Lecture 15: Mixed Logit Model I

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Generalized Extreme Value Models Example in R

Today

- Mixed Logit Model
- Random Coefficients Logit
- Mixed Logit Substitution Patterns
- Mixed Logit and Panel Data
- Mixed Logit Estimation

Upcoming

- Reading for next time
  - Optional: Revelt and Train (1998)
- Problem sets
  - Problem Set 3 is posted, due October 31

Discrete Choice Models

Logit model

- Strong assumption that unobserved components of utility are i.i.d.
- Simple closed-form expressions for choice probabilities
- Taste variation can only be represented by observable data
- Panel data applications are limited by the i.i.d. assumption

Nested logit model

- Correlation between unobserved components of utility can be modeled
- Choice probabilities are more complex but still closed-form
- Taste variation can only be represented by observable data
- Panel data applications are limited by an i.i.d. assumption

What if we want a richer representation of taste variation, more flexible substitution patterns, and an ability to use panel data?

- Mixed logit

Mixed Logit Model

Mixed Logit Model

The mixed logit model overcomes three limitations of the logit model

- Random taste variation
- Unrestricted substitution patterns
- Correlations in unobserved factors over time

What is the catch?

- Mixed logit does not have a closed-form solution
- Estimation requires simulation

Mixed Logit Coefficients

How does the mixed logit model achieves this level of flexibility?

- It does not take a set of coefficients, β
- It takes distributions of coefficients, $f(\beta \mid \theta)$

Distributions of coefficients can

- Model distributions of unobserved taste among the sample of decision makers
- Impose correlations in unobserved utility among alternatives
- Represent individual preferences over time

Mixed Logit Choice Probabilities

The choice probability for the mixed logit model is

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta) d\beta$$

where $L_{ni}(\beta)$ is the logit probability at a given set of coefficients, β,

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$

and $f(\beta \mid \theta)$ is a density function of coefficients β, which depends on a vector of parameters, θ

If we assume representative utility is linear, $V_{ni} = \beta' x_{ni}$,

$$P_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} f(\beta \mid \theta) d\beta$$

Random Coefficients Logit

Random Coefficients

The utility that decision maker n obtains from alternative j is

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$$

- x_{nj}: data for alternative j and decision maker n
- β_n: individual-specific vector of coefficients
- ε_{nj}: i.i.d. extreme value error term

Just as we do not observe ε_{nj}, we also do not observe β_n

- We must use the data we observe to infer how β varies throughout the population of decision makers

We model β as a random variable with density $f(\beta \mid \theta)$

- θ is a set of parameters that describes the density of β
  - For example, the means and variance-covariance of β in the population

Random Coefficients Logit

Each decision maker knows their own β_n and ε_{nj}, so the choice is deterministic from their perspective

- n chooses i if and only if $U_{ni} > U_{nj}\ \forall j \neq i$

We (the researchers) do not observe β_n or the ε_{nj} terms

- If we did know β_n, then the choice probability would be standard logit, and we can write down this *conditional* probability as

$$L_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj}}}$$

But we do not know β_n, so we have to integrate over the density of the random coefficients

$$P_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} f(\beta \mid \theta) d\beta$$

Distributions of Random Coefficients

We previously used the terms“parameter” and “coefficient” interchangeably, but they mean different things in the mixed logit model

- β: Coefficients that appear in the utility expression
  - We will not actually estimate these coefficients in the mixed logit model
- θ: Parameters that define the density of random coefficients
  - We will estimate these parameters in the mixed logit model

Some examples of distributions of random coefficients that we can model

- Normal: $\beta \sim N(b, W)$, we estimate $\theta = (b, W)$
- Log-normal: $\ln \beta \sim N(b, W)$, we estimate $\theta = (b, W)$
- Uniform: $\beta \sim U(b, s)$, we estimate $\theta = (b, s)$
- Triangular: $\beta \sim Tri(b, s)$, we estimate $\theta = (b, s)$
- Many other distributions to choose from
- We can also model covariances between random coefficients, or model each coefficient as being independent

Mixed Logit Substitution Patterns

Alternative Motivation for Mixed Logit

The mixed logit model can also be motivated as a way to represent correlations in the utilities of alternatives

Let the utility of alternative j be expressed as

$$U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj}$$

- x_{nj}, z_{nj}: data for alternative j and decision maker n
- α: vector of fixed coefficients
- μ_n: vector of random coefficients with mean zero
- ε_{nj}: i.i.d. extreme value error term

Correlated Alternatives in Mixed Logit

With this representation of utility, the random or unobserved portion is

$$\eta_{nj} = \mu_n' z_{nj} + \varepsilon_{nj}$$

which creates correlations among utilities

$$Cov(\eta_{ni}, \eta_{nj}) = z_{ni}' W z_{nj}'$$

where W is the variance-covariance matrix of μ_n

This representation generalizes the logit and GEV models

- Logit: $z_{nj} = 0 \ \forall j$
- Nested logit: z_{nj} is a vector of nest dummy variables
- Paired combinatorial logit: z_{nj} is a vector of pairwise dummy variables
- Heteroskedastic logit: z_{nj} is a vector of alternative dummy variables

Random Coefficients or Correlated Utilities?

We have motivated the mixed logit model in two different ways, but the models are identical

Starting from the random coefficients expression, we have

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$$

We can express β_n as $\beta_n = \alpha + \mu_n$ by decomposing β_n into a mean, α, and a deviation from the mean, μ_n, which gives

$$U_{nj} = \alpha' x_{nj} + \mu_n' x_{nj} + \varepsilon_{nj}$$

Let z_{nj} be the subset of data with random coefficients, so utility is

$$U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj}$$

Your motivation will affect which coefficients you model as random and whether to allow correlations between the coefficients

Mixed Logit Substitution Patterns

The mixed logit model does not exhibit independence from irrelevant alternatives

- P_{ni}/P_{nj} depends on all the data of all alternatives

The elasticity of alternative i with respect to the mth attribute of alternative j is

$$E_{nix_{nj}^m} = -\frac{x_{nj}^m}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nj}(\beta) f(\beta \mid \theta) d\beta$$

- x_{nj}^m: data for the mth attribute of alternative j
- β^m: mth element of β
- This expression depends on the covariance between $L_{ni}(\beta)$ and $L_{nj}(\beta)$ as you integrate over the values of β, which is determined by which parameters you specify as random and which have covariances

Mixed Logit and Panel Data

Mixed Logit and Panel Data

The structure of the mixed logit model allows for more flexibility in representing how a single decision maker makes many choices over time, so it provides a better model for most panel data settings

- We still treat ε as an i.i.d. error term that represents utility not captured by the rest of the model
- But the mixed logit model allows for unobserved taste heterogeneity through random coefficients, which yields correlations in utility over time
- In the logit and nested logit models, all unobserved factors are represented by ε, which is i.i.d. so those models cannot accommodate correlations over time

Mixed Logit Model with Panel Data

We add at time index, t, to our representation of utility but still assume the coefficients, β_n, are constant for an individual

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt}$$

and we consider the vector of alternatives that decision maker n chooses over the T time periods

$$i = (i_1, \ldots, i_T)$$

Then the conditional logit probability (conditional on a set of coefficients) for a sequence of choices is

$$L_{ni}(\beta) = \prod_{t=1}^{T} \frac{e^{\beta_n' x_{ni_t t}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj_t t}}}$$

and integrating over the density of coefficients gives the choice probability

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta) d\beta$$

Dynamics in a Mixed Logit Model

Some "dynamics" can be represented in a mixed logit model using panel data

- Past and future exogenous variables can be included to model lagged or anticipatory behavior
- Lagged dependent variables can be included to represent state dependence

This approach is a relatively naive way of incorporating dynamics into a discrete choice model

- We are essentially modeling a sequence of static choices
- A fully "dynamic discrete choice model" would model how every choice affects all subsequent choices
- More on this later in the semester...

Mixed Logit Estimation

Mixed Logit Estimation

Mixed logit choice probabilities do not have a closed-form expression

$$P_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} f(\beta \mid \theta) d\beta$$

- We cannot estimate a mixed logit model using MLE, as we did for logit and nested logit

Instead, we can approximate choice probabilities through simulation and estimate the simulated maximum likelihood estimator (MSLE)

- More on MSLE (and other simulation-based estimation methods) next week...

The `mlogit` package in R has the functionality to estimate a mixed logit model

- More on estimating a mixed logit model with `mlogit` next time

Mixed Logit with Market-Level Data

The mixed logit model can also be estimated from market-level data

- You observe the price, market share, and characteristics of every cereal brand at the grocery store, and you want to estimate the structural parameters of consumer decision making that explain those purchases

When aggregated over many consumers, choice probabilities become market shares

$$S_i = \int \frac{e^{\beta' x_i}}{\sum_{j=1}^{J} e^{\beta' x_j}} f(\beta \mid \theta) d\beta$$

- Because of the integral, mixed logit market shares do not reduce to a linear model as they did for logit and nested logit
- Demand estimation using random coefficients logit is known as "BLP" after Berry, Levinsohn, and Pakes (1995)
- More on BLP later in the semester...

Announcements

Reading for next time

- Optional: Revelt and Train (1998)

Office hours

- Reminder: 2:00–3:00 on Tuesdays in 218 Stockbridge

Upcoming

- Problem Set 3 is posted, due October 31