ECO372H1F Assignment 03

Patrick Blanchenay

Due 2021-03-23, 6pm

1 Assignment presentation

This assignment tests your comprehension and application of the Instrumental Variables approach. 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
- Stata How-To: Instrumental Variables using 2SLS
- 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 `ECO372_Assignment3_2020Nov.zip`, from Quercus in Assignments > Assignment 3, and save it in the working directory.
3. **Extract** the archive `ECO372_Assignment3_2020Nov.zip` in your working directory. This should extract the files and create the following folder structure:

```
[Working directory]
ECO372_Assignment3_2020Nov.pdf
EC0372_Assignment3_SURNAME_FirstName.do
|- datasets
    Angrist_etal_Columbia2002_1.dta
    Angrist_etal_Columbia2002_2.dta
|- papers
    Angrist_etal_2002_ColumbiaVouchers.pdf
```

4. Rename the do-file template `EC0372_Assignment3_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_Assignment3_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_Assignment3_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_Assignment3_SURNAME_FirstName.pdf`. For instance, mine would be called `EC0372_Assignment3_BLANCHENAY_Patrick.pdf`.

5.2 Stata Do-file

Filename: `EC0372_Assignment3_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_Assignment3_[SURNAME]_[FirstName].do`, e.g. `EC0372_Assignment3_BLANCHENAY_Patrick.do`.

5.3 Stata log file

Filename: `EC0372_Assignment3_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_Assignment3_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_Assignment3_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_Assignment3_SURNAME_FirstName.log`.

6 Submission instructions

By 2021-03-23, 6pm, 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 10 points for Code formatting and commenting, and 10 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	5pts per file
[No penalty if Quercus adds a number to the filename in case of multiple submission.]	
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 : Angrist *et al.* (2002)

In the mid-90s, Colombia distributed vouchers which partially covered the cost of private secondary school for students who maintained satisfactory academic progress. Because of oversubscription to the programme, the allocation of vouchers was randomized using lotteries. Lottery winners received the voucher, while lottery losers did not. In their 2002 article,[1] Joshua Angrist and co-authors examine the effect of these vouchers on several educational outcomes.

A subsample of participants took standardized tests in math, reading, and writing.

a. (4 pts) Load the `Angrist_etal_Columbia2002_1.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 whether a student was allocated a voucher?
   (iv) What variable records their math, reading, writing score, and total score?

b. (4 pts) For each of the math, reading, and writing score, use Stata to find the sample average and standard deviation? What does that tell you about these variables?

c. (10 pts) Using carefully chosen regressions, reproduce the results of column (1) in Table 5 Panel A (p. 1546), using robust standard errors. Pay attention to footnote 13 in the paper. Use **`esttab`** to export these regressions results to a single table; you can put the control variables in the **`drop()`** option so that they do not appear in the table. Compare the estimates you obtain to those in the tables? Do you find large differences?

d. (6 pts) Now reproduce the estimate in column (2) in Table 5 Panel A, for Total points only. Pay attention to footnote 13 when choosing the regression to run. Use **`esttab`** to export these regressions results to a single table, again **`drop()`**-ing the control variables from the table. Do you find a large difference with the estimate from the paper? Can you suggest why *your* estimate could be so different?

e. (8 pts) Provide a full interpretation of the number 0.217 in column (1) in Table 5 Panel A (on p. 1546); be clear on causality and magnitude. Is this estimate interesting/meaningful?

We now focus on schooling length: are these vouchers an effective way of making students stay in school? More precisely: what is the highest grade that students achieve? Are students still in school by the time of the survey (3 years later)? We want to know whether those things differ based on voucher use. Contrary to test scores above, data on schooling length is available for more participants, so we now use a larger dataset. This dataset is still a subsample of the study, this time limited to voucher distributed in Bogota in 1995. Note that the data does not exactly match that of the paper, so results that you will obtain will only approximate those of the papers.

f. (4 pts) Load the `Angrist_etal_Columbia2002_2.dta` dataset. For each of these questions, answer in one sentence:

   (i) What is the unit of observation in the dataset, and how many observations are there?
   (ii) What variable records whether a student was allocated a voucher?
   (iii) What variable records whether a student used the voucher?
   (iv) What variable records the highest grade achieved by the student, and whether they are still in school by the time of the survey?

The authors estimate

$$\text{Outcome}_i = \alpha + \beta \cdot \text{WonVoucher}_i + \varepsilon_i \tag{1}$$

using OLS, where WonVoucher_i is a dummy variable equal to 1 if the student was allocated a voucher, and 0 otherwise. The outcomes of interest are 1/ the highest grade achieved by the student, and 2/ whether they are still in school at the time of the survey. Both outcomes indicate longer schooling.

g. (8 pts) For each of the two outcomes, estimate Equation (1), controlling for

   - whether the survey was old/new;
   - whether the survey was conducted in person or not;
   - the respondent's age, and sex;
   - strata fixed effects (nb: this is not simply controlling for `strata` number);
   - month of survey fixed effects (nb: this is not simply controlling for `month` number).

   You do not need to control for city, year of application, or phone access. Using **`esttab`**, create a single table containing the OLS estimates for both dependent variables. Drop the estimates of the control variables by using the **`drop()`** option[2] when

[1] Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment ." *American Economic Review*, 92(5): 1535-1558.

[2] Fixed effects based on variable `var` and used in a regression via **`i.var`** will be called 1.var, 2.var, etc. So they can be dropped with `*.var` in the **`drop()`** option.

using **esttab**.
Compare your estimates and standard errors to those in Table 3 of the paper; in each case, be clear about the column and line you are comparing to.

h. (6 pts) Not all students allocated vouchers actually used the vouchers. Does estimating

$$\text{Outcome}_i = \alpha + \beta \cdot \text{UsedVoucher}_i + \varepsilon_i \tag{2}$$

estimate the causal effect of actually using the voucher? Explain why or why not.

i. (10 pts) The authors decide to instrument voucher use by voucher allocation. Write out the causal chain they have in mind. Critically examine whether the 3 requirements of an IV approach are satisfied.

j. (8 pts) Assume the outcome of interest is the highest grade completed. Estimate Equation (2), first using OLS, then using voucher allocation as an instrument for voucher use. You should control for the same variables as in question g.. Using **esttab**, create a table containing both the OLS and the IV estimates. You should drop the estimates of the control variables by using the **drop()** option when using **esttab**. Compare your estimates and standard errors to those in Table 7; be specific about which column and line of Table 7 you are comparing to.

k. (6 pts) Is there an effect of school vouchers on the highest grade achieved?

l. (6 pts) Is there an effect of school vouchers on the likelihood of being in school by the time of the survey?