Syllabus for ECON 488 A, Causal Inference, Winter 2023
University of Washington, Department of Economics

Lectures: M/W 5:30-7:20 PM, DEM 126
Discussion Sessions: W 7:30-8:20 PM, DEM 126

Team. **Instructor**: Melissa Tartari (mtartari@uw.edu). **Melissa's Office Hours**: M 7:30-8:20 PM, DEM 126. **Teaching Assistant (TA)**: Resemyur Makan (resmakan@uw.edu). **TA's Office Hours**: TU 4:00-5:00 PM, SAV 319F.

Content, Objectives, and Target Audience. In this course we study statistical methods for causal inference. Causal inference focuses on uncovering, i.e. learning about, causal relationships: the scientist is interested in quantifying the effect of a cause (also called a treatment, an action, or an intervention) on one or more outcome variables of interest. For example, we may want to establish whether: (a) classroom training increases the likelihood of finding employment, for a group of unemployed individuals; (b) a particular type of advertisement increases sales of the advertised product; or (c) tax credits that reduce tax liabilities in proportion to earned income (up to a point) encourage tax payers to work/earn more.

Methods of causal inference are widely used both in academia and industry to *ex-post* assess the effect of policy interventions, where the term policy is broadly understood to include any intervention of interest by public or private agents, or by nature. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021 was awarded to three scholars, Joshua Angrist, Guido Imbens, and David Card for their contributions to the analysis of causal relationships and empirical labor economics. In this course we learn many of the inferential methods that they pioneered and read the foundational articles.

The methods covered are: treatment/control simple difference in averages, regression adjustment, matching (exact, nearest-neighbor, caliper, block, with regression adjustment, etc.), difference-in-difference, synthetic controls, instrumental variables and local average treatment effect (LATE), and regression discontinuity design (sharp and fuzzy). At different points during the course we mention how machine learning has recently been used to enrich the classical methods.

By the end of the course, students will be able to approach problems of causal inference that are routinely considered at public and private research institutions and agencies, economic consulting companies, as well as in major technology companies and retailers. This includes the ability to set up, run, and interpret the findings of the methods learned in class. The instructor develops the theory during lectures and guides students to discover results for themselves following a teaching style termed "guided discovery," which is based on a series of hands-on applications.

This course is designed for upper-lever undergraduate, or master-level, students in economics and related fields. This course is a natural companion of ECON 484 "Econometrics and Data Science".

Prerequisites: ECON 482 or equivalent (ECON 382 does not suffice).

Recommended Readings Ahead of the Course: Prior to the start of the quarter, students are encouraged to read two reports made available by the The Royal Swedish Academy of Sciences in conjunction with the 2021 "Nobel Prize in Economics": 1) Popular science background: Natural experiments help answer important questions; and, 2) Scientific Background: Answering causal questions using observational data. Another good pre-class reading is Abadie and Cattaneo (2018)'s article "Econometric Methods for Program Evaluation", Annual Review of Economics, 10:465-503 (Link). It provides an accessible and comprehensive overview of the methods covered during the course. Students are encouraged to read Gentzkow and Shapiro's tutorial "Code and Data for the Social Sciences: A Practitioner's Guide" (Link) – HTML version available here, to learn how to organize their code and their research projects.

Lecture and Discussion Sessions: The instructor holds fifteen 1 hour and 50 minute long lectures with a ten minute break midway. The TA holds eight 50 minute long discussion sessions. Typically, during the discussion sessions the TA goes over problem sets (henceforth, pset) solutions, and gives students hints for upcoming psets. Students should take note of the following salient dates, cancellations, and repurposing:

1st Lecture	W January 4
1st Discussion Session (R Refresher, LaTeX)	W January 4
2nd Discussion Session (LaTeX, LyX, UW-IT Jupyter Notebooks, Markdown)	W January 11
Cancelled Lecture (MLK Day)	M January 16
Midterm Exam (in place of lecture)	W February 8
Cancelled Discussion Session	W February 8
Cancelled Lecture (Presidents Day)	M February 20
Last Discussion Session	W March 1
Last Lecture	M March 6
Final Exam (in place of lecture)	W March 8

Communications and Administration: We use Canvas at `https://canvas.uw.edu/courses/1612452/` as a centralized location for posting announcements, psets, solutions to psets, grades, lecture notes, and tutorials. We maintain a calendar on Canvas ▷ Calendar for all events (lectures, discussion sessions, office hours) and due dates (problem sets, exams). We email students from Canvas exclusively to handle emergencies. Students shall check Canvas regularly.

Online Discussion: We use Piazza at `https://piazza.com/washington/winter2023/econ488` to foster communication and avoid duplications. On Piazza's discussion forum students converse with classmates and the instructor/TA. The forum makes questions asked by some viewable by all - so students won't miss out on interesting topics raised by classmates. Folders are used to filter questions/comments by a particular pset or lecture. Students are strongly encouraged to use this forum *while* they work on psets, as opposed to waiting to ask questions and seek clarification during office hours. Students shall not email the instructor/TA with questions that can be addressed at the Piazza's discussion forum. Students shall address all questions regarding installing software (e.g. R, LyX, LaTeX, see below) to TA and make sure to use her office hours to resolve any technical issue related with using software. Technical/software problems are **not** an accepted excuse for not submitting or submitting with delay solutions to a pset.

Textbooks: The lecture notes are the foundational reference for the course: see Canvas ▷ Files ▷ Lecture_Slides. We recommend adopting this textbook: Scott Cunningham (2021), Causal Inference: The Mixtape, Yale University Press (C21). The online version is available at `http://mixtape.scunning.com/`. The paper version can be purchased on Amazon or from Yale Press. Class lectures will also borrow material and examples from three books: 1) Angrist and Pischke (2009) Mostly Harmless Econometrics, Princeton University Press (AP09); 2) Angrist and Pischke (2014) Mastering Metrics: The Path from Cause to Effect, Princeton University Press (AP14); and, 3) Imbens and Rubin (2015), Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Cambridge University Press (IR15). AP14 requires an introductory knowledge of statistics and probability. AP09 and IR15 require intermediate and respectively advanced-level familiarity with statistics and micro-econometrics.

Course Outline with Textbook References:

The Rubin Causal Model and RCTs	C21 Ch 4, AP09 Ch 2, Lecture Notes
Matching	C21 Ch 5, AP09 Ch 3, Lecture Notes
Difference in Differences	C21 Ch 9, AP09 Ch 5, Lecture Notes
Panel Data Methods	C21 Ch 8, AP09 Ch 5, Lecture Notes
Synthetic Controls	C21 Ch 10, Lecture Notes
IV-2SLS and LATE	C21 Ch 7, AP09 Ch 4, Lecture Notes
Regression Discontinuity	C21 Ch 6, AP09 Ch 6, Lecture Notes
Machine Learning for Causal Inference	Lecture Notes

Statistical Software: Students should have some familiarity with programming in R, e.g. from using it in at least one previous course. R is a language and environment for statistical computing and graphics. Students can download it free of charge from `https://cran.r-project.org` and install on their personal computer. Afterwards, students should install R-Studio Desktop available at `https://rstudio.com/products/rstudio/download/`. An old version of R (R for Windows Version 3.5.1) is available at all student computers at all research libraries (Link). The TA provides a refresher to programming in R during the first week of classes, in place of the weekly discussion session (see salient dates above). The tutorial material is available at Canvas ▷ Files ▷ Lab_Sessions ▷ R_Tutorial. Students who want to catch up on R before the start of the quarter may start off with Emmanuel Paradis's tutorial "R for Beginners" (Link) and Trevor Martin's tutorial "The Undergraduate Guide to R" (Link). For a more hands-on approach, students may choose to take DataCamp's free "R Tutorial" (Link): this site provides interactive lessons that get the user write real code in minutes. Student will then want to progress to the free book "R for Data Science" (Link) to learn more about selected topics such as plotting, data manipulation, and fitting linear regression models. Scripting for psets must abide by the style guidelines available at `http://jef.works/R-style-guide/`.

Journal Articles: Students are asked to read about 4 journal articles which will be made available at Canvas ▷ Files ▷ Literature. In particular students will study (and at times replicate) some of the articles cited in the 2021 Nobel Prize in economics award.

Problem Sets: There are 7 psets, they test and expand upon material covered during lectures. Pset #1 is to be submitted individually. Students are encouraged to submit group-work (2 or 3 students) for psets #2 through #7, one submission per group. Groups may change over time, though it may be more productive if they do not. Group work is allowed to foster cooperation amongst students. Detailed submission instructions are posted at Canvas ▷ Pages. Late submissions are not accepted, no matter the reason. Students who want their pset re-graded must make a request in writing to TA within 1 week upon the return of your pset. Psets are graded based on a 0 to 100 scale. Psets will take quite a bit of your time and are rewarded accordingly: the total weight given to the psets is **30**%. We understand that sometimes emergencies creep up and that a student may not be able to turn in a pset on its due date. For this reason we drop a student's lowest pset score (out of the 7 psets) when calculating the total score. Therefore, the best 6 psets each count for **5**%. If a student fails to submit a pset by the due date it automatically receives a 0 score out of 100. Typewritten answers to psets provide from 0 to 2 points to be added to the course final score as a bonus. Type-writing points are assigned in proportion to how many psets are typewritten out of 7 (e.g., if a student typewrites 6 psets out of 7 they harnesses $2 \times \frac{6}{7} = 1.714286$ bonus type-writing points). Thus, typewriting psets is rewarded but **not** required. Students find detailed instructions concerning how to submit solutions to pset at Canvas ▷ Pages (e.g., which file format to use, how to include the R scripts, and their log files, etc.). Due dates of psets are provided below.

Exams: The midterm is a written exam: it lasts 1 hour and 50 minutes. The final exam is also a written exam: it lasts 2 hours. All exams are closed-book, closed-notes, no calculators, no phones. The TA or the instructor proctors both the midterm and the final exams. All exams are graded based on a 0 to 100 scale. The weights are as follows: the best of the two scores is given a weight of **40**%, the worst of the two scores is given a weight of **30**%. If a student does not take the midterm exam (e.g., because they are sick or have to travel out of the country on that day), the midterm exam's weight is **0**% and the final's weight is **70**%.

Psets and Exam Schedule:

Pset 1	W January 11
Pset 2	W January 18
Pset 3	W January 25
Pset 4	W February 1
MIDTERM	W February 8
Pset 5	W February 15
Pset 6	W February 22
Pset 7	W March 1
FINAL	W March 8

Make up of the Midterm and Final Exams: There is no make-up for the midterm exam. If a student does not take the midterm exam the weight associated with their final exam is **70**%, see above. A word of advice: not taking the midterm exam for reasons other than a sudden illness is highly discouraged and highly risky because the final exam automatically carries a large weight. Requests for a make-up final exam must be accompanied by an email to the instructor and Cced to Ahna Kotila (akotila@uw.edu) that explains the grounds (e.g. health emergency) for the request and must be sent before the day of the exam.

Typewriting: We encourage students to typewrite answers to psets as this leads to much more professionally looking work, pushes students to acquire valuable skills, and is standard expectation for jobs in data science. We recommend any of the following solutions (see Table below as well):

1. write a self-standing LYX document for the text of your solutions;
2. write a self-standing LATEX document for the text of your solutions;
3. combine code, output, and text with the R Markdown package; or,
4. combine code, output, and text in a UW-IT Jupyter Notebook (details at Canvas ▷ Pages).

LYX is an open source front-end for LATEX, which means that you can write complex mathematical expressions *without* the need to learn LATEX. If you choose options 1) or 2) you will also have a separate R script containing your code, and a separate file displaying your output, or a Jupyter notebook containing your code *and* output. If you choose options 3) or 4) you will have everything (code, output, text) in one file. Students can find a tutorial for LYX here. The TA provides a brief tutorial to LYX and LATEX (complete of template .TEX and .LYX files for submitting solutions to psets), Markdown (but not R Markdown), and JupyterHub and Notebook during the first two weeks of classes, at the weekly discussion section (see salient dates above) — we leave out R Markdown because it i too time consuming. The tutorial material is available at Canvas ▷ Files ▷ Lab_Sessions ▷ LyX_LaTeX_Tutorial and Canvas ▷ Files ▷ Lab_Sessions ▷ Jupyter_Tutorial. Creating tables in LATEX can be unpleasant, LATEX Table Generator makes it easier. You must use LATEX to write equations in R Markdown files (tutorial here) and in Jupyter Notebook files (tutorial here).

Solution	**File**	**Handles Text?**	**Handles Text & Code?**	**Up-front Cost**	**Portability of Skill**
LYX document	.lyx	YES	NO	Low	Low
LATEX document	.tex	YES	NO	Medium	High
R Markdown document	.Rmd	YES (LATEX for math)	YES	Medium	Very High
Jupyter Notebook	.ipynb	YES (LATEX for math)	YES	Medium	Very High

Grading: At Canvas ▷ Files ▷ Scores students find files containing statistics for psets and the midterm exam in tabular and graphical form. Both files are updated as new pets are graded. These statistics do not affect a student's grade: a student's grade is solely based on their total score, irrespective of what other students' scores are. That is, there is no grade "curve". A student's weighted sum of the scores in the best 6 psets, the midterm exam, and the final exam, plus the bonus points for type-writing psets produce a *numerical score* that ranges from 0 to 102 (where the 2 extra points come from type-writing psets (1)). The *numerical grade* is based on the following mapping from numerical scores (right) to numerical grades (left):*

*The equivalence scale between numerical grades and letter grades accords with `http://www.washington.edu/students/gencat/front/Grading_Sys.html`.

0.0 (E)	< 30
0.8 (D-)	$[30, 35)$
1.0 (D)	$[35, 40)$
1.3 (D+)	$[40, 45)$
1.7 (C-)	$[45, 50)$
2.0 (C)	$[50, 55)$
2.3 (C+)	$[55, 60)$
2.7 (B-)	$[60, 65)$
3.0 (B)	$[65, 70)$
3.3 (B+)	$[70, 80)$
3.7 (A-)	$[80, 90)$
4.0 (A)	$[90, 102]$

If a student wants to withdraw from the class they have to communicate their intention to do so by the day before the final exam. An Incomplete is granted only for medical reasons as supported by a doctor's note and a justification from the student's counselor.

Students Joining the Course Late, Adding/Dropping the Course and Withdrawing: For organizational purposes we take attendance at each lecture and each lab session. Students who join the course late are recorded as absent on the day of the lecture that they have not attended. Un-submitted psets are automatically assigned a zero score. For rules and regulations about adding/dropping courses, or to withdraw from a course read the instructions at this link.

Academic Honesty: As a University of Washington student, each student the course has agreed to abide by the University's academic honesty policy. All academic work must meet the standards of conduct described in Student Governance and Policies, Chapter 209 (Link), in particular Section 5.B.1. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation.