Week 9: Generalized Extreme Value Models

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last week

- Generalized method of moments

This week's topics

- Nested logit
- Two-step nested logit
- Nested logit properties
- Empirical considerations
- Other generalized extreme value models
- Nested logit R example

This week's reading

- Train textbook, chapter 4

Nested Logit

Course Recap

Structural econometric models

- Random utility model
- Logit model

Estimation methods

- Maximum likelihood estimation
- Generalized method of moments

Now that we know these estimation methods, we will look at discrete choice models that allow for richer representations of individual decision making

- Nested logit: unobserved correlations between alternatives
- Mixed logit: preference variation due to unobserved factors

Nested Logit

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

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

The nested logit model groups alternatives into "nests"

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

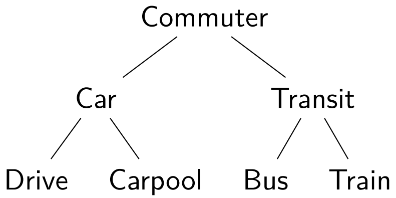
These correlations relax the substitution patterns of the logit model

- 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 "private 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

Nested Logit GEV 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, $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nJ})$, has cumulative distribution

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

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

- The marginal distribution of ε_{nj} is extreme value
- $Cov(\varepsilon_{ni}, \varepsilon_{nm}) = 0$ if $i \in B_k$ and $m \in B_\ell$ with $k \neq \ell$
- $Cov(\varepsilon_{ni}, \varepsilon_{nm}) \geq 0$ if $i \in B_k$ and $m \in B_k$
- λ_k is a measure of independence in nest k
  - $\lambda_k = 1\ \forall k$ gives the logit model

Two-Step Nested Logit

Two-Step 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 that decision maker n chooses nest k in the first step
- $P_{ni|B_k}$ is the probability that decision maker n chooses alternative i in the second step conditional on having chosen nest k in the first step

Representative Utility of Two-Step Nested Logit

We first generalize our original random utility model by decomposing representative utility into nest attributes and alternative attributes

We express the utility of alternative $j \in B_k$ as

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

where

- W_{nk} is utility that depends on observed attributes of nest k
- Y_{nj} is utility that depends on observed attributes of alternative j

Example: consumer product choice

- If products are nested by brands, W_{nk} could include a variable for brand quality or an indicator for brand preferences

Choice Probabilities of Two-Step Nested Logit

The second-step choice probability is a logit choice probability among the alternatives in nest k using alternative-specific representative utility, Y_{nj}, scaled by λ_k (just as ε_{nj} was scaled by λ_k in the joint distribution of $\boldsymbol{\varepsilon}_n$)

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

The first-step choice probability is a logit choice probability among the nests using the expected utility of each nest, $W_{nk} + \lambda_k I_{nk}$

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

where I_{nk} is the log-sum term of nest k's second step, known as the inclusive value of nest k

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

Nested Logit Properties

Nested Logit Choice Probabilities

Multiplying the two logit choice probabilities of the two-step nested logit gives the nested logit choice probability

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

This nested logit choice probability depends on the representative utility of each alternative and the degree of unobserved utility correlation within each 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 Across Nests

The ratio of the choice probability of alternative i in nest k to the choice probability of alternative m in nest ℓ is

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

This ratio depends on all alternatives in nests k and ℓ, but no alternatives in other nests

- Independence of irrelevant nests (IIN) holds across nests

This IIN property follows from the nested logit first step being equivalent to a logit model of the nests

Nested Logit Substitution Patterns Within a Nest

The ratio of the choice probabilities simplifies when the two alternatives are in the same nest ($k = \ell$)

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

This ratio depends on only alternatives i and m

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

IIA within a nest follows from the nested logit second step being equivalent to a logit model of the alternatives within the nest

Nested Logit Elasticities

If we assume representative utility is linear, $V_{nj} = \beta' \boldsymbol{x}_n j$, own elasticity of alternative i in nest k with respect to its attribute z_{ni} is

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

and cross elasticity of alternative i in nest k with respect to attribute z_{nm} of alternative $m \neq i$ is

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

where P_{nB_k} and $P_{ni|B_k}$ are calculated as

$$P_{nB_k} = \sum_{j \in B_k} P_{nj} \quad \text{and} \quad P_{ni|B_k} = \frac{P_{ni}}{P_{nB_k}} = \frac{P_{ni}}{\sum_{j \in B_k} P_{nj}}$$

Empirical Considerations

Nested Logit Estimation

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

- We have more complex choice probabilities in our estimation
- Maximum likelihood is the standard method
- Log-likelihood function is not globally concave, so numerical optimization can be more difficult than for logit

It is not recommended to estimate a nested logit model as two sequential steps

- You must bootstrap to get the correct variance estimator

We will use the `mlogit()` function in R to estimate nested logit models

- We specify our nesting structure using the `nests` argument in the `mlogit()` function

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 attributes 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, $V_j = \beta' \boldsymbol{x}_j$

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

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

$$\ln(S_i) - \ln(S_0) = \beta'(\boldsymbol{x}_i - \boldsymbol{x}_0) + (1 - \lambda_k) \ln(S_{i|B_k}) + \omega_i$$

Nested Logit with Panel Data

If we observe panel data for a discrete choice problem, we can add a time index to our random utility model and nested logit choice probabilities

$$U_{njt} = V_{njt} + \varepsilon_{njt} \quad \Rightarrow \quad P_{nit} = \frac{e^{V_{nit}/\lambda_k} \left(\sum_{j \in B_k} e^{V_{njt}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K} \left(\sum_{j \in B_\ell} e^{V_{njt}/\lambda_\ell} \right)^{\lambda_\ell}}$$

We can estimate this model just as in the cross-section

- We can include lagged or future variables to capture "dynamics"
- We can include previous choices as explanatory variables to represent behavioral factors like habit formation

The i.i.d. assumption still has to hold across decision makers and across time periods

- But the unobserved characteristics of a decision maker that affect choice are unlikely to be independent over time

Other Generalized Extreme Value Models

Generalized Extreme Value Models

The cumulative distribution of the nested logit model is a special case of a generalized extreme value (GEV) distribution

Many discrete choice models can be created from the GEV distribution

- Nested logit
- Three-level nested logit
- Paired combinatorial logit
- Generalized nested logit
- Heteroskedastic 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

This model yields correlations within nests and within subnests

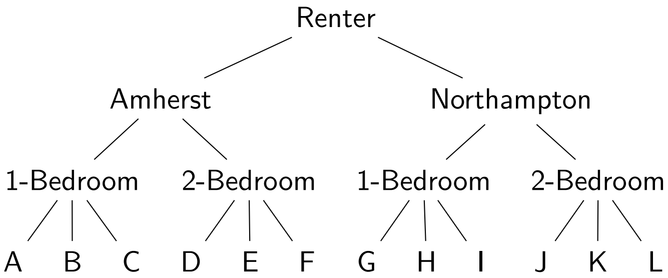
- λ_k defines the level of independence within nest k
- σ_{mk} defines the level of independence within subnest m in nest k
- $1 - \lambda_k \sigma_{mk}$ is an indicator of correlation within subnest m in nest k
- This model is consistent with the random utility model when $\lambda_k \in (0, 1]\ \forall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

We can even model more than three levels of nesting

- Or use another GEV model for more complex correlation structures

Three-Level Nested Logit Example

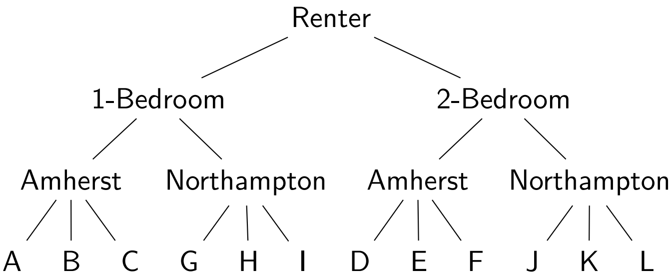
Apartments in Amherst and Northampton



- All 1-bedroom apartments in Amherst are correlated
- 1-bedroom apartments in Amherst are also correlated—but less so—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 correlated—but less so—with 1-bedroom apartments in Northampton
- 1-bedroom apartments in Amherst are uncorrelated with any 2-bedroom apartment

Paired Combinatorial Logit

If we relax the definition of a nest to allow for overlaps, then we can create a separate nest for every pairwise combination of alternatives

- $J(J-1)/2$ nests with two alternatives in each nest
- Each alternative appears in $J-1$ nests

We more flexibly estimate the correlations between pairs of alternatives

- 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 $\lambda_{ij} \in (0,1]\ \forall i,j$

See the Train textbook for choice probabilities and more details

Generalized Nested Logit

We can generalize the nested logit model by allowing an alternative to be in multiple nests and to varying degrees

- Construct nests $B_1, B_2, \ldots, B_K$ by assigning each alternative to one or more nests
- Estimate two sets of parameters related to the nests
  - λ_k: independence within nest k
  - α_{jk}: "weight" or proportion of alternative j in nest k

Example: apartment choice in Amherst and Northampton

- Four nests: Amherst, Northampton, one-bedroom, two-bedroom
- A one-bedroom apartment in Amherst would be in both the Amherst nest and the one-bedroom nest, and we would estimate its "weight" or proportion in each of these nests

See the Train textbook for choice probabilities and more details

Heteroskedastic Logit

We can use the GEV distribution to allow for heteroskedasticity among the 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

Nested Logit R Example

Nested Logit Model Example

We are again studying how consumers make choices about expensive and highly energy-consuming systems in their homes

- We have (real) data on 250 households in California and the type of HVAC (heating, ventilation, and air conditioning) system in their home. Each household has the following choice set, and we observe the following data

Choice set

- `ec`: electric central
- `ecc`: electric central with AC
- `er`: electric room
- `erc`: electric room with AC
- `gc`: gas central
- `gcc`: gas central with AC
- `hpc`: heat pump with AC

Alternative-specific data

- `ich`: installation cost for heat
- `icca`: installation cost for AC
- `och`: operating cost for heat
- `occa`: operating cost for AC

Household demographic data

- `income`: annual income

Random Utility Model of HVAC System Choice

We model the utility to household n of installing HVAC system j as

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} depends on the data about alternative j and household n

The probability that household n installs HVAC system i is

$$P_{ni} = \int_{\varepsilon} \mathbb{1}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \; \forall j \neq i) f(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n$$

Under the logit assumption, these choice probabilities simplify to

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

Representative Utility of HVAC System Choice

What might affect the utility of the different HVAC systems?

- Installation cost
- Annual operating cost
- HVAC system technology
  - Systems with cooling might be preferred to heating-only systems
  - Gas systems might be preferred to electric systems
  - Central systems might be preferred to room systems
- Anything else?

We model the representative utility of HVAC system j to household n as

$$V_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj}$$

Load Dataset

```r
## Load tidyverse and mlogit
library(tidyverse)
library(mlogit)
## Load dataset from mlogit package
data('HC', package = 'mlogit')
```

Dataset

```
## Look at dataset
tibble(HC)
## # A tibble: 250 x 18
##    depvar ich.gcc ich.ecc ich.erc ich.hpc ich.gc ich.ec ich.er  icca
##    <fct>    <dbl>   <dbl>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl> <dbl>
##  1 erc       9.7     7.86    8.79   11.4    24.1   24.5   7.37  27.3
##  2 hpc       8.77    8.69    7.09    9.37   28     32.7   9.33  26.5
##  3 gcc       7.43    8.86    6.94   11.7    25.7   31.7   8.14  22.6
##  4 gcc       9.18    8.93    7.22   12.1    29.7   26.7   8.04  25.3
##  5 gcc       8.05    7.02    8.44   10.5    23.9   28.4   7.15  25.4
##  6 gcc       9.32    8.03    6.22   12.6    27.0   21.4   8.6   19.9
##  7 gc        7.11    8.78    7.36   12.4    22.9   28.6   6.41  27.0
##  8 hpc       9.38    7.48    6.72    8.93   26.2   27.9   7.3   18.1
##  9 gcc       8.08    7.39    8.79   11.2    23.0   22.6   7.85  22.6
## 10 gcc       6.24    4.88    7.46    8.28   19.8   27.5   6.88  25.8
## # ... with 240 more rows, and 9 more variables: och.gcc <dbl>,
## #   och.ecc <dbl>, och.erc <dbl>, och.hpc <dbl>, och.gc <dbl>,
## #   och.ec <dbl>, och.er <dbl>, occa <dbl>, income <dbl>
```

Clean Dataset

```r
## Combine heating and cooling costs into one variable
hvac_clean <- HC %>%
  mutate(id = 1:n(),
         ic.gcc = ich.gcc + icca, ic.ecc = ich.ecc + icca,
         ic.erc = ich.erc + icca, ic.hpc = ich.hpc + icca,
         ic.gc = ich.gc, ic.ec = ich.ec, ic.er = ich.er,
         oc.gcc = och.gcc + occa, oc.ecc = och.ecc + occa,
         oc.erc = och.erc + occa, oc.hpc = och.hpc + occa,
         oc.gc = och.gc, oc.ec = och.ec, oc.er = och.er) %>%
  select(id, depvar, starts_with('ic.'), starts_with('oc.'), income)
```

Cleaned Dataset

```
## Look at cleaned dataset
tibble(hvac_clean)
## # A tibble: 250 x 17
##        id depvar ic.gcc ic.ecc ic.erc ic.hpc ic.gc ic.ec ic.er oc.gcc
##     <int> <fct>   <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl>
##  1      1 erc      37.0   35.1   36.1   38.6  24.1  24.5  7.37   5.21
##  2      2 hpc      35.3   35.2   33.6   35.9  28    32.7  9.33   3.93
##  3      3 gcc      30.1   31.5   29.6   34.3  25.7  31.7  8.14   4.46
##  4      4 gcc      34.5   34.3   32.6   37.5  29.7  26.7  8.04   5.32
##  5      5 gcc      33.5   32.5   33.9   36.0  23.9  28.4  7.15   5.29
##  6      6 gcc      29.2   28.0   26.2   32.5  27.0  21.4  8.6    4.67
##  7      7 gc       34.2   35.8   34.4   39.4  22.9  28.6  6.41   4.18
##  8      8 hpc      27.5   25.6   24.8   27.0  26.2  27.9  7.3    5.37
##  9      9 gcc      30.6   30.0   31.4   33.7  23.0  22.6  7.85   4.74
## 10     10 gcc      32.0   30.6   33.2   34.0  19.8  27.5  6.88   4.32
## # ... with 240 more rows, and 7 more variables: oc.ecc <dbl>,
## #   oc.erc <dbl>, oc.hpc <dbl>, oc.gc <dbl>, oc.ec <dbl>,
## #   oc.er <dbl>, income <dbl>
```

Convert Dataset to dfidx Format

```r
## Convert cleaned dataset to dfidx format
hvac_dfidx <- dfidx(hvac_clean, shape = 'wide',
                    choice = 'depvar', varying = 3:16)
```

Dataset in dfidx Format

```
## Look at data in dfidx format
tibble(hvac_dfidx)
## # A tibble: 1,750 x 6
##        id depvar income    ic    oc idx$id1 $id2
##     <int> <lgl>   <dbl> <dbl> <dbl>   <int> <fct>
##  1      1 FALSE      20 24.5   4.09       1 ec
##  2      1 FALSE      20 35.1   7.04       1 ecc
##  3      1 FALSE      20  7.37  3.85       1 er
##  4      1 TRUE       20 36.1   6.8        1 erc
##  5      1 FALSE      20 24.1   2.26       1 gc
##  6      1 FALSE      20 37.0   5.21       1 gcc
##  7      1 FALSE      20 38.6   4.68       1 hpc
##  8      2 FALSE      50 32.7   2.69       2 ec
##  9      2 FALSE      50 35.2   4.32       2 ecc
## 10      2 FALSE      50  9.33  3.45       2 er
## # ... with 1,740 more rows
```

Multinomial Logit Model

We model the representative utility of HVAC system j to household n as

$$V_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj}$$

```
## Model choice using alternative intercepts and cost data
model_1 <- mlogit(formula = depvar ~ ic + oc | 1 | 0,
                  data = hvac_dfidx,
                  reflevel = 'hpc')
```

Multinomial Logit Model Summary

```
## Summarize model results
summary(model_1)
##
## Call:
## mlogit(formula = depvar ~ ic + oc | 1 | 0, data = hvac_dfidx,
##     reflevel = "hpc", method = "nr")
##
## Frequencies of alternatives:choice
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## nr method
## 7 iterations, 0h:0m:0s
## g'(-H)^-1g = 1.94E-05
## successive function values within tolerance limits
##
## Coefficients :
##                  Estimate Std. Error z-value  Pr(>|z|)
## (Intercept):ec  -6.305515   1.201159 -5.2495 1.525e-07 ***
## (Intercept):ecc  2.142493   0.944478  2.2684    0.0233 *
## (Intercept):er  -7.779634   1.369346 -5.6813 1.337e-08 ***
## (Intercept):erc  0.944165   1.337210  0.7061    0.4801
## (Intercept):gc  -6.120762   0.964956 -6.3430 2.253e-10 ***
## (Intercept):gcc  2.997228   0.395830  7.5720 3.664e-14 ***
## ic              -0.225519   0.043764 -5.1531 2.562e-07 ***
## oc              -1.937800   0.373606 -5.1867 2.140e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -198.57
## McFadden R^2:  0.11831
## Likelihood ratio test : chisq = 53.291 (p.value = 2.6789e-12)
```

Nested Logit Model of HVAC System Choice

We model the utility to household n of installing HVAC system j as

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} depends on the data about alternative j and household n

We model the representative utility of HVAC system j to household n as

$$V_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj}$$

How might unobserved utility, ε_{nj}, be correlated for a decision maker?

- Cooling vs. heating-only
- Electric vs. gas
- Any other nest structures?

Nested Logit Choice Probabilities for HVAC System Choice

The nested logit choice probabilities depend on our model of representative utility and our nests

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

We will use maximum likelihood to find the set of parameters—α_j, β_1, β_2, and λ_k—that maximize the likelihood of generating the choices that we observe

We will use the `mlogit()` function in R to estimate nested logit models

- We specify our nesting structure using the `nests` argument in the `mlogit()` function

Nested Logit Models in R

```r
## Help file for the mlogit function
?mlogit
## Arguments for mlogit GEV functionality
mlogit(formula, data, reflevel, nests, ...)
```

`mlogit()` arguments for nested logit

1. `formula, data, reflevel`: same as a multinomial logit model
2. `nests`: named list of character vectors that defines nests

Nested Logit Model with Cooling Vs. Heating-Only Nests

We model the representative utility of HVAC system j to household n as

$$V_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj}$$

and we have two nests for correlation of unobserved utility, ε_{nj}

- Cooling: `ecc`, `erc`, `gcc`, and `hpc`
- Heating-only: `ec`, `er`, and `gc`

```
## Model choice with cooling vs. heating-only nests
model_2 <- mlogit(formula = depvar ~ ic + oc | 1 | 0,
                  data = hvac_dfidx,
                  reflevel = 'hpc',
                  nests = list(cooling = c('ecc', 'erc', 'gcc', 'hpc'),
                               heating_only = c('ec', 'er', 'gc')))
```

Nested Logit Model Summary

```
## Summarize model results
summary(model_2)
##
## Call:
## mlogit(formula = depvar ~ ic + oc | 1 | 0, data = hvac_dfidx,
##     reflevel = "hpc", nests = list(cooling = c("ecc", "erc",
##         "gcc", "hpc"), heating_only = c("ec", "er", "gc")))
##
## Frequencies of alternatives:choice
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## bfgs method
## 16 iterations, 0h:0m:0s
## g'(-H)^-1g = 0.145
## last step couldn't find higher value
##
## Coefficients :
##                   Estimate Std. Error z-value  Pr(>|z|)
## (Intercept):ec   -3.750232   0.761515 -4.9247 8.449e-07 ***
## (Intercept):ecc   0.606353   0.307990  1.9687  0.048982 *
## (Intercept):er   -5.127597   1.269186 -4.0401 5.344e-05 ***
## (Intercept):erc   0.421811   0.352893  1.1953  0.231972
## (Intercept):gc   -3.913846   0.813040 -4.8138 1.481e-06 ***
## (Intercept):gcc   0.624893   0.262557  2.3800  0.017311 *
## ic               -0.121913   0.046663 -2.6126  0.008984 **
## oc               -0.565040   0.221976 -2.5455  0.010912 *
## iv:cooling        0.203909   0.084303  2.4188  0.015574 *
## iv:heating_only   0.189574   0.066831  2.8366  0.004559 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -189.83
## McFadden R^2:  0.1571
```

Interpreting Parameters

```
## Display model coefficients
coef(model_2)
##   (Intercept):ec (Intercept):ecc  (Intercept):er (Intercept):erc
##       -3.7502320       0.6063529      -5.1275965       0.4218114
##   (Intercept):gc (Intercept):gcc              ic              oc
##       -3.9138460       0.6248927      -0.1219130      -0.5650405
##       iv:cooling iv:heating_only
##        0.2039089       0.1895738
```

How do we interpret these coefficients?

- Alternative-specific intercepts
  - `ecc` and `gcc` provide more utility, *ceteris paribus*, than `hpc`
  - `erc` provides the same utility, *ceteris paribus*, as `hpc`
  - `ec`, `er`, and `gc` provide less utility, *ceteris paribus*, than `hpc`
- An additional $100 of installation cost reduces utility by 0.12
- An additional $100 of annual operating cost reduces utility by 0.57
- Each nest is highly correlated
  - $1 - \lambda_k$ gives a measure of within-nest correlation

Test of Nest Correlations

We can test if the within-nest correlations are statistically different from those implied by the multinomial logit model

- The multinomial logit model assumes that all alternatives are independent, or $\lambda_k = 1 \; \forall k$

We will test the null hypothesis

$$H_0\colon \; \lambda_{cooling} = \lambda_{heating} = 1$$

We talked about three tests we can use for MLE

- Likelihood ratio test, Wald test, and Lagrange multiplier test
- We will use the likelihood ratio test

Likelihood Ratio Test

`lrtest()` conducts a likelihood ratio test on the two models we specify as arguments

```
## Conduct likelihood ratio test of models 1 and 2
lrtest(model_1, model_2)
## Likelihood ratio test
##
## Model 1: depvar ~ ic + oc | 1 | 0
## Model 2: depvar ~ ic + oc | 1 | 0
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   8 -198.57
## 2  10 -189.83  2 17.472  0.0001607 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nested Logit Model with Electric Vs. Gas Nests

We model the representative utility of HVAC system j to household n as

$$V_{nj} = \alpha_j + \beta_1 IC_{nj} + \beta_2 OC_{nj}$$

and we have two nests for correlation of unobserved utility, ε_{nj}

- Electric: `ec`, `ecc`, `er`, `erc`, and `hpc`
- Gas: `gc` and `gcc`

```
## Model choice with electric vs. gas nests
model_3 <- mlogit(formula = depvar ~ ic + oc | 1 | 0,
                  data = hvac_dfidx,
                  reflevel = 'hpc',
                  nests = list(electric = c('ec', 'ecc', 'er',
                                                  'erc', 'hpc'),
                               gas = c('gc', 'gcc')))
```

Nested Logit Model Summary

```
## Summarize model results
summary(model_3)
##
## Call:
## mlogit(formula = depvar ~ ic + oc | 1 | 0, data = hvac_dfidx,
##     reflevel = "hpc", nests = list(electric = c("ec", "ecc",
##         "er", "erc", "hpc"), gas = c("gc", "gcc")))
##
## Frequencies of alternatives:choice
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## bfgs method
## 17 iterations, 0h:0m:0s
## g'(-H)^-1g = 3.29E-07
## gradient close to zero
##
## Coefficients :
##                   Estimate Std. Error z-value  Pr(>|z|)
## (Intercept):ec  -12.69931    4.41999 -2.8732 0.0040640 **
## (Intercept):ecc   2.12355    2.05964  1.0310 0.3025277
## (Intercept):er  -14.46701    4.16332 -3.4749 0.0005111 ***
## (Intercept):erc  -1.22655    4.95891 -0.2473 0.8046435
## (Intercept):gc  -11.39903    3.13007 -3.6418 0.0002708 ***
## (Intercept):gcc   3.97214    1.11935  3.5486 0.0003873 ***
## ic               -0.39486    0.11812 -3.3428 0.0008293 ***
## oc               -2.92921    0.66656 -4.3945  1.11e-05 ***
## iv:electric       2.41263    1.51806  1.5893 0.1119972
## iv:gas            2.08988    0.85145  2.4545 0.0141082 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -195.76
## McFadden R^2:  0.13078
```

Interpreting Parameters

```
## Display model coefficients
coef(model_3)
##  (Intercept):ec (Intercept):ecc  (Intercept):er (Intercept):erc
##     -12.6993138       2.1235486     -14.4670051      -1.2265464
##  (Intercept):gc (Intercept):gcc              ic              oc
##     -11.3990294       3.9721450      -0.3948641      -2.9292077
##     iv:electric          iv:gas
##       2.4126254       2.0898825
```

How do we interpret these coefficients?

- Alternative-specific intercepts
  - `gcc` provides more utility, *ceteris paribus*, than `hpc`
  - `ecc` and `erc` provide the same utility, *ceteris paribus*, as `hpc`
  - `ec`, `er`, and `gc` provide less utility, *ceteris paribus*, than `hpc`
- An additional $100 of installation cost reduces utility by 0.39
- An additional $100 of annual operating cost reduces utility by 2.93
- λ parameters are not easily interpretable when they are outside the $(0, 1]$ range

Likelihood Ratio Test

We can use a likelihood ratio test to test if these within-nest correlations are statistically different from those implied by the multinomial logit model

$$H_0\colon\ \lambda_{electric} = \lambda_{gas} = 1$$

```
## Conduct likelihood ratio test of models 1 and 3
lrtest(model_1, model_3)
## Likelihood ratio test
##
## Model 1: depvar ~ ic + oc | 1 | 0
## Model 2: depvar ~ ic + oc | 1 | 0
##   #Df  LogLik Df Chisq Pr(>Chisq)
## 1   8 -198.57
## 2  10 -195.76  2 5.613    0.06042 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```