R & Stata for Econometric Analysis Syllabus

Erick F. Molina

Spring, 2023

1 Purpose

The purpose of this course is to prepare you to handle econometric tasks, which are the final stage of the long data journey that starts with data collection. For this, the most important aspect is not to teach Econometric theory, but to teach how Econometrics is done in practice. During the course, you will learn to program in Stata and R (more Stata than R). Additionally, you will learn to manage the data flow of a research project.

2 Rules

Communication is the most important thing. Attendance to this course is mandatory. You may register class hours as part of your weekly hours. You are welcome to ask me any questions whenever you want (within reasonable hours).

3 Structure

The course will follow the syllabus presented below. During the class, we will work through examples. There will be a homework assignment after each class, covering the material discussed. For this, each of you will have your own working folder. The deadline for submitting assignments will be two days before the next class to allow time for grading. Homework will also count as reportable working hours.

4 Mini Project (tentative)

As a form of "evaluation" an individual mini-project will be completed. The project will consist of a question that I will provide for you to answer, with no single correct answer. The goal is for you to use your creativity and go through the research process. The project is expected to take around 15-20 hours, which you can report in your work hours. If the work reflects more than 15 hours, a bonus may be granted at Joyce's discretion. More information will be provided as soon as it becomes available.

5 Syllabus

1. The structure of a research project folder.
    - 1.1 Raw data.
    - 1.2 Processed data.
    - 1.3 Scripts.
    - 1.4 Final data.
    - 1.5 Results: graphs, tables, and numbers.
2. STATA.
    - 2.1 Introduction to STATA.
        - 2.1.1 Installation and login.
        - 2.1.2 STATA windows.
            - 2.1.2.1 Do-file editor.
            - 2.1.2.2 Console.
            - 2.1.2.3 Viewer.
            - 2.1.2.4 Data editor.
            - 2.1.2.5 Help search.
        - 2.1.3 Basic command syntax.
        - 2.1.4 STATA structures.
            - 2.1.4.1 Databases.
            - 2.1.4.2 Number types.
            - 2.1.4.3 Matrices.
            - 2.1.4.4 Temp files.
    - 2.2 Basic commands.
        - 2.2.1 Local and global variables: creating a directory.
        - 2.2.2 Arithmetic operations.
        - 2.2.3 Loading datasets.
        - 2.2.4 Exporting data.
    - 2.3 Data cleaning.
        - 2.3.1 Logical operators.
        - 2.3.2 Generating columns (gen, egen).
        - 2.3.3 Replacing values (replace).
        - 2.3.4 Data organization (sort, bysort, order).
        - 2.3.5 Keeping and dropping variables (keep, drop).
        - 2.3.6 Column names and labels (rename, label).
        - 2.3.7 Formatting variables (destring, tostring, dates).
        - 2.3.8 Preserve and restore.

- 2.4 Data combination.
  - 2.4.1 Append.
  - 2.4.2 Merge.
- 2.5 Descriptive statistics.
  - 2.5.1 Count.
  - 2.5.2 Summarize and tabulate.
  - 2.5.3 Percentiles.
  - 2.5.4 Correlate.
  - 2.5.5 Linear regressions.
- 2.6 Graphs.
  - 2.6.1 Histograms.
  - 2.6.2 Scatter plots.
  - 2.6.3 Graph formatting.
- 2.7 Loops.
  - 2.7.1 foreach.
  - 2.7.2 forvalues.

3. LaTeX.
   - 3.1 Introduction to LaTeX (Overleaf, MiTex, etc.).
   - 3.2 Language.
   - 3.3 Document creation.

4. R.
   - 4.1 Introduction to R.
     - 4.1.1 Installation and login.
     - 4.1.2 R windows.
     - 4.1.3 RProjects.
     - 4.1.4 Basic command syntax.
     - 4.1.5 Pipelines.
   - 4.2 Structures in R.
     - 4.2.1 Values.
     - 4.2.2 Vectors.
     - 4.2.3 Matrices.
     - 4.2.4 Data frames.