Lecture 13: Generalized Extreme Value Models I

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Generalized Method of Moments Example in R

Today

- Nested Logit
- Three-Level Nested Logit
- Logit with Overlapping Nests
- Heteroskedastic Logit

Upcoming

- Reading for next time
  - Optional: Train et al. (1987)
- Problem sets
  - Problem Set 3 is posted, due October 31

Course Recap

Discrete choice models

- Random utility model
- Logit model

Estimation methods

- Maximum likelihood estimation
- Nonlinear least squares
- Generalized method of moments

Now that we know these estimation methods, we will jump back to discrete models

- Correlations between alternatives
- Taste variation due to unobserved factors

Nested Logit

Nested Logit

The nested logit model relaxes the (sometimes overly) strong assumption of the logit model

- Nested logit allows for the unobserved components of utility (ε_{nj}) to be correlated between alternatives for the same decision maker

The nested logit model groups alternatives into "nests"

- $Cov(\varepsilon_{nj}, \varepsilon_{nm}) = 0$ if j and m are in different nests
- $Cov(\varepsilon_{nj}, \varepsilon_{nm}) \geq 0$ if j and m are in the same nest

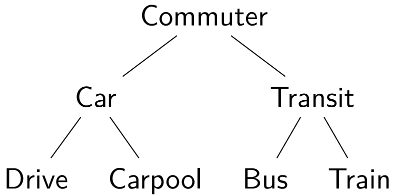
These correlations relax the substitution patterns of the logit model

- IIA holds for two alternatives within the same nest
- IIA does not necessarily hold for alternatives in different nests

Nested Logit Example

Commuters in Boston have four alternatives to commute to work

- Drive alone
- Carpool
- Bus
- Train



Commute travel mode nests

- Drive alone and carpool might belong in a "car" nest
- Bus and train might belong together in a "public transit" nest

Can you rationalize any other sets of nests?

- Choosing nests is more an art than a science

Generalized Extreme Value Distribution

We partition the J alternatives into K nonoverlapping subsets denoted $B_1, B_2, \ldots, B_K$ and called "nests"

The utility for each alternative is again $U_{nj} = V_{nj} + \varepsilon_{nj}$ where the vector of unobserved utility, $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nJ})$, has cumulative distribution

$$F(\varepsilon_n) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right)$$

which is a type of generalized extreme value (GEV) distribution

- The marginal distribution of ε_{nj} is extreme value
- $Cov(\varepsilon_{nj}, \varepsilon_{nm}) = 0$ if $j \in B_k$ and $m \in B_\ell$ with $k \neq \ell$
- $Cov(\varepsilon_{nj}, \varepsilon_{nm}) \geq 0$ if $j \in B_k$ and $m \in B_k$
- λ_k is a measure of independence in nest k

Nested Logit Choice Probabilities

The general form for the choice probability of alternative i is

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \; \forall j \neq i) f(\varepsilon_n) d\varepsilon_n$$

The distributional assumption of the nested logit model yields a closed-form solution for the choice probability of alternative $i \in B_k$

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K} \left(\sum_{j \in B_\ell} e^{V_{nj}/\lambda_\ell} \right)^{\lambda_\ell}}$$

Degree of correlation within a nest

- λ_k is a measure of independence in nest k that we will estimate
- $1 - \lambda_k$ is an indicator of correlation within nest k
- The nested logit model is consistent with the random utility model when $\lambda_k \in (0, 1] \; \forall k$

Nested Logit Substitution Patterns

The ratio of nested logit choice probabilities for alternative $i \in B_k$ and $m \in B_\ell$ is

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_\ell} \left(\sum_{j \in B_\ell} e^{V_{nj}/\lambda_\ell}\right)^{\lambda_\ell - 1}}$$

When $k = \ell$, this ratio simplifies to

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_\ell}}$$

The first ratio depends on all alternatives in nests k and ℓ

- Independence from irrelevant nests (IIN) holds for different nests

This second ratio is irrelevant of all other alternatives

- Independence from irrelevant alternatives (IIA) holds within a nest

Decomposition of Nested Logit

We can decompose the nested logit model into a two-step logit model

1. The decision maker chooses a nest
2. The decision maker chooses an alternative within that nest

Then the choice probability of alternative $i \in B_k$ is

$$P_{ni} = P_{nB_k} P_{ni|B_k}$$

where

- P_{nB_k} is the probability of choosing nest k
- $P_{ni|B_k}$ is the probability of choosing alternative i conditional on choosing nest k

Choice Probabilities of Decomposed Nested Logit

Express the utility of alternative $j \in B_k$ as

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj}$$

where

- W_{nk} depends on characteristics of nest k
- Y_{nj} depends on characteristics of alternative j

Then the choice probabilities from the previous slide are

$$P_{nB_k} = \frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{\ell=1}^{K} e^{W_{n\ell} + \lambda_\ell I_{n\ell}}}$$

$$P_{ni|B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{nj}/\lambda_k}}$$

where

$$I_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}/\lambda_k}$$

Decomposed Nested Logit Intuition

Putting everything together

$$P_{ni} = P_{ni|B_k} P_{nB_k}$$

$$P_{nB_k} = \frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{\ell=1}^{K} e^{W_{n\ell} + \lambda_\ell I_{n\ell}}}$$

$$P_{ni|B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{nj}/\lambda_k}}$$

$$I_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}/\lambda_k}$$

- I_{nk} is called the inclusive value of nest k
- $W_{nk} + \lambda_k I_{nk}$ is the expected utility of nest k
- P_{nB_k} is equivalent to the logit choice probability between nests
- $P_{ni|B_k}$ is equivalent to the logit choice probability within a nest

Nested Logit Elasticities

The nested logit model allows for a richer representation of elasticities than does the logit model

Own elasticity of alternative i in nest k

$$E_{iz_{ni}} = \beta_z z_{ni} \left(\frac{1}{\lambda_k} - \frac{1-\lambda_k}{\lambda_k} P_{ni|B_k} - P_{ni} \right)$$

Cross elasticities of alternative $i \in B_k$ with respect to alternative $m \neq i$

$$E_{iz_{nm}} = \begin{cases} -\beta_z z_{nm} P_{nm} \left(1 + \frac{1-\lambda_k}{\lambda_k} \frac{1}{P_{nB_k}} \right) & \text{if } m \in B_k \\ -\beta_z z_{nm} P_{nm} & \text{if } m \notin B_k \end{cases}$$

Nested Logit Estimation

Estimation of the nested logit model is similar to the logit model

- We have more complex choice probabilities in our estimation
- MLE is the standard method, but you could use NLS or GMM
- Log-likelihood function is not globally concave, so numerical optimization can be more difficult than for logit

It is not recommended to estimate the sequential nested logit model of the previous slides

- You have to bootstrap to get the correct variance-covariance matrix

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

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

Nested Logit with Market-Level Data

The nested 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

$$S_i = \frac{e^{V_i/\lambda_k}\left(\sum_{j\in B_k} e^{V_j/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K}\left(\sum_{j\in B_\ell} e^{V_j/\lambda_\ell}\right)^{\lambda_\ell}}$$

If we assume representative utility is liner, $V_j = \beta' x_j$

$$\ln(S_i) - \ln(S_m) = \beta'(x_i - x_j) + (1-\lambda_k)\ln S_{i|B_k} - (1-\lambda_\ell) S_{m|B_\ell}$$

Set one alternative to be your reference in its own nest (usually the outside option) and estimate the linear regression for the other alternatives

$$\ln(S_i) - \ln(S_0) = \beta'(x_i - x_0) + (1-\lambda_k)\ln(S_{i|B_k}) + \omega_i$$

Three-Level Nested Logit

Three-Level Nested Logit

We can model a richer set of correlations between alternatives by including multiple levels of nests and subnests

- As in the (two-level) nested logit model, we first group alternatives into nests
- Then, within each nest, we further group alternatives into subnests

Correlations within nests and subnests

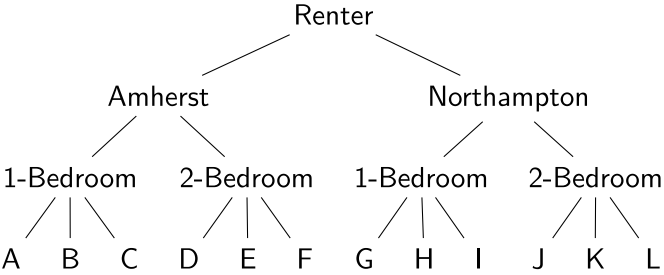
- λ_k defines the level of independence within nest k
- σ_{mk} defines the level of independence within subnest m in nest k
- $1 - \lambda_k \sigma_{mk}$ is an indicator of correlation within subnest m in nest k
- This model is consistent with the random utility model when $\lambda_k \in (0, 1]\ \forall k$ and $\sigma_{mk} \in (0, 1]\ \forall m$

We can even model more than three levels of nests

- There are also other ways to model these more complex correlation structures. . .

Three-Level Nested Logit Example

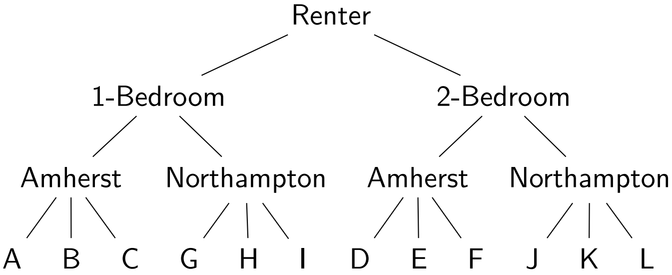
Apartments in Amherst and Northampton



- All 1-bedroom apartments in Amherst are correlated
- 1-bedroom apartments in Amherst are also (less) correlated with 2-bedroom apartments in Amherst
- 1-bedroom apartments in Amherst are uncorrelated with any apartment in Northampton

Three-Level Nested Logit Example

Or maybe the nesting should be in the other order?



- All 1-bedroom apartments in Amherst are correlated
- 1-bedroom apartments in Amherst are also (less) correlated with 1-bedroom apartments in Northampton
- 1-bedroom apartments in Amherst are uncorrelated with any 2-bedroom apartment

Logit with Overlapping Nests

Logit with Overlapping Nests

The nested logit model requires that nests (and subnests) are exhaustive and mutually exclusive

- We have to decide which alternatives belong in nests (or subnests) together
- This structure is more flexible than the logit model
- But it still requires us to make assumptions about which alternatives might be correlated and which are assumed to be uncorrelated

Is there a way to represent correlations even more flexibly?

- Yes, if we allow nests to overlap
- At the extreme, we can let every pair of alternatives form its own nest

Paired Combinatorial Logit

Each pair of alternatives is a nest (note that this does not comply with our prior definition of a "nest")

- We have $J(J-1)/2$ nests
- Each nests has only two alternatives
- Each alternative is in $J-1$ nests

Correlation between alternatives i and j

- λ_{ij} defines the independence between alternatives i and j
- $1-\lambda_{ij}$ is an indicator of the correlation between alternatives i and j
- We have $J(J-1)/2$ λ parameters to estimate
    - We can only identify $J(J-1)/2-1$ covariance parameters, so we have to normalize at least one $\lambda = 1$
- This model is consistent with the random utility model when $\lambda_{ij} \in (0,1]\ \forall i,j$

Paired Combinatorial Logit Choice Probabilities

The choice probability for the paired combinatorial logit model is

$$P_{ni} = \frac{\sum_{j \neq i} e^{V_{ni}/\lambda_{ij}} \left(e^{V_{ni}/\lambda_{ij}} + e^{V_{nj}/\lambda_{ij}} \right)^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{\ell=k+1}^{J} \left(e^{V_{nk}/\lambda_{k\ell}} + e^{V_{n\ell}/\lambda_{k\ell}} \right)^{\lambda_{k\ell}}}$$

The denominator is analogous to the nested logit denominator

- Sum over all nests
- Within a nest, sum over all elements within the nest

The numerator is analogous to the nested logit numerator

- We now have to sum over all nests that contain alternative i

Heteroskedastic Logit

Heteroskedastic Logit

We can also use the GEV distributional assumption to allow for heteroskedasticity of alternatives

- The variance of unobserved utility can be different for each alternative

Utility is specified as $U_{nj} = V_{nj} + \varepsilon_{nj}$ with

$$Var(\varepsilon_{nj}) = \frac{(\theta_j \pi)^2}{6}$$

- We have J variance parameters to estimate
- We can only identify $J-1$, so we have to normalize at least one $\theta = 1$

The choice probabilities for heteroskedastic logit do not have a closed-form expression

- We have to use simulation methods to get choice probabilities and subsequently estimate model parameters

Announcements

Reading for next time

- Optional: Train et al. (1987)

Office hours

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

Upcoming

- Problem Set 3 is posted, due October 31