

SPA 613, Conduct of Inquiry II, Spring 2019

Department of Justice, Law and Criminology
& School of International Service
American University

1 Instructor Information

Dr. Joseph K. Young
Phone: 202.885.2618
Office: Kerwin 266
Email: jyoung@american.edu

Office Hours: 10:00-12:00 Wed
or by appointment
Class Location: Kerwin 201
Course Time: 2:30-5:20 PM Wed

2 Course Description

This course is the second part of a methods sequence that aims to provide you the tools necessary to be social researchers. The primary focus of the course is bivariate and multiple regression analysis. We explore the assumptions underlying the statistical model, what happens when the assumptions are violated, and how to deal with these problems. Specifically, we examine what happens to the standard regression model in the context of multicollinearity, nonlinearity, heteroscedasticity, non-additivity, measurement error and specification error. Beyond what is often called OLS regression, we discuss regression techniques related to limited dependent variables and times series/panel data.

While some courses on this topic are primarily devoted to mathematical exposition of these techniques, our course is for the applied researcher. We will occasionally delve into math, but the main goal is to help you develop the skills necessary to evaluate hypotheses in a large N context. Additionally, we will spend considerable time learning and utilizing statistical software. While there are a host of different software packages, we will use Stata as it is powerful, flexible, and is used by many quantitative political scientists, criminologists and economists. I will provide parallel versions of exercises in R, the powerful, free, and widely used package, as well. Finally, we will write our assignments and final paper in the typesetting program, $\text{\LaTeX}$.

3 Course Objectives

I have three main objectives for this course. First, the course should help you gain the skills necessary to produce an original quantitative project. Second, you should gain the skills necessary to read and critique quantitative work that appears in peer-reviewed academic journals. Third, the course should serve as a basis for the acquisition of more advanced quantitative methods as your professional needs require.

4 Learning Outcomes

- Test hypotheses using multiple regression analysis
- Understand the strengths and limitations of regression analysis
- Accurately interpret regression coefficients

- Know the basic assumptions of the OLS regression model
- Understand the consequences of violating these assumptions
- Given the data generating process of a dependent variable, know the appropriate statistical model and method
- Produce an original research paper applying the techniques learned in the course
- Learn how to use statistical software
- Use L^AT_EX to produce a publishable paper

5 Requirements

Required Books

Both of these books have been around for a long time. You can use an older or newer version depending on your budget.

1. Gujarati, Damodar. 2003. *Basic Econometrics 4th ed.* New York: McGraw-Hill. ISBN 0-07233542-4.
2. Kennedy, Peter. 2003. *A Guide to Econometrics 5th ed.* Cambridge: MIT Press. ISBN 0-262-61183-X.

Recommended Books

1. Baum, Christopher F. 2006. *An Introduction to Modern Econometrics Using Stata.* College Station: Stata Press. 1-59718-013-0
2. Berry, William and Mitchell Sanders. 2000. *Understanding Multivariate Regression.* Boulder: Westview Press. ISBN 0-8133-9971-8.
3. Greene, William. 2003. *Econometric Analysis 5th ed.* New York: Pearson. 0-13513-245-2
4. Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multi-level/Hierarchical Models* New York: Cambridge University Press. 0-521-68689-X
5. Kohler, Ulrich and Frauke Kreuter. 2005. *Data Analysis Using Stata* College Station: Stata Press. 1-59718-007-6
6. Knoke, David, George Bohrnstedt and Alisa Potter Mee. 2002. *Statistics for Social Data Analysis* Belmont: Thomson/Wadsworth. 0-87581-448-4
7. Pollock, Philip H. 2006. *A Stata Companion to Political Analysis* Washington DC: CQ Press. ISBN 0-87289-305-7
8. Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and causal inference.* Cambridge University Press.
9. Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

All of the books are available at the bookstore and online. You can find them on Amazon.com or Addall.com at a reduced price. Since advanced statistics is not always easy, it is helpful to have multiple treatments of the same topic. I find that some books are more useful for understanding certain topics than others. The recommended books provide either more sophisticated or more simplistic treatment of most of the topics we cover. Buying them and using them in conjunction with the required texts would be helpful but not necessary. If you plan to pursue a dissertation that utilizes quantitative methods, then purchasing some/all of these books may be something to consider.

Many of the articles for the class will be on blackboard. You can also find most of the journal articles on JSTOR (www.jstor.org). I will also place readings on blackboard that are not available on JSTOR regardless, and these are noted in the syllabus (*).

Assessment

1. **Attendance and Participation (20%)**—I expect that you will be prepared to discuss all of the readings each week. I will assign a grade for your participation each week based on the following scale:
 - A (outstanding) your comments were insightful and contributed to collective understanding of the material across the readings
 - A- (strong) your comments were sometimes insightful and sometimes average but were not consistent across the readings
 - B+ (good) your comments demonstrated that you understood the material but did not extend the discussion or offer new insights
 - B (average) you participated but did not seem to fully grasp the material
 - B- (poor) you spoke, but your comments were not germane to the material
 - C (unacceptable) you didn't say anything...
2. **Problem Sets (10%)**—You will have problem sets that relate to applied research throughout the semester. You will use Stata for all of these assignments. You may also use R.
3. **Exams (30%)**—The midterm and final exams are take-home exams that ask essay questions and involve applied regression techniques in Stata or R.
4. **Conference Presentation (10%)**—At the end of the term, you will give a 10-15 minute presentation on your research. You will be on a panel of 3-4 papers with a chair and discussant. The presentation must include a powerpoint or Beamer presentation. Later in the term, we will discuss how to present your research effectively. You will also be graded on how well you respond to questions from the discussant and audience.
5. **Research Paper (30%)**—Your research paper must be an original, quantitative paper that incorporates material from the course.

The research paper must be 20-25 pages (including references, tables, graphs, etc). The quality of the paper, however, is more important than the quantity. The paper should have the format as outlined by Barry Weingast here:

http://www.stanford.edu/~weingast/caltech_rules.html

While this is not the *only* way to write a great paper in Social Science, it is a *great* way.

These papers also makes useful suggestions for writing a graduate paper that is potentially publishable:

- (a) Wolfinger, Raymond E. 1993. "Tips for Writing Papers." *PS: Political Science* Vol. 26, No. 1. 87-88.
- (b) Van Cott, Donna Lee. 2005. "A Graduate Student's Guide to Publishing Scholarly Journal Articles," *PS: Political Science & Politics*, 38(4):741-743.
- (c) Thunder, David. 2004. "Back to Basics: Twelve Rules for Writing a Publishable Article," *PS: Political Science & Politics* 37(3): 493-495.
- (d) King, Gary. 2006. "Publication, Publication." *PS: Political Science and Politics* 39(1): 119-125.

This paper by King is also useful for quantitative researchers:

Gary King. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(3): 444-452.

6 *Tentative* Course Schedule

Intro and Review

Wed. Jan. 16—Introduction, Review of Statistical Concepts

- Introductions
- Discussion
- Gujarati, Appendix A
- Kennedy, Chapters 1 & 2, pgs. 257-258
- Angrist, Joshua D and Pischke, Jörn-Steffen. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press. Chapter 2.
- Intro to R, Stata, and Latex

The Regression Model

Wed. Jan. 23—The Bivariate Regression Model

- Gujarati, Chapters 2-5, pgs. 164-175 (Regression basics)
- *Berry & Sanders, Chapter 2
- BRING SOME CONCEPTS**—Bring a list of concepts that are of interest to you. For example, you may be interested in the relationship between democracy and terrorism or party ideology and economic growth.

Multiple Regression: Estimation

- Gujarati, pp. 202-23, 229-33 (Multiple Regression, Problem of Estimation, Partial Correlation Coefficients)
- Kennedy, Chapter 3 (Classic Linear Regression Model)

Stats Lab #1

- Introduction to Stata & Data Manipulation
- Make sure to have software and laptop

Wed. Jan. 30— Multiple Regression: Inference and Dummy Variables

- Gujarati, Chapter 8 (Multiple Regression: The Problem of Inference)
 - Kennedy, Chapter 4 (Interval Estimation and Hypothesis Testing)
 - Angrist, Joshua D and Pischke, Jörn-Steffen. pgs. 28-67. (Regression Fundamentals)
- Dummy Variables
- Gujarati, pp.297-306 (Dummy Variables)
 - Kennedy Chapter 14 (Dummy Variables)

Stats Lab #2

Correlation, Linear Regression

Violations of Regression Assumptions

Wed. Feb. 6—Heteroskedasticity

- Gujarati, Chapter 11 (Heteroskedasticity)
- Kennedy, Chapter 8 (Violating Assumption Three: Nonspherical disturbances)
- Downs, George and David Locke. 1979. "Interpreting Heteroskedasticity." *American Journal of Political Science* 23: 816-828.

Stats Lab #3

Heteroskedasticity

Multicollinearity

- Gujarati, Chapter 10 (Multicollinearity)
- Kennedy, Chapter 11 (Violating Assumption Five: Heteroskedasticity)
- Lemieux, Peter H. 1978. "A Note on the Detection of Multicollinearity." *American Journal of Political Science* 22 (1): 183-186.

Wed. Feb 13—Specification

- Gujarati, pp. 215-217, 506-524
- Kennedy, Chapter 5 (Specification)
- Achen, Christopher. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5: 423-450.
- Clarke, Kevin. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management & Peace Science* 22(4): 341-352.
- Ray, James Lee. 2005. "Constructing Multivariate Analyses (of Dangerous Dyads)," *Conflict Management & Peace Science* 22(4): 277-292.

Stats Lab #4

- Multicollinearity/Specification Lab

Wed. Feb. 20—Interactions

- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis* 14(1):63-82.
- Friedrich, Robert J. 1982. In Defense of Multiplicative Terms in Multiple Regression Equations. *American Journal of Political Science* 26 (November): 797-833.
- Braumoeller, Bear. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58(4): 807-820.

Stats Lab #5

- Interactions
- Guided exercise

Wed. Feb. 27—Autocorrelation and Measurement Error

- Gujarati, Chapter 12 (Autocorrelation), pp. 524-528
- *Carmines and Zeller, pp. 1-16, 29-32, 37-51
- Kennedy, Chapter 9 (Violating Assumption Four: Autocorrelation and Measurement Error)

Project/Paper Workshop

- Bring research question, outline, and data** for paper

Limited Dependent Variables, Forecasting, Displaying Results, & Missing Data

Wed. March 6—Panel Data

- Gujarati, Chapter 16 (Panel Data Regression Models)
- Kennedy, Chapter 17 (Panel Data)
- Beck, Nathaniel and Jonathan Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data." *American Political Science Review* 89: 634-647.
- Angrist, Joshua D and Pischke, Jörn-Steffen, Chapter 5 (Parallel Worlds).
- Achen, Christopher H. "Why lagged dependent variables can suppress the explanatory power of other independent variables." *Ann Arbor* 1001.2000 (2000): 48106-1248.

Stats Lab #6—Panel Data

- Estimating time-series cross-sectional models in Stata

SPRING BREAK—March 10–17, NO CLASS

Wed. March 27—Limited Dependent Variables

- Gujarati, Chapter 15 (Qualitative Response Regression Models)
- Kennedy, Chapters 15–16 (Qualitative and Limited Dependent Variables)
- Long and Freese, Limited Dependent Variables, Chapter 2.

Stats Lab #7—Logit Models

- Read Logistic Regression Diagnostics

-International Studies Association Conference

Wed. April 3—Visual display of quantitative information

- 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.
- Kastellec, Jonathan P., and Eduardo Leoni. 2007. Using Graphs Instead of Tables to Improve the Presentation of Empirical Results in Political Science. *Perspectives on Politics* 5(4):755–771.
- Tufte, Edward. Visual Display of Quantitative Information, Chapter 1.
- Paper workshop

Stats Lab #8

- Extracting and Displaying Quantities of Interest
- Learning Clarify and SPOST

Mon. April 10—Missing Data

- Schafer, Joseph L., and Maren K. Olsen. “Multiple imputation for multivariate missing-data problems: A data analyst’s perspective.” *Multivariate behavioral research* 33.4 (1998): 545–571.
- Horton, Nicholas J., and Ken P. Kleinman. “Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models.” *The American Statistician* 61.1 (2007): 79–90.
- Rubin, Donald B. “Multiple imputation after 18+ years.” *Journal of the American statistical Association* 91.434 (1996): 473–489.
- King, Gary, et al. “Analyzing incomplete political science data: An alternative algorithm for multiple imputation.” *American Political Science Review* 95.1 (2001): 49–69.
- Honaker, James, and Gary King. “What to do about missing values in time-series cross-section data.” *American Journal of Political Science* 54.2 (2010): 561–581. (skim)

Stats Lab #9

- Multiple Imputation

Wed. April 17—Forecasting and Cross-Validation

- Ward, Michael, Brian Greenhill, and Kristin Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4): 363-375.

- Schrodt, Philip. 2002. "Forecasts and Contingencies: From Methodology to Policy." Paper presented at the American Political Science Association annual meeting, September.

- O'Brien, Sean. 2002. "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis." *Journal of Conflict Resolution* 46(6): 791-811.

Stats Lab #10

- Cross-Validation

Stats Lab #11

- Final Help with Model Estimation/Presentation

- Visualizing data

- Bring data, do-files, graphs

Wed. April 24—Extensions to and Critiques of Classic Regression Modeling

- Western, Bruce, and Simon Jackman. "Bayesian inference for comparative research." *American Political Science Review* 88.2 (1994): 412-423.

- Silver, Nate. *The Signal and the Noise*, Chapter 8.

- Mahoney, James, and Gary Goertz. "A tale of two cultures: Contrasting quantitative and qualitative research." *Political Analysis* 14.3 (2006): 227-249.

- Schrodt, Philip A. "Seven deadly sins of contemporary quantitative political analysis." *Journal of Peace Research* 51.2 (2014): 287-300.

- Kennedy, Chapter 13 (The Bayesian Approach)

- Final Presentations #1, Final Papers Due

Wed. May 1—Final Presentations #2