Syllabus for *Quantitative Political Analysis*

David Hendry & Thomas J. Leeper
Department of Political Science and Government
Aarhus University

December 8, 2014

Individuals working in the public and private sectors are often tasked with analyzing quantitative data or making use of analyses performed by others. The purpose of this course is to expand and improve significantly participants' ability to perform quantitative analyses of political science data and, further, to better evaluate the use of quantitative research results. Several advanced statistical and analytic techniques will be introduced and applied to research questions from political science. Participants will leave the course with the ability to better assess published research, better perform their own analyses of quantitative data, and better describe and understand the results of such analyses.

Among others, the course will discuss the following techniques and topics:

- Linear regression analysis
- Interpretation of interactions in regression models
- Regression with binary, categorical, and count outcomes
- Analysis of data gathered over time and across geographies
- Research design for causal inference
- Using Stata for statistical analysis

In addition, topics such as hypothesis generation and research design will be discussed throughout the course. The course does not have a unified theoretical or empirical focus, but we will touch upon and read empirical literature from most areas of political science including comparative politics, public administration, and international relations.

Each week, the course will consist of two sessions:

- Lecture
- Hands-on lab session

1 Objectives

The learning objectives for the course are as follows. By the end of the course, students should be able to:

1. Describe politically relevant research questions and hypotheses
2. Evaluate and deduce observable implications from political science theories
3. Explain statistical procedures and their appropriate usages
4. Apply statistical procedures to relevant research problems
5. Synthesize results from statistical analyses into well-written and well-structured essays
6. Demonstrate how to use Stata for statistical analysis

2 Exam and Weekly Assignments

2.1 Exam (Only for Master/KA)

The final exam is a seven-day written assignment analyzing a topic outlined by the instructors using quantitative data. In addition, the students must write four essays during the course on topics defined by the instructors as a prerequisite for the exam

The essays (3-5 pages) must be written in English and are due via email to the appropriate instructor at 12:00 on the dates listed below. The students will receive feedback on the essays and they must all be approved before taking the final exam.

1. Essay 1 due February 27 to David
2. Essay 2 due March 20 to Thomas
3. Essay 3 due April 10 to David
4. Essay 4 due May 8 to Thomas

2.2 For PhD Students

Students can choose from the following course elements:

1. "Regression with Matrices and MLE" (1 essay, 2 ECTS)
2. Sessions 1–5 (1 essay, 7 ECTS)
3. Sessions 1–5, plus one additional session (2 essays, 9 ECTS)
4. Sessions 1–5, plus two additional sessions (3 essays, 11 ECTS)
5. Sessions 1–5, plus three additional sessions (3 essays, 13 ECTS)
6. Entire course (4 essays, 15 ECTS)

PhD students must notify the course instructors of their choice by the end of Session 5.

3   Reading Material

The assigned material for the course includes about 2400 pages including several textbooks and empirical research articles, all of which are available online or in the printed course packet. All readings should be completed for the day they are described. *There is reading assigned on the first day.* The textbooks for the course are as follows:

– A. Colin Cameron and Pravin K. Trivedi. *Microeconometrics Using Stata.* Stata Press, College Station, TX, revised edition, 2010.

– Kim Mannemar Sønderskov. *Stata: En Praktisk Introduktion.* Hans Reitzel, Copenhagen, 2014.

– Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, NJ, 2009.

– J. Scott Long. *Regression Models for Categorical and Limited Dependent Variables.* Sage, London, 1997.

– Paul D. Allison. *Fixed Effects Regression Models.* Sage, Newbury Park, CA, 2009.

– William D. Berry. *Understanding Regression Assumptions.* Sage, Newbury Park, CA, 1993.

Note: An English-language translation of the Sønderskov book, with identical content, is also available from the same publisher.

PhD students are additionally required to obtain:
– John Fox. *Mathematical Primer for Social Scientists.* SAGE Publications, Newbury Park, CA, 2009.

4   Course Website

All information about the course will be posted on Blackboard. Any changes to the syllabus or additional notes will be made available there.

5 Schedule

The general schedule for the course is as follows. Details on topics covered and the readings for each week are provided on the following pages.

5.1 Introduction and Research Design I (Week 6)

What topics will we cover in this course? How do we think about causality for the purposes of research design? How can experiments help us understand causal relationships?

Instructor: Thomas

Lecture

- Course overview
  - Readings and textbooks
  - Exam
  - Four essays
  - Plan for the course
- Asking good research questions
- Research design
- Philosophies of causality
- Experiments and matching

Readings

– Ch.2 from Angrist and Pischke.

– Kosuke Imai, Gary King, and Elizabeth A. Stuart. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):481–502, April 2008.

– Martin Gilens. Political ignorance and collective policy preferences. *American Political Science Review*, 95(2):379–396, 2001.

– John Henderson and Sara Chatfield. Who matches? propensity scores and bias in the causal effects of education on participation. *The Journal of Politics*, 73(3):646–658, 2011.

See Also:
– Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236, December 2007.

– Donald B. Rubin. Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Servicess & Outocmes Research Methodology*, 2:169–188, 2001.

Lab

- Basics in Stata

Readings for lab

– Ch. 1–3 from Sønderskov.

5.2 Research Design II (Week 7)

What does it mean to make a causal inference? Can we make causal inferences from non-experimental data? Can we make causal inferences from non-experimental data? What are the necessary requirements for causal inference?

Instructor: David

Lecture

- Defining causality in a regression framework
- Model-building
- Reporting regression results

Readings

– Ch.3 (up to p.69) from Angrist and Pischke.

– Ch.3 from Gary King, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton University Press, Princeton, NJ, 1994.

– pp.17–25 from Damodar N. Gujarati. *Basic Econometrics.* McGraw Hill/Irwin, New York, 2002.

– Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

See Also:

– Scott Ashworth, Joshua D. Clinton, Adam Meirowitz, and Kristopher W. Ramsay. Design, inference, and the strategic logic of suicide terrorism. *American Political Science Review*, 102(2):269–273, 2008.

– Markus Prior. News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, 49(3):577–592, 2005.

Lab

- Further basics in Stata

Readings for lab

– Ch. 4–6 from Sønderskov.

5.3 Ordinary Least Squares Regression (Week 8)

How do we estimate causal effects using regression analysis? How do we interpret linear regression coefficients for different types of variables? What do goodness-of-fit measures tell us?

Instructor: Thomas

Lecture

- OLS method
- Interpretation of coefficients
- Standard errors, t-tests, and p-values
- Goodness-of-fit measures

Readings

– pp.1–67 from Berry.

– Ch.3 (from p.69) from Angrist and Pishke.

– Ch.3 (up to p.103) from Cameron and Trivedi.

– Pierre Englebert. Pre-colonial institutions, post-colonial states, and economic development in tropical africa. *Political Research Quarterly*, 53(1):7–36, 2000.

See Also:

Lab

- OLS in Stata
- Reporting regression results
- Stata factor variables

Readings for lab

– Ch. 8–9 from Sønderskov.

5.4 Ordinary Least Squares Regression II (Week 9)

How do we state and test hypotheses about heterogeneous effects? How do we interpret those effects in OLS interaction terms? When do we need to estimate alternative standard errors for OLS estimates?

Instructor: David

Lecture

- Effect heterogeneity and interaction terms
- Standard errors
- Heteroskedasticity

Readings

– pp.293–315 from Angrist and Pischke.

– p.67 to end from Berry.

– Robert J. Friedrich. In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, 26(4):797–833, 1982.

– Thomas Brambor, William Roberts Clark, and Matt Golder. Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63–82, 2006.

– Sara McLaughlin Mitchell and Will H. Moore. Presidential uses of force during the cold war: Aggregation, truncation, and temporal dynamics. *American Journal of Political Science*, 46(2):438–452, 2002.

See Also:

– William Rhodes. Heterogeneous treatment effects: What does a regression estimate? *Evaluation Review*, 34(4):334–361, 2010.

– Garrett Glasgow, Matt Golder, and Sona Golder. Who 'wins'? determining the party of the prime minister. *American Journal of Political Science*, 55(4):937–954, 2011.

Lab

- Estimating and interpreting interaction terms
- Heteroskedasticity-consistent standard errors
- Clustered standard errors
- The `margins` command

Readings for lab

– Ch. 10 from Sønderskov.

5.5 Practical Data Issues (Week 10)

How do we construct, tidy, and transform empirical observations into an organized rectangular dataset for use in our analyses? What problems do we encounter when dealing with real-world data? How do we address those challenges?

Instructor: Thomas

Lecture

- Data cleaning and transformations
- Multivariate scaling and reliability
- Variable transformations
- Missing data, case deletion, and imputation

Readings

– TBD from Sharon Lohr. *Sampling: Design and Analysis.* Advanced (Cengage Learning). Cengage Learning, 2009.

See Also:
– Paul D. Allison. *Missing Data.* Sage, Newbury Park, CA, 2002.
– Roderick J.A. Little. Regression with missing x's: A review. *Journal of the American Statistical Association*, 87(420):1227–1237, 1992.

Lab

- Scaling and reliability
- Variable transformations and regression interpretation
- Missing data imputation

Readings for lab

– Ch. 2 from .
Cameron and Trivedi

5.6 Research Designs for Causal Inference (Week 11)

What kinds of naturally occurring variation can we utilize to make meaningful causal inferences? What sorts of research designs (and associated analyses) that do not involve researcher-created randomization allow us to infer causal relationships?

Instructor: Thomas

Lecture

- Instrumental Variables
- Regression Discontinuity Designs
- Difference-in-Differences

Readings

– Ch.4,6 from Angrist and Pischke.

– Daron Acemoglu, Simon Johnson, and James A. Robinson. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5):1369–1401, December 2001.

– Clifford J. Carrubba. The electoral connection in European Union politics. *The Journal of Politics*, 63(01):141–158, July 2008.

– David Dreyer Lassen and Søren Serritzlew. Jurisdiction size and local democracy: Evidence on internal political efficacy from large-scale municipal reform. *American Political Science Review*, 105(2):238–258, 2011.

See Also:

– Paul R Rosenbaum. *Observational studies*. Springer, 2002.

– Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York, 2007.

– William R. Shadish, Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton-Mifflin, Boston, MA, 2001.

Lab

- Regression Discontinuity Design analysis

Readings for lab

– Ch. 6 from Cameron and Trivedi.

5.7 Panel Analysis for Continuous Outcomes (Week 12)

What is panel data? How does the nature of panel data relate to the assumptions of the classic linear regression model? How can we leverage time trends to discuss causation? How can we account for unit heterogeneity in a panel setting?

Instructor: David

Lecture

- Difference-in-differences, continued
- First-differences and fixed effects
- Random effects models

Readings

– Ch.5 from Angrist and Pischke.

– Ch.1,2 from Allison.

– Ch.16 fromDamodar N. Gujarati. *Basic Econometrics*. McGraw Hill/Irwin, New York, 2002.

– Steven E. Finkel and Amy Erica Smith. Civic education, political discussion, and the social transmission of democratic knowledge and values in a new democracy: Kenya 2002. *American Journal of Political Science*, 55(2):417–435, 2011.

– Alan S. Gerber and Gregory A. Huber. Partisanship, political control, and economic assessments. *American Journal of Political Science*, 54(1):153–173, 2010.

See Also:

– James A. Stimson. Regression in space and time: A statistical essay. *American Journal of Political Science*, 29(4):914–947, 1985.

– Nathaniel Beck. Time-series–cross-section data: What have we learned in the past few years? *Annual Review of Political Science*, 4:271–293, 2001.

– David Dreyer Lassen and Søren Serritzlew. Jurisdiction size and local democracy: Evidence on internal political efficacy from large-scale municipal reform. *American Political Science Review*, 105(2):238–258, 2011.

– Jeffrey Pickering and Emizet F. Kisangani. Diversionary despots? comparing autocracies' propensities to use and to benefit from military force. *American Journal of Political Science*, 54(2):477–493, 2010.

– Thomas Plümper, Vera E. Troeger, and Philip Manow. Panel data analysis in comparative politics: Linking method to theory. *European Journal of Political Research*, 44(2):327–354, 2005.

Lab

- Data in "wide" and "long" formats
- Panel regression

Readings for lab

– Ch.8 from CameronTrivedi2010.

5.8 Multi-level Modeling (Week 13)

What do we do when we have data at more than one level of analysis? How does the nature of multilevel data differ from the assumptions of the classical linear regression model? Can we simultaneously summarize relationships at different levels of analysis?

Instructor: David

Lecture

- Multi-level data structures
- The logic of multi-level models
- The hierarchical linear model and its interpretation
- Random effects versus fixed effects
- Panel models as special cases of multilevel models

Readings

– (51–83; 91–130) from Sophia Rabe-Hesketh and Anders Skrondal. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, College Station, TX, second edition, 2008.

– Ch.1–3 from Stephen W. Raudenbush and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. London, Sage, second edition, 2002.

See Also:

– Kim Mannemar Sønderskov. Does generalized social trust lead to associational membership? unravelling a bowl of well-tossed spaghetti. *European Sociological Review*, 27(4):419–434, 2011.

– Edward Fieldhouse, Mark Tranmer, and Andrew Russell. Something about young people or something about elections? electoral participation of young people in europe: Evidence from a multilevel analysis of the european social survey. *European Journal of Political Research*, 46(6):797–822, 2007.

– Marco R. Steenbergen and Bradford S. Jones. Modeling multilevel data structures. *American Journal of Political Science*, 46(1):218–237, 2002.

Lab

- Estimation
- Testing for fixed versus random effects
- Interpretation

Readings for lab

– (Review previous sections) from Sophia Rabe-Hesketh and Anders Skrondal. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, College Station, TX, second edition, 2008.

– (56–58; 305–316) from Cameron and Trivedi.

5.9 Maximum Likelihood Estimation (Week 14)

What is the likelihood theory of statistical inference? What do we do when our dependent variable is binary? What is the relationship between classical linear regression models and generalized linear models?

Instructor: David

Lecture

- Probability distributions
    - Discrete, continuous
- Maximum Likelihood Estimation
- Generalized Linear Models
- Logistic regression

Readings

– Ch.1–3 from Long.

– Sara Binzer Hobolt. Taking cues on europe? voter competence and party endorsements in referendums on european integration. *European Journal of Political Research*, 46(2):151–182, 2007.

– Vincent Price and John Zaller. Who gets the news? alternative measures of news reception and their implications for research. *Public Opinion Quarterly*, 57(2):133–164, 1993.

See Also:

– (13–26) from G. S. Maddala. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York, 1986.

– Ch.1–3 from Gary King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Cambridge University Press, New York, 1989.

– Ch.1–4 from Jeff Gill. *Generalized Linear Models: A Unified Approach*. Sage, London, 2000.

– Lyn Ragsdale. The politics of presidential speechmaking, 1949–1980. *American Political Science Review*, 78(4):971–984, 1984.

– John D. McCarthy, Clark McPhail, and Jackie Smith. Images of protest: Dimensions of selection bias in media coverage of washington. *American Sociological Review*, 61(3):478–499, 1996.

Lab

- Estimation of maximum likelihood models
- Interpretation of logistic regression coefficients
- Interpret Wald tests and likelihood ratio tests

Readings for lab

– Ch.3–4 (skim sections on probit) from J. Scott Long and Jeremy Freese. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press, College Station, TX, second edition, 2005.

5.10 Interpretation of GLMs (Week 15)

How do we interpret the results of models involving binary outcomes? How can we translate regression results for these models into meaningful quantities of interest and interpretable graphical results?

Instructor: Thomas

Lecture

- Logit versus Probit
- Heterogeneous effects and interaction terms
- Predicted probabilities and marginal effects
- Interpretation in Generalized Linear Models

Readings

– Ch. 4 from Long.

– Ch.14 from Cameron and Trivedi.

– Thomas Brambor, William Roberts Clark, and Matt Golder. Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1):63–82, 2006.

– David J. Houston. 'walking the walk' of public service motivation: Public employees and charitable gifts of time, blood, and money. *Journal of Public Administration Research and Theory*, 16(1):67–86, 2006.

See Also:
– Sara Binzer Hobolt. Taking cues on europe? voter competence and party endorsements in referendums on european integration. *European Journal of Political Research*, 46(2):151–182, 2007.

Lab

- `margins` for generalized linear models
- `marginsplot`

Readings for lab

No class (Week 16)

5.11 GLMs for Ordered, Multinomial, and Count Outcomes (Week 17)

What do we do when our dependent variable is qualitative, categorical, or a count? How can we make predictions from qualitative, categorical, or count models? What are the similarities between linear and generalized linear models when it comes to inference?

Instructor: David

Lecture

- Ordered logit and probit
- Multinomial logit
- Count outcomes
    - Poisson regression
    - Dispersion and alternative count models

Readings

– Ch.5,6,8 from Long.

– Guy D. Whitten and Harvey D. Palmer. Heightening comparativists' concern for model choice: Voting behavior in great britain and the netherlands. *American Journal of Political Science*, 40(1):231–260, 1996.

See Also:

– Todd G. Shields and Chi Huang. Presidential vetoes: An event count model. *Political Research Quarterly*, 48(3):559–572, 1995.

– Melissa R. Michelson. The corrosive effect of acculturation: How mexican americans lose political trust. *Social Science Quarterly*, 84(4):918–933, 2003.

– R. Michael Alvarez and Jonathan Nagler. Economics, entitlements, and social issues: Voter choice in the 1996 presidential election. *American Journal of Political Science*, 42(4):1349–1363, 1998.

Lab

- Estimation of ordered, multinomial, and count models
- Interpretation of results
- Presentation of results
- Testing for overdispersion in Poisson models

Readings for lab

– Ch.5,6,8 from J. Scott Long and Jeremy Freese. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press, College Station, TX, second edition, 2005.

5.12   Survival and Duration Analysis (Week 18)

What do we do if the quantity that we want to model is time? What if the phenomenon that we want to examine is political change? How can we study the duration and timing of political events? What are the similarities between duration models and other regression models when it comes to estimation and interpretation?

Instructor: David

Lecture

- The logic of survival/duration analysis
- Maximum likelihood survival models
- The Cox model
- The proportional hazards assumption
- Model diagnostics

Readings

– Janet M. Box-Steffensmeier and Bradford S. Jones. Time is of the essence: Event history models in political science. *American Journal of Political Science*, 41(4):1414–1461, 1997.

– Ch.3,4,8 from Janet M. Box-Steffensmeier and Bradford S. Jones. *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press, New York, 2004.

– Janet M. Box-Steffensmeier and Christopher J. W. Zorn. Duration models and proportional hazards in political science. *American Journal of Political Science*, 45(4):972–988, 2001.

See Also:

– Janet M. Box-Steffensmeier, Dan Reiter, and Christopher J. W. Zorn. Nonproportional hazards and event history analysis in international relations. *Journal of Conflict Resolution*, 47(1):33–53, 2003.

– D. Scott Bennett. Testing alternative models of alliance duration, 1816–1984. *American Journal of Political Science*, 41(3):846–878, 1997.

– John S. Ahlquist. Policy by contract: electoral cycles, parties, and social pacts, 1974–2000. *Journal of Politics*, 72(2):572–587, 2010.

– Kathleen Gallagher Cunningham. Divide and conquer or divide and concede: How do states respond to internally divided separatists? *American Political Science Review*, 105(2):275–297, 2011.

– Beth A. Simmons and Zachary Elkins. The globalization of liberalization: Policy diffusion in the international political economy. *American Political Science Review*, 98(1):171–189, 2004.

Lab

- Estimation of maximum likelihood survival models
- Model testing and comparison
- Estimation of Cox models
- Testing the proportional hazards assumption
- Interpretation of coefficients

Readings for lab

– (47–77, 129–144, 203–211, 245–268) from Mario A. Cleves, William W. Gould, and Roberto G. Gutierrez. *An Introduction to Survival Analysis Using Stata*. Stata Press, College Station, TX, revised edition, 2003.

5.13   Panel Analysis for Discrete Outcomes (Week 19)

How do we analyze data involving both repeated observation of the same units and non-continuous outcomes?

Instructor: Thomas

Lecture

- Panel binary outcome models
- Panel count models

Readings

– Ch.3,4 from Allison.

– pp.615–641 from Cameron and Trivedi.

See Also:

– Ch.5 from .

– Ch.17.4 from .

Lab for lab

- Estimating panel models for binary outcomes
- Model interpretation

Readings

5.14 Conclusion and Wrap-up (Week 20)

What have we learned? What didn't we learn?

Instructor: David

Instructor: Thomas

Lecture

- Wrap-up
- Course evaluations
- Questions about the exam

Readings

– Beth A. Simmons and Daniel J. Hopkins. The constraining power of international treaties: Theory and methods. *American Political Science Review*, 99(4):623–631, 2005.

See Also: