Department of Technology Management and Innovation
Center for Urban Science and Progress

MG-GY 9753 Selected Topics in Management
CUSP-GX 9113E Special Topics in Urban Science & Informatics
"From Correlation to Causation: Data Science for Decision Making"

Fall 2024

Professor: Takahiro Yabe

Contact Details: takahiroyabe@nyu.edu

Office/Hours: Wednesdays 3:30pm – 5pm. Please email instructor to schedule a meeting
370 Jay St., #1308

Class Schedule: Thursdays 2pm – 4:30pm

Course Pre-requisites: None. Knowledge of Python and basic statistics are preferred.

Course Description:

While machine learning models are capable of exploiting correlations within high-dimensional data to perform predictions, uncovering the causal mechanisms — *understanding if and how an intervention X causes an outcome Y* — is vital for informed decision-making in business and policy. This course builds upon a foundation of basic statistics and programming to explore the essential principles of causal inference within data science. It provides practical training in applying causal inference techniques. First, the course will introduce tools for understanding causal structures, including graphical causal models. The course will then cover key methodologies in causal inference, such as propensity score matching, difference-in-differences, synthetic control methods, and more advanced techniques like instrumental variable (IV) estimations and causal machine learning models. This course serves as an introduction to the cutting-edge field of causal inference, with a focus on project-based and hands-on learning approach.

About the Instructor:

Dr. Takahiro 'Taka' Yabe is an Assistant Professor in the Department of Technology Management and Innovation (TMI) and the Center for Urban Science + Progress (CUSP) at the Tandon School of Engineering, New York University. His research develops data science tools and computational models for analyzing large-scale human behavior data to better understand collective social dynamics during disruptions, and to improve the resilience of communities and cities to urban shocks (e.g., disasters, pandemics, and disruptive technology). His research lies in the intersections of urban science, computational social science, and complex systems. Previously, he was a Postdoctoral Associate at the MIT Institute for Data, Systems, and Society (IDSS) and the MIT Media Lab, and a

Data Science Consultant for the World Bank, developing data science pipelines for urban disaster risk management and resilience analytics.

Course Objective:

This course offers an introduction to the field of causal inference, with a focus on project-based and hands-on learning approach. Using basic concepts that students have learned in statistics (e.g., linear regression, logistic regression, and basic machine learning algorithms) and computer coding, the course will introduce methods and principles of causal inference applied to real world problems including decision making in business and policy making for urban challenges.

Upon completion of this course, students will be able to:

- Understand the conceptual foundations related to causal inference.
- Formulate causal inference problems using structural causal models, and apply causal inference techniques to analyze and evaluate critical questions for decision-making.
- Communicate the insights obtained from causal inference effectively for decision making in business and policy scenarios through discussions, reports, and presentations.

Course Structure:

The course is structured into three sections. Part I “Preparation for causal inference” will introduce key concepts and foundational techniques to tackle causal inference problems, including randomized experiments and graphical causal models. Assignment #1 is aimed to help students understand these foundational concepts. Part II “Essential methods for causal inference” will cover the key techniques that are often used to solve causal inference problems. These include propensity score matching, inverse propensity score weighting, difference-in-differences method, synthetic control, structural time series, regression discontinuity design, and instrumental variables. Assignment #2 is are aimed to help the students’ understanding of these methods’ assumptions, strengths, and weaknesses. Finally, Part III “Advanced methods for causal inference” will delve into machine learning and AI based methods to estimate the heterogeneous causal effects.

In parallel to the coursework and homework assignments, students will be assigned a Paper Share assignment and a final project. The aim of the Paper Share assignment is to encourage students to take a deep dive into a study and to provide an opportunity to critically think about real-world causal inference problems. The exercise also aims to provide students with an opportunity to practice explaining complex data science concepts and experimental designs to a broader audience. The goal of the final project is to provide students with real-world experience on a causal inference problem using actual or synthetic data, together with a group of 2-3 students who are ideally from different backgrounds and at different stages of their career.

Readings:

Required Texts:

- [CIP] Matheus Facure, “Causal Inference in Python: Applying Causal Inference in the Tech Industry”, O'Reilly Media, 2023
- [CISP] Judea Pearl, “Causal Inference in Statistics - A Primer”, Wiley, 2016
- + academic papers assigned in each lecture

Optional Texts:

- Judea Pearl and Dan Mackenzie, “The Book of Why: The New Science of Cause and Effect”, Basic Books, 2018

- Aleksander Molak, "Causal Inference and Discovery in Python: Unlock the secrets of modern causal machine learning with DoWhy, EconML, PyTorch and more", Packt Publishing, 2023
- Stephen Morgan and Christopher Winship, "Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)", Cambridge University Press, 2014
- Joshua Angrist and Jorn-Steffen Pischke, "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University Press, 2009

Course Assignments:

Attendance and Quizzes

Attendance is mandatory, and punctuality is expected. Advanced notification of your absence via email is preferred. For any excused and pre-authorized absences, you are required to provide proof of absence (doctor's note, emergency, etc.) to the Office of Student Advocacy. If you cannot make a class, you are still responsible for the material covered. Please review the materials before the next class. In the first 10 – 15 minutes of every class, we will have a short quiz to refresh your memory from the last class.

Assignments

There will be 3 assignments throughout the semester. Each assignment is designed to help students review the concepts covered in the previous weeks.

- Assignment #1 (assigned on Week 4, due on Week 5): Problem set covering randomized experiments, graphical causal models, regression analysis
- Assignment #2 (assigned on Week 6, due on Week 7): Problem set covering matching methods
- Assignment #3 (assigned on Week 8, due on Week 9): Problem set covering causal inference for panel data

Paper Share Presentation

- Students will each select 1 paper from the paper list and give a 10-minute presentation to the class explaining the key points of the research. Students will complete this exercise in pairs.
- The aim of the *Paper Share* is to encourage students to take a deep dive into a study and to provide an opportunity to critically think about real-world causal inference problems. The exercise also aims to provide students with an opportunity to practice explaining complex data science concepts and experimental designs to a broader audience.
- The presentation should clearly communicate the contents of the research paper aimed at a broad audience, by answering the following questions: 1) what was the causal statement that was tested and why was it important? 2) what data was used? 3) what methods were used? 4) what was the result/finding?

Final Project

The goal of the final project is to provide students with real-world experience on a causal inference problem using actual or synthetic data.

- Students will work in groups of 3. Ideally, the team should be a mix of students from different backgrounds and at different stages of their career.
- The final project will be evaluated based on 3 components: Project Idea Presentation (on week 9), the Final Project Presentation (on week 14), and the Project paper (due on week 15).

- Teams are required to book an office hour to consult their proposed research topics with the course instructor between October 15 – 20, 2024 (week 8), at least a week before the Project Idea Presentation.
- Each project shall include data analysis, causal inference, and interpretation for decision making. Teams need to choose a dataset to conduct their analysis on. Datasets used for the project may be open source or proprietary. Here are some open-source dataset repositories:
  - Zenodo: https://zenodo.org/
  - Awesome causality data (GitHub): https://github.com/rguo12/awesome-causality-data
  - Kaggle: https://www.kaggle.com/search?q=causal+inference+in%3Adatasets

Final Exam

The goal of the final exam is to evaluate your understanding of 1) how to formulate causal inference problems, 2) deciding appropriate methods to approach a given problem, and 3) how to evaluate the causal effects. The final exam will be an in-class, open-note written exam. Students are allowed to bring 1 letter-size sheet of paper with their notes.

The final exam will consist of a series of scenario-based questions in business and/or policy settings where you are the data scientist working with colleagues (who have obviously not taken this class and are confusing correlations with causations) to make business and/or policy decisions. You will be responsible for giving expert input and recommendations to the decision maker based on what you have learned in this course to help the team make informed decisions.

Grading Policy:

- Attendance and quizzes: 15%
- Assignments: 15% (5% * 3 assignments)
- *Paper Share* presentation: 10%
- Final project: 40%
  - Idea presentation: 10%
  - Final project presentation: 20%
  - Project paper: 10%
- Final exam: 20%

Letter Grades:

Letter grades for the entire course will be assigned as follows:

Letter Grade	Points	Percent
A	4.00	92.50% and higher
A-	3.67	87.50% – 92.49%
B+	3.33	84.50% – 87.49%
B	3.00	80.00% – 84.49%
B-	2.67	72.50% – 79.99%
C+	2.33	65.50% – 72.49%
C	2.00	60.00% – 65.49%
F	0.00	59.99% and lower

Method of Communication:
NYU Brightspace (https://brightspace.nyu.edu/) is the principal repository for all information, files, lectures, instructions, and submissions. You are expected to log into the course site regularly to learn about any developments related to the course as well as to upload assignments, so please check Brightspace before each class for any updates. For all technical questions about Brightspace, email AskIT@nyu.edu.

Policy on the Use of Generative AI:
Generative AI are software that can generate content in response to user prompting. There are many kinds of generative tools for generating text, audio, images, video, and code (e.g., ChatGPT). Students are discouraged but allowed to use AI in coursework under two circumstances: 1) contributions from any source, including AI sources, must be appropriately quoted and cited every time they are used, and 2) the use does not violate NYU's Academic Integrity Policy, which forbids "submitting work (papers, homework assignments, computer programs, experimental results, artwork, etc.) that was created by another, substantially or in whole, as one's own." Failure to do so constitutes an academic integrity violation, and I will follow the institution's policy to the letter in those instances.

You are responsible for the information submitted based on an AI query (for example, that it does not violate intellectual property laws or contain misinformation or unethical content). Your use of AI tools must be appropriately documented and cited (for example, *in lines 30-50 on page 3, GPT-4 was used to write a sentence about the common usage of IDW Interpolation technique. "Text of your query." Generated using OpenAI. https://chat.openai.com/*). Therefore, generative AI could be used only for exploratory tasks, and then you will guide, verify, and craft the ultimate answers, so please don't just copy and paste without understanding.

NYU Important Dates for Fall 2024:
- Sep 3, 2024: Fall 2024 classes begin
- Sep 16, 2024: Fall 2024 Add/Drop deadline
- Sep 17, 2024: Withdrawal Period Begins. Classes dropped on or after this date result in a "W" grade on the student transcript.
- Oct 14, 2024: Fall Break
- Nov 28 – 29, 2024: Thanksgiving Recess, no classes scheduled.
- Dec 12, 2024: Last day of Fall 2024 classes
- Dec 16 – 20, 2024: Final Exam Period

Course Topic Outline

Class Date	Topic	Readings	Assignments & Exams
Part I. Preparation for causal inference			
Week 1 Sep 5, 2024	Course overview *Why causal inference?* – some motivating examples Course roadmap and logistics	• CIP chapter 1 • CISP sections 1.1,1.2 • Luca, M. (2021) "Leaders: Stop Confusing Correlation with Causation", *Harvard Business Review* • Popper, N. (2022) "Causal Inference: A Guide for	**Pre-course survey (in-class)**

		Policymakers", *Simons Institute White Paper, UC Berkeley* ● Using Causal Inference to Improve the Uber User Experience	
Week 2 Sep 12, 2024	Class discussion on correlation and causation Randomized experiments – the gold standard for causal inference? Recap of statistics	● CIP chapter 2 ● CISP section 1.3 ● Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, *4*, 3895-3962. ● Gallo, A. (2007) "A Refresher on A/B Testing", *Harvard Business Review*	**Quiz #2**
Week 3 Sep 19, 2024	Introduction to graphical causal models – a tool to organize causal mechanisms	● CIP chapter 3 ● CISP sections 1.4 – 2.4	**Quiz #3**
Week 4 Sep 26, 2024	Graphical causal models continued Using linear regression to adjust for confounding bias	● CIP chapter 4 ● CISP chapter 2.1	**Quiz #4** **Assignment #1 assigned**
Part II. Essential methods for causal inference			
Week 5 Oct 3, 2024	Matching methods ● Propensity score matching	● CIP chapter 5 ● O'Shaughnessy, E., Barbose, G., Wiser, R., Forrester, S., & Darghouth, N. (2021). The impact of policies and business models on income equity in rooftop solar adoption. *Nature Energy*, *6*(1), 84-91.	**Quiz #5** **Assignment #1 due**
Week 6 Oct 10, 2024	Matching methods, continued ● Inverse propensity weighting	● CIP chapter 5 ● Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in	**Quiz #6** **Assignment #2 assigned**

	● Doubly robust estimation	observational studies for causal effects." *Biometrika 70*, 1 (1983): 41-55.	
Week 7 Oct 17, 2024	Causal inference using panel (time series) data ● Difference-in-differences method	● CIP chapter 8 ● Gao, J., Jun, B., Pentland, A. S., Zhou, T., & Hidalgo, C. A. (2021). Spillovers across industries and regions in China's regional economic diversification. *Regional Studies*, *55*(7), 1311-1326. ● Leng, Y., Dong, X., Moro, E., & Pentland, A. (2023). Long-range social influence in phone communication networks on offline adoption decisions. *Information Systems Research*.	**Quiz #7** **Assignment #2 due**
Week 8 Oct 24, 2024	Causal inference using panel (time series) data, continued ● Synthetic control ● Structural time series	● CIP chapter 9 ● Yabe, T., Zhang, Y., & Ukkusuri, S. V. (2020). Quantifying the economic impact of disasters on businesses using human mobility data: a Bayesian causal inference approach. *EPJ Data Science*, *9*(1), 36.	**Quiz #8** **Assignment #3 assigned**
Week 9 Oct 31, 2024	**Project Idea Presentations**		**Quiz #9** **Assignment #3 due**
Week 10 Nov 7, 2024	Designing experiments for causal inference ● Regression discontinuity design	● CIP chapter 10 ● García Bulle Bueno, B., Horn, A. L., Bell, B. M., Bahrami, M., Bozkaya, B., Pentland, A., & Moro, E. (2024). Effect of mobile food environments on fast food visits. *Nature Communications*, *15*(1), 2291.	

Week 11 Nov 14, 2024	Using instrumental variables (IVs) to estimate causal effects	● CIP chapter 11 ● Abbiasov, T., Heine, C., Sabouri, S., Salazar-Miranda, A., Santi, P., Glaeser, E., & Ratti, C. (2024). The 15-minute city quantified using human mobility data. *Nature Human Behaviour*, 1-11.	**Quiz #10**
Part III. Advanced methods for causal inference			
Week 12 Nov 21, 2024	Dealing with heterogeneous causal effects ● Conditional average treatment effects (CATE)	● CIP chapter 6	**Quiz #11**
Nov 28, 2024	No classes (Thanksgiving Recess)		
Week 13 Dec 5, 2024	ML for causal inference ● Meta-learners ● Causal forests	● CIP chapter 7 ● Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the National Academy of Sciences 116*, 10 (2019): 4156-4165. ● Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.	**Quiz #12**
Week 14 Dec 12, 2024	**Final Project Presentations**		
Week 15 Dec 19, 2024	**Final Exam**		**Final Project papers due**

This syllabus and schedule will be followed as closely as possible. However, course content and assignments are subject to revision at the instructor's discretion as the course needs to arise. Revisions will be announced in class and on Brightspace.

Academic Integrity:

All students are responsible for understanding and complying with the NYU Statement on Academic Integrity.

Academic Integrity for Students at NYU

This policy sets forth core principles and standards with respect to academic integrity for students at New York University. Each school at New York University may establish its own detailed supplemental guidelines for academic integrity, consistent with its own culture, and consistent with the University-wide general guidelines described in this document.

At NYU, a commitment to excellence, fairness, honesty, and respect within and outside the classroom is essential to maintaining the integrity of our community. By accepting membership in this community, students take responsibility for demonstrating these values in their own conduct and for recognizing and supporting these values in others. In turn, these values will create a campus climate that encourages the free exchange of ideas, promotes scholarly excellence through active and creative thought, and allows community members to achieve and be recognized for achieving their highest potential.

In pursuing these goals, NYU expects and requires its students to adhere to the highest standards of scholarship, research and academic conduct. Essential to the process of teaching and learning is the periodic assessment of students' academic progress through measures such as papers, examinations, presentations, and other projects. Academic dishonesty compromises the validity of these assessments as well as the relationship of trust within the community. Students who engage in such behavior will be subject to review and the possible imposition of penalties in accordance with the standards, practices, and procedures of NYU and its colleges and schools. Violations may result in failure on a particular assignment, failure in a course, suspension or expulsion from the University, or other penalties.

Faculty are expected to guide students in understanding other people's ideas, in developing and clarifying their own thinking, and in using and conscientiously acknowledging resources - an increasingly complex endeavor given the current environment of widely available and continually emerging electronic resources. In addition, students come to NYU from diverse educational contexts and may have understandings regarding academic expectations that differ from those at NYU. NYU values and respects all academic traditions; however, while at NYU, students are expected to adhere to the norms and standards of academic integrity espoused by the NYU community and will be assessed in accordance with these standards. Students should ask their professors for guidance regarding these standards as well as style guide preferences for citation of sources for assignments in their courses.

Following are examples of behaviors that compromise the academic and intellectual community of NYU. The list is not exhaustive. Students should consult the websites and guidelines of their individual schools for an extended list of examples and for further clarification.

1. Plagiarism: presenting others' work without adequate acknowledgement of its source, as though it were one's own. Plagiarism is a form of fraud. We all stand on the shoulders of others, and we must give credit to the creators of the works that we incorporate into products that we call our own. Some examples of plagiarism:

- a sequence of words incorporated without quotation marks
- an unacknowledged passage paraphrased from another's work
- the use of ideas, sound recordings, computer data or images created by others as though it were one's own

2. Cheating: deceiving a faculty member or other individual who assess student performance into believing that one's mastery of a subject or discipline is greater than it is by a range of dishonest methods, including but not limited to:

- bringing or accessing unauthorized materials during an examination (e.g., notes, books, or other information accessed via cell phones, computers, other technology or any other means)
- providing assistance to acts of academic misconduct/dishonesty (e.g., sharing copies of exams via cell phones, computers, other technology or any other means, allowing others to copy answers on an exam)
- submitting the same or substantially similar work in multiple courses, either in the same semester or in a different semester, without the express approval of all instructors
- submitting work (papers, homework assignments, computer programs, experimental results, artwork, etc.) that was created by another, substantially or in whole, as one's own
- submitting answers on an exam that were obtained from the work of another person or providing answers or assistance to others during an exam when not explicitly permitted by the instructor
- submitting evaluations of group members' work for an assigned group project which misrepresent the work that was performed by another group member
- altering or forging academic documents, including but not limited to admissions materials, academic records, grade reports, add/drop forms, course registration forms, etc.

3. Any behavior that violates the academic policies set forth by the student's NYU School, department, or division.

Moses Center Statement of Disability

If you are student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 2nd floor.