ISOM 7000M: IS Doctoral Seminar

Empirical Studies of Innovation and Digitization – Part I Methods

2022 Spring

Ohchan Kwon
ohchankw@ust.hk

Logistics

- Meets Wednesday 7-10pm
- Location: TBD

Overview

The goal of this course is to teach you how to read and write empirical papers that use observational data to make causal claims (“what causes what and why”). As a consumer of these papers, you need to be able to evaluate under what conditions one can interpret some regression results as causal evidence. As a future producer of these papers, you need to be able to identify research opportunities in which you can establish some interesting causal relationships and execute them. Our goal is to develop you as a critical consumer and able producer of these empirical papers.

There are already good econometrics (e.g., SOSC 5090 Quantitative Methods for Social Science Research, ECON 5300 Econometrics) and causal inference courses (e.g., SOSC 5340 Econometric Approaches to Social Science Research II). Students planning to do empirical research should certainly take these courses. However, it has been my experience that these methods courses still leave some important skillsets uncovered. Typical econometrics courses focus more on estimation and inference issues, that is, how to establish relationships among variables, and less on causal identification issues. Causal inference courses discuss identification issues, but because many “classical” examples are from papers published in disciplinary journals a long time ago, students often do not learn some implicit rules and patterns to write a publishable paper in their respective fields as of now.

To fill the gap, this course uses recent papers in the economics of IT/digitization area. We will learn about causal inference techniques by carefully examining and replicating examples from recent papers. Compared to other typical quantitative methods courses, this course will be focusing on a few specific techniques (e.g., difference-in-differences) and discussing those techniques in much greater depth. Moreover, this course tries to cover more practical tools and techniques side of social science research, often described as “hidden curriculum” in doctoral education.

Anyone interested in empirical research and/or digitization is welcome to take this course. We assume your familiarity with undergraduate-level statistics and econometrics. We also assume some familiarity with at least one statistical programming language (e.g., R, Stata, Python). Students without such background can still take the course as long as they are willing to spend considerable time to catch up.

Class Structure

Each class will start by reviewing weekly assignment. A typical weekly assignment will include 1) each student presenting one paper using a specific research method, and 2) discussing coding assignment. Then we move on to learn new concepts and toolkits. We will emphasize intuition and application over mathematical proof. We will first establish key intuition behind each research method, then discuss a few papers examples with code references. We will then go through some etiquettes in communicating the method and results to an audience without overselling.

This course is a seminar, and not a lecture. We learn these techniques only by doing, and thus expect everyone to come to class well prepared. We will be happy to answer any questions that may arise; however, we will not be lecturing on the content of the background readings in detail. Because of the learning-by-doing nature of the course, the workload will likely end up being on the higher side. Therefore, you should plan accordingly.

Attendance is required. Using electronic devices during the class will not be allowed.

Requirements

- **Weekly Assignments (30%)**: There will be weekly homework assignments. The assignments consist of 1) picking up a paper using a specific research method for summary, and 2) some data analysis questions. No late submission will be accepted. We will have a total of seven weekly assignments, but we will drop your lowest grades. We encourage students to work together on the assignments, but you need to write your own response. You must not simply copy and paste someone else's answers or computer code. We also ask that you write the names of your collaborators on your assignments.

- **Participation (20%)**: Students are strongly encouraged to ask questions and actively participate in discussions during lectures.

- **Final Project (50%):** Students will give presentations during the regular class time. Presentations should be approximately 10 minutes in length (determined based on the class size, but time limits will be strictly enforced) and should be similar quality to presentations at major academic conferences. See below for more details.

Final Project

The final project will be a short empirical paper. You are free to choose any question of your interest as long as you are ***asking a clear causal question that you can address using the tools covered in the course***. It is ideal if you use your on-going research projects, for instance with your advisors. You are welcomed to write a replication paper, but the replication paper should extend the original analysis by applying techniques covered in the course. You are also encouraged to collaborate with other students.

You should start thinking about the final project as early as possible, especially if you do not have any on-going research project suitable for causal inference. Please familiarize yourself with all the causal inference toolkits as soon as possible so that you will understand what toolkits you can use. You can also skim some of the empirical papers contained in this syllabus to get some sense. Please be mindful that acquiring new datasets is uncertain and time-consuming. You are strongly encouraged to ask our advice and feedback before deadlines.

At the minimum, the final paper should be around 10 pages in length and look like an empirical journal article minus literature review and theory development. Having said that, I strongly encourage students to use this course as a platform to develop papers for conference submission or degree requirements.

Here is a tentative timeline that you will follow:

- **Week 5:** Turn in **research proposal**. The proposal should contain your research question and its importance ("why should we care?"), methodology, and data sources.

- **Week 9:** Prepare **materials** for the "paper-in-a-day workshop." The materials should include abstract, figures and tables, and the list of references. You should upload the materials to the shared folder at least one day before the workshop.

- **Week 10**: Turn in a **first draft** of your paper after "paper-in-a-day workshop." The draft should contain abstract, introduction, data and methods, results, and all figures and tables. The writing should be relatively polished (e.g., no incomplete sentences). Your draft will be shared to other students so that you will receive feedback from your peers.

- **Week 12**: Present your paper in the class. The maximum time allowed is ten minutes.

- **By the end of the semester**: Turn in the **final version of the paper**.

Statistical Computing

We teach this course in R. R is an open-source statistical computing environment that is increasingly used in social science. Compared to Stata, R is more versatile in that it allows you to collect (e.g., web scraping) and analyze the data at the same time. Compared to Python, R provides more specialized functionalities related to data analysis. You can download it for free from www.r-project.org. For assignments you are welcome to use any statistical programming language. However, we have limited ability to support you using other languages.

We assume some familiarity with R. That said, we are happy to point to R resources and provide advice on how to improve your programming skills. Here are a few resources we recommend:

- R for Data Science: https://r4ds.had.co.nz

- Using R for Introductory Econometrics: http://www.urfie.net/read/index.html

- Introduction to Econometrics with R: https://www.econometrics-with-r.org

Books

Other book chapters and journal articles will be also assigned as required readings. We will provide either the scanned copies or links to electronic versions on Canvas.

- **Required books:** We use the following books to cover key concepts and techniques. We strongly recommend that you purchase these books for your own reading. They are also available in the library.
  - Békés, G. and Kézdi, G., 2021. *Data Analysis for Business, Economics, and Policy*. Cambridge University Press.
  - Cunningham, S., 2020. *Causal Inference: The Mixtape.* Yale University Press. Freely available at https://mixtape.scunning.com.
- **Recommended books:**
  - Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
  - Dunning, T., 2012. *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press.

General Readings

These are the books and writings I found useful when I am trying to develop and pursue research ideas.

- ANGRIST, J.D., BLAU, D.M., Falk, A., Robin, J.M. and Taber, C.R., 2006. How to do empirical economics. *investigaciones económicas*, *30*(2), pp.179-206.
- Claus O. Wilke. How to develop a research question. https://clauswilke.com/blog/2014/06/15/how-to-develop-a-research-question/
- Brynjolfsson, E. and Kahin, B. eds., 2002. *Understanding the digital economy: data, tools, and research*. MIT press.
- Brynjolfsson, E. and Saunders, A., 2009. *Wired for innovation: How information technology is reshaping the economy*. Mit Press.
- McAfee, A. and Brynjolfsson, E., 2017. *Machine, platform, crowd: Harnessing our digital future*. WW Norton & Company.

- Tirole, J., 2017. *Economics for the common good*. Princeton University Press.

- Goldfarb, A. and Tucker, C., 2019. Digital economics. *Journal of Economic Literature*, *57*(1), pp.3-43.

- Schultze, C.L. and Newlon, D.H., 2011. Ten years and beyond: Economists answer NSF's call for long-term research agendas (compendium). *TEN YEARS AND BEYOND: ECONOMISTS ANSWER NSF'S CALL FOR LONG-TERM RESEARCH AGENDAS, Charles L. Schultze and Daniel H. Newlon, eds., American Economic Association*.

Topics and Readings

Required readings are marked with a (*) and are in **bold**.

Week 1: Introduction

Readings:

- **Békés and Kézdi: Chapter 19. (*)**
- **Cunningham: Potential Outcomes Causal Model. (*)**
- Hernan, M. A. 2018. The C-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health* 108(5) 616–619.
- Forman, C., A. Goldfarb, S. Greenstein. 2012. The Internet and Local Wages: A Puzzle. *American Economic Review* **102**(1) 556–575.

Readings: Writing Empirical Papers

- **Imai, Kosuke. "How to Write an Empirical Social Science Paper." (*)**
- Chaubey, Varanya. 2017. *The Little Book of Research Writing*. CreateSpace Independent Publishing Platform.
- **Gelman, Andrew, and Guido Imbens. Why ask why? Forward causal inference and reverse causal questions. No. w19614. National Bureau of Economic Research, 2013. (*)**
- **Little, Andrew T. "Three Templates for Introductions to Political Science." (2016). (*)**
- **Cachon, G.P., 2012. What is interesting in operations management? *Manufacturing & Service Operations Management*, *14*(2), pp.166-169. (*)**
- Goldfarb, A., C. Tucker. 2014. Conducting Research with Quasi-Experiments: A Guide for Marketers. 1–38.

Week 2: Field Experiment

Readings: Theory

- **Békés and Kézdi: Chapter 20. (*)**

- Angrist and Pischke: Chapters 1 and 2.

Readings: Applications

- Bloom, N., J. Liang, J. Roberts, Z. J. Ying. 2015. Does Working from Home Work? Evidence from a Chinese Experiment. *The Quarterly Journal of Economics* 130(1) 165–218.

- Bertrand, M. and Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), pp.991-1013.

- **Edelman, B., M. Luca, D. Svirsky. 2017. Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics* 9(2) 1–22. (*)**

- Cui, R., J. Li, D. J. Zhang. 2019. Reducing Discrimination with Reviews in the Sharing Economy: Evidence from Field Experiments on Airbnb. *Management Science*.

Readings: Managing Empirical Projects

- **Wickham, H., 2014. Tidy data. *Journal of statistical software*, *59*(1), pp.1-23. (*)**

- Gandrud, C., 2018. *Reproducible research with R and RStudio*. Chapman and Hall/CRC.

- **Gentzkow, M. and Shapiro, J.M., 2014. Code and data for the social sciences: A practitioner's guide. *Chicago, IL: University of Chicago*. (*)**

- Knittel, C.R. and Metaxoglou, K., 2018. Working with data: Two empiricists' experience. *Journal of Econometric Methods*, *7*(1).

Programming Requirements

- Complete the Datacamp chapters "Introduction to R" and "Data Manipulation with dplyr." Learn to conduct t-test for means comparison.

Week 3: Matching

Readings: Theory

- **Békés and Kézdi: Chapter 21. (*)**
- Angrist and Pischke: Chapters 3.
- Rubin, D.B., 2008. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, *2*(3), pp.808-840.
- Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* **25**(1) 1–21.
- **Ho, D. E., K. Imai, G. King, E. A. Stuart. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15(3) 199–236. (*)**
- Iacus, S. M., G. King, G. Porro. 2012. Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20 1–24.
- Roberts, M. E., B. M. Stewart, R. A. Nielsen. 2020. Adjusting for Confounding with Text Matching. *American Journal of Political Science* **64**(4) 887–903.

Readings: Applications

- **Gunarathne, P., H. Rui, A. Seidmann. 2021. Racial Bias in Customer Service: Evidence from Twitter. *Information Systems Research*. (*)**

Programming Requirements

- Complete the Datacamp chapters “Introduction to Regression in R” and “Intermediate Regression in R.” Learn to make publication-quality tables using ‘modelsummary’ and other packages.

Week 4: Difference-in-Differences (1)

Readings: Theory

- **Békés and Kézdi: Chapters 22 and 23. (*)**

- Angrist and Pischke: Chapters 5.

- Bertrand, M., E. Duflo, S. Mullainathan. 2004. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics* **119**(1) 249–275.

Readings: Applications

- **Card, D., A. B. Krueger. 1994. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania.** ***American Economic Review*** **84(4) 772–793. (*)**

Programming Requirements

- Complete the Datacamp chapters “Joining Data with dplyr” and “Data Visualization with ggplot2.” Learn to aggregate data by time and unit level, create time-series plots, and conduct regression analysis in the context of context of difference-in-differences.

Week 5: Difference-in-Differences (2)

Readings: Theory

- **Goldfarb, A. and Tucker, C.E., 2014. Conducting research with quasi-experiments: A guide for marketers. *Rotman School of Management Working Paper*, (2420920). (*)**

- Dunning, T., 2012. *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press. Chapters 7 and 8.

Readings: Applications

- **Nagaraj, Abhishek. "Information seeding and knowledge production in online communities: Evidence from OpenStreetMap." *Management Science* (2021). (*)**

- Zhang, L. 2018. Intellectual Property Strategy and the Long Tail: Evidence from the Recorded Music Industry. *Management Science* 64(1) 24–42.

- Sen, A., C. Tucker. 2020. Product Quality and Performance in the Internet Age: Evidence from Creationist-Friendly Curriculum. *Journal of Marketing Research*.

- Hwang, E. H., S. Lee. 2021. A Nudge to Credible Information as a Countermeasure to Misinformation: Evidence from Twitter. *Available at SSRN*.

Week 6: Event Studies (1)

Readings: Theory

- **Békés and Kézdi: Chapters 24. (*)**
- **Cunningham: Difference-in-Differences. (*)**
- Athey, S. C., G. W. Imbens. 2022. Design-based analysis in Difference-In-Differences settings with staggered adoption. *Journal of Econometrics* 226(1) 62–79.
- Goodman-Bacon, A. 2019. Difference-in-Differences with Variation in Treatment Timing. 1–48.
- Liu, L., Y. Wang, Y. Xu. 2021. A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *arXiv preprint arXiv:2107.00856*.

Readings: Applications

- **Seamans, R. C., F. Zhu. 2013. Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. *Management Science* 60(2) 476–493. (*)**
- Williams, H. L. 2013. Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy* 121(1) 1–27.

Readings: Platform Contexts

- Park, J., M.-S. Pang, J. Kim, B. Lee. 2021. The Deterrent Effect of Ride-Sharing on Sexual Assault and Investigation of Situational Contingencies. *Information Systems Research*.
- Hasan, S. and Kumar, A., 2019. Digitization and divergence: Online school ratings and segregation in America. *Available at SSRN 3265316*.
- Chan, J., A. Ghose. 2014. Internet's Dirty Secret. *MIS Quarterly* **38**(4) 955–976.
- Liu, J., A. Bharadwaj. 2020. Drug Abuse and the Internet: Evidence from Craigslist. *Management Science* **66**(5) 2040–2049.

Programming Requirements

- Learn to create dynamic treatment plot.

Week 7: Event Studies (2)

Readings: Theory

- Imai, K., I. S. Kim, E. H. Wang. Matching Methods for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*.

- Cengiz, D., A. Dube, A. Lindner, B. Zipperer. 2019. The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics* **134**(3) 1405–1454. **– Appendix D for "Stacked Design".**

- Keele, L., 2010. An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. *White Paper. Columbus, OH*, *1*, p.15.

- Oster, E. 2017. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 40(2) 1–18.

- Cinelli, C., C. Hazlett. 2020. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(1) 39–67.

Readings: Applications

- **Wang, J., G. Li, K.-L. Hui. 2021. Monetary Incentives and Knowledge Spillover: Evidence from a Natural Experiment. *Management Science*. (*)**

- **Pattabhiramaiah, A., E. M. Overby, L. Xu. 2021. Spillovers from Online Engagement: How a Newspaper Subscriber's Activation of Digital Paywall Access Affects Her Retention and Subscription Revenue. *Management Science*. (*)**

- Xu, Y., A. Ghose, B. Xiao. 2021. Mobile payment adoption: An empirical investigation on Alipay. *Available at SSRN 3270523*.

- Aral, S., P. S. Dhillon. 2020. Digital Paywall Design: Implications for Content Demand and Subscriptions. *Management Science*.

- Chu, J., Y. Duan, X. Yang, L. Wang. 2020. The Last Mile Matters: Impact of Dockless Bike Sharing on Subway Housing Price Premium. *Management Science*.

- Dube, S. and Zhu, C., 2021. The Disciplinary Effect of Social Media: Evidence from Firms' Responses to Glassdoor Reviews. *Journal of Accounting Research*, *59*(5), pp.1783-1825.

- Azoulay, P., J. Graff Zivin, G. Manso. 2011. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics* 42(3) 527–554.

Week 8: Instrumental Variables

Readings: Theory

- **Cunningham: Instrumental Variables. (*)**
- Angrist and Pischke: Chapters 4.

Readings: Applications

- **Sun, M., F. Zhu. 2013. Ad Revenue and Content Commercialization: Evidence from Blogs. *Management Science*. (*)**
- **Seiler, S., S. Yao, W. Wang. 2017. Does Online Word of Mouth Increase Demand? (And How?) Evidence from a Natural Experiment. *Marketing Science* 36(6) 838–861. (*)**
- Todri, V. 2021. Frontiers: The Impact of Ad-Blockers on Online Consumer Behavior. *Marketing Science*.

Week 9: Recent Studies on Human-Algorithm Interaction

- Guest lecture by Professor Zhitao Yin.

Readings:

- Logg, J. M., Minson, J.A., & Moore, D.A. (2019). Algorithm Appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103. **(Required reading)**
-
- "Exposure to opposing views on social media can increase political polarization" by Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, Alexander Volfovsky, *Proceedings of the National Academy of Sciences* Sep 2018, 115 (37) 9216-9221. **(Required reading)**

Week 10: "Paper-in-a-Day" Workshop

We will lock the door, turn off our cell phone and internet connection, and spend uninterrupted three hours to write a paper.

http://deevybee.blogspot.com/2019/02/the-paper-in-day-approach.html

- Two weeks before ICIS deadline. April 20…?
- Submission Deadline: May 3, 2022, 11:59PM EDT (New York time)

Before this class.

- Prepare 1) abstract + introduction. 2) tables/figures. 3) list of related literature.
- This will be shared.

After this class:

- Peer comments. (see requirements above.)

Readings:

- Imai.
- Writing like a Robot.
- Tucker and Goldfarb.
- Something about "how to provide constructive feedback"

Week 11: Causal ML

- Guest lecture by Professor Carlos Fernandez-Loria.

Readings: To be announced.

Week 12: Final Presentation

Readings

- **Jesse Shapiro, How to Give an Applied Micro Talk. (*)** **https://www.brown.edu/Research/Shapiro/pdfs/applied_micro_slides.pdf**

- **Shane Greenstein, How to Present a Paper. (*)**