

ECO372H1F: Data Analysis and Applied Econometrics in Practice

LEC0101 — Summer 2024

Department of Economics, University of Toronto

Instructor

Name: Sean Elliott

Email: sean.elliott@mail.utoronto.ca

Class information

Lecture: Tuesday & Thursday from 10 AM to 12 PM in SS1070

Tutorial: Friday from 1 PM to 3 PM in SS1070

Office Hours: Tuesday from 1PM to 2PM & Thursday from 9 AM to 10 AM (location TBD)

Website: Quercus ([Link](#))

Course description

ECO372 is an intermediate-level course in econometrics which equips students with a modern approach to data analysis and econometrics focusing on the use of data to answer causal questions. Students will learn about different empirical techniques that economists use to do so: random assignment, linear regression, difference-in-differences, regression discontinuity design, and instrumental variables. Students will learn about applications of these techniques in academic research. Econometric methods will be illustrated using the application of regressions to a wide variety of economic questions and data sources, including the use of statistical software (R).

Prerequisites: ECO200Y1/ ECO204Y1/ ECO206Y1; ECO220Y1/ ECO227Y1/ (STA237H1, STA238H1)/ (STA247H1, STA248H1)/ (STA257H1, STA261H1).

Exclusion: ECO351H1.

References

We will be following Scott Cunningham's book, Causal Inference: The Mixtape, quite closely albeit, in a slightly different order than that which is presented in the book. It is freely available online or a physical copy may be purchased from Amazon or other retailers. We will also be looking at journal articles which illustrate the concepts we are learning in class. I will post these articles on Quercus. I have also listed a few additional textbooks which serve as supplementary reading for the topics in Cunningham's book. However, they are not required.

- **Required:** Causal Inference: The Mixtape by Scott Cunningham (<https://mixtape.scunning.com/>)
- **Supplementary:**
 - Mastering 'Metrics by Joshua Angrist and Jörn-Steffen Pischke (Princeton University Press, ISBN:978-0-691-15284-4)
 - Introduction to Econometrics, 4th Edition by James H. Stock and Mark W. Watson
 - Introductory Econometrics: A Modern Approach, 7th Edition by Jeffrey M. Wooldridge

Software

We will be using the programming language R and the integrated development environment RStudio in this course. No prerequisite programming knowledge is required. Both R and RStudio can be downloaded and installed by following the instructions at this link: <https://posit.co/download/rstudio-desktop/>.

Weekly schedule

I will post all materials on our course webpage ([Quercus](#)) on the **Sunday evening** of the upcoming week. This includes the required readings, lecture slides, practice problems, code, and data. This is a condensed six-week course and, as such, we will move through the material rather quickly in terms of calendar time. I encourage all of you to come to the lecture prepared by going through the readings *ahead* of time. Do not fall behind – there will be no time to catch up.

- **Lectures:** In our two weekly lecture time slots I will introduce the econometric theory and show you examples, by way of economic research, of the theory being used in practice. You will want some way to take notes and you should also bring a device which you can use to access R/RStudio. I will often show examples of code in class and for you to get the most out of this you should follow along in real time.
- **Tutorials:** In the tutorial time slot we will be working through the practice problems. These problems will include some theory but mostly it will involve working with data in R. As such, you need some way to access R/RStudio during this session. It is likely that this will not take up the entire session most weeks – the remaining time will be use for office hours and open Q&A.
- **Office Hours:** I will have two weekly in-person office hours where you can either ask questions about the theory or get help with R. I will also schedule additional office hours before exams.

Course coverage

This is the tentative plan for the material.

Topic	Textbook chapter(s)
1 Introduction, probability, and regression	Chapters 1 & 2
2 Causality, potential outcomes, and selection bias	Chapters 3, 4, & 5
3 Regression discontinuity design (RDD)	Chapter 6
4 Panel data, difference-in-differences (DiD)	Chapters 8 & 9
5 Instrumental variables (IV)*	Chapter 7

* Time permitting.

The textbook chapters will be supplemented with journal articles which I will post on Quercus with each week's lecture materials. Also, I will let you know each week which sections of the textbook are particularly relevant for us.

Course evaluation

Your final grade will be based on the following:

Assessment	Weight	Due Date
In-class participation	10%	NA
Assignment 1	10%	May 24 (at 11:59 PM)
Midterm exam	25%	May 28 (in class)
Assignment 2	10%	June 14 (at 11:59 PM)
Final exam	45%	TBA (Possible dates are June 19-24)

Some notes about how the various components will work:

- **In-class participation:** Here you will be assessed by how much you contribute to the classroom discussion and your participation in various polls and questionnaires I will run during the lectures. This grade will be posted at the end of the course along with your final grade. I will exclude one lecture—out of our 12 meetings—when calculating this grade.
- **Assignments:** You will have two assignments each weighted equally. These will be a combination of theory and code. We will require you to hand in both typed answers and a code replication package. The submission will be online and you will have until 11:59 PM of the day it is due to submit. However, try not to submit too close to the deadline – see the late assignment submission policy below. You are free to collaborate with one another on these but you must hand in your own version of the answers written in your own words.
- **Midterm/Final** An hour-long midterm will be held on May 28 during the first hour of class. The final will be 2 hours long with the date TBA. Both exams will consist of short-answer questions. Practice exams with solutions will be provided for both.

You may use **generative AI** to help you debug your code or to answer questions about high-level concepts covered in the course. However, you should be careful with this. I have noticed that AI can often give you misleading information regarding causal inference and statistics, in general. Additionally, any written work you submit must be your own and may not include any content from generative AI. Including text (not code) generated by AI in your assignment submissions constitutes the use of an unauthorized aid and is considered an academic offence. I want to encourage you to use these tools for what they are good at—generating code syntax—and discourage you from using them for what they are bad at—explaining model results.

Remark requests

There will be an MS Form posted on the course webpage where you can go to request a remark of any submitted work. The entire submission will be remarked – your mark could go up or down. Any remark requests must be made within one week of the graded assessment being returned to you.

Missed work/late submission

Any **missed work** (assignment or exam) automatically gets a grade of zero. Special considerations can be made for *legitimate* reasons as long as they are communicated to me by e-mail in **advance** of the due date. In the case of a missed midterm exam, there will be an opportunity to write a make-up exam within one week of the original exam date. **Late assignment submissions** are also considered “missed” work and will be subject to the zero grade penalty.

Accessibility

If you require any additional accommodations or have any accessibility concerns, please reach out to Accessibility Services (<http://www.studentlife.utoronto.ca/as>) as soon as possible.