



Data Analytics & Security

INTA 4803/6450

Instructor Info —



David Muchlinski



By Appointment



Habersham 147



www.davidmuchlinski.com



david.muchlinski@inta.gatech.edu

Course Info —



Preq: At least one seminar in probability theory and OLS regression highly recommended

About —

The vast majority of social science data do not meet the assumptions of the classic linear model and thus necessitate alternative regression models for the accurate estimation of regression coefficients and standard errors. This is especially true in the fields of International Relations and Comparative Politics where experimentation is difficult or unethical for variables of interest and researchers must rely on observational data. This seminar will introduce students to the practice and estimation of Maximum Likelihood Models, a class of regression models for non-normally distributed dependent variables.

Overview

Most social science data do not fit the assumptions of the classical linear regression model. Social science data are often categorized into discrete categories, or counts of events, or may be censored at the low or high ends of the distribution. Such variables pose problems for traditional Ordinary Least Squares. This course will introduce students to the estimation and derivation of maximum likelihood regression models for non-normally distributed variables. Topics covered will include models for binary outcomes, count variables, variables where time or duration is an inherent component of estimation, ordinal variables, and nominal variables. Additionally, this seminar will cover advanced topics in statistical analysis, time permitting.

As in previous seminars, students will gain experience in data analysis using the R statistical analysis software. By the end of this seminar, students will be proficient in the basics of regression modeling for social science data, able to clearly and effectively communicate their results, and have the necessary statistical knowledge to take additional seminars on more advanced topics not offered by the Nunn School.

Material

Required Texts

Scott Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, 7. Sage Publications

Eliason, S. R. (1993). *Maximum Likelihood Estimation: Logic and Practice* (No. 96). Sage.

King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press.

Recommended Additional Texts

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical Distributions* (Vol. 4). New York: Wiley.

Other

Any required reading not explicitly covered in Eliason and Long, including articles and book chapters, can be found using Google Scholar.

Grading Scheme

50% Homework (3 assignments)

50% Research Project

Grades will follow the standard scale: A = 89.5-100; B = 79.5-89.4; C = 69.5-79.4; D = 60-69.4; F <60. Curving is at the discretion of the professor and will only be utilized to maintain a normal grade distribution. "Rounding" of grades will only occur if a student has no outstanding assignments.

FAQs

? What are the Prerequisites for this Course?

! While no prerequisites are formally required, students should have a solid background in probability and statistics with previous coursework covering INTA 6003 and INTA 8001 or departmental equivalents. This course is an advanced course in regression modeling and assumes students are familiar with the basics of probability theory, matrix and linear algebra, and with the normal linear regression model, Ordinary Least Squares.

? What Programming Language is Used?

! This seminar will utilize R exclusively. However, if students have a preference for another programming language such as Python or STATA, that is okay.

? Why Take Another Stats Seminar?

! As I am fond of saying in my other seminars, political science, and the world more generally, has become increasingly data filled, and data driven. Knowing how to analyze these data and make data-driven recommendations based on empirical scientific research is necessary both to understand and to succeed in the modern world. Skills in MLE are essential for those looking to apply as analysts as the vast majority of variables you will encounter do not meet the assumption of the OLS model. Thus understanding the topics to be covered in this seminar are critical for analysis of almost all interesting social science variables.

Learning Objectives

- Students will be able to apply basic statistical skills to include quantitative and qualitative methodologies in academic and professional contexts within the field of international affairs.
- Students will be able to apply research skills to address problems in the field of international affairs.

Research Project

Graduate students will complete a semester-long research project culminating in a standard journal length article (8-10k words, standard APSA formatting rules) on a topic of their choosing. Graduate students must submit to me by week 5 a detailed prospectus detailing their research topic. They may collect their own data or utilize existing data in a novel way. Undergraduate students will complete a research project prospectus. This prospectus is to be a detailed proposal for “future” research and must identify a source of data and appropriate methodology to potentially analyze this data. Undergraduate students may present preliminary findings, but a fully fleshed out research project is not necessary. As co-authorship is common in the field, students may form groups of no more than 3 co-authors to submit a single co-authored final paper. These teams should be homogeneous with respect to graduate and undergraduate students except in circumstances negotiated with the professor.

Homework There are three homework assignments that will be assigned. Students are encouraged to work continuously on all homework assignments as they are challenging. All homework assignments are to be submitted on time in .html format using R Markdown. Late homework will not be accepted. All homework analysis is to be one's own, but collaboration for the purposes of problem solving or coding is encouraged. This means that students may work cooperatively in completing all homework assignments, but all submitted work should be unique to each individual. Code should not be strictly copied from other's work when avoidable, and especially, all analysis must be one's own. The use of AI like Chat GPT for *assistance* in writing code is acceptable, but students should seek to solve all problem sets without the use of AI first. Due dates for each assignment will appear on Canvas.

Make-up Policy and Late Work

Make-up assignments and exams will not be permitted unless in case of legitimate medical or other concerns which should be discussed privately with the professor to determine legitimacy. If an extension is granted, work must be submitted by that time. If a student submits late work without notifying the professor of any change in circumstances, such work will not be accepted and receive a score of zero. Undergraduates will be penalized one letter grade per day late, for up to 3 days after which the assignment will receive a score of zero.

Diversity and Inclusivity Statement

The Institute does not discriminate against individuals on the basis of race, color, religion, sex, national origin, age, disability, sexual orientation, gender identity, or veteran status in the administration of admissions policies, educational policies, employment policies, or any other Institute governed programs and activities. The Institute's equal opportunity and non-discrimination policy applies to every member of the Institute community. The Institute's affirmative action program, Title IX program, and related policies are developed in compliance with applicable law. Pursuant to Title IX, the Institute does not discriminate on the basis of sex in its education programs and activities. As such, the Institute does not tolerate any kind of gender-based discrimination or harassment, which includes sexual violence, sexual harassment, and gender-based harassment. Inquiries concerning the Institute's application of or compliance with Title IX may be directed to the Title IX Coordinator, Burns Newsome, burnsnewsome@gatech.edu, 404-385-5151. Additionally, inquiries concerning the application of applicable federal laws, statutes, and regulations (such as Title VI, Title IX, and Section 504) may be directed to the U.S. Department of Education's Office of Civil Rights at www2.ed.gov/ocr.

Accommodations for Students with Disabilities

Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Office of Disability Services at Suite 123, Smithgall Student Services Building, 353 Ferst Drive, 404-894-2563 (Voice); 404-894-1664 (TDD). For more information on Georgia Tech's policy on working with students with disabilities, please see review the Office of Disability Service's web page at <https://policies.ncsu.edu/regulation/reg-02-20-01/>. The Office of Disability Services collaborates with students, faculty, and staff to create a campus environment that is usable, equitable, sustainable and inclusive of all members of the Georgia Tech community. Disability as an aspect of diversity that is integral to society and Georgia Tech. If students encounter academic, physical, technological, or other barriers on campus, the Disability Services team is available to collaboratively find creative solutions and implement reasonable accommodations.

Academic Integrity

Academic dishonesty in the form of cheating or plagiarism will not be tolerated. In brief, plagiarism is defined, for the purposes of this class, as: copying, borrowing, or appropriating another entity's work and presenting it as your own in a any submitted assignment, deliberately or by accident. Acts of plagiarism will be reported in accordance with the Honor Code. In order to avoid being charged with plagiarism, if you use the words, ideas, phrasing, charts, graphs, or data of another person or from published material, then you must either: 1) use quotation marks around the words and cite the source, or 2) paraphrase or summarize acceptably using your own words and cite the source. The plagiarism policy is not restricted to books, but also applies to video and audio content, websites, blogs, wiki's, AI-generated content like Chat-GPT, and podcasts. Plagiarism includes putting your name on a group project to which you have minimally contributed. For information on Georgia Tech's Academic Honor Code, please visit <http://www.catalog.gatech.edu/policies/honor-code/> or <http://www.catalog.gatech.edu/rules/18/>. Any student suspected of cheating or plagiarizing on a assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations. The student will also receive a grade of zero on the assignment at the professor's discretion and be referred to the Department Chair, Associate Chair, or Graduate Committee Chair.

Class Schedule

SECTION 1: Into to MLE

Jan 9	Welcomes and Introductions	No Required Reading
Jan 11	Professor away for Conference	No Class Meeting
Jan 16-18	Intro to MLE	King Chs. 1-2
Jan 23-25	Intro to GLMs	Eliaison Ch. 1 Long Ch. 1
J 30-Feb 1	Simulation	TBD Problem Set 1 Released
Feb 6-8	MLE and the Normal Distribution	Elison Ch. 2-3 King Chs. 3-4 Long Ch. 2
Feb 13-15	Binary Variable Models: Logit and Probit	King Ch. 5-6 Long Ch. 3-4 Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, Insurgency, and Civil War. <i>American Political Science Review</i> , 97(1), 75-90. Collier, P., & Hoeffler, A. (2004). Greed and Grievance in Civil War. <i>Oxford Economic Papers</i> , 56(4), 563-595. Hegre, H. (2001). Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816–1992. <i>American Political Science Review</i> , 95(1), 33-48. Fjelde, H., Knutsen, C. H., & Nygård, H. M. (2021). Which Institutions Matter? Re-considering the Democratic Civil Peace. <i>International Studies Quarterly</i> , 65(1), 223-237. Stewart, M. A. (2018). Civil War as State-Making: Strategic Governance in Civil War. <i>International Organization</i> , 72(1), 205-226.
Feb 20-22	Models for Ordered Outcomes	Long Ch. 5 Bagozzi, B. E., Hill Jr, D. W., Moore, W. H., & Mukherjee, B. (2015). Modeling Two Types of Peace: The Zero-Inflated Ordered Probit (ZiOP) Model in Conflict Research. <i>Journal of Conflict Resolution</i> , 59(4), 728-752. Birch, S., & Muchlinski, D. (2018). Electoral Violence Prevention: What Works?. <i>Democratization</i> , 25(3), 385-403. Sarwari, M. (2021). Impact of Rebel Group Ideology on Wartime Sexual Violence. <i>Journal of Global Security Studies</i> , 6(2)

Feb 27-29 Models for Unordered Outcomes

Long Ch. 6

Bagozzi, B. E. (2016). The Baseline-Inflated Multinomial Logit Model for International Relations Research. *Conflict Management and Peace Science*, 33(2), 174-197.

Cook, S. J., Niehaus, J., & Zuhlke, S. (2018). A Warning on Separation in Multinomial Logistic Models. *Research & Politics*, 5(2).

DeTommaso, M. M., Schulz, M., & Lem, S. B. (2017). Choices of Justice: Effects of Civil War Termination on Postconflict Justice Mechanisms Implemented by the State. *International Journal of Transitional Justice*, 11(2), 218-238.

Humphreys, M., & Weinstein, J. M. (2008). Who Fights? The Determinants of Participation in Civil War. *American Journal of Political Science*, 52(2), 436-455.

Problem Set 2 Released

Mar 5-7 Professor away for Conference

No Class

Mar 12-14 Models for Counts

King Ch. 8

Long Ch. 8

Daxecker, U. E. (2014). All Quiet on Election Day? International Election Observation and Incentives for pre-Election Violence in African Elections. *Electoral Studies*, 34, 232-243.

Heldt, B. (2018). Atrocity Crimes as a Disease: A Statistical Approach to Early Detection. In *Preventing Mass Atrocities* (pp. 40-59). Routledge.

Bagozzi, B. E. (2015). Forecasting Civil Conflict with Zero-Inflated Count Models. *Civil Wars*, 17(1), 1-24.

Mar 19-21 Spring Break

Professor away for Conference

Mar 26-29 Survival/Duration Models

Box-Steffensmeier, J. M., & Jones, B. S. (1997). Time is of the Essence: Event History Models in Political Science. *American Journal of Political Science*, 1414-1461.

Box-Steffensmeier, J. M., Reiter, D., & Zorn, C. (2003). Non-proportional Hazards and Event History Analysis in International Relations. *Journal of Conflict Resolution*, 47(1), 33-53.

Greenstein, Claire, and Muchlinski, David. Reducing Mass Atrocities through Transnational Justice. *Working Paper*

Optional Reading

Box-Steffensmeier, J. M., & Jones, B. S., (2004). *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press.

April 2-4 Other MLE Models

King Ch. 9

Long Ch. 9

Balcells, L., & Kalyvas, S. N. (2014). Does Warfare Matter? Severity, Duration, and Outcomes of Civil Wars. *Journal of Conflict Resolution*, 58(8), 1390-1418.

Muchlinski, D. (2021). Swords and Plowshares: Property Rights, Collective Action, and Nonstate Governance in the Jewish Community of Palestine 1920–1948. *American Political Science Review*, 115(4), 1373-1387.

Problem Set 3 Released

Section 2: Additional Topics: Time Permitting

April 9-11 Missing Data

Little, R. J., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. John Wiley & Sons.

April 16-19 Hierarchical Modeling and Panel Data

TBD RESEARCH NOTE/PAPER

Due 22:00
