SOSC 5340: Econometric Approaches to Social Science Research

Spring 2024

Lecture Time: Monday 09:00 - 11:50

Lecture Room: 5508 (Lift 25 and 26)

This version prepared on Jan 30, 2024

	Instructor	Teaching Assistant
	ZHANG, Han	**Li, Wei**
Office	Room 2379, Academic Building, Lift 15	Room 3001, Academic Building, Lift 4
Email	zhangh@ust.hk	wlick@connect.ust.hk
Office Hour	Monday 16:00 - 17:00	TBD

Prerequisite

This course is the second of the two-semester graduate-level applied social science statistics sequence (after SOSC 5090). It is designed heavily toward training students for academic research.

- Students are expected to be familiar with the materials covered in SOSC 5090 or its equivalent.
- Students should also be able to use at least one statistical programming language. We will use R in lecture and tutorials. All course contents and exercises can be implemented in other languages such as Stata, but we won't answer your detailed questions about how to implement something in Stata.

Goal

Upon finishing the course, students should be able to:

1. Understand the core math concepts of statistical estimation and inference
2. Be familiar with applied regression modeling
3. Be able to articulate the challenges in establishing causal arguments in social sciences research, and understand some of the statistical methods used to address these challenges.
4. Build a solid, reproducible research pipeline to go from raw data to the final paper.
5. Gain hands-on experience of writing and presenting a research paper that aims to eventually be publishable in an academic journal.

Course Outline (Tentative)

Week	Date	Topic
1	*[2024-02-05 Mon]*	Review of Statistical Inference
2	*[2024-02-19 Mon]*	Linear regression
3	*[2024-02-26 Mon]*	Logistic/probit
4	*[2024-03-04 Mon]*	Generalized linear models
5	*[2024-03-11 Mon]*	Causal inference: counterfactual framework
6	*[2024-03-18 Mon]*	Causal inference in experiments
7	*[2024-03-25 Mon]*	Causal inference in observational studies
8	*[2024-04-08 Mon]*	Diff-in-diff and fixed effect model
9	*[2024-04-15 Mon]*	Instrumental variables
10	*[2024-04-22 Mon]*	Regression discontinuity
11	*[2024-04-29 Mon]*	Fuzzy regression discontinuity
12	*[2024-05-06 Mon]*	Presentation

Grading Policy

Your score will be accessed based on the following five components (no mid-term and final exams):

Attendance	10%
Assignments	30%
Presentation of a published research (15 min)	10%
Presentation of your final paper (20 min)	15%
Write-up of your final paper	35%

1. Homework assignment: short coding homework to make sure that you know how to run models we covered in the lectures. Our TA will hold 4-5 tutorial sections to teach you how to run these models before assignments.
    - All assignments are due at the beginning of the class. You can turn it in online through Canvas.
2. Presentation of a published article: during the semester, you will select one article from a list of articles provided by the instructor and present it to the class. Each student has to choose a different article. This exercise prepares you a chance to practice giving a presentation at an academic conference.
3. Final paper:
    - Preliminary draft of the final paper: double spaced, 4-6 pages. Describe the background, hypotheses, data, and methods you plan to use. We will give you feedback
    - Presentation of your final paper: Your presentation will be on **May 6**, the last day of the class.
    - The final paper is due three weeks after the final class, on **May 27, 11:59 PM**. Double spaced; at least 20 pages including everything, i.e., title, abstract, texts, tables, figures, and references.

Late submission policy

Late delivery of due items will be marked down 75% if received within 1 day of the due date, and 50% if received within 3 days of the due date; you will receive zero credit if the due item is not delivered within 3 days of the due date. Contact the instructor if there are rare unforeseen circumstances.

Re-grade policy

If you want to dispute a grade, please submit your argument in writing along with your assignment. We will evaluate the merit of your argument as well as perform a full reassessment of your entire assignment. This means that your grade may end up lower than it was originally.

Course Materials

There is no single text book. Some of my lecture contents are drawn from the three books below:

- Angrist, J. D. & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. It's available for free download at here.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press. My weekly schedule is most similar to this book, but examples are not. It can be freely viewed at `https://mixtape.scunning.com/index.html`
- Aronow, P. M. & Miller, B. T. (2019). *Foundations of Agnostic Statistics*. New York: Cambridge University Press. A "modern" interpretation of statistics without relying on regressions. Particularly useful for statistical foundations.
- Hansen, Bruce (2020). *Econometrics*. Free at the author's website `https://www.ssc.wisc.edu/~bhansen/econometrics/`

Other recommended textbooks if you need to read on a specific topic

- Applied regression modeling:
  - Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: SAGE Publications, Inc, 3 edition. A classical (and traditional) treatment of applied regression modeling.
  - Gelman, A. & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge ; New York: Cambridge University Press, 1 edition edition. More on hierarchical models.
  - Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning, 6 edition edition. Classical econometric textbook on regressions.
- Econometrics:
  - Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. More advanced topics.
- Causal Inference:
  - Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press, 1 edition edition. A detailed treatment causal inference topics in social sciences.

Programming Resources

We will use R lectures and tutorials. Using Stata or other languages (e.g., Python or Julia) is acceptable. R is freely available for download and runs on Macintosh, Windows, and Linux computers. Students are strongly encouraged to use Rstudio, another freely available software package that has numerous features to make data analysis easier.

Additional resources for learning R:

- Download R: `https://www.r-project.org/`
- Download Rstudio: `https://www.rstudio.com/`
- R syntax:
  - R for beginners. Good overview.
  - Software Carpentry course on R

  - Princeton's Computing for Data Analysis in the Social Sciences workshops

- Using R for regression and causal inference:

  - Princeton R Tutorial. Quick start on using R for regression modeling.
  - UCLA IDRE. Detailed resources for regression modeling with R.

- Visualization:

  - R graphics with ggplot2 workshop notes at Harvard
  - Healy, K. (2018). *Data Visualization: A Practical Introduction.* Princeton, NJ: Princeton University Press, 1 edition edition
  - Andrew Gelman: Lets Practice What We Preach: Turning Tables into Graphs

- Reproducibility:

  - Gentzkow and Shapiro, "Code and Data for the Social Sciences: A Practitioners Guide". Detailed advices on organizing your codes and data to enhance reproducibility.

- If you already know Stata:

  - A Quick Introduction to R (for Stata Users)