission instructions

By 2021-04-07, 11.59pm, you should have uploaded all three documents on Quercus. Failure to include one or more of these documents, or submit after the deadline will count as late submission.

No submission will be accepted on paper, or by email, regardless of any technological problem.

7 Grading

7.1 Expectations & Grading

The results PDF should be properly spaced, easy to read. Sentences should be grammatically correct, short, and to the point, using arguments from the course, and results from your data analysis. Stata results that directly help answering the questions should be included in the PDF file (either numbers, or whole tables, as appropriate). Contrary to the first assignment, some questions ask you to print a nicely formatted table using **esttab**.

The do-file should be based on the template provided; it must be executable in one pass by only changing the working directory, without generating errors. The code must be commented to indicate the question you are answering, or the type of manipulation/analysis you are doing. The code must be properly indented if necessary.

The assignment is worth 100 points; see the points allocated to each question. In addition, there is 8 points for Code formatting and commenting, and 8 points for PDF formatting.

Questions are graded as a whole, based on the quality of the answers to the questions. Emphasis will be put on clear and concise answers that address specifically the question, and show your understanding of the topic and the statistical issues it raises. Appropriate use of the Stata output in the answer will also be taken into account: use what is necessary, leave out the irrelevant.

All results file will be checked, and all do-files will be run for errors, some log-files will be checked at random.

7.2 Penalties

Note the penalties below, as they can quickly lower your grade:

| Problem | Penalty |
|---|-----------------|
| Late submission (starting immediately at deadline) | 10pts per 24hrs |
| File names do not follow the prescribed pattern <i>[No penalty if Quercus adds a number to the filename in case of multiple submission.]</i> | 5pts per file |
| Do-file generates errors after modifying working directory | 10pts |
| Do-file does not run in one go after modifying working directory | 10pts |
| Log file does not correspond to do-file | 10pts |
| Results are used that are not reproducible with the do-file | 10pts |

8 Questions

Exercise : Agan & Starr (2018)

This exercise uses data from the Agan & Starr (2018) paper, who examined the results of a policy to ban employers from asking Criminal record information to job applicants. The paper demonstrates that in New Jersey and New York City, the policy increased the racial gap in call back between (fictitious) Black and White applicants.

Note: In all the questions there can be minor differences between your results and the paper due to the data cleaning process, but these differences should be minimal.

- a. (7 pts) Load the `AganStarr2018.dta` dataset. For each of these questions, answer in one sentence:
 - (i) What is the unit of observation in the dataset?
 - (ii) How many observations are there?
 - (iii) What variable records an applicant received a positive response to their application? And an interview request?
 - (iv) What proportion of interviews receiving a positive response also received an interview request? (No need to show the Stata output.)
 - (v) What variables record whether an applicant had a criminal record or not, whether they are Black or White applicants, whether they have an employment gap? Whether they hold a GED?
 - (vi) What variable records whether an application took place before or after the ban?
 - (vii) What variable records whether the job application asked about the criminal record of the applicant?
- b. (8 pts) Use **table** to produce a small table of summary statistics about the characteristics of the applicants, similar to panels “Characteristics” and “Results” of Table I in the paper. You may omit the column “Combined”, and everything in the panel entitled “Callback rate by characteristics”. If it makes formatting easier, you may present in two separate tables the numbers corresponding to “Pre-BTB”, and “Post-BTB”.
- c. (8 pts) Are there any differences in the characteristics of the applicants Pre and Post-BTB? (Perform any statistical analyses as necessary.) Why is this the case? And why is this important?
- d. (8 pts) Summarize in one sentence the main finding of Table III column (1) in the paper. Summarize in one sentence the main finding of Table III column (2) in the paper. Re-obtain columns (1) and (2) of table III using judiciously chosen regressions, and use **esttab** to output those results in a single table.
 - Using the **drop()** option from the **esttab** command, make sure the estimates for fixed effects are NOT included in the table.
 - You may omit from your output the lines “Sample”, “Chain FE”, and “Center FE” (although you should take those into considerations when deciding on what regressions to run).
- e. (6 pts) Replicate columns (1) and (2) of table III, but this time showing results about interview requests (as opposed to callback); and use **esttab** to output those results in a single table. (The formatting instructions are the same as in d.)
- f. (8 pts) As in any panel data analysis, it is possible that some stores may be present in one time period, but not in the next, thereby changing the composition of stores between periods. Explain why this would be a problem for the main estimation of the paper.
- g. (3 pts) What variable denotes where an application is sent to a store present in both time periods? How applications were sent to such stores?
- h. (12 pts) Now turning to the “main” results of the paper, replicate column (1), (2), (3) and (5) from Table IV in the paper. Output the regression results to a single table using **esttab**.
 - Using the **drop()** option from the **esttab** command, make sure the estimates for fixed effects are NOT included in the table.
 - You may omit from your output all the rows below “N”, that is: “Controls” . . . “Sample” (although you pay careful attention to those when deciding on what regressions to run).
- i. (10 pts) Replicate column (1), (2), (3) and (5) from Table IV in the paper, but this time about interview requests. (The formatting instructions are the same as in h.)
- j. (6 pts) Given your results in questions e. and i., are the paper’s conclusions similar for interview requests?
- k. (8 pts) Suppose that estimates of interest would have been negative and statistically significant in column (5). How would this have been a problem for the estimation strategy in this paper?