

# PPD 716: Econometrics for PPM I

Nicolas Duquette

Fall 2022

Thursday – 2PM to 5:20 PM

Location: RGL 103

Instructor: Nicolas Duquette

Office: RGL 234

Office Hours: 10 am to noon on Tuesdays, or by appointment.

Email: nduquett@usc.edu

## Course Description

This course is designed to expand students' knowledge of econometric methods. In particular, the course will develop econometric reasoning and skills through practical applications. The course builds on the foundations of PPD558 and PPDE668 and prepares students for Econometrics for Policy, Planning and Management II. The course is appropriate for Ph.D. students in Price School and other students in the social sciences with the appropriate econometrics background.

## Learning Objectives

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

- implement a wide range of microeconomic techniques
- critically assess applied quantitative studies
- identify empirical strategies appropriate for a given research question and data structure
- perform a variety of data preparation, programming and analytic tasks

## Prerequisites

Students should take PPDE668 or an equivalent course before enrolling in PPD 716. In addition to this course prerequisite, students not currently enrolled in a PhD program must submit a brief statement when requesting a D-clearance that explains the reason for their interest in PhD-level coursework.

## Office Hours

Office hours are an opportunity to “stop by” and talk about course materials or any other questions you may have. It’s also OK to stop by without a particular question in mind. You do not need to tell me in advance that you will be coming to scheduled office hours—just stop in.

My scheduled office hours will be 10 am to noon on Tuesdays. There will not be scheduled office hours on USC holidays, during study days, or during the final exam period.

If you’d like to meet outside that time, I’m happy to make appointments. To make an office hours appointment, please fill out this Google Form.<sup>1</sup> This form will make sure I see your request promptly and that I have all the information I need for us to schedule a date and time.

## Description and Assessment of Assignments

There will be eight problem sets in this class, each requiring students to examine and replicate a portion of a published paper or otherwise apply class techniques to real data.

You are allowed and encouraged to work in teams of three or fewer on the first seven problem sets, with the following restrictions:

- You may not work with the same problem set partner on more than three of the first seven problem sets.
- You may not work with the same partner on consecutive problem sets. (So if you work with someone on problem set #1, you cannot work with them again on problem set #2.)
- Problem set #8 allows you to work in larger groups and does not count toward these limits.

---

<sup>1</sup><https://forms.gle/qmiK6ioLsSpUd9nJ6>

Answers to each problem set must be submitted in PDF format, with supplementary files included in a single compressed directory format (.TGZ, .7Z or .ZIP). Each group needs to submit only one copy of its problem set.

Submit your final version by email to nduquett@usc.edu, and CC all group members on the submission.

Please use filenames indicating the problem set number followed by all group member's names, e.g. PS1 Ali Singh Wang.pdf and accompanying files PS1 Ali Singh Wang.7z. If your files are too big for email after compression, you can share a Google Drive folder containing the relevant files instead.

Each problem set is worth 12 points.

<b>Graded items by date and weight</b>		
Assignment	Due date	Weight
<i>Problem sets</i>		
1 - Warmup	September 8	12
2 - Difference-in-difference	September 22	12
3 - Regression discontinuity	October 6	12
4 - Synthetic control	October 20	12
5 - Event study	November 3	12
6 - Multilevel & selection	November 17	12
7 - Quantiles & Lasso	December 1	12
8 - Students' choice	December 14	12
<i>Total points</i>		<i>96</i>

## Course grades

Total points are rounded to the nearest tenth of a point and then converted to a letter grade as follows:

<b>Grading by points</b>	
Total Points	Letter Grade
$\geq 91$ out of 96	A
88-90.9	A-
85-87.9	B+
80-84.9	B
77-79.9	B-
$< 77$	$\leq C+$

## Technology

You are expected to attend each class meeting with a computer equipped with Stata, a modern web browser, and a working internet connection. I do not recommend using an iPad or similar tablet that cannot function as a traditional computer.

If you need resources to successfully participate in your classes, such as a laptop or internet hotspot, you may be eligible for the university's equipment rental program. To apply, please submit an application online.

Stata is statistical software commonly used in academic and think-tank policy research, and the most-used econometrics software within the Price School. You can purchase Stata at Stata-Corp's web site.<sup>2</sup> Stata/BE (the cheapest but most limited option) should be sufficient for almost everything we do in this course. You can access virtual machines with Stata/SE (the fancy version) installed at <https://cloudapps.usc.edu>. See also the list of software available to USC students at <https://software.usc.edu/>. If you have an older version of Stata already, version 14+ should be sufficient for most exercises, though you may need to use the USC cloud version or alternative commands for some of the advanced topics at the end of the semester.

In the final weeks of the semester, we will use Python, a general-purpose programming language increasingly used for data analytics, particularly in non-academic jobs, and particularly for machine learning methods. We will do this using online "notebooks" which will run Python code on a remote server, so you will not be required to install Python directly on your personal machine if you don't want to.

If you are already very familiar with Stata, you are welcome to try using a different software system to complete the assignments, with the warning that the instructor does not provide support for other software. Python can be used for assignments designed for Stata, too; if you decide to do this, I recommend installing Anaconda Python on your computer. Alternatively, the statistics software R is open-source, free of charge, and widely used, and would also be a good choice if you want to learn a second econometric toolkit.

## Course Web Sites

Lecture slides, assignments, data sets, and other useful resources will be posted to the class Google Drive. You must be logged into your USC account to access the materials.

**USC prohibits sharing of any synchronous and asynchronous course content** outside of the learning environment. Sharing of course materials will lead to academic integrity sanctions. Specifically, SCampus Section 11.12(B) states

---

<sup>2</sup><https://www.stata.com/order/new/edu/gradplans/student-pricing/>

*Distribution or use of notes or recordings based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study is a violation of the USC Student Conduct Code. This includes, but is not limited to, providing materials for distribution by services publishing class notes. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the Internet or via any other media. (See Section C.1 Class Notes Policy).*

## Readings Overview

Each class meeting has an associated list of suggested readings. No specific reading is required, but all have potential to be useful.

Where possible, I have placed readings on digital reserve. I definitely recommend studying from reserve copies before buying to make sure you have the right text for your future research needs.

The textbooks I'll refer you to the most this semester are:

- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press, New Haven. Approachable overview of causal inference with workhorse estimators in economics. Affordable paperback with a free online version at this link. Copious practical examples Stata, R, and (online only) Python.
- Huntington-Klein, N. (2022). *The Effect: An Introduction to Research Design and Causality*. CRC Press, Boca Raton, Fla. Very accessible and informal text on research design and causal inference for beginners. Copious code examples in Stata, Python and R. Another affordable paperback, also has a free version online at this link.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press. Works through the matrix algebra for almost all major econometric techniques, with examples.
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. Accessible yet rigorous exploration of most common causal identification strategies.

You may also find useful some of the more general research resources I've liked to on my professional web page.

## Course Schedule: A Weekly Breakdown

Below is a list of topics by weeks and associated methodological readings. In addition to these readings, I will present applications of each technique from real research in class. Bibliographies for all examples will be included in the reference section of the slide decks.

### August 25

Description, prediction, and explanation; Statistics review; Potential outcomes; Randomized social experiments; Coding workflows and research best practice

- Cunningham, “Potential outcomes causal model”
- Angrist and Pischke, chapter 2.
- Cameron and Trivedi, chapter 2.
- Morgan, S. L. and Winship, C. (2015). *Counterfactuals and Causal Inference*. Cambridge University Press, second edition. Chapters 1 and 2. Emphasizes application of causal inference to econometric methods. The authors are sociologists and the examples come from both sociology and economics. LOC H62.M646.
- Gentzkow, M. and Shapiro, J. M. (2014). *Code and Data for the Social Sciences: A Practitioner’s Guide*. Available at <https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf>
- <https://xkcd.com/2180/>

### September 1

Least-squares regression; Probit and logistic regression

- Cameron and Trivedi, chapters 4 and 14.
- Cunningham, “Properties of regression”
- Huntington-Klein, “Regression”
- Angrist and Pischke, chapter 3.
- Long, J. S. and Freese, J. (2014). *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press, College Station, TX, third edition. Chapters 5–6. Excellent coverage of binary, categorical, and count dependent variable models. Includes authors’ own code for straightforward calculation of marginal effects. Does not cover new features in Stata v16, such as `cmsset` commands. LOC QA 278.2 L848r

- Hansen, B. (2020). *Econometrics*. University of Wisconsin. Chapter 25, “Binary choice.” Free and available at <https://www.ssc.wisc.edu/~bhansen/econometrics/>. An econometrics textbook used in the first year of the economics PhD at Wisconsin. Useful if you would like to see the probability-theoretic underpinnings of particular estimators.

## September 8

Propensity score matching and weighting; 2x2 Difference-in-differences; Difference-in-difference-in-differences

- Cunningham, “Differences-in-differences” and “Matching and subclassification”
- Huntington-Klein, “Matching” and “Difference-in-differences”

## September 15

Causal graphs; Instrumental variables

- Cunningham, “Directed acyclical graphs” and “Instrumental Variables”
- Huntington-Klein, chapters 6–8 and 19
- Morgan and Winship, chapters 3 and 9.
- Pearl, J. (2009). *Causality*. Cambridge University Press, second edition. Pearl’s *magnum opus*. USC has access to a cumbersome (but free-of-charge) e-book version through the university libraries.
- Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley, first edition. Again, USC has access to a cumbersome (but free-of-charge) e-book version through the university libraries.
- Angrist and Pischke, chapter 4.
- <https://xkcd.com/2560/>

## September 22

Regression discontinuity; Differences-in-regression discontinuity; Notches and kinks

- Cunningham, “Regression Discontinuity”

- Huntington-Klein, chapter 20
- Angrist and Pischke, chapter 6.
- Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2018). *A Practical Introduction to Regression Discontinuity Designs*. Forthcoming, Cambridge University Press. Includes Stata and R code throughout. Two volumes, available as preprint PDFs at <https://rdpackages.github.io/>.
- Skovron, C. and Titiunik, R. (2015). A practical guide to regression discontinuity designs in political science. *American Journal of Political Science*, 2015:1–36

## September 29

Time series basics; Interrupted time series; Introduction to panel data; Synthetic control

- Pickup, M. (2014). *An Introduction to Time Series Analysis*. SAGE Publications
- Beckett, S. (2013). *Introduction to Time Series using Stata*. Stata Press, College Station, Texas, chapters 2 and 5.
- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425
- Morgan and Winship, chapter 11.
- Cunningham, “Synthetic control”

## October 6

Fixed effects; Panel diff-in-diff; Event study models

- Cunningham, “Panel data.”
- Huntington-Klein, chapters 16 and 17
- Cameron and Trivedi, chapters 21–22.
- Angrist and Pischke, chapter 5.
- Hatfield, L. and Zeldow, B. (2019). Difference-in-differences. Free guide online at <https://diff.healthpolicydatascience.org/>. An approachable exposition of DD methods with illustrative gifs.



- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199
- Goodman-Bacon, A., Goldring, T., and Nichols, A. (2019). `bacondecomp`: Stata module for decomposing difference-in-differences estimation with variation in treatment timing. Statistical Software Components S458676, Boston College Department of Economics

## October 13

Standard errors; Clusters; Bootstrap; Randomization inference

- Cameron and Trivedi, chapters 11 and 24.
- Angrist and Pischke, chapter 8.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427
- Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research
- Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal*, 17(3):630–651
- Kerwin, J. (2017). Randomization inference vs. bootstrapping for p-values. Blog post, <https://jasonkerwin.com/nonparibus/2017/09/25/>
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249
- Correia, S. (2014). `reghdfe`: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components S457874, Boston College Department of Economics, revised 10 Mar 2019
- Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using `boottest`. *The Stata Journal*, 19(1):4–60
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372

## October 20

Generalized least squares; Random effects; Multi-level modeling; Dynamic panel models; Linear systems; Seemingly unrelated regression; Simultaneous equation systems

- Rabe-Hesketh, S. and Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, College Station, TX, third edition. Needs an update to more recent Stata syntax, but extremely thorough exposition of the method and its applications.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press. Deep dive into panel methods. LOC HB139.W663.
- Angrist and Pischke, chapter 5.
- Cameron and Trivedi, chapters 2.4–2.5 and 6.
- Hansen, chapter 11.8.

## October 27

Censored and truncated regression; Sample selection models; Hurdle models; Measurement error

- Cameron and Trivedi, chapters 26, 27.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161

## November 3

Heterogeneity; quantile regression; distribution regression

- Cameron and Trivedi, chapter 4.6.
- Hao, L. and Naiman, D. Q. (2007). *Quantile Regression*. Sage Publications

## November 10

Introduction to Python; notebooks; Pandas; web scraping

- McKinney, W. (2018). *Python for Data Analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc., second edition. Available for free with a USC email address at <https://learning.oreilly.com/home/>.
- Daniel Sullivan, “Stata to Python equivalents.” [http://www.danielsullivan.com/pages/tutorial\\_stata\\_to\\_python.html](http://www.danielsullivan.com/pages/tutorial_stata_to_python.html)

## November 17

Predictive modeling; basic machine learning; lasso regression

- Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11:685–725
- Burkov, A. (2019). *The Hundred-page Machine Learning Book*. <http://themlbook.com/>. Available as a pay-what-you-want e-book. Short but clear.
- Efron, B. and Hastie, T. (2021). *Computer Age Statistical Inference*. Cambridge University Press, chapter 16. A 2016 edition (with corrections) is available as a PDF at [https://web.stanford.edu/~hastie/CASI\\_files/PDF/casi.pdf](https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf).
- Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. Chapters 1–3 and 7. Clearly written explanation by the inventors of many ML techniques. Plenty of math. Presents coding examples in R. PDF versions free of charge at <https://web.stanford.edu/~hastie/ElemStatLearn/>.
- StataCorp (2019b). *Stata Lasso Reference Manual*. Stata Press. <https://www.stata.com/manuals/lasso.pdf>

## November 24

THANKSGIVING DAY (NO CLASS)

## December 1

Ordered logit; Multinomial logit; Decision trees; Random forests

- <https://xkcd.com/2494/>

- Athey and Imbens
- Cameron and Trivedi, chapter 15.
- Long and Freese, chapters 7–8.
- Efron and Hastie, chapters 8 and 17.
- StataCorp (2019a). *Stata Choice Models Reference Manual*. Stata Press, release 16 edition. <https://www.stata.com/manuals/cm.pdf>.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press. The decision tree OG, I haven't found a clearer exposition of the math behind the algorithm (in greater detail than Burkov) than this. Available as a free PDF through the USC libraries.
- Burkov.
- Hastie et al., chapter 15.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242

## Statement on Academic Conduct and Support Systems

### Academic Conduct

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” <http://policy.usc.edu/scampus--part--b>. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <http://policy.usc.edu/scientific--misconduct>.

### Students and Disability Accommodations

USC welcomes students with disabilities into all of the University’s educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at [osas.usc.edu](http://osas.usc.edu). You may contact OSAS at (213) 740-0776 or via email at [osasfrontdesk@usc.edu](mailto:osasfrontdesk@usc.edu).

### Support Systems

*Counseling and Mental Health - (213) 740-9355 – 24/7 on call*

<http://studenthealth.usc.edu/counseling>

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

*National Suicide Prevention Lifeline – 1 (800) 273-8255 – 24/7 on call*

<http://suicidepreventionlifeline.org>

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

*Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press “0” after hours – 24/7 on call*

<http://studenthealth.usc.edu/sexual-assault>

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

*Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086*

<http://eeotix.usc.edu>

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

*Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298*

[http://usc-advocate.symplicity.com/care\\_report](http://usc-advocate.symplicity.com/care_report)

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

*USC Campus Support and Intervention - (213) 821-4710*

<http://campussupport.usc.edu>

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

*Diversity, Equity and Inclusion - (213) 740-2101*

<http://diversity.usc.edu>

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

*USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*

<http://dps.usc.edu>, [emergency.usc.edu](http://emergency.usc.edu)

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

*USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call*

<http://dps.usc.edu>

Non-emergency assistance or information.

*Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)*

<http://ombuds.usc.edu>

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

*Occupational Therapy Faculty Practice - (323) 442-3340 or [otfp@med.usc.edu](mailto:otfp@med.usc.edu)*

<http://chan.usc.edu/otfp>

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

*Last updated: August 11, 2022*