** APPLIED MICROECONOMICS **

FALL 2023

Instructor:
Professor Petra Moser
pmoser@stern.nyu.edu
OH: M 12-1.15 (on Zoom, please sign-up through link Brightspace by 8am)
Lecture: Tuesday and Thursday, 11-12.15 (ROOM KMC 4-110)

Teaching Fellow:
Ludovica Ciasullo
lc4177@nyu.edu
Recitation: Friday 3.30-4.45 pm (Room KMC 4-110)
OH: F 2-3.15 (office 617 Econ Building, email me for appointment by 8am)

Course Description:

The primary purpose of this class is to teach you to use data to answer interesting economic questions. You will learn to analyze data and identify causal effects using R, a statistical programming tool widely used in economics, finance, data science, and many other fields that require quantitative skills.

Class work alternates between lectures and hands-on laboratory work programming in R.

R is very similar to Python. If you have some experience in Python already R will be a straightforward and useful related language to learn. R is quickly gaining ground as an analytical tool in data science, industry, finance, in addition to academia. So, whether you want to get a job in industry or go to graduate school, R will be an incredibly useful and marketable skill. Comparing Python with R,

- Main Applications: Python is mostly built for websites, games and data analysis, while R is especially useful for data analysis and visualization. Both R and Python can do anything in data science, statistics, or economics that you'll ever need but R will do it with nice intuitive commands
- Both are free
- Python is fast with large data sets, while R is unbelievably fast.
- Applications for R are endless

Practical steps for programming in R:

Please install R *before* our first class on Tuesday October 31st 2023.

Mac

Go to https://cran.r-project.org/bin/macosx/, click on R-3.6.3.nn.pkg (signed) (scroll down to see it).

Windows
Go to https://cran.r-project.org/bin/windows/base/old/3.6.3/ . Click on Download R 3.6.3 for Windows.

Next, install the free version of the program RStudio on your computer by going the following website: https://www.rstudio.com/products/rstudio/download/. R is a programming language, and Rstudio is an application that allows you to write code in R.

To check that everything worked, simply open Rstudio. Everything worked correctly if you see a message that starts with the following:

```
R version 3.6.3 (2020-02-29) – "Holding the Windsock"
Copyright I 2020 The R Foundation for Statistical Computing
```

If you previously worked in R, you don't have to update to the most recent version. Any version above R version 3.5 will work just fine.

Data Camp access:

To help you learn R, we will be using datacamp.com. It offers short explanations of how R works, followed by interactive exercises. We have free access for our course, you will only have to **register with your nyu.edu email**. I will send an e-mail with an invitation to register.

Course Requirements:

The success of this class depends on a high level of preparation to ensure high quality discussions and participation. Course requirements are designed to encourage such preparation.

Your grade for this class is based entirely on participation, your lab reports plus a few opportunities for extra credit. There are no exams. But you will work hard and continuously throughout the term, learning about coding, data analysis, and hypothesis testing.

Lab reports will count for 85% of your grade, the remaining 15% is attendance and participation. We will have a total of six lab reports. There are no late submissions. But we will only count your top five reports to your grades. This means that you get to drop the lowest grade, but you still have to submit all six of them.

In lab sessions, we work together on basic hypothesis testing. We'll identify a research question, set up a hypothesis, and try to falsify it using real-world data.

Lab work can be done alone or in group. In my experience, groups of two to three people are most productive. It is perfectly acceptable for all members in your group to submit the same program (R script), as well as a set of identical figures and tables, but you need to each submit an independent write-up of your results. In other words, you can all have the same tables and figures, but the write-up and the description of the results must be your own.

Guidelines for reports

Please submit a concise report of around 4-5 pages in length, including tables and figures (12-point Times New Roman or Calibri, 1.5-line spacing, 1-inch margin throughout).

Please save your report as a **pdf**.

Submit your report along with the R script that you used to generate all figures and tables. You should upload them on Brightspace.

State your name at the top of the lab report and state the names of your team on the R script. You are graded for the quality of your reports, not the scripts themselves. However, we will be happy to answer any questions you may have about writing scripts.

Deadlines for lab reports are strict as reported on the schedule below. Reports are due at the beginning of lecture. No late submissions are accepted!

Point allocations and criteria for grading (maximum is 20):

- Tables and Figures (8 points).
    1. If there is a datacamp.com exercise assigned in a given week, you get 2 points for completing it.
    2. Create a self-contained set of tables and figures that is accessible and easy to read. Include an informative title for each table or figure.
    3. Label the axes of a figures legibly. Make a legend if necessary.
    4. Provide notes to explain what the graph shows.
    5. List your sources of data.
- Text (12 points).
    1. Explain clearly what you are testing, how you are testing it, and what you find.
    2. Introduction (2 points)

       What is research question? What is the hypothesis that you are testing? Why is it important? What is the setting of the empirical analysis? What is the data? Why is it novel and important?
    3. Empirical evidence (6 points)

       Please discuss the empirical evidence that you have examined. Describe all tables and figures. Do they reject your hypothesis? Do they bring up other questions, beyond the original research question?
    4. Discussion and Conclusion (4 points)

       Based on the empirical evidence, do you reject or fail to reject your main hypothesis? Do your findings have policy implications? What are they? What are potential extensions and what data you would need for them?

<u>Extra credit (5 points each):</u> These are additional questions in the assignments that are not compulsory.

Course policies:

Late work does not count for course credit. It gets a grade of zero. No exceptions. Please keep in mind though that we will drop the lowest grades from your lab reports.

Errors in grading occur, if very rarely; this section lays out rules and procedures for requesting a correction. Students who have been graded incorrectly should petition for a correction in writing to the Professor. Requests should occur within a week of the work being returned.

The entire graded work (problem set or examination) should be resubmitted; there is no guarantee that grades will rise. If the request arises because you think different students have been graded differently, all the affected students should submit their work as a group (there is no guarantee that only upward adjustments will occur).

Schedule of Readings

* are required readings. Other readings are optional background readings; I will summarize those briefly in class.

Textbooks

Joshua D. Angrist, Joern-Steffen Pischke, *Mastering Metrics, The Path from Cause to Effect*. 2015. The libraries has copies of this book. Please let us know if you have any trouble getting a copy.

Christoph Hanck, Martin Arnold, Alexander Gerber and Martin Schmelzer, *Introduction to Econometrics with R*. 2019. This is a free econometrics textbook that teaches you how to do statistical analysis in R. It is available here: https://www.econometrics-with-r.org/

Lecture 1 – Course Introduction & Basics of R

Tuesday, October 31st 2023

We will start with a course overview. We will review the syllabus, course policies and deliverables.

In the second part of class, we will start to learn the basics of R. To be as effective as possible and use our time well, please install R and RStudio on your computer before this first class. Please email your TA if you need any help with installing R before the class. We will learn how to write and save R scripts and some basics of the language.

Lecture 2: Aggregating data in R

Thursday, November 2nd 2023

In our first data lab, we will learn to aggregate data in R by specific categories to take a stance on a hotly debated policy question: Should living organisms be patentable? To investigate this question, we will use dramatic change in historical patent laws – the introduction of plant patents in 1930 – to examine whether intellectual property in plants has encouraged innovation.

Data Source: Moser, Petra and Paul Rhode: "Did Plant Patents Create the American Rose?" in Joshua Lerner and Scott Stern (eds). *The Rate and Direction of Technological Change*, Chicago University of Chicago Press, 2012, pp. 413-41.

Learn to aggregate data by years or by classes and to create new columns in your data. Loading additional packages into R.

Data lab 1 – data manipulation + basic plots in R

Friday, November 3rd 2023
Create time-series plots in R, using data on rose registration and patents from Moser and Rhode (2012)

Complete the Datacamp chapters "Introduction to R" and "Data Manipulation with dplyr" (parts 1-4). Remember to use you nyu.edu email for Datacamp.

Readings:
1. Angrist and Pischke, Mastering Metrics, Chapter 1 "Randomized Trials"
2. https://www.econometrics-with-r.org/, Chapters 1.1 (Introduction to R)

Lecture 3 – Stastistical significance, t-tests, linear regressions

Tuesday, November 7th
Report 1 due at 11 (before class!) Please submit via Brightspace

In moves towards "patent harmonization" under TRIPS and other international agreements, industrialized countries push to extend their patent laws to other countries. To minimize the negative welfare effects of price increases and reductions in product variety because of patents, these treaties typically include a compulsory licensing provision, which allows countries to issue licenses to domestic firms – against the wishes of the patent owner. Do you think this policy encourages or discourages innovation? This data lab will answer this question, using data on US patents.

Readings and data source:
Moser, Petra and Alessandra Voena. "Compulsory Licensing: Evidence from the Trading with the Enemy Act" *American Economic Review*, 102, Issue 1 – February 2012. Last working paper version available at SSRN http://ssrn.com/abstract=1313867.

Data lab 2 - statistical significance, t-tests, linear regressions

Thursday, November 9th 2023

Today we will revise the concept of statistical significance. T-test for means comparison, confidence intervals, type I and type II errors.

Friday, November 10th

Using data from Moser and Voena (2012) we will learn to set up treatment and controls groups, and perform t-tests and regressions.

Complete "Data Visualization with ggplot2 (Part 1)" on Datacamp.

Readings

https://www.econometrics-with-r.org chapter 3.5 (performing a t-test)

Lecture 4 – Difference-in-differences analysis

Tuesday, November 14th
Submit Report 2 before class.

Today, we examine the effects of economic incentives – and specifically the way in which universities "tax" innovations by faculty - on the creation of start-ups and on innovation.

Readings and Data Source: Jones, Ben and Hans Hvide. 2018. "University Innovation and the Professor's Privilege" *American Economic Review, VOL. 108, NO. 7, JULY 2018, pp. 1860-98*

Data lab 3 – Regression tables and DiD

Thursday, November 16th 2023

A fun 11-minute lecture by Josh Angrist on differences-in-differences on Marginal Revolution University. And an application to Wikipedia.

Friday, November 17th 2023

Using data on patents and start-up firms in Norway from Jones and Hvide (2018), we'll learn how to perform regression analysis in the context of difference-in-differences. Running and reporting regression results in R. Making pretty regression tables

Readings:

1. Angrist and Pischke, Mastering Metrics, Chapter 2 "Regression" and Chapter 5 "Difference-in-Differences"
2. https://www.econometrics-with-r.org chapter 4 (linear regression with one regressor)

Lecture 5 – Standard errors in difference-in-differences analysis

Tuesday, November 21st, 2023

Report 3 due before class

Today we will look a little deeper at calculating standard errors, clustering, and how to deal with situations when the number of observations is small. We will use these techniques to ask whether copyrights, which protect creative content, such as books, music, and film, have encouraged creativity, using data on Italian operas.

Reading and Data Source: Giorcelli, Michela and Petra Moser "Copyrights and Creativity. Evidence from Italian Operas" Short film on Youtube.

Data lab 4 – Standard errors in two-way-fixed effects (TWFE) analyses

Tuesday, November 28th, 2023

We will review standard errors and compare the costs and benefits of robust versus clustered standard errors. Using data from Giorcelli and Moser (2020), we'll investigate alternative ways of clustering, including the bootstrap, to compute standard errors in R.

We'll talk about the Two Way Fixed Effect design (here is a primer), recent discussion on its issues (see here), and recent developments in the econometrics literature about how to address them (Sun and Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects", 2021)

Lecture 6 – Propensity score matching

Thursday, November 30th, 2023

We will learn context of another change in copyright policy, the US WWII Book Republication program, which made books with enemy-owned property rights available to US publishers. Did this policy change benefit US science?

Reading and Data Source: Biasi, Barbara, and Petra Moser "Effects of Copyright on Science. Evidence from the WWII Book Replication Program" http://ssrn.com/abstract=2542879

Data lab 5 – Propensity Score Matching

Friday, December 1st 2023 Report 4 due before class

Tuesday, December 5th2023

Chapter 3.3 from Mostly Harmless Econometrics (Angrist and Pischke, 2009)

Using data on book prices and scientific citations from Biasi and Moser (2018), we will learn how to set up propensity score matching in R

Lecture 6 – Instrumental Variables

Thursday, December 7th, 2023

Instrumental Variables: Why do we need them? How do they work? And how do we estimate them? We will answer these questions in lecture and illustrate them at the example of the IV in Moser et al (2015).

Reading and Data source: Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German Jewish Émigrés and U.S. Invention" *American Economic Review*, 2015 http://ssrn.com/abstract=1910247.

Data lab 6 – OLS with Selection

Friday, December 8th 2023
Report 5 due before class

We will replicate the OLS analyses in Moser, Voena and Waldinger and discuss issues of selection.

Lecture 7 – Instrumental Variables

Tuesday, December 12th 2023

Readings:
Chapter 3 in Angrist and Pischke, Mastering Metrics, "Instrumental Variables," pp. 98-142.
A fun short 12-minute lecture by Josh Angrist on Instrumental Variables.

Data lab 7 – Instrumental Variables

Thursday, December 14th 2023

Using data from Moser, Voena, and Waldinger (2014) we'll implement IV in R

Friday, December 15th 2023: Last day of the term.
OH
Report 6 due before midnight.