POLS W4291
Advanced Topics in Quantitative Research: Limited and Qualitative Dependent Variables

Fall 2008
Tuesdays and Thursdays, 9:10–10:25am
253 Engineering Terrace

Prof. Gregory Wawro 212-854-8540
741 International Affairs Bldg. gjw10@columbia.edu
Office Hours: Tuesdays, 3–5pm and by appt.

Course Overview

This course covers methods for empirical models that have dependent variables that are not continuous. These models include dichotomous and polychotomous response models, models for censored and truncated data, sample selection models, duration models, and models for count data.

Prerequisites: Students wishing to take this course should have taken an introductory course in probability, a course on regression analysis, and have basic knowledge of calculus and matrix algebra.

Course Requirements:The grading for the course is based on a mid-term and final exam (30% and 40%, respectively), and homework assignments (30%).

Lecture Notes: PDF versions of my lecture notes are available from the course web site (the URL is `http://www.columbia.edu/~gjw10/w4291.html`). Students should download and print up a copy of the notes for themselves so that they can follow along with the lectures. Students who do not have a copy of the notes will be at a severe disadvantage.

Texts: The following books are available at Book Culture on 112th St.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Cambridge University Press.

Cameron, A. Colin and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*, Cambridge University Press (also available in electronic format through CLIO).

Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice*, Sage Publications.

Fox, John. 2002. *An R and S-Plus Companion to Applied Regression*, Sage Publications.

Greene, William H. 2008. *Econometric Analysis*, Sixth Edition, Pearson Prentice Hall.

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

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

All of these books are useful and are strongly recommended. For several sections in the course, I have assigned overlapping readings that give somewhat different treatments of the material, usually varying by level of difficulty. Although there will be some redundancy, students should read all assignments in order to get the most out of the material.

All reading assignments from books are on reserve at Lehman. All reading assignments from journals are available through JSTOR (http://www.jstor.org/). I have assigned readings from, but have not ordered, the following book, which is available online (see the course web site for a link):

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*, Cambridge University Press.

Computing/Software: We will use `R` and `Stata` for computational exercises, with more emphasis on `R`. The latter is free and available for download from `http://www.r-project.org/`. Students are welcome to use other software packages that they are familiar with, but we will provide support only for `R` and students will be required to program in `R` for problem sets and exams. Be prepared to devote a good deal of time to programming, which is essential to becoming proficient in the methods covered in this course.

Course Outline

Based on prior experience with courses of this nature, I have not included dates to allow for maximum flexibility in covering the topics listed below. I will announce in class what readings students should do each week.

1. Course Motivation
    - Long, Chs. 1 and 2.
    - Maddala, Ch. 1.
2. Review of Fundamentals of Probability
    - Greene, Appendix B.
3. Maximum Likelihood Estimation
    - Greene, Ch. 16
    - Eliason, *Maximum Likelihood Estimation.*
4. Models derived from a common structure: the normal regression model

    4.1 Introduction
    - Long, Ch. 2.

    4.2 Probit

- Long, Ch. 3.
- Maddala, Ch. 2.1–2.5.

4.3 Ordered Probit

- Long, Ch. 5.

4.4 Censored Regression

- Long, Ch. 7.
- Maddala, 6.1–6.6.

4.5 Truncated Regression

- Maddala, Ch. 6.9–6.10.

4.6 Sample Selection Models

- Maddala, Ch. 9.

Take-home midterm about here

5. Probabilistic Choice Models

5.1 Introduction

- Review Long, Ch. 3 for binary logit.

5.2 The Multinomial Logit Model

- Long, Ch. 6.
- Maddala, Ch. 2.10–2.12, Ch. 3.

5.3 The Conditional Logit Model

- Long, Ch. 6.
- Maddala, Ch. 3.

5.4 The Nested Logit Model

- Greene, pp. 847–850.
- Maddala, pp. 67–70.
- Train, pp. 80–96.

5.5 The Multinomial Probit Model

- Greene, pp. 850–851.
- Train, pp. 101–114.
- Alvarez, R. Michael and Jonathan Nagler. 1995 "Economics, Issues, and the Perot Candidacy: Voter Choice in the 1992 Election," *American Journal of Political Science* 39:714–744.

5.6 Mixed Logit

- Train, Ch. 6

6. Duration Models

   6.1 Introduction

   - Box-Steffensmeier and Jones, Ch. 1.

   6.2 Functions for Analyzing Duration Data

   - Box-Steffensmeier and Jones, Ch. 2.

   6.3 Parametric Models

   - Box-Steffensmeier and Jones, Ch. 3.

   6.4 The Cox Proportional Hazard Model, Semi-parametric and Nonparametric Approaches

   - Box-Steffensmeier and Jones, Ch. 4.

   6.5 Diagnostics

   - Box-Steffensmeier and Jones, Ch. 8.

   6.6 Frailty and Split-population models

   - Box-Steffensmeier and Jones, Ch. 9.

7. Event Count Models

   - Long, Ch. 8.