

CHOICE MODELING IN R

CLASS 1

Wiktor Budziński

AIM OF THE CLASS

Provide a structured overview of modern methods for modeling individual choice behavior

- With emphasis on economic foundations and interpretation

Develop both:

- Theoretical understanding of choice models
- Practical skills to implement and estimate models in R (Apollo)

Enable students to critically interpret estimation results and conduct applied research

Offer teaching comparable in scope to specialized international choice modeling courses

WHO AM I?

Adjunct professor at the University of Warsaw with over 10 years of experience in choice modeling

- I have published over 20 articles related to choice modeling and presented them on multiple international conferences

Worked in multiple research projects when I used choice modeling for non-market valuation, demand predictions and market simulations

My website: [Wiktor Budziński – ACEP](#)

I also co-developed a package for estimation of choice models in MATLAB:

- [czaj/DCE: Models for Discrete Choice Experiments](#)

CLASS ORGANIZATION

Room A102, Thursdays, 9:45-11:15

- In principle the class is stationary. In some special cases the meeting may take place online on Zoom. We will discuss it beforehand if needed.
- 14 meetings in total

This is a workshop-type class that will consist of lecture, software demonstration (case studies), and individual exercises

- Exercises will not be graded, but I encourage you to try doing them to better understand the material and familiarize yourself with the software

We will be mostly relying on the Apollo package in R

- <https://www.apollochoicemodelling.com/>
- <https://www.sciencedirect.com/science/article/abs/pii/S1755534519300703?via%3Dihub>

CLASS ORGANIZATION

All the materials for the class will be posted on the subpage of my website

- <https://www.acep.uw.edu.pl/budzinski/choicemodeling/>

In-person consultations will take place on Thursdays, at 3 PM in room B307

- Please let me know beforehand by email that you plan to attend
- If you would want to meet on a different day, then we can arrange an online consultation through Zoom

The grading will be based on the conducted research project

- More details will be provided later in class
- Basically, it will require you to conduct an analysis of choice data using Apollo and writing a short report based on it
- The submission deadline will be announced later, but it will be sometime in the exam session

CLASS ORGANIZATION

This is **the first iteration** of this class, so some stuff I will have to figure out as we go

- Sorry in advance
- This also means that if you are interested in some particular aspect of choice modeling, then you may let me know, and I will try to cover it

WHAT WILL BE COVERED (WIP)

1. Introduction to choice modeling and Apollo package
2. Random utility model and multinomial logit (MNL)
3. MNL extensions
 - Controlling scale differences
 - Nested logit
4. Unobserved preference heterogeneity – mixed logit
5. Unobserved preference heterogeneity – latent class logit
6. Random regret minimization and decision process heterogeneity
7. Attribute non-attendance
8. Hybrid choice models
9. Dealing with endogeneity in choice models
10. Multiple discrete-continuous choice
 - Combination of discrete choice and count data models
 - MMDCEV model
11. Simulations and programming your own choice model

CHOICE MODELING

INTRODUCTION



ROLE OF CHOICE IN ECONOMICS

Microeconomic theory is built upon the concept of utility

- Consumer is assumed to have stable, well-behaved preferences that can be summarized by a sort of well-being index that we call utility
- Consumer is then assumed to maximize this index (utility) as they go through life and make decisions
- Maximization is subjected to constraints: income, time, etc.

Utility function cannot be directly observed

- We can only observe the effect of the utility maximization
 - Individual choices
 - Aggregated choices (e.g., aggregated demand)

As such, choices are a fundamental building block of an applied microeconomic analysis

- Choices reveal consumers' preferences

TYPES OF CHOICES

Discrete choice

- This is the main focus of the class
- Consumer chooses 1 out of J mutually exclusive alternatives
 - Examples: Presidential elections, mode of transportation to work, brand choice (e.g., phone), residential location choice, university choice

Multiple discrete choice

- Consumer chooses multiple alternatives out of J available alternatives
 - Examples: Streaming services, grocery basket choice, transportation portfolio

Multiple discrete-continuous choice

- Consumer chooses continuous quantity for each out of J alternatives
 - Examples: Grocery choice – how much of different types of meat do you purchase, how many times did you visited national parks
- Sometimes it can be represented as a repeated discrete choice

TYPES OF CHOICES

- Different types of choices lead to different types of dependent variables
- They require different econometric models

Discrete choice

- This is the main focus of the class
- Consumer chooses 1 out of J mutually exclusive alternatives
 - Examples: Presidential elections, mode of transportation to work, brand choice (e.g., phone), residential location choice, university choice

Multiple discrete choice

- Consumer chooses multiple alternatives out of J available alternatives
 - Examples: Streaming services, grocery basket choice, transportation portfolio

Multiple discrete-continuous choice

- Consumer chooses continuous quantity for each out of J alternatives
 - Examples: Grocery choice – how much of different types of meat do you purchase, how many times did you visited national parks
- Sometimes it can be represented as a repeated discrete choice

CHOICE MODELS VS. MACHINE LEARNING

Why use choice models in the first place

- They are rooted in the microeconomic theory
 - Or other, more “behavioral” theory (i.e., random regret model)
- Allow for microeconomic inference:
 - Welfare analysis (calculating consumer surplus, willingness to pay, etc.)
 - Theory-driven counterfactual policy simulations
 - Provide insights about consumers’ behavior

In contrast, machine learning models are mainly optimized for prediction

- Abstract from the theory
 - Coefficients do not have a direct interpretation

In the end, the choice of the method depends on your goal

- There is some effort in research to combine the two

BASIC DISCRETE CHOICE MODEL

Vast majority of discrete choice models are conceptualized using the Random Utility Theory

$$U_{ij} = V(\mathbf{X}_{ij}, \boldsymbol{\beta}) + \varepsilon_{ij}$$

Utility is assumed to be additively separable with respect to deterministic (systematic) portion of the utility $V()$ and the stochastic term ε_{ij}

The error term captures

- Unobserved attributes of the choice
- Unobserved preference heterogeneity
- Measurement error
 - For example, consumers do not take into account actual value of the attribute, but its perception

The error term does not necessarily imply irrationality

- Individuals are assumed to take it into account when making choices
- We just do not observe what it consists of

BASIC DISCRETE CHOICE MODEL

The deterministic part of the utility $V(\mathbf{X}_{ij}, \boldsymbol{\beta})$ is a function of alternative's attributes

The discrete choice differs from the standard microeconomic formulation of utility which is defined over quantities consumed $V(Q_{i1}, Q_{i2}, \dots, Q_{iJ})$

- This is because in DCMs consumer can only select one unit of a single alternative

Instead, DCM is operationalized by Lancaster's Theory of Value

- Consumers derive utility not from goods themselves, but from the characteristics those goods embody
- We will call these characteristics attributes of the alternative

Example: mode choice

- Attributes could be travel cost, travel time, comfort, risk of being late

BASIC DISCRETE CHOICE MODEL

Usually, the choice models have linear specification of the utility

$$V(\mathbf{X}_{ij}, \boldsymbol{\beta}) = \mathbf{X}_{ij}\boldsymbol{\beta}$$

This implies that model's coefficients can be interpreted as marginal utilities

- Kind of – we will get back to it

$$\frac{\partial U_{ij}}{\partial \mathbf{X}_{ij}} = \boldsymbol{\beta}$$

As such, we can easily calculate marginal rates of substitution as a ratio of coefficients

- If we have a monetary attribute, it can be interpreted as willingness to pay

BASIC DISCRETE CHOICE MODEL

Linking utility to choices is done by assuming utility maximization

$$P(y_{ij} = 1) = P(U_{ij} = \max(U_{ik})) = P\left(\bigwedge_{k \neq j} U_{ij} > U_{ik}\right)$$

The specific model will depend on the assumptions regarding error term

- If we assume iid Gumbel distribution (extreme value) the probability formula will have a closed-form of **multinomial logit**

$$P(y_{ij} = 1) = \frac{\exp(V(\mathbf{X}_{ij}, \boldsymbol{\beta}))}{\sum_{k=1}^J \exp(V(\mathbf{X}_{ik}, \boldsymbol{\beta}))}$$

BASIC DISCRETE CHOICE MODEL

Important characteristics of the discrete choice models

- Utility is level-free, so we cannot estimate the constant term. It is normalized to 0
- Utility is scale-free, so we cannot estimate the variance of the error term
 - In MNL it is normalized to $\frac{\pi^2}{6}$

APOLLO

INTRODUCTION



APOLLO

Apollo is an R package created by Stephane Hess and David Palma (*University of Leeds / Choice Modelling Centre*)

- Created by choice modeling experts for choice modeling experts
- The main philosophy of the package is flexibility and free-access
- On the website there is available a manual ([Apollo.pdf](#)) as well as multiple code examples

Software versions I am running for this class:

- R: 4.5.2
- R studio: 2026.01.0+392
- Apollo: v0.3.6

APOLLO

With the flexibility comes a cost - running a model in Apollo is pretty code-intensive

- Requires specification of multiple blocks of code

Main blocks of code:

- *apollo_control*
 - A list of settings for Apollo
- *apollo_beta*
 - A list of coefficients with their names and starting values
- *apollo_inputs*
 - An object with data, betas, controls etc.
- *apollo_probabilities*
 - A function that specifies the choice model. This is the most important part of the code
- *apollo_estimate()*
 - Function that runs the model
- *Apollo_modelOutput()*
 - Function that prints model's results

EXERCISE

Analyze consumers' preferences for the air travel using MNL

Pick Your Preferred Flight

Three flight options are described for your trip from Chicago to San Diego . These are options that might be available on this route or might be new options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that everything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

FEATURES	Non-Stop (Option 1)	1 Stop (Option 2)	1 Stop (Option 3)
Departure time (local)	6:00 PM	4:30 PM	6:00 PM
Arrival time (local)	8:14 PM	8:44 PM	9:44 PM
Total time in air	4 hr 14 min	4 hr 44 min	4 hr 44 min
Total trip time	4 hr 14 min	6 hr 14 min	5 hr 44 min
Legroom ☐	typical legroom	2-in more of legroom	4-in more of legroom
Airline [Airplane]	Depart Chicago Continental Airlines [B737] to San Diego	Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego	Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego
Fare	\$565	\$485	\$620

1. Which is MOST attractive? ☐ Option 1 ☐ Option 2 ☐ Option 3

2. Which is LEAST attractive? ☐ Option 1 ☐ Option 2 ☐ Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. ☐ Yes ☐ No

EXERCISE

Analyze consumers' preferences for the air travel using MNL

Pick Your Preferred Flight

Three flight options are described for your trip from Chicago to San Diego . These are options that might be available on this route or might be new options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that everything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

FEATURES	Non-Stop (Option 1)	1 Stop (Option 2)	1 Stop (Option 3)
Departure time (local)	6:00 PM	4:30 PM	6:00 PM
Arrival time (local)	8:14 PM	8:44 PM	9:44 PM
Total time in air	4 hr 14 min	4 hr 44 min	4 hr 44 min
Total trip time	4 hr 14 min	6 hr 14 min	5 hr 44 min
Legroom ☐	typical legroom	2-in more of legroom	4-in more of legroom
Airline [Airplane]	Depart Chicago Continental Airlines [B737] to San Diego	Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego	Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego
Fare	\$565	\$485	\$620

1. Which is MOST attractive?

Option 1

Option 2

Option 3

2. Which is LEAST attractive?

Option 1

Option 2

Option 3

Let's go to  