Discrete Choice Modeling
William Greene
Stern School of Business, New York University

Course Outline

This is a course in econometric analysis of discrete data. There are a huge variety of models that are used in this context. We will focus on four that arguably comprise the foundation for the area, the fundamental model of binary choice (and a number of variants), models for ordered choices, models for count data and the most basic model for multinomial choice, the multinomial logit model. The course will consist of a series of sessions of discussions of models and methods followed by laboratory sessions which will apply the techniques to 'live' data sets. Discussions will cover the topics listed below. Lab sessions will apply the techniques discussed in the preceding sessions. Practicals will consist of directed exercises and student assignments to be completed singly or in groups.

The main textbook references for the course are

Applied Choice Analysis by Hensher, D., Rose, J. and Greene, W., (Cambridge University Press, 2005).
Econometric Analysis, 6th ed., by Greene, W. (Prentice Hall, 2008). Five specific chapaters: Chapter 14, Estimation Methods; Chapter 16, Maximum Likelihood; Chapter 17, Simulation Based Estimation; Chapter 23, Discrete Choice Models, Chapter 25, Models for Count Data and Duration are included with the readings for the course.

I will also draw on three survey papers:

"*Modeling Ordered Choices*," W. Greene (with D. Hensher), 2010, Cambridge University Press.
"Functional form and Heterogeneity in Models for Count Data," W. Greene, *Foundations and Trends in Econometrics*, 2007
"Discrete Choice Modeling," W. Greene, *Palgrave Handbook of Econometrics*, Vol. 2, 2009 (forthcoming)

The received literature on discrete choice models is vast – one could easily compose a list of thousands of articles. Your course materials include a small handful of articles, some by your instructor with David Hensher, mainly for the purpose of illustrating particular techniques or ideas. These are:

"Economic Choices," *American Economic Review*, McFadden, D. (2001). (Nobel Prize lecture)
"Mixed MNL Models for Discrete Response," McFadden, D. and Train, K., *Journal of Applied Econometrics*, 2000.
"Convenient Estimators for the Panel Probit Model: Further Results," Greene, W., *Empirical Economics*, 2003
"Non-parametric regression for binary dependent variables," Frolich, M. *Econometrics Journa*l, 2006
"The Mixed Logit Model: The State of Practice," Hensher, D. and Greene, W., 2002.
"Choosing Between Conventional, Electric and LPG/CNC Vehicles in Single-Vehicle Households," Hensher, D. and Greene, W., 2000.
"Deriving Willingness to Pay Estimates from Observation Specific Parameters," Hensher, D., Greene, W. and Rose, J., 2004.
"Combining Sources of Preference Data," IATBR: Louviere. J. and Hensher, D., 2000
"A Latent Class Model for Discrete Choice Analysis: Constrast with Mixed Logit," Greene, W. and Hensher, D., 2003.
"Finishing High School and Starting College," Evans, W. and Schwab, R., QJE, 1995.

The software for the course is NLOGIT (Econometric Software, Inc., Version 4.0, 2007). The course materials include a short guide to using the program, "Introduction to NLOGIT."

Course Agenda

Session Topics

---

0 Introduction to the course
1 Methodology, software, modeling concepts, regression basics

2 Standard models for binary choice
Lab 1: Software. Using NLOGIT, Regression Computations

3 Analysis of binary choice. marginal effects, fit measures, prediction, hypothesis tests
Lab 2: Binary Choice Models, Estimation, Testing Hypotheses, Prediction, Analysis

4 Panel data models for binary choice, random effects, fixed effects, Mundlak formulation, incidental parameters problem, dynamic probit model
Lab 3: Binary Choice Panel Data Models

5 Choice model extensions: endogenous variables, bivariate probit models, simultaneous equations, sample selection, multivariate probit model, marginal effects, prediction and analysis
Lab 4: Model Building; Extended Binary Choice Models

6 Panel data, heterogeneity, simulation and latent class models,
Lab 5: Heterogeneity in Binary Choice Models

7 Ordered choice models, ordered outcomes, estimation and inference, generalized models, recent developments
Lab 6: Model Building for Ordered Choices

8 Models for count data – applications in health econometrics
Lab 7: Poisson and Negative Binomial Models for Count Data

9 The multinomial logit model, random utility models, IIA, nested logit modeling, fit and prediction, marginal effects, model simulation, heteroscedasticity, use of the MNL model
Lab 8: Multinomial logit models and multinomial choice models

10,11 Extensions of the MNL model, heteroscedasticity, multinomial probit, nested logit
12,13,14 Heterogeneity in multinomial choice models, latent class and mixed logit models
Lab 9: Multinomial probit, random parameters, latent class models

15 Repeated observations, panel data, revealed vs. stated preference data
Lab 10: Stated and Revealed Preference Data

Conclusion: Topics in discrete choice models; discussion, closing remarks
Lab 11: Student Projects

The course materials also include displays (Powerpoint presentations), a set of assignments, and a large number of data sets. Each of the numbered sessions is accompanied by a set of Powerpoint slides, **Part#-title.ppt**. There are electronic copies of these on the computer you are using, and you have a hard (paper) copy of them in your course materials. (Each of these is provided with a **.pdf** format counterpart in case you would like a more portable, durable file version.)

The course will include numerous hands on sessions using the NLOGIT computer program. Your course materials include a short introduction to NLOGIT that will get you started using the program. I will also give instruction on using the program during the labs.

The lab sessions will each cover a specific part of the course. Labs will consist of two parts. To begin, I will discuss particular kinds of models, and how to analyze them with NLOGIT. To help things along, we have presentation slides for these as well, all of the name **DCLab#.ppt**. Your course materials also contain hard copies of these slides. The commands that I use in these presentations are replicated in NLOGIT command files, so that you can do the computations easily without having to type them yourself – you need only upload the files, **DCLab#.lim** into NLOGIT, and the instructions can be executed immediately. The second part of each lab session will consist of your carrying out a short assignment using NLOGIT. These will show you different aspects of discrete choice model building with specific computations. Each of these will also provide an easy way for you to experiment with new features of the models and of NLOGIT. The assignments are in presentation files **LabAssignment-#.ppt** on your computers, and also in hard copy in your course materials. Since the assignments involve large amounts of NLOGIT computation, I have replicated the commands you need to carry them out in NLOGIT command files, **LabAssignment-#.lim** so that you can easily upload them to carry out the assignments. You can also easily change these files so that you can experiment with the models and the program.

We will use a large number of data sets in this course. The data files are in the form of 'limdep project files,' **name.lpj**. The file type .LPJ is 'registered with Windows on your computer, so you can launch NLOGIT and upload the data file by double clicking the project file name. Each file is also accompanied by a companion .LIM file, which contains text that describes the variables that are in the file. .LIM is also a Windows registered file type, so double clicking a .LIM file name in Windows launches NLOGIT and uploads the .LIM file into a text editing window. The data files that we will be analyzing are the following:

AmEx	American Express data base, 13,444 observations. Contains data on several binary and count variables.
brandchoicesSP	Discrete choice simulated data. Panel data set on brand choices. 800 people, 4 choices, 8 replications. Use for multinomial choice analysis
Burnett	Data on women's studies and gender economics. Use for study of a simultaneous equations binary choice model
clogit	Survey data on mode choice for travel between Sydney and Melbourne. 210 individual choices. Used for study of multinomial choice models
dairy	Panel data set on production, outputs and inputs. 247 Spanish dairy farms, 6 years. Used to illustrate how to use NLOGIT.
healthcare	Downloaded survey. Unbalanced panel data, 7209 individuals, 1 to 7 observations. Many health care variables. Binary choice and count data and a self administered ordered choice on health satisfaction.
innovation	Binary choice data on innovation by German manufacturing firms. Transformed form of panelprobit data set ussed to study binary logit models
labor	Mroz's 753 observations on women's labor supply. Use to study binary choice and sample selection.
panelprobit	Panel data, 1270 firms, 5 years. Binary choice data on innovation behavior. From published study on FDI, Imports and innovation.
Spector	Small sample on binary choice by high school students – whether grade on a test improved after an economics course. Used to introduce binary choice models.
SpRp	Large unbalanced panel on stated and revealed preference for travel mode. Used to study nested logit model and mixture of stated and revealed preference data.