CEE 6650: DISCRETE CHOICE MODELING

Spring 2024

Instructor

Dr. Laurie Garrow
Office: SEB 326
Email: laurie.garrow@ce.gatech.edu
Tel: 404-385-6634
Office Hours: By appointment
URL: http://garrowlab.ce.gatech.edu/ and https://airmobility.gatech.edu and https://ce-atlatgt.gatech.edu

Introduction

This course presents the theory and practice underlying the formulation and estimation of models of individual discrete choice behavior. The course will provide students with an understanding of the theory, methods, application and interpretation of multinomial logit (MNL), nested logit (NL) and other members of the Generalized Extreme Value (GEV) family of models. It will also include an introduction to mixed logit models, latent class models, and advanced modeling concepts (e.g., methods to account for price endogeneity). The models are used in many transportation applications. Within aviation, these applications include passenger no show and cancellation models, itinerary choice (used for schedule planning), and product selection with substitution effects (used in revenue management).

Required Background

This course will use concepts from probability, statistics, will require you to complete homework assignments and projects using Larch, an open-source Python library that provides tools for the estimation and application of logit-based discrete choice models. . I have designed this course to be accessible to students who have had an undergraduate probability and statistics course, but I am not expecting you to have prior programming experience in Python. As part of the course, you will learn about different discrete choice models and learn how to apply these techniques to ground and air transportation applications.

Practical Skill Sets You Will Acquire in Dr. Garrow's DCA Course

There are several goals that I had in mind when designing this course. The first was to expose you to the wide range of discrete choice models that are used in practice including advanced GEV and mixed logit models. The second is to introduce you to software that is commonly used in practice. To this end, we will be using Larch, developed by Jeff Newman, who works part-time as a research engineer with Dr. Laurie Garrow at Georgia Tech. Larch has been downloaded more than 100,000 times and is commonly used by the travel demand modeling community to estimate MNL and NL models for urban transportation applications. Multiple agencies, including FDOT and the Activity consortium, use Larch for estimating discrete choice models. Thus, as part of this course, you will gain practical experience in how to use this open-source tool to estimate discrete choice models.

Course Evaluation

The course grade is based on performance in several areas.

1 Midterm Exam	25%
Homework Assignments (equally weighted)	50%
Two Mini Projects (equally weighted)	20%
Class Attendance and Participation	5%

The instructor may provide opportunities throughout the course to earn extra credit (see the individual assignments for details). Extra credit is designed to improve the course for future students.

Course Organization
The course is organized into eight teaching modules.

Module 1: Introduction to Choice Theory
This module provides an introduction to choice theory including decision-making frameworks, utility functions, and general derivation of choice probabilities.

Module 2: Introduction to Python
This module provides an introduction to Python and will show you how to use training resources available through GT's LinkedIn Learning subscription.

Module 3: Binary and Multinomial Logit Models
This module covers fundamentals of discrete choice models including utility functions, generic versus alternative-specific variables, log-likelihoods and rho-square calculations and discrete choice data formats. We will also cover the derivation of the MNL model and discuss the independence of irrelevant alternatives (IIA) and other key properties of the MNL models, including adjustment procedures for choice-based samples.

Module 4: Model Specification, Development and Testing
Now that we know want a MNL models is, let's estimate some and interpret the results! This module covers different ways to interpret the results of discrete choice models including value of time calculations, constraining values, model comparison tests, model segmentation tests, odds ratios, and elasticities. We will also cover maximization methods used to find the parameter estimates for MNL and other discrete choice models.

Module 5: Nested Logit Models
Although the MNL model is arguably the most widely used model in practice, it does suffer from several limitations. In this module, we begin to look at ways we can relax the assumptions of the MNL model and estimate more general models. The nested logit model relaxes the IIA assumption by grouping alternatives into common "nests" and estimating a logsum parameter for the nest. In this module, we've cover the derivation of NL models, learn about the independence of irrelevant nests (IIN) and other properties of NL models, and normalized and non-normalized formulations for NL models.

Module 6: GEV and Count Models
Ready for more ways to generalize the MNL model? Get ready for GEV models that allow alternatives to be allocated to multiple nests! This is particularly relevant for airline itinerary choice models where competition across alternatives occurs along multiple dimensions (carrier, departure time of day, nonstop vs. connecting service, etc.). In this module, we'll cover the nested logit, ordered logit, network GEV, and other models belonging to the general GEV class of models. We'll also derive network GEV models and look at how these advanced GEV models have been applied to airline itinerary choice applications. Finally, we'll look at count models that are particularly relevant for modeling infrequent events, such as accident data.

Module 7: Mixed Logit and Latent Class Models
MNL, NL, and GEV models are convenient in the sense that they are "easy" and usually "fast" to estimate because no integration is required to compute the probability an individual selects a particular alternative. However, these models also assume that the estimated parameters represent the average "value" that the population places on a given attribute. But what if the value of time is distributed normal in the population? How can we incorporate distributions for values of time or other attributes? We can do this using mixed logit and/or latent class models! This module covers the random parameter and error component formulation for mixed logit models, looks at how to specify and interpret results from random parameter models, and covers

variance reduction methods that are important for speeding up the time required to estimate mixed logit models.

Module 8: Price Endogeneity and Advanced Topics

One of the most vexing problems we encounter in modeling demand for products and services is that prices tend to increase as demand increases. If we simply put demand and price into the same model, we may find that the model recommends raising prices to increase demand! In this module, we'll learn about techniques to control for this price endogeneity problem. Specifically, we'll cover instrumented variable approaches for discrete choice models and explore other advanced topics related to methods for controlling for spatial and/or temporal correlation for discrete choice model applications.

Learning Objectives

By the end of this course, you will be able to

- Identify the strengths and limitations of different discrete choice models and justify why a particular discrete choice model(s) should be used to analyze a research question.
- Describe how assumptions used to derive a multinomial logit model can be relaxed to incorporate flexible substitution patterns and correlation across alternatives and observations.
- Combine theory, judgment and statistical analysis to specify discrete choice models.
- Be comfortable programming in Python and using Larch to estimate discrete choice models
- Have the knowledge you need to pursue an internship or career in travel demand modeling

Course Textbook and Other Resources

Students taking the course are not required to purchase a textbook. However, several online references and texts will be used in the course.

Primary Text:

1. *Discrete Choice Modeling and Air Travel Demand: Theory and Application* by Laurie Garrow (2010 edition), available from Ashgate Publishing (not required to purchase).
2. *Discrete Choice Methods with Simulation* by Kenneth Train (2nd edition) published by University Press. Available through online bookstores, or via Kenneth Train's website at http://elsa.berkeley.edu/books/choice2.html.

Reference Texts:

Students who will be doing research in this area may want to purchase the following:

Discrete Choice Analysis: Theory and Application to Travel Demand (DCA) by Ben-Akiva and Lerman (1985 edition), available from The MIT Press, 55 Hayward Street, Cambridge, MA 02142, (1-800-356-0343 or mitpress-orders@mit.edu).

In addition, the following publication is a useful guide for modeling:

A Self Instructing Course in Mode Choice Modeling (2006). Prepared For U.S. Department of Transportation, Frank S. Koppelman and Chandra R. Bhat. Available online at http://www.caee.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf

In addition, I have pre-recorded my lectures and will make these available through a private YouTube channel. Note that I do not intend to run this course virtually – the reason I prerecord lectures is so that you can review material while doing homework assignment or preparing for exams and so that you can manage your workload and meet course deadlines. I realize that things come up – job interviews, being sick, etc. I have designed the course so that if you need to you

can work ahead or catch up when unexpected things happen (all HW assignments and lectures are available at the start of the semester).

Course Website
All materials are available on Canvas. However, to facilitate training of industry experts, I have also created a course website that contains all of the resources for the course, located here: https://sites.gatech.edu/ce-atlatgt/courses/. The site is password protected and the password will be distributed in class. Course communication will be through Canvas.

Policies on Homework and Exams
Please note that all assignments must be turned in via Canvas. on the due date at the beginning of class. Only medical reasons will be considered for late assignments. Only in extreme cases will late homework will be accepted with a penalty.

Collaboration and Group Work
Collaboration on homework assignments and projects is strongly encouraged. I realize that we have a wide variety of skill sets and background coming into the course, and I am a firm believer that one of the best ways to learn new material is by collaborating with others, particularly when you have complementary backgrounds. Students may work in groups of two or three for the homework assignments and submit a single assignment. Also, if you are particularly proud of the way you completed an assignment, let me know! As part of this collaborative learning environment, I ask for volunteers to present their solutions so that other students can learn and see how others approached the problem (and get programming tips!) You will have the opportunity to demonstrate how well you have mastered the course concepts on your own as part of the midterm exam.

Requests for Regrade
All requests for regrade must be made in writing using the "request for regrade form." I reserve the right to regrade the entire exam, homework, or project, thus please be aware that requests for regrades may result in raising OR lowering your grade OR having it stay the same. All requests for regrades must be submitted in writing within one week of when the instructor returns the HW/Exam/Project. There are no exceptions to this policy. I will not consider regrades at the end of the semester.

Academic Integrity
Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on GT's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/. Any student suspected of cheating or plagiarizing on an exam or homework will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities
If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404) 894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Course Outline

Week	Date	Assignment Due Dates	Mini Project Due Dates	Topics
1	1/8/2023			Introduction to the Course and Choice Theory
	1/10/2023			Python
2	1/15/2023			MLK - No class
	1/17/2023	HW 1		DCM Fundamentals Part 1
3	1/22/2023	HW 2		
	1/24/2023			
4	1/29/2023	HW 3		DCM Fundamentals Part 2
	1/31/2023			
5	2/5/2023	HW 4		MNL models
	2/7/2023			
6	2/12/2023	HW 5		Numerical Maximization Methods
	2/14/2023			
7	2/19/2023	HW 6		NL Models
	2/21/2023		Project 1	
8	2/26/2023	HW 7		GEV models
	2/28/2023			
9	3/5/2023			Ordered and Count Models
	3/7/2023		Project 2	
10	3/12/2023			Review for Midterm
	3/14/2023			Midterm Exam
11	3/19/2023			Spring Break - No class
	3/21/2023			
12	3/26/2023			Estimation with Simulation / Variance Reduction
	3/28/2023			
13	4/2/2023	HW 8		Mixed Logit
	4/4/2023			
14	4/9/2023	HW 9		Latent Class
	4/11/2023			
15	4/16/2023	HW 10		Price Endogeneity
	4/18/2023			
16	4/23/2023			Slack

DISCLAIMER
The instructor reserves the right to amend this syllabus as necessary. Any changes will be announced in class. (Updated: July 30, 2023)