ECO 751–Research Methods I: Introduction to Econometric Analysis

Edo Navot

Spring 2019

E-mail: enavot@jjay.cuny.edu
Office Hours: By appointment only
Office: New Bldg 9.63.23

Class Hours: M 06:00-08:00 p.m.
Class Room: New Bldg 9.63.24

Course Description

This course is designed to provide an introduction to quantitative research methods in economics. It will cover working with data, measurement of economic variables, methods of data collection, data presentation, hypothesis testing, interpreting the results of various statistical procedures and teach regression-based research methods. We will begin with a review of probability and statistics, move on to simple and multiple regression, then build upon that base after gauging progress during the course of the semester. By the end of this course, students will recognize which data types require which analytical techniques, will be able to formulate a quantitative research question, obtain data, structure the data in preparation for analysis, and produce professional quantitative policy briefs utilizing regression analysis.

Textbooks and required readings

```
Wooldridge, Jeffrey. Introductory Econometrics: A Modern Approach.
    6th Edition. Cengage Learning.

Wickham, Hadley and Grolemund, Garrett. R for Data Science.
    OReilly Media.
```

Introductory Econometrics will be our main textbook and guide to statistical concepts and analysis. While the John Jay book store should have some copies, you may find a lower price through the online behemoth of your choice. Please try to obtain the 6th Edition. However, only purchase the "standalone book" or get a used copy; we will not utilize the online supplementary materials. If you have more than one class in which a textbook published by Cengage has been assigned, be aware that they have special "bulk" pricing which they refer to as "Cengage Unlimited." I have no experience with this business model and cannot recommend for or against it.

R for Data Science (R4DS) is an introduction to the R language for statistical programming. It covers the logic, syntax, and structure of the R language. It covers data manipulation, visualization, and basic analysis as well as many other important topics. While the book is required, you

do not have to purchase it. You can either purchase the book in hard copy from the vendor of your choice or use the free online version, available at https://r4ds.had.co.nz. You can also access the files used to produce the book on GitHub at https://github.com/hadley/r4ds.

Additional required readings will be provided as pdf files. These files will be accessible through the course folder on my DropBox site. I suggest you bookmark this folder as I will use to provide you with pdf readings, homework assignments, and supplementary materials.

Software

One of the most difficult initial challenges facing both the instructor and students of econometrics is the choice of software package to learn. Each statistical analysis software suite (Stata, SAS, R, EVeiws, SPSS, etc.) has its own trade-offs: when choosing software one must consider monetary costs, time costs of learning the language, the uses for which one aspires apply the software, and the community of users. Practitioners often find themselves locked into a particular language because of the opportunity costs of learning a new language, even as learning principles of coding in one language makes learning the next language somewhat easier.

In our case, we will learn econometrics using R. This choice has been made for us largely by the fact that R is free and CUNY does not provide licenses to any other statistical analysis software suites. R is an open source statistical programming language with a highly active and deep developer base. As is often the case, R's advantages and drawbacks emerge largely from the fact that it is open source. You should know the particular trade-offs unique to R up front: R's main weakness is that it is one of the more difficult languages to learn. This is compounded by the fact that R leverages its open source primarily by allowing developers to create custom packages which can be easily downloaded and installed within R. However, this means that one must memorize not only a larger number commands and functions relative to other languages, but also packages. Furthermore, R's coding syntax – not easy to begin with – is complicated by the fact that different developers may not adhere to the same syntactic principles in their packages.

The good news is that there has been a concerted effort within the R developer community to rectify the above-described challenges. A group of developers is currently working on a unified and simplified set of syntactic principles, largely borrowed from the computer programming world, for R code. These efforts consist of a group of packages gathered under the title "The tidyverse." The book by Wickham and Grolemund (two of the central developers within the tidyverse) expounds upon the principles of tidy R coding and data analysis and will be our principle guide to learning the R language.

In exchange for the extra effort required to learn R, you will obtain access to the most flexible and most advanced statistical capabilities available, easily the most powerful and compelling data visualization engine, as well as lifetime free access that will never require a license.

I will provide a resource list for R which will be distributed in class.

Important note: We will also utilize Excel early in the semester to emphasize our hands-on approach. If you do not have Excel or another similarly capable software suite available to you, please notify me.

Pedagogical Approach

Econometrics is challenging: it requires understanding the principles and mathematics of probability in combination with a craft-like knowledge of many tools which are appropriate only in specific contexts and if certain assumptions hold. Learning econometrics while also learning a (potentially new) statistical programming language is more than doubly challenging.

With these challenges in mind, this course is structured to attempt to minimize, as much as possible, the most demanding aspects of the usual introduction to econometrics curriculum. The core pedagogical approach utilized for this course will be to substitute statistical programming for algebra in teaching, homework, and exams. Of course, we will still examine the algebraic expressions but algebraic manipulation will not be required in this course. Instead, we will learn to translate algebraic expressions into English as well as computer code, while emphasizing the actual operations being performed using simple data examples. In other words, this course aims to instill an understanding of statistical processes and analysis usually gained from deep engagement with algebra with translation of the algebra into more intuitive and concrete expressions. We will also emphasize the importance of applying every concept to actual data to make the concepts and operations clear. As we build our statistical intuitions and toolbox, we will begin to analyze larger and larger data, slowly building the complexity of concepts and applications.

Class Structure

Each class will be structured into two complimentary parts: The first half of the course will consist mainly of lectures, with a focus on translating statistical concepts from the current reading into English and computer code, as described in the pedagogical approach above. These lectures will include utilizing R's helpful *notebook* format to demonstrate how the R code is built, what it is doing, how it expresses the relevant statistical concept, and how to apply it. I will share these notebooks after each class.

The second half of each class will consist of a more open discussion of supplementary readings or individual/group work to practice statistical programming and applied analysis.

Computers

Please bring your laptops with R + R Studio installed to class.

Be aware, however, that Research strongly suggests that using your laptop to take notes during a lecture diminishes your overall comprehension of the material. While we will utilize our computers when doing applied work in R Studio, I urge you to take notes using pen and paper. Since my lecture notes will consist of R notebooks with code and since I will share these notes with you, there's no inherent need to use your computer to take notes.

While we're on the topic, its also suggested that you put your phone away when doing homework and research.

Course Policies

Grading Policy

- **50%** of your grade will be determined by homework assignments. There will be ten homework assignments, each worth five percent of your total, final grade.
- **25%** of your grade will be determined by a 24-hour take-home midterm exam.
- **25%** of your grade will be determined by a final paper. The paper will be a policy brief that demonstrates your ability to analyze data, visualize and describe analytical results, and cogently discuss a quantitative policy topic.

All assignments will be completed within R Studio, utilizing the R Markdown facility to generate reports that include both statistical code and narrative descriptions in a single document. A condition of the assignments you hand in is that they must be fully reproducible. I will run the R Markdown documents on my computer and they must work in order for you to get credit.

Attendance Policy

Attendance is not only mandatory but also necessary for comprehension of the material. While it is not my habit or desire to actually record your attendance, as already stated, it is mandatory. If you are unable to attend a class please notify me in advance. While you will obtain the relevant lecture notes for each class regardless of attendance, you should make an effort to check in with your classmates to review materials you missed in class.

Policy Regarding Academic Integrity

Don't cheat. This course is structured in such a way that if you cheat, you will only be cheating yourself out of a learning opportunity. However, working collaboratively will be helpful to you and is encouraged.

John Jay's policy on academic integrity can be found here:
http://www.jjay.cuny.edu/web_images/Policyand_Procedures.pdf

Disabilities Policy

Federal law mandates the provision of services at the university-level to qualified students with disabilities. Students with disabilities will be provided reasonable accommodations if they are determined eligible by the Office of Accessibility Services (OAS). Prior to granting disability accommodations in this course, the instructor must receive written verification of a student's eligibility from the OAS (phone # 212-237-8031). It is the student's responsibility to initiate contact with that office and to follow the established procedures for having the accommodation notice sent to the instructor.

Jay/ECO 751/correlation.png

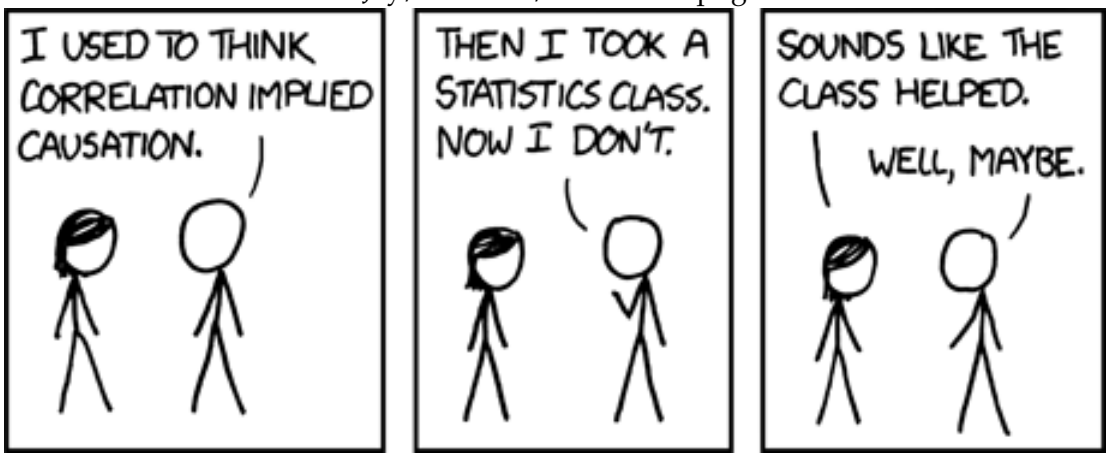

Figure 1: Correlation/Causation

Class Schedule

Important: class readings are subject to change, contingent on mitigating circumstances and the progress we make as a class.

Students must complete the reading prior to each class session. Readings marked as (pdf) will be provided via DropBox:

`https://www.dropbox.com/sh/m552bzybfa5jlt6/AABDN4dw3Rr7nKt8JDywN_Rca?dl=0`

Class 01, 01/28: Introductions

- Review Syllabus
- Introduction to R + R Studio, R Markdown and R Notebooks
- Introduction to statistical computing
- Discussion of statistical concepts in economics and political economy

Class 02, 02/04: Probability / Classical Methods

- Ch. 1 & 2, *Introduction to Econometrics*. By Stock and Watson. 3rd Edition, updated. (pdf)
- Introduction, The *Grundrisse*. By Karl Marx. (pdf)

Excel Exercise 1 due.

Class 03, 02/11: Review of Statistics / Probabilistic thinking in classical Political Economy I

- Ch. 3, *Introduction to Econometrics*. By Stock and Watson. 3rd Edition, updated. (pdf)
- Ch. 1, *Unholy Trinity: Labor, Capital, and Land in the New Economy*. By Duncan Foley. (pdf)

Excel Exercise 2 due.

Class 04, 02/18: Get hands dirty with R / Probabilistic thinking in classical Political Economy II

- Ch. 2, 3, 4 & 21 *R4DS*.
- Forward and Introduction. *Laws of Chaos: A Probabilistic Approach to Political Economy*. By Emmanuel Farjoun and Moshe Machover. (pdf)

R Exercise 1 due.

Class 05, 02/25: Exploratory Data Analysis with R / Introduction to Major Cross-Sectional Data Sources

- Ch. 1, 5, 7 & 8 *R4DS*.
- Orientation to IPUMS, CPS, ACS, and other data. (pdf)

R Exercise 2 due.

Class 06, 03/04: Simple Regression

- Ch. 1 & 2, *Wooldridge*.
- Ch. 9 & 12 *R4DS*.

R Exercise 3 due.

Class 07, 03/11: Multiple Regression

- Ch. 3, *Wooldridge*.

R Exercise 4 due.

Class 08, 03/18: Inference

- Ch. 4, *Wooldridge*.

R Exercise 5 due.

Class 09, 03/25: Transformations, Scaling, and Special Controls

- Ch. 6, *Wooldridge*.
- Ch 14 & 15, *R4DS*.
- , Jessica and Baker, Dean. (2018) "Does tax deductability affect CEO pay? The case of the health insurance industry." EPI & CEPR policy brief. (pdf)

Choose a project topic for final paper.

Midterm Exam

Class 10, 04/01: Binary Variables

- Ch. 7, *Wooldridge*.
- Ch. 16 & 17, *R4DS*.
- Rosnick, David and Baker, Dean. (2017) "The Wealth of Households: An analysis of the 2016 Survey of Consumer Finances." CEPR policy brief. (pdf)

R Exercise 6 due.

Class 11, 04/08: Limited Dependant Variable Models

- Ch. 17, pgs. 524-536. *Wooldridge*.
- Bucknor, Cherrie. (2016) "Black Workers, Unions, and Inequality." CEPR policy brief. (pdf)

R Exercise 7 due.

Class 12, 04/15: Post-Estimation / Conducting an Empirical Project

- Ch. 8 & 19 *Wooldridge*.
- Hirsch, Barry and Schumacher, Edward. (2001) "Private sector union density and the wage premium: Past, present, and future." Journal of Labor Research, 22(3), 487-518. (pdf)

R Exercise 8 due.

Begin work on your final paper.

Class 13, 04/22: Spring Recess, no class

Class 14, 04/29: Special topics dependent on time I

- Cunningham, Scott and Shah, Manisha. (2017) "Decriminalizing indoor prostitution: Implications for sexual violence and public health." The Review of Economic Studies, 85(3), 1683-1715. (pdf)
- Additional readings TBD.

Class 15, 05/06: Special topics dependent on time II

- Glover, Dylan; Pallais, Amanda; and Pariente, William. (2017) "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores." The Quarterly Journal of Economics, qjx006. (pdf)
- Additional readings TBD.

Class 16, 05/13: Special topics dependent on time III

- Bertrand, Marianne; Kamenica, Emir; Pan, Jessica. (2015) "Gender identity and relative income within households." Quarterly Journal of Economics. (pdf)
- Additional readings TBD.