17.804: Quantitative Research Methods III

Fall 2014

Instructor: Teppei Yamamoto
TA: Elissa Berwick

Department of Political Science
MIT

1 Contact Information

	Teppei	Elissa
Office:	E53–401	E53–422
Phone:	617–253–6959	781-718-1365
Email:	teppei@mit.edu	eberwick@mit.edu

2 Logistics

- Lectures: Mondays and Wednesdays 9:30 – 11:00am, E53–438

- Recitations: Fridays, 10 am (tentatively) E53-485

- Teppei's office hours: Make an appointment

- Elissa's office hours: TBD

3 Course Description

This course is the third course in the quantitative research methods sequence at the MIT political science department. Building on the first two courses of the sequence (17.800 and 17.802), this class covers advanced statistical tools for empirical analysis in modern political science. Our focus in this course will be on techniques for *model-based inference*, including various regression models for cross-section data (e.g., binary outcome models, discrete choice models, sample selection models, event count models, survival outcome models, etc.) as well as grouped data (e.g., mixed effects models and hierarchical models). This complements the methods for *design-based inference* primarily covered in the previous course of the sequence. This course also covers basics of the fundamental statistical principles underlying these models (e.g., maximum likelihood theory, theory of generalized linear models, Bayesian statistics) as well as a variety of estimation techniques (e.g., numerical optimization, bootstrap, Markov chain Monte Carlo). The ultimate goal of this course is to provide students with adequate methodological skills for conducting cutting-edge empirical research in their own fields of substantive interest.

4 Prerequisites

There are three prerequisites for this course:

1. Mathematics: Basic college-level calculus and linear algebra.

2. Probability and statistics covered in 17.800 and 17.802, including linear regression and basic causal inference.

3. Statistical computing: familiarity with at least one statistical software. We will use R and JAGS in this course (more on this below).

For 1 and 3, we expect the level of background knowledge and skills equivalent to what is covered in the department's Math Camp II; see

https://stellar.mit.edu/S/project/mathcamp2/

5 Course Requirements

The final grades are based on the following items:

- **Problem sets** (50%): Six problem sets will be given throughout the semester. Problem sets will contain analytical, computational, and data analysis questions. Each problem set will be counted equally toward the calculation of the final grade. The following instructions will apply to all problem sets unless otherwise noted.

  - Neither late submission nor electronic submission will be accepted unless you ask for special permission from the instructor in advance of the deadline. (Permission may be granted or not granted, with or without penalty, depending on the specific circumstances.)
  - Working in groups is encouraged, but each student must submit their own writeup of the solutions. In particular, you should not copy someone else's answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions.
  - For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. All results should be presented so that they can be easily understood.

- **Final project** (40%): The final project will be a short research paper which typically applies a method learned in this course to an empirical problem of your substantive interest.

  We encourage you to co-author a paper with another student. By co-authoring you will (1) learn how to effectively collaborate with someone else on your research, which is very important in political science where most cutting-edge research is collaborative (see any recent issue of *APSR* or *AJPS*!) and (2) more likely have a good, potentially publishable paper (multiple brains are usually better than one).

  Unless you already have a concrete research project suitable for this course (e.g., from your dissertation project), we recommend that you start with replicating the results in a published article and then improve the original analysis using the methods learned in this course (or

elsewhere). Oftentimes, the most time-consuming part of a research project is data collection (which is not the focus of this course) and using data someone has already archived for their publication and made publicly available gets around this problem.

Students are expected to adhere to the following deadlines:

- October 1: Turn in a **brief description of your proposed project**. By this date you need to have found your coauthor, acquired the data you plan to use, and completed a descriptive analysis of the data (e.g. simple summary statistics, crosstabs and plots). Meet with the instructor to discuss your proposal during his office hours. You may be asked to revise and resubmit the proposal in two weeks from the meeting.

- November 24 and 26: Students will give **presentations in front of the class** based on the semi-final versions of their papers. Each presentation should last about 10 minutes (may vary depending on class size) and will be followed by a short Q&A session. Students should prepare electronic slides to accompany their presentation. Performance on this presentation will be counted as part of the class participation grade (see below). Make final revisions to your paper based on the feedback.

- December 10: **Paper due.** Please turn in one printed copy of your paper by the end of the day, and email electronic copies to the instructor and TA.

- **Participation and presentation** (10%): Students are strongly encouraged to ask questions and actively participate in discussions during lectures and recitation sessions.

In addition, there will be recommended readings for each section of the course which students are strongly encouraged to complete prior to the lectures in order to get the most out of them.

6 Course Website

You can find the Stellar website for this course at:

http://stellar.mit.edu/S/course/17/fa14/17.804/

We will distribute course materials, including readings, lecture slides and problem sets, on this website.

7 Questions about Course Materials

In this course, we will utilize an online discussion board called *Piazza*. Below is an official blurb from the Piazza team:

Piazza is a question-and-answer platform specifically designed to get you answers fast. They support LaTeX, code formatting, embedding of images, and attaching of files. The quicker you begin asking questions on Piazza (rather than via individual emails to a classmate or one of us), the quicker you'll benefit from the collective knowledge of your classmates and instructors. We encourage you to ask questions when you're struggling to understand a concept ... See this New York Times article to learn more about their founder's story:
http://www.nytimes.com/2011/07/04/technology/04piazza.html

In addition to recitation sessions and office hours, please use the Piazza Q & A board when asking questions about lectures, problem sets, and other course materials. You can access the Piazza course page either directly from the below address or the link posted on the Stellar course website:

https://piazza.com/mit/fall2014/17804

Using Piazza will allow students to see other students' questions and learn from them. Both the TA and the instructor will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. A student's respectful and constructive participation on the forum will count toward his/her class participation grade. *Do not email your questions directly to the instructors or TAs* (unless they are of personal nature) — we will not answer them!

8 Recitation Sessions

Weekly recitation sessions will be held in E53-485 on days and times to be determined in the first week of class. Sessions will cover a review of the theoretical material and also provide help with computing issues. The teaching assistant will run the sessions and can give more details. Attendance is strongly encouraged.

9 Notes on Computing

In this course we use R, an open-source statistical computing environment that is very widely used in statistics and political science. (If you are already well versed in another statistical software, you are free to use it, but you will be on your own.) Each problem set will contain computing and/or data analysis exercises which can be solved with R but often require going beyond canned functions and writing your own program.

In addition to the materials from the department's math prefresher (see above), there are many resources for R targeted at both introductory and advanced levels, including:

- Fox, John and Sanford Weisberg. 2010. *An R Companion to Applied Regression.* Sage Publications. (focused on regression analysis)

- Venables, W. N. and B. D. Ripley. 2002. *Modern Applied Statistics with S*, 4th ed. Springer. (general statistics)

- For specific questions about R, searching the CRAN website or Stack Overflow with appropriate keywords will often yield satisfactory results.

- As a last resort, you can post your question to the R help e-mail list, but be sure to read the posting guidelines before doing so, and follow exactly what they say. The list is run by a very busy group of people (you will frequently get answers from R Core team members) and they can be nasty if you are not respectful of the norms.

For Bayesian statistical modeling, we also use JAGS, a cross-platform, open-source software for Markov chain Monte Carlo (MCMC) via Gibbs sampling. JAGS uses syntax similar to R and comes with an easy-to-use interface with R.

10   Books

- Recommended books: We will read chapters from these books throughout the course. We strongly recommend that you at least purchase (1) either one of the first two books, (2) Jackman and (3) Gelman and Hill. These books will be available for purchase at COOP and online bookstores (e.g. Amazon) and on reserve in the library.

  - Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press. (A standard reference for applied econometricians for most topics covered in the first part of the course.)
  - Cameron, Colin and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press. (Slightly less standard, but covers most of the topics throughout the course.)
  - Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press. (A standard, non-technical textbook for Bayesian hierarchical models.)
  - Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences.* Wiley. (Introduction to Bayesian statistical modeling with political science applications. Good mix of theory and practice.)

- Optional books: These books are standard references for specific topics covered in this course. We will assign a chapter or two from them. Those chapters will be on electronic reserve. Nice books to have for advanced students, but no need to purchase only for this course.

  - McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. Chapman and Hall/CRC. (Generalized linear models)
  - Efron, Bradley and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap.* Chapman and Hall/CRC. (Bootstrap)
  - Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*, 2nd ed. Chapman and Hall/CRC. (A standard textbook on applied Bayesian statistics.)
  - Kalbfleisch, John D. and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*, 2nd ed. Wiley. (Survival analysis)

11   Tentative Course Outline

11.1   Generalized Linear Models and Extensions

Binary Outcome Models

1. Binary Logit and Probit Models

   Recommended:

   - Wooldridge Ch.15 or Cameron & Trivedi Ch.14

2. Theory of Maximum Likelihood Estimation

   Recommended:

- Wooldridge Ch.13 or Cameron & Trivedi Ch.5, 7.2–7.4
- Buse, A. 1982. "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note." *The American Statistician*, 36(3), 153–157.

3. Numerical Optimization

Recommended:

- Wooldridge Ch.12.7 or Cameron & Trivedi Ch.10

4. Bootstrap and Monte Carlo Approximation

Recommended:

- Efron & Tibshirani, Ch.6
- King, Gary, Mike Tomz and Jason Wittenberg, 2000, "Making the Most out of Statistical Analysis: Improving the Interpretation and Presentation." *American Journal of Political Science*, 44(2), pp.341–355.

Optional:

- Wooldridge Ch.12.8.2 or Cameron & Trivedi Ch.11

Discrete Choice Models

1. Multinomial Logit and Probit Models

2. Ordered Logit and Probit Models

Recommended:

- Wooldridge Ch.16 or Cameron & Trivedi Ch.15

Optional:

- Alvarez, R. Michael and Jonathan Nagler, 1995, "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science*, 39(3), 714–744.

Event Count Models

1. Theory of Generalized Linear Models

Recommended:

- McCullagh & Nelder, Ch.2
- Gelman & Hill, Ch.6

Optional:

- McCullagh & Nelder, Ch.9.1, 9.2

2. Event Count Models

   Recommended:

   - Wooldridge Ch.18 or Cameron & Trivedi Ch.20
   - Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review*, 95(4), 793–810.

Models for Panel and Multilevel Data

1. Fixed and Random Effects Models

   Recommended:

   - Wooldridge, Ch.10 or Cameron & Trivedi, Ch.21
   - Green, Donald P., Soo Yeon H. Kim and David Yoon, 2001, "Dirty Pool," *International Organization*, 55(2), 441–468.

2. Mixed Effects Models

   Recommended:

   - Gelman & Hill, Ch.11
   - Zorn, Christopher J.W., 2001, "Generalized Estimating Equation Models for Correlated Data: A Review with Applications," *American Journal of Political Science*, 45(2), 470–490.

   Optional:

   - Cameron & Trivedi, Ch.22.8, 24.6

11.2 Bayesian Statistical Modeling

Introduction to Bayesian Statistics

1. Basic Concepts of Bayesian Statistics

   Recommended:

   - Jackman, Ch.1, 2

2. Markov Chain Monte Carlo

   Recommended:

   - Jackman, Ch.4 (skim), 5 and 6

   Optional:

   - Jackman, Simon, 2000, "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo," *American Journal of Political Science*, 44(2), 375–404.

- Casella, George and Edward I. George, 1992, "Explaining the Gibbs Sampler," *The American Statistician*, 46(3), 167–174.
- Chib, Siddhartha and Edward Greenberg, 1995, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

Bayesian Statistical Modeling

1. Hierarchical Linear and Nonlinear Models

   Recommended:

   - Gelman & Hill, Ch.12, 13

   Optional:

   - Jackman, Ch.7
   - Gelman & Hill, Ch.14, 15

2. Missing Data

   Recommended:

   - Gary King, James Honaker, Anne Joseph and Kenneth Scheve. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. American Political Science Review , Vol. 95, No. 1 (Mar., 2001), pp. 49-69

3. Measurement and Item Response Theory

   Recommended:

   - Jackman, Ch.9.

   Optional:

   - Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science*, 52(1): 201–217.
   - Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review*, 98(2): 355–370.

11.3 Survival Analysis (Time Permitting)

1. Basic Concepts of Survival Analysis

   Recommended:

   - Box-Steffensmeier, Janet M. and Bradford S. Jones, 1997, "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science*, 41(4), 1414–1461.

   Optional:

   - Freedman, David A., 2008, "Survival Analysis: A Primer," *The American Statistician*, 62(2), 110–119.

- Wooldridge Ch.20 or Cameron & Trivedi Ch.17, 19

2. Parametric Regression Models

   Recommended:

   - King, Gary, James E. Alt, Nancy Burns and Michael Laver, 1990, "A Unified Model of Cabinet Dissolution in Parliamentary Democracies," *American Journal of Political Science*, 34(3), 346–871.
   - Warwick, Paul and Stephen T. Easton, 1992, "The Cabinet Stability Controversy: New Perspectives on a Classic Problem," *American Journal of Political Science*, 36(1), 122–146.

3. Semiparametric and Competing Risks Models

   Recommended:

   - Warwick, Paul, 1992, "Rising Hazards: An Underlying Dynamic of Parliamentary Government," *American Journal of Political Science*, 36(4), 857–876.
   - Diermeier, Daniel and Randy T. Stevenson, 1999, "Cabinet Survival and Competing Risks," *American Journal of Political Science*, 43(4), 1051–1068.