

ECON 167: Econometrics with Linear Algebra

General Information

Professor: Augusto Gonzalez-Bonorino
Room: CA Room 110 (Carnegie Building)
Times: Mondays from 06:50 to 09:50 PM
Email: agxa2023@pomona.edu
Office hours: Tuesday & Thursday 11 AM - 1 PM @ Carnegie 217
Resources: [GitHub repository](#)

Helpful pre-reqs: Calculus, Linear Algebra, Statistics, and Data Analysis

Books (recommended):

1. Mostly Harmless Econometrics: An Empiricist's Companion - Joshua D. Angrist and Jorn-Steffen Pischke (conceptual with good data-driven takes)
2. Econometric Analysis - William H. Greene (the canonical reference)
3. A Guide to Econometrics - Peter Kennedy (data sciency flavor)

Overview

This course is about foundations of econometric theory, particularly applied to cross-sectional dataset, with a computational touch. My aim is to introduce all topics in the regular curricula from three perspectives: 1) Conceptual (i.e., understanding the underlying ideas), 2) Mathematical (i.e., doing the math), and 3) Simulated (i.e., testing implications of the theory with R).

There is a computational requirement, but not on what programming language you use. In-class coding sessions will be conducted in R, but I am comfortable with Stata and Python as well. Feel free to use any of these for your homework submissions. Given recent changes in skills demanded by the labor market, I recommend either Python or R since they are more general programming languages. Stata is convenient, but limited (and proprietary). Moreover, translating from one language to another has been made trivial with the advent of code large language models. I expect you to leverage them, but it is your responsibility to ensure understanding of the concepts underlying the code.

The semester can be, broadly, organized into three parts:

1. **Basic Statistics and Core Econometrics:** We will briefly review essential concepts of statistics and unconstrained optimization. These techniques will be applied to derive the OLS estimator and dive deeper into the Linear Regression model.

2. **Performance Metrics and Inference:** OLS is every economist's favorite tool, but its applicability and performance hinders on key assumptions. The second part of the course will focus on testing the Gauss Markov Assumptions (GMA) with data, performance metrics, and proper statistical inference.
3. **Addressing Common Biases:** Like any instrument, our regressions are subject to biases. Part II gives you the tools to identify potential biases, so Part III will teach you some techniques to correct these whenever possible. We will leverage computer simulations to test the sensibility of our models to minor changes in the parameters. If time permits, the course will conclude with an introduction to Instrumental Variables (IV) estimation and some comments on time-series econometrics.

Expect this to be a hands-on class. There are no exams in this class but homeworks will be long and challenging. Your grading will be a weighted average of the quizzes, homeworks, and final project. Each quiz is meant to test your conceptual recall and mathematical skills, these will be 25 minutes long each. Refer to homework sheets for more information about specific assignments, these will allow you to develop your mathematical and programming skills. In general, each assignment will have two parts: One for the math and one for the coding. The final project is the most important component of the class, you should see it as an opportunity to build something for your portfolio that you are proud to show your peers, job interviewers, or other faculty. It can be a theory or empirical paper, we will discuss the difference between these two in more detail towards the end of the semester.

Learning Outcomes

1. Concepts of Empirical Analysis
2. Data science and Data visualization
3. Ordinary Least Squares (OLS) Estimation
4. Uni- and Multi-variate Linear Regression
5. Gauss-Markov Assumptions (GMAs)
6. Goodness of Fit Measures
7. Hypothesis Testing and Statistical Inference
8. Model Selection

Evaluation

Quizzes will be held every two weeks, starting on Week 4 (February 10), at the beginning of class, and no calculators will be needed (but you can use one if you want). The quizzes will test you on concepts or math (*never programming!*) of topics covered during the weeks prior to the quiz. Each class accumulates on the concepts of previous ones, so the quizzes are designed to help you stay on track and apply skills learnt along the way. The topic and date for each quiz is:

Quiz	Topic	Date
1	Basic Stats I	February 10
2	OLS estimator	February 24
3	Basic Stats II	March 10
4	Hypothesis Testing	March 24
5	Model Selection	April 7
6	Heteroskedasticity	April 21

All assignments are due by 11:59 PM on Sunday. You will have 3 weeks for each assignment. You must format your assignments in the following manner. A PDF of the mathematical exercises, ideally generated with LaTeX, and a code file, ideally a notebook (.ipynb or .rmd), of the programming exercises. I am aware that this may be your first time using LaTeX. You are free to type in your math in a word or google document, but I highly recommend you spend the time to pick it up. You will need it for graduate school, writing academic papers, and your work if you follow a research role. Please get set up in <https://www.overleaf.com/> before the first day of class.

Grading

6 quizzes $\sim 18\%$

3 homeworks $\sim 30\%$

Final Project $\sim 45\%$

Participation $\sim 7\%$

The following letter grade scheme applies for the final grade weighted average calculation:

A	≥ 93
A-	≥ 90
B+	≥ 85
B	≥ 80
C+	≥ 75
C	≥ 70
D+	≥ 65
D	≥ 60
F	< 60