

Economics 20: Econometrics

Group project: overview

Mauricio Fernández-Duque

Dartmouth College

Fall 2025

Group project

- ▶ You will conduct an original analysis using one of the identification strategies we have seen in class
- ▶ You will write up your findings in the style of a (short) economics research paper, with some group and some individual components
- ▶ The goal is to figure out what causally-identified questions can be answered using a non-experimental dataset
- ▶ You will be given a dataset and will know the underlying data-generating process; your task is to use that information to design a credible non-experimental analysis
- ▶ Data: Records from a Mexican NGO
- ▶ Final product will be due at 8pm on Monday, November 17

Step 0: form a group

- ▶ I will let you form groups, subject to my oversight/approval
- ▶ Should have 3-4 students each
 - ▶ This project will involve a lot of Stata coding, so good to make sure that at least someone is fairly comfortable with it
 - ▶ It is good to have a discussion early on about what role(s) each group member will play
 - ▶ In general, group members should be in the same section (10A vs. 2A)
 - ▶ If you have a group of 3, I may need to add one more student
- ▶ Please email me the names of who is in your proposed group by the end of the day on Friday 10/31
- ▶ Please reach out to me (the sooner the better) if you don't have a group (or a big-enough group) and would like me to be a matchmaker

Step 1: Identify your research question and identification strategy

- ▶ First, formulate a causal research question using one of the identification strategies covered in class
- ▶ Your grade depends on the strength of your identification strategy
- ▶ Possible approaches: Multivariate regression, Difference-in-Differences, Instrumental Variables, or Regression Discontinuity
- ▶ Clearly state your key variables and justify the identification assumptions underlying your chosen strategy (I will provide further details)
- ▶ Due at 11:59pm on Thursday, November 6, along with Step 2
- ▶ You will need to get approval from me

Step 2: Identify the outcome of interest and relevant subgroups for analysis

- ▶ As you define your treatment and control groups, also decide:
 - ▶ What outcome variable you will analyze, and
 - ▶ Which subgroups, if any, your analysis will focus on
- ▶ Focus on groups whose outcomes you are especially interested in or who are more likely to be affected by the treatment
- ▶ I will provide guidance on this, though some trial and error may be involved
- ▶ Due at 11:59pm on Thursday, November 6, along with Step 1
- ▶ You will need to get approval from me

Step 3: Download the datasets for your analysis and produce a new, clean dataset

- ▶ We will look at the data together during class time on Thursday, November 6
- ▶ You may need to make sample limitations and/or construct variables
- ▶ You should have a do-file for data construction
 - ▶ Should do every step from raw dataset to analysis dataset (100% replicable)
- ▶ By the end of class on Tuesday, November 11, you must show that you have downloaded and constructed the dataset you will use for your analysis

Step 4: Preliminary analyses

- ▶ Your analysis should include the following elements:
 - ▶ Table of summary statistics
 - ▶ Plots depending on your identification strategy (I will provide further details)
 - ▶ Regression tables for the chosen identification strategy
- ▶ You should have a do-file devoted to analysis
 - ▶ Input can be the *output* from the data construction do-file, or it can be all one do-file
 - ▶ Should also be 100% replicable: no interactive Stata for final results
- ▶ Due date: 11:59pm on Friday, November 14 (by email)

Step 5: Write up your findings

- ▶ The final write-up will include the following elements:
 1. Introduction (state research question, motivation for examining your outcome and subgroup(s))
 2. Data (Explain data source, sample, and construction of variables)
 3. Graphs and tables (numbered so that you can easily refer to them)
 4. Empirical Strategy (identification strategy, regression equation, assumptions)
 5. Results (Present and discuss your findings)
 6. Conclusion
 7. Statement on AI use (At most half-a-page explaining how you used AI in your work)
 8. Do- and log files for data construction and analysis
- ▶ First three sections: about 3–4 pages of text (excluding graphs and tables)
- ▶ Last three sections: about 3–4 pages as well, though you may need more space
- ▶ Due date: 8pm on Monday, November 17 by email (one submission per group)

Step 6: Individual quiz on your project

- ▶ On Tuesday, November 18, there will be an in-class written quiz about your project
- ▶ The questions will focus on what your project did and how to interpret the findings
- ▶ You will not have your paper in front of you during the quiz
- ▶ The goal is not to test specific details, but your understanding of what you did, why you did it, what you found, and what your results mean

Additional notes

- ▶ I will be providing information and support on carrying out specific parts of this project
- ▶ We will devote several class meetings to “OH in class” and group project work

Recommendations on group work

- ▶ Use Dropbox (or something similar) to share files; can have separate 'cd' command for each user in do-file
- ▶ With separate data construction and data analysis do-files, can separate tasks more easily
- ▶ Make use of group size: check each others' work, or both do the same task as a check (within-group replication – I do this for research projects); can also add robustness checks
 - ▶ If you are going to specialize, agree on how beforehand
 - ▶ E.g. one person could write do-files, another could write the intro, another the data section
- ▶ Early on, make a plan on how you are going to work together; useful to write it down
 - ▶ You may want to designate a “group leader” who organizes work and is the one to turn it in
 - ▶ You are encouraged to include a note in the group writeup explaining broadly what each person did – many natural sciences journals do this
- ▶ You are encouraged to discuss the empirical strategy and results with others in your group, even though these will be tested individually