MARK 8397 Modern Choice Modeling Spring 2011 Professor James Hess

Office Hours: W 12:00-4:00
375H Melcher, 713 743-4175
Home phone 713 663-7012
JHESS@UH.EDU

Modern Choice Modeling: Simulation Methods for Maximum Likelihood and Bayesian Estimation

This course focuses on the modern statistical analysis of discrete choices of the form, "There are K options and you must choose one." Inevitably, an empirical model of such discrete choices revolves around the computation of Probability[$\varepsilon_1 > \max\{\varepsilon_2, \ldots \varepsilon_K\}$]. In first-generation choice models, very special probability distributions were forced on us to make the multi-dimensional integration of this probability computational feasible (as in the multinomial logit model). With the recent advent of Markov chain Monte Carlo simulation procedures, the new generation of choice models are not restricted to such special distributions. Not only can we estimate the maximum likelihood of more satisfying models, but the simulation capability has dramatically increased the applicability of Bayesian estimation. This course will develop skills in applying modern modeling tools to real empirical data of discrete choices, both through homework exercises and a term project. In addition, the student will gain experience using the statistical programming language R, which has become a *de facto* standard among statisticians for the development of statistical software (R is the free version of the commercial S+ language).

Prerequisite: A doctoral level statistics or econometrics course.

Textbook: Train, Kenneth (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press. This book may be bought from online sellers such as Amazon, paperback about $33, but Train makes pdf versions of the chapters available for free on his webpage, http://www.econ.berkeley.edu/~train/distant.html. At the same location are his videotaped lectures based upon the book.

Additional Readings:

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

Lancaster, Tony (2004), *An Introduction to Modern Bayesian Econometrics*, Blackwell Publishing.

Kennedy, Peter (208), *A Guide to Econometrics*, 6th edition, Wiley-Blackwell.

Homeworks: To develop expertise in empirical analysis of choice data, you must experience data analysis. This will also allow you to become proficient with R. I will therefore ask you to do multiple homework assignments. The problem sets can sometimes be very challenging. To promote efficient learning, follow "Hess's Rule": if you have not been able to answer a problem after working on it for three hours, you must call me - at the office, at home, in the morning, or late at night - to get a hint or two. Seriously! My office and home phone numbers are listed above, but I am also in the telephone directory.

How good are your skills at typing mathematical and statistical equations? To enhance them, I'd like you to write-up your homework exercises answers using MS Word, using the Equation Editor to express mathematical formulas. There is a nice tutorial with videos at http://www.ist.uwaterloo.ca/ec/equations/equation.html. You can cut and paste output from R into Word. Submit hardcopy in class, **not** electronically online or via Blackboard Vista.

R Software: The statistical software language R is freely available at the Comprehensive R Archive Network (CRAN) http://www.cran.r-project.org . You should download and install R on your own computer. In session 2, we will begin to use R.

Computer Labs: Almost every class meeting I will be doing data analysis with R to repeatedly demonstrate that the abstract analysis of the textbook becomes real when data is confronted. In addition, several class sessions will be devoted to computer analysis. Please bring your laptop computer to class on those days so that you can participate in the computer lab.

Term Project: There are several hundred datasets on choice models available in R and textbooks; a list is included in our Blackboard bulletin board. You should select one of these that interests you and develop R software that implements a modern choice model. There is no rush to decide on the dataset and topic; on March 21, please turn in a one page statement of your research plans. I would like to meet with you later that week to discuss your project. You will get to communicate your analysis to the class as you would at an academic conference (15 minute presentation followed by 5 minutes Q&A) and in a paper (10 page maximum, double-spaced, excluding figures and tables).

Course Grading: Like in all doctoral courses, our goal is for you to do research that is publishable in the top academic journals (whose editorial boards are also receiving submissions from the best "veteran" researchers in the world). If your statistical analysis looks inexpert, how will you convince the editor that you have truly uncovered surprising but valid linkages between variables? I really want you to succeed in this endeavor, so I will provide evaluations of problem sets and project from the perspective of a journal editor. Personal improvement is all that matters to you and to me, but because the university requires that I give you a grade from a standardized set of options, I will combine my evaluations using the following weights: homework problems 50%, project presentation 10%, and project paper 40%.

Blackboard Vista : We will use the Blackboard Vista as a bulletin board to facilitate electronic communication. On our MARK 8349, I will post datasets, SPSS simpleton's guides, lecture notes for some topics and respond to your questions about statistical research. You can log onto Blackboard Vista from any computer that has Web access to http://uh.edu/webct/index.html**.**

Academic Honesty: The University of Houston Academic Honesty Policy is strictly enforced by the C. T. Bauer College of Business. No violations of this policy will be tolerated in this course. A discussion of the policy is included in the University of Houston Student Handbook, http://www.uh.edu/dos/hdbk/acad/achonpol.html. Students are expected to be familiar with this policy.

Accommodations for Students with Disabilities: The C. T. Bauer College of Business would like to help students who have disabilities achieve their highest potential. To this end, in order to receive academic accommodations, students must register with the Center for Students with Disabilities (CSD) (telephone 713-743-5400), and present approved accommodation documentation to their instructors in a timely manner.

Session	Date	Topic	Readings
1	Jan 19	Introduction to Discrete Choice	Skim for now McFadden's Nobel Prize Lecture in (2001) *AER*
2	Jan 24	**Computer Lab**: CRAN, R commander, data i/o, object manipulation, array and matrix operations, functions, if & for, graphs, shaping choice data	Kuhnert and Venables, "An Introduction to R." pp. 11-96, Croissant "Explanation of mlogit package"
3	Jan 26	Theory of Discrete Choice	Train 2
4	Jan 31	Probability and Statistics via R	Greene Appendix B
5	Feb 2	Regression Analysis via R	Long 2, Kennedy 1-3, Farnsworth, "Econometrics in R"
6	Feb 7	Normally Distributed Latent Utility: Probit Choice Model	Train 5.1-5.4, Amemiya (1981) *JEL*
7	Feb 9	**Computer Lab**: Programming in R, if & for, vector calculations, random variables, optimization, generalized linear model	Kuhnert and Venables, pp. 97-107
8	Feb 14	**Computer Lab**: Binary Probit Choice Models in R	http://www.ats.ucla.edu/stat/R/dae/probit.htm
9	Feb 16	Binary Logit models	Long 3
10	Feb 21	Maximum Likelihood Estimation	Myung (2003) *J Math Psychol*, Train 8
11	Feb 23	Simulation Based Statistics	Stern (1997) *JEL*
12	Feb 28	Drawing from Densities	Train 9.1-9.3
13	Mch 2	Estimation of Multinomial Probit with simulation	Train 5.6
14	Mch 7	Estimation with simulation	Train 10
15	Mch 9	**Computer Lab**: Multinomial Probit in R	Long 4,
16	Mch 21	Bayesian Statistics	Rossi and Allenby (2003) *MktSci*, Lancaster 1,Train 12
17	Mch 23	**Computer Lab**: Probit via MCMC Bayes	Imai and Dyk, MNP: R Package for Fitting the Multinomial Probit
18	Mch 28	Multinomial Logit models	Train 3
19	Mch 30	Mixed Logit	Train 6
20	Apr 4		
21	Apr 6	GEV/Nested Logit	Train 4
22	Apr 11	Marginal Effect on Probabilities, Moderation effects	Greene 23.4, Ai and Norton
24	Apr 13	Individual-level Parameters	Train 11
25	Apr 18	Endogeniety	Train 13
26	Apr 20	**Computer Lab**: Bayesian choice modeling	
27	Apr 25		
28	Apr 27	Student presentations	
29	May 2	Student presentations	

Supplemental Readings:

Chandukala, Sandeep R., Jaehwan Kim, Thomas Otter, Peter E. Rossi and Greg M. Allenby, "Choice Models in Marketing: Economic Assumptions, Challenges and Trends", *Foundations and Trends in Marketing*: Vol. 2: No 2, pp 97-184, 2007.

D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers of Econometrics*, New York, NY, Academic Press, 1974

D. McFadden, "Quantal Choice Models: A Survey," *Annals Econ and Soc Meas*, 5/4. 363-390, 1976

K. Train, "A Validation Test of a Disaggregate Mode Choice Model," Transportation Research, Vol. 12, pp. 167-174, 1978.

D. McFadden, "Modeling the Choice of Residential Location," in A. Karlquist, *et al.* (eds.), *Spatial Interaction Theory and Planning Models*, Amsterdam, North-Holland Publishing Company, 1978.

K. Train, *Qualitative Choice Analysis*, Cambridge, MA, MIT Press, 1986, Ch. 8

K. Train, D. McFadden, and M. Ben-Akiva, "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *RAND Journal of Economics*, Vol. 18, No. 1, pp. 109-123, 1987.

J. Hausman and D. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences," *Econometrica*, Vol. 48, No. 2, pp. 403-426, 1978.

S. Lerman and C. Manski, "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA, MIT Press, 1981.

M. Ben-Akiva and D. Bolduc, "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure," working paper, 1996, Department d'Economique, Universite Laval, Quebec, Canada.

A. Boersch-Supan and V. Hajivassiliou, "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variables," J*ournal of Econometrics*, Vol. 58, pp. 347-368, 1993.

V. Hajivassiliou, D. McFadden, and P. Ruud, "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives," *Journal of Econometrics*, Vol. 72, No. 1-2, pp. 85-134, 1996.

D. Revelt and K. Train, "Mixed Logit with Repeated Choices," *Review of Economics and Statistics*, Vol. LXXX, No. 4, pp. 647-657, 1998. Preprint version is viewable here , but not the published reprint.

D. Brownstone and K. Train, "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, Vol. 89, No. 1-2, pp. 109-129, 1998/99.

D. McFadden and K. Train, "Mixed MNL Models of Discrete Response,"*Journal of Applied Econometrics*, Vol. 15, No. 5, pp. 447-470, 2000. Preprint version is viewable here , but not the published reprint.

K. Train, "Recreation Demand Models with Taste Variation," *Land Economics*, Vol. 74, No. 2, pp. 230-239, 1998. Preprint version is viewable here , but not the published reprint.

M. Ben-Akiva, Denis Bolduc, and Joan Walker, "Specification, Identification, and Estimation of the Logit Kernel (or Continuous Mixed Logit) Model," DRAFT, February 2001.

D. McFadden, "Lectures on Simulation-Assisted Statistical Inference," [PostScript] presented at EC^2 Conference, Florence, Italy, 1996.

D. McFadden, "A Method of Simulated Moments for Estimation of Discrete Choice Models without Numerical Integration," *Econometrica*, Vol. 57, No. 5, pp. 995-1026, 1989.

A. Pakes and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, Vol. 57, No. 5, pp. 1027-1057, 1989.

L.-F. Lee, "On the Efficiency of Methods of Simulated Moments and Simulated Likelihood Estimation of Discrete Response Models," *Econometric Theory*, Vol. 8, No. 4, pp. 518-552, 1992.

V. Hajivassiliou and D. McFadden, "The Method of Simulated Scores with Application to Models of External Debt Crises," *Econometrica*, Vol. 66, No. 4, pp. 863-896, 1998.

M. Keane, "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, Vol. 62, No. 1, pp. 95-116, 1994.

V. Hajivassiliou and P. Ruud, "Classical Estimation Meethods for LDV Models using Simulation," in *Handbook of Econometrics*, R. Engle and D. McFadden (eds.), New York, NY, Elsevier Science, 1994.

C. Bhat, "Quasi-Random Maxium Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," working paper, Department of Civil Engineering, University of Texas, Austin, 1999, forthcoming in Transportation Research

K. Train, "Halton Sequences for Mixed Logit," working paper no. E00-278, Department of Economics, University of California, Berkeley, 2000.

C. Bhat, "Simulation Estimation of Discrete Choice Models Using Randomized and Scrambled Halton Sequences," working paper, Department of Civil Engineering, University of Texas, Austin, 2001.

D. Revelt and K. Train, "Customer-Specific Taste Parameters and Mixed Logit," working paper no. E00-274, Department of Economics University of California, Berkeley, 2000.

S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, Vol. 49, pp. 327-335, 1995.

J. Huber and K. Train, "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths," working paper no. E00-289, Department of Economics, University of California, Berkeley, 2000. Subsequently published in Marketing Letters, Vol. 12, No. 3, pp. 257-267, 2001.

J. Albert and S. Chib, "Bayesian Analysis of Binary and Polychomotomous Response Data,"*Journal of the American Statistical Association*, Vol. 88, No. 422, June 1993, Theory and Methods.

R. McCulloch and P. Rossi, "An Exact Likelihood Analysis of the Multinomial Probit Model," Journal of Econometrics, Vol. 64, No. 1-2, pp. 207-240, 1994.

G. Allenby and P. Rossi, "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, Vol. 89, No. 1-2, pp. 57-78, 1998/99.