

Introduction to Applied Empirical Methods

Paul Goldsmith-Pinkham

January 14, 2025

Assumed Background

- This class is targeted towards researchers interested in empirical work
- The course assumes exposure to Ph.D.-level econometrics material already
 - E.g. first year sequence here at Yale
- This is not because the material is deeply technical, but because I want to be able to assume some basic fluency in statistical concepts

Requirements

- There are no exams, I will be assigning problem sets on a (quasi) weekly basis.
- Since there is a very large number of students, I will be grading the problem sets automatically. I will be requiring the code in R so I can check the output automatically.
 - See syllabus for details
- There are no required readings, but the papers listed in the syllabus are relevant to the material we will cover in class. I will post lecture notes after class on the material along with the class slides.
- I also highly recommend the following texts:
 - Angrist and Pischke, Mostly Harmless Econometrics
 - Hansen, Econometrics
 - Miller and Aronow, Foundations of Agnostic Statistics
 - Cunningham, Causal Inference: The Mixtape, <https://mixtape.scunning.com/>
 - Imbens and Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences

Important caveat

- In the end, this is a graduate course targetted at making you a good empirical researcher
- My goal is to exposed to a wide range of empirical methods, and understand how they connect.
 - We will not drill down deeply into some material
 - I am happy to discuss it more outside of class
- I will also emphasize how to communicate the econometrics underlying your research ideas
 - This includes good graphic design!

Important caveats (Part II)

- The definition of being a good empirical researcher has expanded significantly
 - Not just about methods
 - Also about reproducibility and replicability
- Cost to do good coding and good coding practice has reduced *dramatically* thanks to LLMs
 - These LLMs will be useful as inputs to research, but also as technologies to improve your research
 - I would encourage you use Copilot in VS Code (or similar coding assistant) as well as various LLMs
 - I will *expect* that you can write good code now – the bar has been raised because gravity is lower now!

https://paulgp.com/2024/06/24/llm_talk.html



Structure of the course

- Six parts, first three are “structure” (12 lectures), second three parts are on different “bespoke” methods (12 lectures)
- We will begin with an overview on the structure for causal inference
- N.B. I am keeping everything on the github repo and will update you via Canvas notifications!

Introductions!

