LECTURE / DISCUSSION

Nested Logit

Generalized Extreme Value (GEV) Models

- Also known as nested logit, structured logit, sequential logit.
- Allows partial relaxation of IIA property.
- Useful when some alternatives are similar to other alternatives in unobserved factors.
- "Natural" modeling process for interrelated choices.
- Examples:
  - service options (e.g., call forwarding, waiting)
  - appliance choice
  - travel mode and destination
  - DSM program participation and implementation of DSM measures

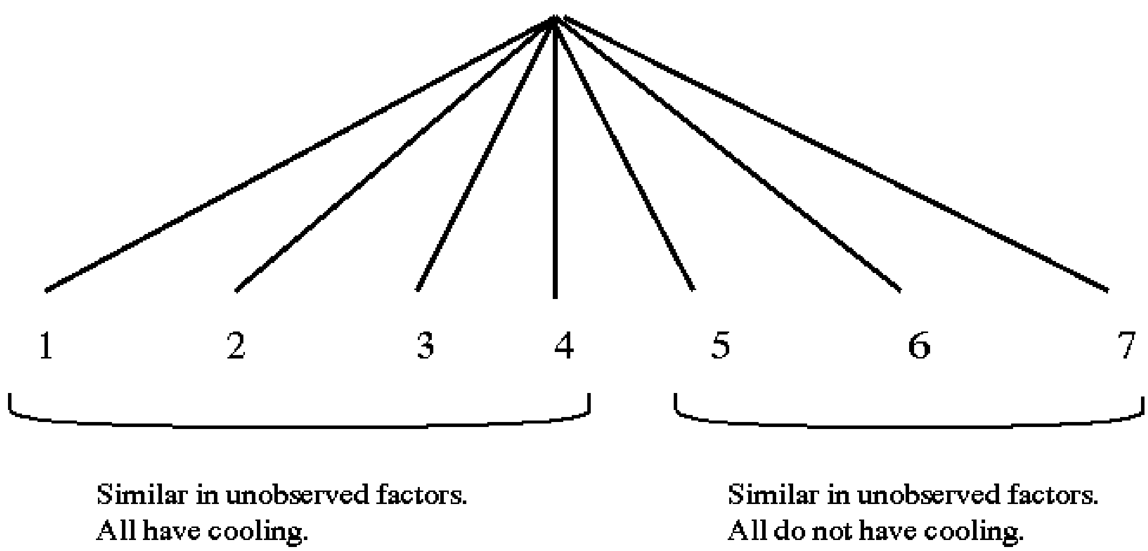
Example: Heating and Central Cooling Choice

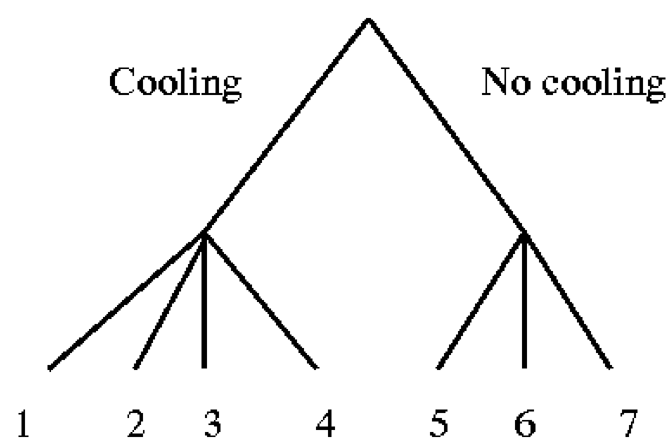
Central cooling: yes
no

Heating: gas central
electric central
electric room
heat pump

Alternatives:

1. Central cooling and gas central heating
2. Central cooling and electric central heating
3. Central cooling and electric room heating
4. Heat pump (central cooling and heating)
5. No central cooling, gas central heating
6. No central cooling, electric central heating
7. No central cooling, electric room heating


1
2
3
4
5
6
7
Similar in unobserved factors.
All have cooling.
Similar in unobserved factors.
All do not have cooling.


Cooling
No cooling
1
2
3
4
5
6
7

GEV Model

$$U_i = V_i + \varepsilon_i \quad , \qquad i = 1,\ldots,7$$

$$\varepsilon_i \sim \text{GEV}$$

with: correlation among ε_1, ε_2, ε_3, and ε_4;
correlation among ε_5, ε_6, and ε_7;
no correlation between ε_i, where $i = 1,\ldots,4$ and ε_j, where $j = 5,\ldots,6$; and
$1 - \lambda$ is a measure of correlation.

Then:

$$i = 1,\ldots,4 \qquad P_i = \frac{e^{V_i/\lambda}\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda-1}}{\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda} + \left(\sum_{j=5}^{7} e^{V_j/\lambda}\right)^{\lambda}}$$

$$i = 5,\ldots,7 \qquad P_i = \frac{e^{V_i/\lambda}\left(\sum_{j=5}^{7} e^{V_j/\lambda}\right)^{\lambda-1}}{\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda} + \left(\sum_{j=5}^{7} e^{V_j/\lambda}\right)^{\lambda}}$$

- IIA holds within nests but not across nests:

$$\frac{P_1}{P_2} = \frac{e^{V_1/\lambda}\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda-1}}{e^{V_2/\lambda}\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda-1}} = \frac{e^{V_1/\lambda}}{e^{V_2/\lambda}}$$

- depends on V_1 and V_2 only.

$$\frac{P_1}{P_5} = \frac{e^{V_1/\lambda}\left(\sum_{j=1}^{4} e^{V_j/\lambda}\right)^{\lambda-1}}{e^{V_5/\lambda}\left(\sum_{j=5}^{7} e^{V_j/\lambda}\right)^{\lambda-1}}$$

- depends on all V_1 ,..., V_7 .

- An improvement in the attributes of one alternative draws proportionately from other alternatives in the nest, but disproportionately from alternatives outside the nest.

Alternative	Original Probabilities	Rebate on Heat Pumps			
		Logit		Nested Logit	
1	.600	.54	(-10%)	.528	(-12%)
2	.055	.0495	(-10%)	.0484	(-12%)
3	.002	.0018	(-10%)	.0017	(-12%)
4	.033	.133		.133	
5	.080	.072	(-10%)	.0744	(-7%)
6	.010	.009	(-10%)	.0093	(-7%)
7	.220	.198	(-10%)	.2046	(-7%)
P_1 / P_2	10.91	10.91		10.91	
P_1 / P_5	7.5	7.5		7.09	

Decomposition of Nested Logit Model

Nested logit probabilities can be expressed as the product of two simple logits.

$$P_i = \text{Prob (nest containing i)} \times \text{Prob (i, given nest containing i)}$$

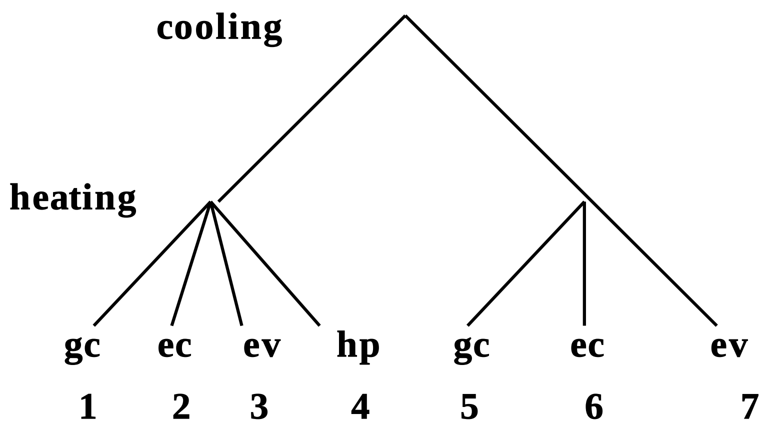
E.g.,

$$P_1 = \text{Prob (central cooling)} \bullet \text{Prob (gas central, given central cooling)}$$

Reasons for decomposition:

1. Assists with interpretation of model.
2. Allows estimation on standard logit software routines.

Schematic



Note: Household is ***not*** assumed to choose sequentially. Diagram simply represents nesting patterns and structure of system of logit models.

Then can rewrite GEV probabilities as

$$P_i = P_n \cdot P_{i|n}$$

$$P_{i|n} = \frac{e^{Y_i}}{\sum_{j \in n} e^{Y_i}}$$

$$P_n = \frac{e^{Z_n + \lambda IV_n}}{\sum_{m} e^{Z_m + \lambda IV_m}}$$

where

$$IV_n = \left(\ln \sum_{j \in n} e^{Y_j} \right)$$

and Y_i are variables that vary **across** nests,
Z_n are variables that vary **within** nests.

IV_n is called the:

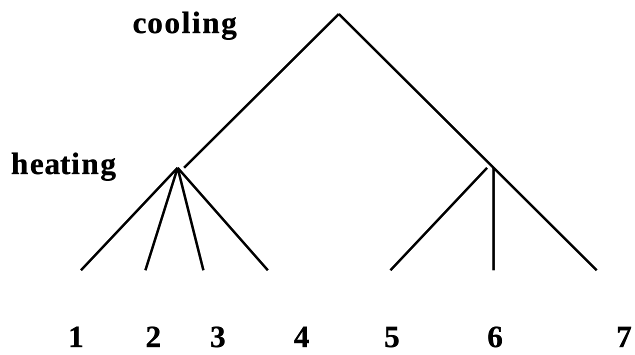
log-sum for nest n
inclusive value of nest n

IV_n is the expected utility for the choice of alternatives within nest n :

$$IV_n = E\left(\max_{j \in n} U_n\right) = E\left(\max_{j \in n} V_j + \varepsilon_j\right)$$

λ is called the log-sum coefficient.

Example:



$i = 1,2,3,4 \qquad V_i = \delta ICCA + \varphi OCCA + \alpha ICH_i + \beta OCH_i$

$i = 5,6,7 \qquad V_i = \underline{\hspace{5cm}} + \underline{\alpha ICH_i + \beta OCH_i}$

Vary across nests. Vary within nests.

Enter model of cooling choice.

Enter model of heating choice given cooling choice.

What is λ ?

Model contains an extra parameter: λ . $(1 - \lambda)$ is a measure of correlation in unobserved factors within each nest.

$\lambda = 1$	No correlation (standard logit)
.9	Low correlation
.5	Moderate correlation
.1	High correlation
0	Perfect correlation

When $\lambda = 1$ (no correlation within nests) the probabilities become simple logit:

$$P_i = \frac{e^{V_i}\left(\sum_{j=1}^{4} e^{V_j}\right)^0}{\left(\sum_{j=1}^{4} e^{V_j}\right)^1 + \left(\sum_{j=5}^{7} e^{V_j}\right)^1}$$

$$= \frac{e^{V_i}}{\sum_{j=1}^{7} e^{V_j}}$$

I. "Lower" Level Model

- logit model of heating system choice given cooling choice
- gives ***conditional*** probability of heating choice (conditional on cooling choice)

$$P(\text{gas central} \mid \text{cooling} = \text{yes}) = \frac{e^{\alpha ICH_1 + \beta OCH_1}}{\sum_{j=1}^{4} e^{\alpha ICH_j + \beta OCH_j}}$$

Similarly for other heating systems without cooling.

$$P(\text{gas central} \mid \text{cooling} = \text{no}) = \frac{e^{\alpha ICH_5 + \beta OCH_5}}{\sum_{j=5}^{7} e^{\alpha ICH_j + \beta OCH_j}}$$

Estimation of "Lower" Level

α and β enter both

P(gas central | yes)

and

P(gas central | no)

So: two "lower" models must be estimated together in this case.

Each customer gets an observation.

For customers with cooling systems, choice is between alts. 1-4.
For customers without a cooling system, choice is between alts. 5-7.

```
set cy = (depvar <= 4)
set cn = (depvar > 4)
mnl dep[depvar] \
    ivalt[ich: ich1 ich2 ich3 ich4 ich5 ich6 ich7 \
      och: och1 och2 och3 och4 och5 och6 och7]\
    censor[cy cy cy cy cn cn cn] \
    coef[beta]
```

II. "Inclusive Value" Term

- expected utility from heating choice given cooling choice
- log of denominator of "lower" model

$$IV_{c=yes} = \ln\left(\sum_{j=1}^{4} e^{\hat{\alpha}ICH_j + \hat{\beta}OCH_j}\right)$$

$$IV_{c=no} = \ln\left(\sum_{j=5}^{7} e^{\hat{\alpha}ICH_j + \hat{\beta}OCH_j}\right)$$

These variables enter the "upper" model of cooling choice.

- in SST, create inclusive value as follows:

```
set v1 = beta [1]*ich1 + beta[2]*och1
.
.
.
set v7 = beta[1]*ich7 + beta[2]*och7
set ivy = log(exp(v1) + exp(v2) + exp(v3) + exp(v4))
set ivn = log(exp(v5) + exp(v6) + exp(v7))
```

III. "Upper" Level Model

- logit model of cooling choice
- gives ***marginal*** probability of cooling choice, marginal over all possible heating systems

$$P(\text{cooling} = \text{yes}) = \frac{e^{\theta ICCA + \psi OCCA + \lambda IV_{yes}}}{e^{\theta ICCA + \psi OCCA + \lambda IV_{yes}} + e^{\lambda IV_{no}}}$$

$$P(\text{cooling} = \text{no}) = \frac{e^{\lambda IV_{no}}}{e^{\theta ICCA + \psi OCCA + \lambda IV_{yes}} + e^{\lambda IV_{no}}}$$

Note:

1. Expected utility from heating choice affects cooling choice.
2. Coefficient of inclusive value is the same for both nests in this case. In general, it need not be.

Estimation of "Upper" Level

```
set dup = 1 if[depvar <= 4]
set dup = 2 if[depvar > 4]
set zero = 0

mnl dep[dup] \
    ivalt[ic:  icca   zero \
          oc:  occa   zero \
          iv:  ivy   ivn]
```

Alternative 1 = cooling; 2 = no cooling.

Dependent variable **dup** identifies whether customer chose cooling (alts. 1-4) or no cooling (alts. 4-7).

The logit model includes the installation cost and operating cost for cooling (which are zero for no cooling) and the inclusive value terms.

SST Program to Run Sequential Estimation of a Nested Logit Model

```
clear

spool file[nest.out]

load file[nested1.sav]

set z = 0
set one = 1
set aa1 = (depvar<=4)
set aa2 = 1-aa1

mnl dep[depvar] \
        ivalt[ich: ich1 ich2 ich3 ich4 ich5 ich6 ich7 \
              och: och1 och2 och3 och4 och5 och6 och7 \
              incr: z z income z z z income] \
        coef[beta] \
        censor[aa1 aa1 aa1 aa1 aa2 aa2 aa2]

set v1 = beta[1]*ich1 + beta[2]*och1
set v2 = beta[1]*ich2 + beta[2]*och2
set v3 = beta[1]*ich3 + beta[2]*och3 + beta[3]*income
set v4 = beta[1]*ich4 + beta[2]*och4
set v5 = beta[1]*ich5 + beta[2]*och5
set v6 = beta[1]*ich6 + beta[2]*och6
set v7 = beta[1]*ich7 + beta[2]*och7 + beta[3]*income

set ivc = log(exp(v1) + exp(v2) + exp(v3) + exp(v4))
set ivr = log(exp(v5) + exp(v6) + exp(v7))

set dup = 1; if[depvar<=4]
set dup = 2; if[depvar>4]

mnl dep[dup] \
        ivalt[ica: icca z \
              oca: occa z \
              cc: one z \
              incc: income z \
              iv: ivc ivr] \
        prob[p2]

set p1 = 1-p2
cova var[p1 p2]

spool off
```

Output

Output is presented on the next two pages.

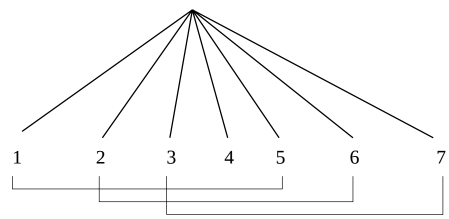
Estimated λ : 0.570

```
SST Spool File:  my.out
Tue Jun 20 10:43:45 1995

load file[nested1.sav]
set z = 0
set aa1 = (depvar<=4)
set aa2 = 1-aa1
mnl dep[depvar] \
        ivalt[ich: ich1 ich2 ich3 ich4 ich5 ich6 ich7 \
              och: och1 och2 och3 och4 och5 och6 och7 \
              incr: z z income z z z income] \
        coef[beta] \
        censor[aa1 aa1 aa1 aa1 aa2 aa2 aa2]


********** MULTINOMIAL LOGIT **********
Dependent variable: depvar

Value        Label         Count        Percent
    1                       186          74.40
    2                         4           1.60
    3                         1           0.40
    4                        26          10.40
    5                        24           9.60
    6                         1           0.40
    7                         8           3.20

ITERATION  1:   OLD LLF =      -3.37080e+002     STEP =             1.73279
                NEW LLF =      -2.13112e+002     GRAD*DIREC =       1.84520e+002

ITERATION  2:   OLD LLF =      -2.13112e+002     STEP =             1.84748
                NEW LLF =      -1.84688e+002     GRAD*DIREC =      33.50359

ITERATION  3:   OLD LLF =      -1.84688e+002     STEP =             2.30230
                NEW LLF =      -1.36743e+002     GRAD*DIREC =      53.68878

ITERATION  4:   OLD LLF =      -1.36743e+002     STEP =             1.09778
                NEW LLF =      -1.35585e+002     GRAD*DIREC =       2.17848

ITERATION  5:   OLD LLF =      -1.35585e+002     STEP =             1.00792
                NEW LLF =      -1.35582e+002     GRAD*DIREC =       5.31008e-003

At convergence grad * dir =       8.28880e-010


Independent        Estimated           Standard             t-
 Variable          Coefficient          Error            Statistic

     ich          -9.64665e-003        9.83426e-004       -9.80922
     och          -1.46792e-002        1.65619e-003       -8.86324
    incr          -0.64825             7.56586e-002       -8.56810

auxiliary statistics              at convergence     initial
log likelihood                       -135.58         -337.08
number of observations                   250
percent correctly predicted             77.6

set v1 = beta[1]*ich1 + beta[2]*och1
set v2 = beta[1]*ich2 + beta[2]*och2
set v3 = beta[1]*ich3 + beta[2]*och3 + beta[3]*income
set v4 = beta[1]*ich4 + beta[2]*och4
set v5 = beta[1]*ich5 + beta[2]*och5
set v6 = beta[1]*ich6 + beta[2]*och6
set v7 = beta[1]*ich7 + beta[2]*och7 + beta[3]*income
set ivc = log(exp(v1) + exp(v2) + exp(v3) + exp(v4))
set ivr = log(exp(v5) + exp(v6) + exp(v7))
```

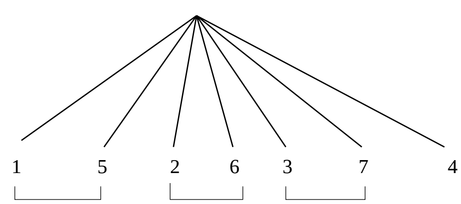
```
set one = 1
set dup = 1; if[depvar<=4]
set dup = 2; if[depvar>4]
mnl dep[dup] \
        ivalt[ica: icca z \
              oca: occa z \
              cc: one z \
              incc: income z \
              iv: ivc ivr] \
        prob[p2]

********** MULTINOMIAL LOGIT **********
Dependent variable: dup

Value        Label        Count        Percent
    1                      217          86.80
    2                       33          13.20

ITERATION  1:   OLD LLF =     -1.73287e+002      STEP =           1.82177
                NEW LLF =     -58.45817          GRAD*DIREC =     1.69974e+002

ITERATION  2:   OLD LLF =     -58.45817          STEP =           1.41577
                NEW LLF =     -45.91881          GRAD*DIREC =    19.94985

ITERATION  3:   OLD LLF =     -45.91881          STEP =           1.57150
                NEW LLF =     -42.76575          GRAD*DIREC =     4.68132

ITERATION  4:   OLD LLF =     -42.76575          STEP =           1.05160
                NEW LLF =     -42.65353          GRAD*DIREC =     0.21703

At convergence grad * dir =       2.31222e-004


Independent       Estimated            Standard             t-
 Variable         Coefficient          Error              Statistic

     ica          -2.24509e-003        1.10144e-003        -2.03832
     oca          -1.06135e-002        1.03061e-002        -1.02983
      cc          -5.83337             4.77673             -1.22121
    incc           0.24431             4.94624e-002         4.93937
      iv           0.57041             0.14902              3.82783

auxiliary statistics               at convergence      initial
log likelihood                       -42.654          -173.29
number of observations                 250
percent correctly predicted            93.6

set p1 = 1-p2
cova var[p1 p2]

Variable:       p1

Mean                    0.86785       Standard deviation       0.24571
Minimum                 2.21914e-002  Skewness                -1.95214
Maximum                 1.00000       Kurtosis                 5.75231
Valid observations       250

Variable:       p2

Mean                    0.13215       Standard deviation       0.24571
Minimum                 5.45636e-008  Skewness                 1.95214
Maximum                 0.97781       Kurtosis                 5.75231
Valid observations       250


spool off
```

Alternative Nesting Patterns

Example: Nest according to heating system:



or, rearranged:



$$P_1 = \frac{e^{V_1/\lambda}\left(\sum_{j=1,5} e^{V_j/\lambda}\right)^{\lambda-1}}{\left(\sum_{j=1,5} e^{V_j/\lambda}\right)^{\lambda} + \left(\sum_{j=2,6} e^{V_j/\lambda}\right)^{\lambda} + \left(\sum_{j=3,7} e^{V_j/\lambda}\right)^{\lambda} + e^{V_4}}$$

Criteria for Nesting Alternatives

1. Based on specific problem, partition choice set into mutually exclusive subsets within which (a) unobserved factors are correlated, and (b) relative odds are independent of other alternatives.

2. Run alternative nested models and examine IV coefficient. May reject nesting structure if value is implausible or not significantly different from one.

3. Apply formal tests for IIA.

More Complex GEV Models

1. Nests within nests.

2. Overlapping nests.

WORKSHOP ON SEQUENTIAL ESTIMATION OF NESTED LOGIT MODELS

Sequential Estimation

1. Run the model that was described in lecture. The file **nest.cmd** will do it. Verify that your results match those given in lecture, and that you understand each step of the program and output.

2. People might evaluate heating systems differently if they have a cooling system than if they do not. Run a nested logit model in which two separate "lower" models are estimated (one for each nest). However, as in **nest.cmd**, require the inclusive value term to have the same coefficient for each nest.

3. The correlation in unobserved factors might be different for the two nests. Run a nested logit model in which inclusive value has a different coefficient for the two nests. (As in step 2, allow each lower model to have different parameters.)

   Which nest has more highly correlated errors? What reason can you give for this result?

   Does it seem that this model is better or worse than those obtained in steps 1 and 2?

4. Think about other nesting structures. Give reasons for and against other structures.

DISCUSSION OF WORKSHOP RESULTS

Sequential Estimation

- Parameter estimates are consistent but not efficient.
- Standard errors in upper level model are biased downward.
- Parameters that enter both levels are not constrained to be equal.

Simultaneous Maximum Likelihood Estimation

- Estimate all levels simultaneously using likelihood estimation for full model.

- Asymptotically efficient; it uses all the information available from data to obtain best parameter estimates at all levels.

- Standard errors are unbiased.

- LIMDEP is only estimation package available. Convergence not guaranteed.

- Advisable to estimate model sequentially to obtain starting values for FIML estimation.