

# PSCI 7095-001: Advanced Political Data Analysis (Data II)

University of Colorado Boulder

Spring 2021

**Time:** Monday, 4:45-7:15  
**Location:** KTCH 1B31 (or virtual)  
**Instructor:** Dr. Andrew Q. Philips  
**Office:** KTCH 144  
**Email:** andrew.philips@colorado.edu  
**Office hours:** Monday 3:00-4:30, or by appointment

**COURSE DESCRIPTION:** Quantitative analysis is an important component of nearly all political science research. This course is designed to introduce you to the fundamental tools used for data analysis. We will review and build on what you have learned in Data I. Most of this course focuses on understanding the approach of ordinary least squares (OLS). We will start by reviewing some topics you may already know, then move onto OLS estimation. Then we will cover violations of OLS assumptions, as well as various approaches to addressing such violations.

While it is crucial for political scientists to master OLS, in reality most research involves using alternative modeling techniques, many of which were developed for use in situations where OLS is inappropriate. Therefore, in the second half in the course, we cover a variety of alternative extensions and issues such as dichotomous dependent variables, endogeneity, as well as spatial and temporal dependence. In addition, we will spend a lot of time discussing how to present and interpret regression results, as well as substantive and statistical significance.

By the end of this course you should be able to:

- Understand what is going on “under the hood” of OLS, and interpret regression results
- Diagnose and address violations of the regression assumptions
- Have a variety of models to add to your “toolkit”
- Apply what you have learned to your research.

**PREREQUISITES:** This is a graduate level course; students should have a background in introductory regression (i.e., Data I). We will be working in matrix algebra notation throughout much of the course, although prior experience with this is not necessary.

**SOFTWARE:** We will use R for most of this course. Although familiarity with R is not necessary, it is a plus. Those unfamiliar with this program may want to purchase or borrow the suggested textbooks that cover working with R, although there are copious amounts of information available for free online. Please download both R (<https://cran.r-project.org/>) and RStudio (<https://www.rstudio.com/>) before the first class session. We will probably also use some Stata, although it is not necessary to purchase it for this course. Although there will not be a substantial amount of writing, students are encouraged to write up any assignments using  $\LaTeX$ .

**GRADES:** Course grades will be based on the following. Participation and homework assignments make up 40% of the final grade. About halfway through the semester, a mid-term exam will be given that is worth 30% of the final grade. At the end of the semester, there will be a take-home final exam worth 30% of the final grade. There are no opportunities for extra credit.

Participation and Homework	40%
Midterm Exam	30%
Final Exam	30%

The following scale will be used to turn numerical grades into letter ones. Note that I will round up a letter should your grade fall on the number (but on or above 0.5) between two letters (e.g., 89.5 up to 90 rounds up to an A-).

#### Grade Scale

A	95-100
A-	90-94
B+	87-89
B	84-86
B-	80-83
C+	77-79
C	74-76
C-	70-73
D+	67-69
D	64-66
D-	60-63
F	0-59

**PARTICIPATION:** Participation is an integral component of graduate courses. Students are expected to come to every class *having already read the assigned readings for that day*, and should be prepared to discuss them. Graduate-level courses only are successful when all students participate actively in the discussion.

**HOMEWORK:** Throughout the semester, there will be various homework assignments. Some of these will be group assignments, others on your own. Most will involve some form of data analysis and interpretation/presentation of regression results. We will discuss more specifics on homework in class.

**MIDTERM EXAM:** About halfway through the semester there will be a midterm exam. This will be closed book.

**FINAL EXAM:** At the end of the semester, there will be a comprehensive final exam. The final exam will be take-home and open book. You may consult textbooks, articles, and your notes for this exam, *but not others* (this will be considered a form of cheating).

**ATTENDANCE AND LATE POLICY:** Attendance is a key component of succeeding in graduate school. I provide slides for each class, but we will have a much more comprehensive discussion than what appears on the slide. Attendance is mandatory. If you have to miss a class, you should let me know in advance so that we can make arrangements.

Assignments are due on the day listed in the syllabus. Late assignments will not be accepted.

**REQUIRED TEXTS:** The following text is required for the course. Any additional readings will be made available to you on the first day of class or as needed. This text is advanced, but will be a helpful reference after the semester is over.

- Greene, William H. *Econometric analysis*. 2017. 8th edition. Pearson.

Note that it is expected to read the week's required readings before coming to class.

**RECOMMENDED TEXTS:** The following texts are not required, but may be helpful to some. In the schedule below there are additional texts in the "suggested readings".

- Gujarati, Damodar N. and Dawn C. Porter. *Essentials of econometrics*. 2009. 4th edition. McGraw-Hill Education. A good introductory econometrics text.
- Angrist, Joshua D. and Jorn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. 2009. Princeton University Press. A quite accessible introduction to causal inference and various other topics.
- Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. 2010. 2nd edition. MIT Press. A Greene-level text that focuses on cross-sectional and cross-sectional time series data.
- Kennedy, Peter. *A guide to econometrics*. 2008. 6th edition. Wiley-Blackwell. A popular introductory econometrics text.
- Monogan, James E. 2015. *Political analysis using R*. Springer.
- Philips, Andrew Q. 2019. "AT<sub>E</sub>X: A brief introduction". Available [here](#).
- Neumayer, Eric and Thomas Plümper. 2017. *Robustness tests for quantitative research*. Cambridge University Press. Good discussion of what constitutes "robustness", and how we should test for and present it.

**TENTATIVE SCHEDULE:** Note that this schedule is subject to change. We will spend as long as we need to on a topic, and many of the topics in the second half of the course may not take a full course day to cover.

## **Week 1: No Class (MLK Day)**

## **Week 2: Course Introduction, Regression Assumptions, Introduction to Matrix Algebra and R**

Required Readings:

- Greene, Appendix A

Suggested Readings:

- Burns, Patrick. 2011. *The R Inferno*. Available at: <http://www.burns-stat.com/documents/books/the-r-inferno/>.
- Philips, Andrew Q. 2019. "R: A brief introduction."
- Monogan, James E. 2015. *Political analysis using R*. Springer.

## **Week 3: Under the Hood: OLS**

Required Readings:

- Greene, Ch. 2 and 3

## **Week 4: Under the Hood: OLS (continued)**

Required Readings:

- Greene, Ch. 4 (up to 4.5)

## **Week 5: OLS in Practice**

Required Readings:

- Greene, Ch. 4 (4.5 on), 5

## **Week 6: OLS in Practice (Continued)**

Required Readings:

- Greene Ch. 6 (6.1-6.2, 6.5)
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political Analysis* 14(1): 63-82.
- Berry, William D., Matt Golder, and Daniel Milton. 2012. "Improving tests of theories positing interaction." *Journal of Politics* 74(3): 653-671.
- Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis* 27:163-192.

Suggested Readings:

- Franzese, Robert, and Cindy Kam. 2009. *Modeling and interpreting interactive hypotheses in regression analysis*. University of Michigan Press.

## **Week 7: Generalized Least Squares**

Required Readings:

- Greene Ch. 9

Suggested Readings:

- White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society*: 817-838.

## **Week 8: Binary Dependent Variables**

Required Readings:

- Greene Ch. 17 (up to 17.6)
- Berry, William D., Jacqueline HR DeMeritt, and Justin Esarey. 2010. "Testing for interaction in binary logit and probit models: Is a product term essential?" *American Journal of Political Science* 54(1):248-266.

## **Week 9: MID-TERM EXAM**

## **POST MID-TERM: Weeks 10-16**

The second half of the course will consist of a variety of topics that build on the first half of the course. Depending on time, we will cover everything below (and maybe some extra topics). There

will not be extensive coverage of each topic; in fact most topics could comprise an entire semester-long course. Instead, it is to help familiarize you with the various methodological tools that are out there for you to use, depending on your research question.

## **Time Series:**

Required Readings:

- Greene Ch. 20 and 21 (skim both)
- Keele, Luke, and Nathan J. Kelly. 2006. "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables." *Political Analysis* 14(2): 186-205.
- DeBoef, Suzanna, and Luke Keele. 2008. "Taking time seriously." *American Journal of Political Science* 52(1): 184-200.
- Philips, Andrew Q. 2018. "Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models." *American Journal of Political Science*. 62(1): 230-244.

## **Spatial Statistics:**

Required Readings:

- Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. "Space is more than geography: Using spatial econometrics in the study of political economy." *International Studies Quarterly* 50(1): 27-44.
- Franzese, Robert J., and Jude C. Hays. 2007. "Spatial-econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data." *Political Analysis* 15(2): 140-164.

## **Endogeneity:**

Required Readings:

- Greene Ch. 8

## **Causality and Causal Inference:**

Required Readings:

- Greene Ch. 6 (6.3-6.4)
- Matthay, Ellicott C., Erin Hagan, Laura M. Gottlieb, May Lynn Tan, David Vlahov, Nancy E. Adler, M. and Maria Glymour. 2020. "Alternative causal inference methods in population health research: Evaluating tradeoffs and triangulating evidence." *SSM-Population Health* 10: 100526.
- Pearce, Neil and Debbie A. Lawlor. 2016. "Causal inference—so much more than statistics." *International Journal of Epidemiology*: 1895-1903.
- Keele, Luke. 2015. "The statistics of causal inference: A view from political methodology." *Political Analysis*: 23:313-335.

Suggested Readings:

- Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. Chapters 1 and 2.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The colonial origins of comparative development: An empirical investigation." *American Economic Review* 91(5):1369-1401.
- Meyersson, Erik. 2014. "Islamic rule and the empowerment of the poor and pious." *Econometrica* 82(1): 229-269.

## Panel Data, Clustering, and Hierarchical Linear Modeling:

### Required Readings:

- Greene Ch. 11
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with time-series cross-section data." *American Political Science Review*: 634-647.
- King, Gary and Margaret E. Roberts. 2015. "How robust standard errors expose methodological problems they do not fix, and what to do about it." *Political Analysis* 23: 159-179.

### Suggested Readings:

- Beck and Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science*: 331-352.
- Arceneaux, K. and Nickerson, D.W., 2009. "Modeling certainty with clustered data: A comparison of methods." *Political Analysis* 17(2):177-190.

## Maximum Likelihood Estimation:

### Required Readings:

- Greene Ch. 14

### Suggested Readings:

- King, Gary. 1998. *Unifying political methodology: The likelihood theory of statistical inference*. University of Michigan Press: Chapters 1-3.

## Resampling, Quantities of Interest, and Presenting Results:

### Required Readings:

- King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the most of statistical analyses: Improving interpretation and presentation." *American Journal of Political Science* 44(2):347-361.
- King, Gary and Langche Zeng. 2006 "The dangers of extreme counterfactuals." *Political Analysis* 14(2): 131-159.
- Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models." *American Journal of Political Science* 57(1):263-277.

### Suggested Readings:

- Carsey, Thomas M., and Jeffrey J. Harden. 2014. *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks, CA: Sage.
- Greene Ch. 15 (up to 15.7)
- Tomz, Michael, Jason Wittenberg, and Gary King. 2001. "Clarify: Software for interpreting and presenting statistical results."
- Imai, Kosuke, Gary King, and Olivia Lau. 2008. "Zelig: Everyone's statistical software."

## Machine Learning

### Required Readings:

- Hindman, Matthew. 2015. "Building better models: Prediction, replication, and machine learning in the social sciences" *The Annals of the American Academy of Political and Social Science* 659(1):48-62.
- Montgomery, Jacob M., and Santiago Olivella. 2018. "Tree-based models for political science data." *American Journal of Political Science* 62(3): 729-744.

#### Suggested Readings:

- Grimmer, Justin. 2015. "We are all social scientists now: How big data, machine learning, and causal inference work together" *PS*: 80-83.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. (JWHT 2013) "An introduction to statistical learning: With applications in R." Springer Series in Statistics. 1st edition. ISBN: 978-1461471370

## Final Exam: TBA (take-home)

### SYLLABUS CHANGES

I reserve the right to make changes to the syllabus during the course of the semester as needed and will make the most updated copy available to you and announce said changes during class.

**Last updated:** January 11, 2021

### UNIVERSITY-MANDATED STATEMENTS

#### Classroom behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the policies on [classroom behavior](#) and the [Student Code of Conduct](#).

#### Requirements for COVID-19

As a matter of public health and safety due to the pandemic, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements, and public health orders in place to reduce the risk of spreading infectious disease. Required safety measures at CU Boulder relevant to the classroom setting include:

- maintain 6-foot distancing when possible,
- wear a face covering in public indoor spaces and outdoors while on campus consistent with state and county health orders,
- clean local work area,
- practice hand hygiene,
- follow public health orders, and
- if sick and you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or if you live on-campus, please alert CU Boulder Medical Services.

Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policies on COVID-19 Health and Safety and classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please see the "Accommodation for Disabilities" statement on this syllabus.

All students who are new to campus must complete the COVID-19 Student Health and Expectations Course. Before coming to campus each day, all students are required to complete the Buff Pass.

Students who have tested positive for COVID-19, have symptoms of COVID-19, or have had close contact with someone who has tested positive for or had symptoms of COVID-19 must stay home. In this class, if you are sick or quarantined, you must contact the instructor and/or TAs. You do not need to disclose the nature of your illness.

## **Accommodation for disabilities**

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the [Disability Services website](#). Contact Disability Services at 303-492-8671 or [dsinfo@colorado.edu](mailto:dsinfo@colorado.edu) for further assistance. If you have a temporary medical condition or injury, see [Temporary Medical Conditions](#) under the Students tab on the Disability Services website.

## **Preferred student names and pronouns**

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

## **Honor code**

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code ([honor@colorado.edu](mailto:honor@colorado.edu)); 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.

## **Sexual misconduct, discrimination, harassment and/or related retaliation**

The University of Colorado Boulder (CU Boulder) is committed to fostering an inclusive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or [cureport@colorado.edu](mailto:cureport@colorado.edu). Information about the OIEC, university policies, [anonymous reporting](#), and the campus resources can be found on the [OIEC website](#).

Please know that faculty and graduate instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

## **Religious holidays**

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, I will try to accommodate your requests, but you must contact me early in the semester.

See the [campus policy regarding religious observances](#) for full details.