# POLS 505: Advanced Maximum Likelihood Estimation

## Introduction to Analyzing Categorical and Longitudinal Data

### 126 Herzstein Hall, Tuesday 9:00a-12:00p

Dr. Justin Esarey
Assistant Professor of Political Science
Rice University

Office: 108 Herzstein Hall
Office Hours: M W 2:30 – 3:30 PM
E-mail: justin@justinesarey.com
Phone: 678-383-9629

Syllabus Version: 1/13/2015

## COURSE OBJECTIVES AND LEARNING OUTCOMES

This course introduces students to statistical techniques and their application to problems in the social sciences.

Students will be able to:

1. understand, implement, and interpret models derived from the Generalized Method of Moments
2. employ and interpret appropriate models for:
    a. endogenous data
    b. multinomial choice data
    c. censored data
    d. duration/survival data
    e. count data
    f. dynamic time series
    g. panel (also known as time-series cross-section, or TSCS) data
3. Use the Stata and R software packages in order to implement and interpret the models above.

## GRADING POLICIES AND ASSIGNMENT DETAILS

### Grade Components:

- Homework: 25%
- Exam 1: 25%
- Exam 2: 25%
- Final Exam: 25%

**Grading Scale:**

| | |
|---|---|
| 100%-97%: A+ | 76.9%-73%: C |
| 96.9%-93%: A | 72.9%-70%: C- |
| 92.9%-90%: A- | 69.9%-67%: D+ |
| 89.9%-87%: B+ | 66.9%-63%: D |
| 86.9%-83%: B | 62.9%-60%: D- |
| 82.9%-80%: B- | >59.5%: F |
| 79.9%-77%: C+ | |

**Exams:** There will be three exams in this class, two midterms and a final. All exams are cumulative, but will focus on material learned since the last exam. You must complete each exam within the allotted time period, and must submit a typed LaTeX answer sheet. The exams are open book and open note, but you may not consult anyone for advice on the exam. The rough timing of the exams is indicated on the course outline, and specific times will be scheduled in consultation with the class.

**Homework:** Homework problem sets will be distributed during class. I encourage collaborative work on problem sets: the goal of a homework problem set is to help you learn the material and enable you to perform well on the (non-collaborative!) exams. With that said, simply copying another student's homework answers is not permitted and will be treated as academic dishonesty.

All homework answers must be typed in LaTeX.

**Attendance:** Attendance is mandatory in this class, and as graduate students I expect that attendance will not be a problem for you. Every class you fail to attend (without an acceptable excuse—see below) will result in a 2.5 percentage point deduction from your final grade. (I expect that this will never happen.)

Attendance penalties may be waived in the event of death in the immediate family (parent, spouse, sibling, or child) within 2 weeks before the due date, in the event of an unforeseeable medical emergency affecting yourself, your spouse, or your child, if you are participating in a pre-approved academic activity (e.g., a conference), or for other unforeseeable exigencies; all waivers are at the discretion of the instructor. Supporting documentation may be required to support an attendance penalty waiver.

**COURSE POLICIES**

**Late Work:** Assignments are due at the date and time I specify for the assignment. Late homeworks will be marked off at 5 percentage points for the first 24 hours late, and an additional 10 percentage points for every subsequent 24 hours late. For exams, the first hour late incurs a 5 percentage point penalty and each additional hour incurs a 10 percentage point penalty.

Late work penalties may be waived in the event of death in the immediate family (parent, spouse, sibling, or child) within 2 weeks before the due date, in the event of an unforeseeable medical emergency affecting yourself, your spouse, or your child, or for other unforeseeable exigencies; all waivers are at the discretion of the instructor. Supporting documentation may be required to support an attendance penalty waiver.

**Honor Code/Academic Misconduct:** All forms of academic misconduct will be handled according to the Rice University Honor Code. Details on the Honor Code are available at http://honor.rice.edu/honor-system-handbook/.

If you ever have any questions about what you should do to stay within the honor code on a particular assignment, PLEASE contact me with your question and I can assist you. I cannot guarantee a timely response unless you contact me at least 24 hours in advance of the time the assignment is due.

**Students with Disabilities:** If you have a disability and require accommodation in this class, please contact me as soon as possible (within the first two weeks of class) to discuss these accommodations. You will also need to contact the Disability Support Services Office (telephone extension: 5841) in the Allen Center.

**Syllabus Change Policy:** The policies of this syllabus (other than absence policies) may be changed by Prof. Esarey with advance notice.

## COURSE MATERIALS

### Required Texts:

- Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Other readings are available on the web or the OWL-Space website.

**Software:** This course will use both the R and Stata statistical software packages. R is free and available from http://cran.r-project.org/. Stata is available on computers furnished by the department of political science.

All homeworks, papers and presentation slides must be typed in LaTeX, but free software exists to make this easy to accomplish. I suggest using LyX (http://www.lyx.org/) in combination with MiKTeX on Windows (http://miktex.org/), MacTeX on Macintosh (http://www.tug.org/mactex/) or TeXLive on Linux (http://www.tug.org/texlive/).

All students must have a valid Rice e-mail address and login (and access to the OWL-space website) to participate in this course.

## COURSE OUTLINE AND ASSIGNED READINGS

### 0) Review of Matrix Algebra/Introduction to R

Questions: How do we use matrix notation to write about data and statistical models? How do we employ matrix methods in R?

Skills and Concepts:

- Matrix and vector notation
- Basic matrix operations
- Implementation of Matrix Algebra in R

Readings:

- None.

### 1) The Generalized Linear Model

Questions: How do we use matrix notation to write about data and statistical models?

Skills and Concepts:

- Writing the Generalized Linear Model (GLM) in matrix notation
- Deriving the ML estimator of the GLM in matrix notation
- Review of GLM for continuous (regression) and dichotomous (probit, logit) models

Readings:

- Cameron and Trivedi, Chapter 4, Sections 4.1-4.7; Chapter 5

**2) Models for Endogenous Data**

Questions: What is endogeneity, and what are its consequences for modeling and inference? How can we use instrumental variable models to overcome problems of endogeneity and confounding to estimate a causal relationship?

Skills and Concepts: inferential problems stemming from endogeneity; solving confounding and endogeneity problems with instrumental variables; two stage least squares analysis

Readings:

- Cameron and Trivedi, Chapter 4, Sections 4.8-4.9.
- Angrist, Joshua D. and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Chapter 4: "Instrumental Variables in Action," pp. 113-220.

**3) The Generalized Method of Moments**

Questions: How can the Generalized Method of Moments (GMM) be used to derive a statistical model? How do GMM and ML estimators compare?

Skills and Concepts: Constructing linear and IV models using GMM; comparing GMM estimators to ML estimators.

Readings:

- Cameron and Trivedi, Chapter 6.

**4) Specification Testing**

Questions: How can we diagnose specification problems with models? How can we choose the best model from a set of candidate models?

Skills and Concepts: m-tests; overidentification tests; Hausman test; BIC and AIC; Vuong likelihood-ratio tests; R-squared and pseudo-R-squared; residual analysis; in- and out-of-sample prediction; added-variable plots; DFFITS and DFBETAs.

Readings:

- Cameron and Trivedi, Chapter 8.
- Kutner, Nachtsheim, Neter, and Li. *Applied Linear Statistical Models, 5th ed.* McGraw Hill-Irwin. Chapters 9 and 10, "Model Selection and Validation" and "Diagnostics."

********** **Cutoff for Exam #1** **********

**5) Models for Multinomial Choice Data**

Questions: When the dependent variable is ordered or categorical, how do we determine the statistical relationship between this DV and various types of independent variables?

Skills and Concepts: the ordered and multinomial logit and probit models; the random utility interpretation of OL/OP and MNL/MNP models.

Readings:

- Cameron and Trivedi, Chapter 15.

**6) Models for Censored Data**

Questions: What are the consequences for GLM models when data is systematically censored (i.e., some observations are excluded from the data set)? How can we appropriately determine a statistical relationship from censored data?

Skills and Concepts: the Tobit and Heckman selection models.

Readings:

- Cameron and Trivedi, Chapter 16.

**7) Models for Duration/Survival Data**

Questions: When the dependent variable is a duration (i.e., a length of time before some relevant event occurs), how do we determine the statistical relationship between this DV and various types of independent variables?

Skills and Concepts: Hazard and survival functions; exponential, Weibull, and Cox proportional hazard models.

Readings:

- Cameron and Trivedi, Chapters 17 and 19.

**8) Models for Count Data**

Questions: When the dependent variable is a count of the number of times that some event occurs, how do we determine the statistical relationship between this DV and various types of independent variables?

Skills and Concepts: Poisson, negative binomial, and zero-inflated negative binomial models.

Readings:

- Cameron and Trivedi, Chapter 20.

********** **Cutoff for Exam #2** **********

**9) Identifying and Modeling Stationary Time Series Data**

Questions: What are the consequences for inference when data are multiple observations of a single unit over time and observations are correlated with one another according to temporal proximity? What procedures are available for drawing accurate inferences from data with this structure?

Skills and Concepts: stationarity; autocorrelation functions; autoregressive moving average (ARMA) models; autodistributed lag/error correction models.

Readings:

- Chatfield, Chris. 2003. *The Analysis of Time Series: An Introduction, Sixth Edition*. Boca Raton, FL: CRC/Chapman and Hall. Chapters 1-4.
- 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.

**10) Models for Panel Data I: Pooled Models with Standard Error Corrections**

Questions: What are the consequences for inference when data are multiple observations of a multiple units over time? What procedures are available for drawing accurate inferences from data with this structure in situations where the pooled estimator is unbiased and/or consistent?

Skills and Concepts: forms of unit heterogeneity; cluster-robust standard errors; cluster-adjusted standard errors; panel-corrected standard errors.

Readings:

- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3): 634-647.
- Cameron, A. Colin and Douglas L. Miller. 2010. "Robust Inference with Clustered Data." Chapter 1 in A. Ullah and D. E. Giles, eds., *Handbook of Empirical Economics and Finance*. Boca Raton, FL: CRC/Chapman and Hall.
- Esarey, Justin, and Andrew Menger. 2015. "Practical and Effective Approaches to Dealing with Clustered Data." Working Paper.

**11) Models for Panel Data II: Fixed and Random Effects Models**

Questions: What procedures are available for drawing accurate inferences from panel data with unit heterogeneity but no temporal dynamics?

Skills and Concepts: fixed and random effects models.

Readings:

- Cameron and Trivedi, Chapter 21.

**12) Models for Panel Data III: Dynamic Panel GMM Models**

Questions: What procedures are available for drawing accurate inferences from dynamic panel data with unit heterogeneity?

Skills and Concepts: Arellano-Bond and Blundell-Bond dynamic panel GMM models.

Readings:

- Cameron and Trivedi, Chapter 22.