

PS2730
MAXIMUM LIKELIHOOD ESTIMATION:
CATEGORICAL AND LIMITED DEPENDENT VARIABLE MODELS
Fall Semester 2021: Thursday 9:30-11:55 AM, 4430 Posvar Hall

Steven E. Finkel
4804 Posvar Hall
finkel@pitt.edu

Office Hours: T 10:30-1:00

Course Description

This course focuses on models for the analysis of categorical and other kinds of non-continuous dependent variables, which are typically estimated via maximum likelihood estimation methods. We will begin with logit and probit models for *binary* outcomes (e.g. individuals voting or abstaining in an election) and link those models to the broader “generalized linear model” framework. We’ll then cover maximum likelihood methods for estimating parameters in logit and probit (and many other) models, and discuss recent issues and problems in the specification, estimation and interpretation of these models. We’ll then move on to models for other common non-continuous dependent variables: *ordinal* outcomes (e.g., regimes being “repressive,” “semi-repressive,” or “free”); *nominal* outcomes with more than two unordered categories (e.g. voting for a “Left”, “Right”, “Populist” or “Green” party); *count* outcomes, e.g. the number of vetoes exercised by Presidents during a given term in office; and *censored* outcomes, where the values of observations above or below a certain point are unknown (e.g. “desired” campaign contributions where the observed amount is legally capped at a certain value). We will then consider models for categorical and limited variables in multilevel or longitudinal data structures (e.g., multiple observations over time for the same individuals, or multiple individuals observed in different countries). The final unit will cover methods that attempt to handle the problems of *endogenous regressor* and *sample-selection* biases in order to estimate causal effects in models for non-continuous outcomes.

The goals of the class are for you to be able to use these methods in your own research, and to understand and critique published works in the discipline that make use of these techniques.

Texts

Core:

Long, *Regression Models for Categorical and Limited Dependent Variables* (Sage, 1997)
Long and Freese, *Regression Models for Categorical Dependent Variables Using Stata, Third Edition* (Stata Press, 2014)

Supplemental:

Allison, *Fixed Effects Regression Models* (Sage, 2009)
Breen, *Regression Models: Censored, Sample Selected or Truncated Data* (Sage, 1996)
Liao, *Interpreting Probability Models: Logit, Probit and Other Generalized Linear Models* (Sage, 1994)
Rabe-Hesketh and Skrondal, *Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts, and Survival, Third Edition* (Stata Press, 2012)

Advanced References:

Gill and Torres, *Generalized Linear Models: A Unified Approach, 2nd Edition* (Sage, 2020)
Ward and Alquist, *Maximum Likelihood for Social Science: Strategies for Analysis* (Cambridge, 2018)

Requirements

Grades will be based on a 20-25 page research paper (40%), two homework exercises which relate to specific statistical methods and problems we will discuss (25% each), and an oral presentation of your research paper on December 9 or 16 (10%). The paper will be a quantitative analysis using methods from this course of data that you will collect or access from social science archives or other sources. The paper should have some substantive interest to you or be relevant to your studies in the graduate program; ideally, you can think of it as the first draft of a convention paper or possible journal publication. The paper will discuss your basic theoretical framework, your hypotheses, statistical models, results, possible problems with the analysis and what you may have done to correct or account for these problems. It will conclude with a discussion of the relevance of your findings for the general topic and for future research.

Course Outline

The course is organized by units and then topics within units. We will maintain a certain amount of flexibility with the schedule, so that we can spend more time on some topics/units and scale back on others as circumstances warrant.

UNIT 1: Models for Dichotomous Dependent Variables (September 2, September 9, September 16, September 23)

1. Logit, Probit and the Generalized Linear Model

Long, chapters 1, 3.

Long and Freese, chapters 4.

Fox, *Applied Regression Analysis and Generalized Linear Models*, chapter 15, pp. 379-385.

2. Maximum Likelihood: Estimation and Interpretation

Long, chapters 2, 4.

Long and Freese, chapters 3, 5.

Fox, *Applied Regression Analysis and Generalized Linear Models*, chapter 15, pp. 402-417.

3. Issues in the Estimation and Interpretation of Logit and Probit Models

Liao, chapters 1-3.

Long and Freese, chapter 6.

Hanmer, Michael, and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models", *American Journal of Political Science* 57(1): 263-277.

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models", *Annual Review of Sociology* 44: 39-54.

Rainey, Carlisle. 2016. "Dealing with Separation in Logistic Regression Models", *Political Analysis* 24: 339-355.

UNIT 2: Models for Ordered, Nominal, Count, and Censored Variables (September 23, September 30, October 7, October 14, October 21)

1. Ordered Outcomes

Long, chapter 5.

Liao, chapter 5.

Long and Freese, chapter 7.

Fullerton, Andrew. 2009. "A Conceptual Framework for Ordered Logistic Regression Models", *Sociological Methods and Research* 38(2): 306-347.

2. Nominal Outcomes

Long, chapter 6.

Liao, chapters 6-7.

Long and Freese, chapter 8.

3. Count Outcomes

Long, chapter 8.

Liao, chapter 8.

Long and Freese, chapter 9.

Rabe-Hesketh and Skrondal, *Multilevel and Longitudinal Modeling Using Stata Volume II: Categorical Responses, Counts and Survival*, chapter 13.

4. Censored Outcomes

Long, chapter 7, p.187-210.

Breen, chapters 1-2.

HOMEWORK 1 HANDED OUT SEPTEMBER 30; DUE OCTOBER 10

UNIT 3: Longitudinal and Multilevel Models (October 28, November 4, November 11)

1. Panel and Multilevel Models for Dichotomous Outcomes

Andress, Golsch and Schmidt, *Applied Panel Data Analysis for Economic and Social Surveys*, pp. 203-248.

Allison, *Fixed Effects Regression Models*, chapters 1-3.

Rabe-Hesketh and Skrondal, *Multilevel and Longitudinal Modeling Using Stata Volume II: Categorical Responses, Counts and Survival*, chapter 10.

2. Panel and Multilevel Models for Ordinal, Nominal and Count Outcomes

Allison, chapter 4.

Rabe-Hesketh and Skrondal, *Multilevel and Longitudinal Modeling Using Stata Volume II: Categorical Responses, Counts and Survival*, chapters 11-13.

**UNIT 4: Endogenous Regressor and Sample Selection Models
(November 18, December 2)**

Angrist and Pischke, "Instrumental Variables in Action", chapter 4 in *Mostly Harmless Econometrics: An Empiricist's Companion* (2009, Princeton University Press).

Breen, chapters 3-5.

Maydeu-Olivares, Shi, and Fairchild, Estimating Causal Effects in Linear Regression Models With Observational Data: The Instrumental Variables Regression Model, *Psychological Methods* 25, 2 (2020), 243-258.

Terza, Basu, and Rathouz, "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling", *Journal of Health Economics* 27 (2008), 531-543.

Stata Extended Regression Reference Manual (Release 17), pp.1 -67.

HOMEWORK 2 HANDED OUT NOVEMBER 18; DUE NOVEMBER 29

PAPERS DUE DECEMBER 8; PRESENTATIONS DECEMBER 9, DECEMBER 16