Lecture 14: Generalized Extreme Value Models II

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman
University of Massachusetts Amherst

Agenda

Last time

- Generalized Extreme Value Models

Today

- Generalized Extreme Value Models Example in R

Upcoming

- Reading for next time
  - Train textbook, Chapter 6
- Problem sets
  - Problem Set 3 is posted, due October 31

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

Generalized Extreme Value Model Example in R

Nested Logit Model Example

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

- GCC: gas central with AC
- ECC: electric central with AC
- ERC: electric room with AC
- HPC: heat pump with AC
- GC: gas central
- EC: electric central
- ER: electric room

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

Load Dataset

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

Dataset

```
## Look at dataset
as_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>
```

Format Dataset in a Long Format

```
### Format dataset
## Gather into a long dataset
hvac_long <- HC %>%
  mutate(id = 1:n()) %>%
  gather(key, value, starts_with('ich.'), starts_with('och.')) %>%
  separate(key, c('cost', 'alt')) %>%
  spread(cost, value) %>%
  mutate(choice = (depvar == alt)) %>%
  select(-depvar)
```

Dataset in a Long Format

```
## Look at long dataset
as_tibble(hvac_long)
## # A tibble: 1,750 x 8
##     icca  occa income    id alt     ich   och choice
##    <dbl> <dbl>  <dbl> <int> <chr> <dbl> <dbl> <lgl>
##  1  17    2.79     60   133 ec    20.3   4.52 FALSE
##  2  17    2.79     60   133 ecc    8.46  4.52 FALSE
##  3  17    2.79     60   133 er     7.7   4.32 FALSE
##  4  17    2.79     60   133 erc    8.16  4.32 FALSE
##  5  17    2.79     60   133 gc    25.3   2.26 FALSE
##  6  17    2.79     60   133 gcc    6.33  2.26 TRUE
##  7  17    2.79     60   133 hpc   11.1   1.63 FALSE
##  8  18.1  2.55     50    14 ec    25.6   5.21 FALSE
##  9  18.1  2.55     50    14 ecc   11.2   5.21 FALSE
## 10  18.1  2.55     50    14 er     9.3   3.8  FALSE
## # ... with 1,740 more rows
```

Clean Dataset

```
## Combine heating and cooling costs into one variable
hvac_clean <- hvac_long %>%
  mutate(cooling = (nchar(alt) == 3),
         ic = if_else(cooling, ich + icca, ich),
         oc = if_else(cooling, och + occa, och)) %>%
  select(id, alt, choice, ic, oc, income)
```

Cleaned Dataset

```
## Look at cleaned dataset
as_tibble(hvac_clean)
## # A tibble: 1,750 x 6
##       id alt   choice    ic    oc income
##    <int> <chr> <lgl>  <dbl> <dbl>  <dbl>
##  1   133 ec    FALSE   20.3  4.52     60
##  2   133 ecc   FALSE   25.5  7.31     60
##  3   133 er    FALSE    7.7  4.32     60
##  4   133 erc   FALSE   25.2  7.11     60
##  5   133 gc    FALSE   25.3  2.26     60
##  6   133 gcc   TRUE    23.3  5.05     60
##  7   133 hpc   FALSE   28.1  4.42     60
##  8    14 ec    FALSE   25.6  5.21     50
##  9    14 ecc   FALSE   29.2  7.76     50
## 10    14 er    FALSE    9.3  3.8      50
## # ... with 1,740 more rows
```

Convert Dataset to `mlogit` Format

```
## Convert cleaned dataset to mlogit format
hvac_mlogit <- mlogit.data(hvac_clean, shape = 'long',
                           choice = 'choice', alt.var = 'alt')
```

Dataset in `mlogit` Format

```
## Look at data in mlogit format
as_tibble(hvac_mlogit)
## # A tibble: 1,750 x 6
##       id alt   choice    ic    oc income
##    <int> <fct> <lgl>  <dbl> <dbl>  <dbl>
##  1   133 ec    FALSE   20.3  4.52     60
##  2   133 ecc   FALSE   25.5  7.31     60
##  3   133 er    FALSE    7.7  4.32     60
##  4   133 erc   FALSE   25.2  7.11     60
##  5   133 gc    FALSE   25.3  2.26     60
##  6   133 gcc   TRUE    23.3  5.05     60
##  7   133 hpc   FALSE   28.1  4.42     60
##  8    14 ec    FALSE   25.6  5.21     50
##  9    14 ecc   FALSE   29.2  7.76     50
## 10    14 er    FALSE    9.3  3.8      50
## # ... with 1,740 more rows
```

Nested Logit Models to Estimate with `mlogit()`

The representative utility of each alternative is

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

Nesting structures to consider

- Air conditioning vs. heating only with unequal within-nest correlation
- Air conditioning vs. heating only with equal within-nest correlation
- Electric vs. gas with unequal within-nest correlation
- Pairwise combinatorial logit

Nested Logit Models in R

```
### Generalized extereme value models using the mlogit package
## Help file for the mlogit function
?mlogit
## Arguments for mlogit GEV functionality
mlogit(formula, data, reflevel, nests, un.nest.el, heterosc, ...)
```

`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, or '`pcl`'
3. `un.nest.el`: `TRUE` forces all nests to have equal correlation, `FALSE` estimates a different correlation for each nest
4. `heterosc`: `TRUE` estimates a heteroskedastic logit model

Nests for Air Conditioning Vs. Heating Only

```
### Model HVAC choice as a nested logit
## Model choice using alternative intercepts, cost data, and nests for
## cooling vs. heating-only with unequal correlations
model_1 <- hvac_mlogit %>%
  mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = 'hpc',
         nests = list(cooling = c('ecc', 'erc', 'gcc', 'hpc'),
                      heating_only = c('ec', 'er', 'gc')),
         un.nest.el = FALSE)
```

Model Results with AC and Heat Nests

```
## Summarize model results
model_1 %>%
  summary()
##
## Call:
## mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = "hpc",
##     nests = list(cooling = c("ecc", "erc", "gcc", "hpc"), heating_only = c("ec",
##         "er", "gc")), un.nest.el = FALSE)
##
## Frequencies of alternatives:
##   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|)
## ec:(intercept)  -3.750232   0.761515 -4.9247 8.449e-07 ***
## ecc:(intercept)  0.606353   0.307990  1.9687  0.048982 *
## er:(intercept)  -5.127597   1.269186 -4.0401 5.344e-05 ***
## erc:(intercept)  0.421811   0.352893  1.1953  0.231972
## gc:(intercept)  -3.913846   0.813040 -4.8138 1.481e-06 ***
## gcc:(intercept)  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
```

Nests for AC Vs. Heat Only with Common Correlation

```
### Model HVAC choice as a nested logit with equal correlations
## Model choice using alternative intercepts, cost data, and nests for
## cooling vs. heating-only with a common correlation
model_2 <- hvac_mlogit %>%
  mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = 'hpc',
         nests = list(cooling = c('ecc', 'erc', 'gcc', 'hpc'),
                      heating_only = c('ec', 'er', 'gc')),
         un.nest.el = TRUE)
```

Model Results with Common Correlation

```
## Summarize model results
model_2 %>%
  summary()
##
## Call:
## mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = "hpc",
##     nests = list(cooling = c("ecc", "erc", "gcc", "hpc"), heating_only = c("ec",
##         "er", "gc")), un.nest.el = TRUE)
##
## Frequencies of alternatives:
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## bfgs method
## 6 iterations, 0h:0m:0s
## g'(-H)^-1g = 0.0404
## last step couldn't find higher value
##
## Coefficients :
##                  Estimate Std. Error z-value  Pr(>|z|)
## ec:(intercept)  -5.11652    1.05845 -4.8340 1.338e-06 ***
## ecc:(intercept)  1.33538    0.66988  1.9935 0.0462093 *
## er:(intercept)  -7.18773    1.39565 -5.1501 2.603e-07 ***
## erc:(intercept)  0.78856    0.80472  0.9799 0.3271286
## gc:(intercept)  -5.40623    1.01328 -5.3354 9.534e-08 ***
## gcc:(intercept)  1.49002    0.59701  2.4958 0.0125677 *
## ic              -0.19748    0.05086 -3.8827 0.0001033 ***
## oc              -1.17564    0.42794 -2.7472 0.0060104 **
## iv               0.47950    0.17612  2.7227 0.0064758 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -192.26
## McFadden R^2:  0.1463
```

Likelihood Ratio Test of Correlation Parameters

Test that the correlation parameters are equal

$$H_0\text{: } \lambda_k = \lambda_\ell$$

```
### Test that the nest correlations are equal
## Conduct a likelihood ratio test of the first two models
lrtest(model_1, model_2)
## Likelihood ratio test
##
## Model 1: choice ~ ic + oc | 1 | 0
## Model 2: choice ~ ic + oc | 1 | 0
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1  10 -189.83
## 2   9 -192.26 -1 4.8641    0.02742 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nests for Electric Vs. Gas

```
### Model HVAC choice as a nested logit
## Model choice using alternative intercepts, cost data, and nests for
## electric vs. gas
model_3 <- hvac_mlogit %>%
  mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = 'hpc',
         nests = list(electric = c('ec', 'ecc', 'er', 'erc', 'hpc'),
                      gas = c('gc', 'gcc')))
```

Model Results with Electric and Gas Nests

```
## Summarize model results
model_3 %>%
  summary()
##
## Call:
## mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = "hpc",
##     nests = list(electric = c("ec", "ecc", "er", "erc", "hpc"),
##         gas = c("gc", "gcc")))
##
## Frequencies of alternatives:
##   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|)
## ec:(intercept)  -12.69931    4.41999 -2.8732 0.0040640 **
## ecc:(intercept)   2.12355    2.05964  1.0310 0.3025277
## er:(intercept)  -14.46701    4.16332 -3.4749 0.0005111 ***
## erc:(intercept)  -1.22655    4.95891 -0.2473 0.8046435
## gc:(intercept)  -11.39903    3.13007 -3.6418 0.0002708 ***
## gcc:(intercept)   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
```

Likelihood Ratio Test of Nesting Structure

Can we test if the cooling vs. heating-only nests are better than the electric vs. gas nests?

- Likelihood ratio test compares a restricted model to its unrestricted counterpart
- Is there any way to write down one model as a restriction of the other?
- If not, then we cannot use a likelihood ratio test in this case

Paired Combinatorial Logit

```
### Model HVAC choice as a paired combinatorial logit
## Model choice using alternative intercepts, cost data, and paired
## combinatorial nests
model_4 <- hvac_mlogit %>%
  mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = 'hpc',
         nests = 'pcl')
```

Model Results for Paired Combinatorial Logit

```
## Summarize model results
model_4 %>%
  summary()
##
## Call:
## mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = "hpc",
##     nests = "pcl")
##
## Frequencies of alternatives:
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## bfgs method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g =  22.2
## last step couldn't find higher value
##
## Coefficients :
##                   Estimate Std. Error z-value Pr(>|z|)
## ec:(intercept)  -5.374944         NA      NA       NA
## ecc:(intercept)  2.236896         NA      NA       NA
## er:(intercept)  -8.015109         NA      NA       NA
## erc:(intercept)  0.961361         NA      NA       NA
## gc:(intercept)  -6.315344         NA      NA       NA
## gcc:(intercept)  3.086513         NA      NA       NA
## ic              -0.232755         NA      NA       NA
## oc              -1.999125         NA      NA       NA
## iv:hpc.ec        0.372322         NA      NA       NA
## iv:hpc.ecc       1.217594         NA      NA       NA
## iv:hpc.er        1.144613         NA      NA       NA
## iv:hpc.erc       0.356555         NA      NA       NA
## iv:hpc.gc        0.993960         NA      NA       NA
## iv:hpc.gcc       1.160793         NA      NA       NA
## iv:ec.ecc       -0.020867         NA      NA       NA
```

Paired Combinatorial Logit with Normalization

```
### Model HVAC choice as a paired combinatorial logit
## Model choice using alternative intercepts, cost data, and paired
## combinatorial nests with normalized parameters
model_5 <- hvac_mlogit %>%
  mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = 'hpc',
         nests = 'pcl', constPar = c('iv:hpc.er', 'iv:hpc.gc',
                                     'iv:ec.erc', 'iv:ec.gcc',
                                     'iv:ecc.er', 'iv:ecc.gc',
                                     'iv:er.gc', 'iv:er.gcc',
                                     'iv:erc.gc', 'iv:erc.gcc'))
```

Model Results for Normalized Paired Combinatorial Logit

```
## Summarize model results
model_5 %>%
  summary()
##
## Call:
## mlogit(formula = choice ~ ic + oc | 1 | 0, data = ., reflevel = "hpc",
##     nests = "pcl", constPar = c("iv:hpc.er", "iv:hpc.gc", "iv:ec.erc",
##         "iv:ec.gcc", "iv:ecc.er", "iv:ecc.gc", "iv:er.gc", "iv:er.gcc",
##         "iv:erc.gc", "iv:erc.gcc"))
##
## Frequencies of alternatives:
##   hpc    ec   ecc    er   erc    gc   gcc
## 0.104 0.004 0.016 0.032 0.004 0.096 0.744
##
## bfgs method
## 6 iterations, 0h:0m:1s
## g'(-H)^-1g = 4.62E+04
## last step couldn't find higher value
##
## Coefficients :
##                     Estimate  Std. Error z-value  Pr(>|z|)
## ec:(intercept)    -5.148636   12.895514 -0.3993  0.689703
## ecc:(intercept)    2.240558    3.249728  0.6895  0.490534
## er:(intercept)    -7.997358    5.332818 -1.4996  0.133705
## erc:(intercept)   -0.423721   65.347447 -0.0065  0.994826
## gc:(intercept)    -6.332982    2.638104 -2.4006  0.016369 *
## gcc:(intercept)    2.974811    1.484866  2.0034  0.045132 *
## ic                -0.232663    0.111786 -2.0813  0.037403 *
## oc                -1.957154    0.753594 -2.5971  0.009402 **
## iv:hpc.ec         -1.849577   66.229822 -0.0279  0.977721
## iv:hpc.ecc         2.031239    4.699801  0.4322  0.665598
## iv:hpc.erc         0.136486    0.042923  3.1798  0.001474 **
## iv:hpc.gcc         0.816480    3.646220  0.2239  0.822816
## iv:ec.ecc         -0.487297   19.439060 -0.0251  0.980001
```

Announcements

Reading for next time

- Train textbook, Chapter 6

Upcoming

- Problem Set 3 is posted, due October 31