PLS 900: Statistical Models for Political Science Data

Fall 2015
M/W: 12:40-2PM
Room: 217 Bessey Hall

Instructor: Prof. Corwin D. Smidt
Office: 320 S. Kedzie Hall
Email: smidtc@msu.edu – I do not reply to messages over weekends.
Phone: 353-3292
Website: d2l.msu.edu

Office Hours: M/W 3-4PM, or by appointment

Course Description

The goal of this course is for you to gain sufficient working knowledge of statistical models that extend beyond basic least squares regression and that are common or have special relevance to contemporary political science research. Our focus is on familiarizing you with the skills and techniques commonly deemed essential for publishing quantitative research in today’s journals.

We will start by developing an understanding of likelihood as a mode of statistical inference and then apply popular maximum likelihood models within political science research. We will then proceed to discuss other methodological issues/practices common to political science and how we address/utilize them. We will mostly focus on generalizations of the linear model to other types of measures, but will potentially discuss: panel data/multilevel modeling, Bayesian estimation, semi/non-parametric estimation, matching, prediction, missing data, and beyond.

Those who successfully complete this course should have a sound working knowledge of maximum likelihood estimation and be able to successfully utilize all practices covered within this class in their own applied research. Students are also expected to gain an awareness of other advanced modeling techniques and topics, especially those that may be more specific to their research interests and which may be featured within advanced methods courses in or outside this department.

Course Prerequisites

Students should have already taken a course on probability and linear regression analysis (i.e., PLS 801 and 802).
Course Materials

Required Book

Recommended Books
You do not have to have these books to take this course.

**Comprehensive Econometric Books:** Cover course topics and much more. Each has its own strengths and weaknesses. Search the Internet, go to the library, or come to my office and read them if you are interested in knowing which one to buy:


**Software-Specific Books:** Most often you have the choice to complete class assignments in R or in Stata. While the code to estimate such models in both programs will be covered in class, you are certain to come across problems and questions with each program during your own estimation. In such cases I suggest you consider one or both of the following:


**In-depth Treatments of Specific Course Topics:**


**Computing Resources**

Most class assignments will allow you to use either R or Stata for completion, however others might require you to use a method only available within one program. If you are unfamiliar with either program I would recommend you familiarize yourself with it in this class. It is your responsibility to make sure you have adequate access to a computer for class assignments.

**Requirements and Grading**

Your grade will be comprised of the following components:

1. **Assignments (40%)**: After a couple weeks, assignments will given on an approximate bi-weekly basis. They will often require you to estimate various models, interpret your findings, present your results, and perform diagnostics using a statistical software
program. Along with answers, students need to include annotated copy of the code they used to develop their answers.

2. Application/Replication Paper (40%): You need a data set to test and write up a research question of your own or you will need to find an article and corresponding data set, preferably within your substantive area of interest, to perform a replication and/or modification testing its research question. Either way, your paper needs to use techniques learned in this class. You are required to write up a description and presentation of your analysis in a paper which will be due at the end of the semester. The paper should showcase your ability to use course methods to address substantive questions and your ability to discuss and present your findings in an accurate and persuasive manner. For data of papers to replicate go to the ICPSR website (http://icpsr.org/) or Harvard’s Dataverse website (http://dvn.iq.harvard.edu/dvn/).

3. Data Analysis Presentations (20%): Once we get into dichotomous dependent variables, each student will give an in-class presentation of his or her efforts to answer a research question using one of the models and presentation methods discussed in class. The goal of discussions is to evaluate how well the statistical method and presentation of results address your substantive research question or test your hypothesis. This presentation will often be made in coordination with your final paper work.

Grading in this class follows typical graduate school conventions. A 4.0 represents very good work, a 3.5 represents adequate completion of the course, and a 3.0 or lower generally indicates less than adequate or worse performance.

Note: For your benefit, I do not favor giving out incompletes. I also do not accept late assignments.

Schedule

Most of the journal articles are found on Jstor or alternative library databases. For those readings not available, electronic copies are available on the class website. We will see how far we get into this.

Core Content

1. Introduction, Notation, Software, and Other Basics

2. Understanding Likelihood Theory and Inference
   - Long Ch. 1-2
   - Pawitan p. 1-48 (on d2l)
   - King Ch. 4 (on d2l)

3. Dichotomous Dependent Variables: Logit/Probit, Extensions (Rare Events, Scobit, Heteroskedastic Models, Bivariate Probit . . . )
• Long Ch. 3

4. Getting the Substance from MLE: Interpretation, Predicted Probabilities, Marginal Effects, and Model Fit

• Long Ch. 4

5. Interactions, Variance Correction, and Bootstrapping in ML Models


6. Panel/Longitudinal/Multilevel Data Models: Between and Within Effects, Fixed Effects, Random Effects, Variance Correction Approaches

• Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models* (selected chapters)
• 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


7. Ordinal Dependent Variables: Ordered Logit and Probit and Extensions (Generalized Ordered Logit)
   • Long Ch. 5

8. Unordered or Semi-Ordered Dependent Variables: Multinomial Logit, Multinomial Probit, Mixed/Random Parameters Logit, Ranked Logit, Stereotype Logit
   • Long Ch. 6

   • Long Ch. 7

10. Counting Distributions: Poissons, Negative Binomial, Hurdle, and Zero-Inflated Models
    • Long Ch. 8

11. Duration Distributions (Event History): Parametric, Semi-Parametric (Cox), Discrete Time


**Additional Content:** Depending on student interest

1. Model dependencies, model comparisons (i.e., Bayes Factor), and model-independent inference (Matching, Regression Discontinuity)

   • Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference* (selected chapters)


   • Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. “Causal Inference Without Balance Checking: Coarsened Exact Matching.” *Political Analysis*


2. Bayesian methods


   • Jackman, Simon. nd. “Statistical Inference: Classical vs. Bayesian.” (on d2l)
3. Nonlinear structural equation and endogenous regressor models (Instrumental Variables, 2SLS, Control Function, Special Regressors, Causal Mediation).


4. Missing Data problems (Multiple Imputation, Ecological Inference): Probably just Gary King’s Stuff

5. Prediction/Variable Selection/Data-Mining (Random Forest; Neural Networks; Lasso; I-score)

**A Couple Last Things**

**Group Work and Academic Misconduct**

I recognize that working in groups is essential to scientific progress. You are allowed to collaborate with others to complete the assignment portion of this class. You can work together but your answers should be submitted individually and reflect your personal understanding of the topic. If you did work with others, please also indicate what other students you worked with on the assignment. Working in groups is not allowed for the paper component of this class.

Academic misconduct will not be tolerated. Cheating or plagiarism is an insult to me, your peers, and yourself; it is not to be tolerated. Instances of cheating will be handled according the school’s policy on integrity of scholarship and grades.

**Electronic Submissions**

As a general rule, students should always submit their work in paper form. If, under special circumstances, you are submitting a document electronically, then you need to submit it in an archival format. This means no modifiable Word/Text documents (.doc, .txt, .rtf) and instead formats where content is fixed (.pdf, .ps).