ECO372 Assignment #2

Patrick Blanchenay

Due 2021/03/07, 6PM

1 Assignment presentation

This assignment tests your comprehension of regressions and non-experimental data. You must answer the questions in Section 7. 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 4.

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.

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 more than 10% of your final grade, I have to report any suspicion of academic offense to the Undergraduate Chair.

3 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 `ECO372_Assignment2_2021Mar.zip`, from Quercus in Assignments > Assignment 2, and save it in the working directory.
3. **Extract** the archive `ECO372_Assignment2_2021Mar.zip` in your working directory. This should extract the files and create the following folder structure:

```
[Working directory]
ECO372_Assignment2_2021Mar.pdf
EC0372_Assignment2_SURNAME_FirstName.do
|- papers
    LaLonde_1986.pdf
|- datasets
    LaLonde_experimental.dta
    LaLonde_PSID3.dta
```

4. Rename the do-file template `EC0372_Assignment2_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_Assignment2_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.

4 Documents to upload

4.1 Results PDF document

Filename: `EC0372_Assignment2_SURNAME_FirstName.pdf`

The Results PDF should be a single document with your answers to all questions. Some 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).

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.

There are points allocated to the formatting of the PDF, see section 6.1.

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 2
- Second line: Student Nb: [Student Nb]
- Answers should be clearly numbered; do not copy the text of the questions.
- Filename should be: `EC0372_Assignment2_SURNAME_FirstName.pdf`. For instance, mine would be called `EC0372_Assignment2_BLANCHENAY_Patrick.pdf`.

4.2 Stata Do-file

Filename: `EC0372_Assignment2_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, see section 6.1. 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_Assignment2_[SURNAME]_[FirstName].do`, e.g. `EC0372_Assignment2_BLANCHENAY_Patrick.do`.

4.3 Stata log file

Filename: `EC0372_Assignment2_SURNAME_FirstName.log`

If you followed the steps in Section 3, your log file should be created automatically when you run your do-file. And it will be automatically named `EC0372_Assignment2_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_Assignment2_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 3).

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_Assignment2_SURNAME_FirstName.log`.

5 Submission instructions

By 2021/03/07, 6PM, you should have uploaded all three documents, and only those. Failure to include any of the three required files will count as late submission.

Do not include the datasets in your submission. Do not group files in a zip file.

No submission will be accepted on paper, or by email, regardless of any technological problem. Late penalties start applying immediately at deadline; there is no leniency for "my submission is only 1 minute late" requests.

6 Grading

6.1 Expectations

The results PDF should be properly spaced, easy to read, and have a professional appearance. 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).

Note that your explanations are as important as your statistical analyses, and therefore proper English is crucial to get your point across. You will lose points if what you write is unclear. Proper English helps clarity.

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.

6.2 Rubric

The assignment is worth 100 points, graded according to the following rubric.

Item	Points	Item	Points
Question a	10	Question g	4
Question b	6	Question h	8
Question c	8	Question i	10
Question d	5	Question j	7
Question e	8	Question k	6
Question f	6	Question l	8
Code formatting & commenting	8	PDF formatting	6

For exercise questions, you will be graded on the quality of the answers to the questions. Each exercise subquestion a, b, c, ... is graded as a whole. 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. Some do-files and log-files will be checked at random.

6.3 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	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

7 Questions

Exercise 1: LaLonde (1986) This exercise uses data from LaLonde (1986), who examined the results of the National Supported Work Demonstration (NSW), a temporary employment program in the US, designed to help disadvantaged workers lacking basic job skills move into the labour market by giving them work experience and counselling in a sheltered environment.[1] Unlike other federally sponsored employment and training programs at the time, the NSW program assigned qualified applicants to training positions randomly. We are considering here NSW male participants only.

Load the `LaLonde_experimental.dta` dataset. Examine the variable list. Most of them should be self-explanatory. Variables `re75` and `re78` denote revenues in 1975 (pre-experiment) and in 1978 (post-experiment). All the others are measured pre-experiment.

a. Do individuals who were assigned into the training and those who were not differ in their pre-experiment observables characteristics? Use regressions and perform the relevant statistical analyses to conclude. Why is this important?

b. In Table 5 of the paper, looking at the row Controls, locate the number $886 in column (4). Write the regression model you would run to re-obtain that number. Then run that regression. How do your results compare with that of the original paper?

c. In Table 5 of the paper, looking at the row Controls, locate the number $798 in column (5). Write the regression model you would run to re-obtain that number. Then run that regression (you may need to create additional variables). How do your results compare with that of the original paper?

d. Re-run the same regression you did in Question c., this time using classic instead of robust standard errors. Did the original paper use classic or robust standard errors?

e. Under what assumptions would differences in labour market outcomes, as measured by differences in their 1978 revenues between those who received training and those who didn't, identify the causal effect of the training program? Are they likely to be satisfied in the NSW program? Explain.

f. Is the difference in 1978 revenues between the two groups significant at the 5% level? What do you conclude about the effectiveness of the training program?

The point of LaLonde (1986) article was to show that estimates from the experimental design of NSW might differ from estimates drawn from non-experimental (observational) data. To do this, he dropped the "control" group in the original experimental data, and created various alternative control groups drawn from non-experimental data (namely from the Current Population Survey, and Panel Study of Income Dynamics). Dataset `LaLonde_PSID3.dta` is constructed by replacing the original control group with a control group drawn from the PSID.

g. For NSW male participants, how is the PSID-3 control group defined?

h. In this "quasi-observational" data, are there differences in observable characteristics between treatment and control groups, that could affect labour market outcomes?

i. What is the estimate of the effect of the training program on 1978 revenues? Does it differ from the estimate drawn from the experimental design above? Explain why or why not.

A researcher argues that if treatment and control groups differ in their pre-experiment earnings, we should look at their growth in revenue between 1975 and 1978. Construct a new variable that capture revenue change (in dollars) between 1975 and 1978.

j. Based on question h., which group do you expect to experience the strongest wage change in the absence of a treatment effect?

k. Write the regression model you would run to re-obtain the estimate from column (7); then run the regression.

l. In the light of questions h. and j., how do you interpret the difference between the estimates in column (7) between the "Controls" row and the "PSID-3" row?

[1] LaLonde, R. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *The American Economic Review*, 76(4), 604-620.