Lecture 5
Multiple Choice Models
Part I – MNL, Nested Logit

1

DCM: Different Models

• Popular Models:

1. Probit Model
2. Binary Logit Model
3. Multinomial Logit Model
4. Nested Logit model
5. Ordered Logit Model

• Relevant literature:

- Train (2003): **Discrete Choice Methods with Simulation**
- Franses and Paap (2001): **Quantitative Models in Market Research**
- Hensher, Rose and Greene (2005): **Applied Choice Analysis**

Multinomial Logit (MNL) Model

• In many of the situations, discrete responses are more complex than the binary case:

- Single choice out of more than two alternatives: Electoral choices and interest in explaining the vote for a particular party.

- Multiple choices: "Travel to work in rush hour," and "travel to work out of rush hour," as well as the choice of bus or car.

• The distinction should not be exaggerated: we could always enumerate travel-time, travel-mode choice combinations and then treat the problem as making a single decision.

• In a few cases, the values associated with the choices will themselves be meaningful, for example, number of patents: $y = 0, 1, 2,...$ (count data). In most cases, the values are meaningless.

Multinomial Logit (MNL) Model

• In most cases, the value of the dependent variable is merely a coding for some qualitative outcome:

- Investment in stocks: we code "yes" as 1 and "no" as 0 (qualitative choices)

- Occupational field: 0 for economist, 1 for engineer, 2 for lawyer, etc. (categories)

- Opinions are usually coded with scales, where 1 stands for "strongly disagree", 2 for "disagree", 3 for "neutral", etc.

• Nothing conceptually difficult about moving from a binary to a multi-response framework, but numerical difficulties can be big.

• A simple model to generalized: The Logit Model.

Multinomial Logit (MNL) Model

• Now, we have a choice between J (greater than 2) categories

• Dependent variable y_j = 1, 2, 3, …., J.

• Explanatory variables

- z_n: different across individuals, not across choices (*standard MNL* model). The MLN specifies for choice 1, 2, 3, …., J:

$$P[y_n = j \mid z_n = z] = \frac{exp(z'\alpha_j)}{1 + \sum_{l=1}^{J} exp(z'\alpha_l)}$$

- x_n: different across (individuals and) choices (*conditional MNL* model). The conditional logit model specifies for choice j:

$$P[y_n = j \mid x_n] = \frac{exp(x_n'\beta_j)}{\sum_{l=1}^{J} exp(x_n'\beta_l)}$$

• Both models are easy to estimate.

Multinomial Logit (MNL) Model

• The MNL can be viewed as a special case of the conditional logit model. Suppose we have a vector of individual characteristics Z_i of dimension k, and J vectors of coefficients α_j, each of dimension k. Then define,

$$X_{i1} = \begin{pmatrix} Z_i \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots\ldots \quad X_{iJ} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ Z_i \end{pmatrix}, \text{ and } X_{i0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and define the common parameter vector β as $\beta' = (\gamma_1', \ldots, \gamma_J')$.

• We are back in the conditional logit model.

MNL – Link with Utility Maximization

• The modeling approach (McFadden's) is similar to the binary case.

- Random Utility for individual n, associated with choice j:

$U_{nj} = V_{nj} + \varepsilon_{nj} = \alpha_j + \boldsymbol{z}_n{}' \delta_j + \boldsymbol{w}_n{}' \gamma_j + \varepsilon_{nj}$ - utility from j

Then, if $y_n = j$ if (U_{nj} - U_{ni}) > 0 (n selects j over i.)

<u>Note</u>: The utility parameters/weights (δ_j, γ_j) are common across consumers (same for all n). This is a strong assumption. This can be relaxed. The constant picks heterogeneity among choices, which cannot be attributed to $\boldsymbol{z}_n$ & $\boldsymbol{w}_n$, for example "image."

• If we have a panel, we can allow ε_{nj} to be autocorrelated. We think of ε_{nj} as picking unobserved attributes of choice j that change over time. For example, time-varying tastes, that can be correlated.

MNL – Link with Utility Maximization

• Like in the binary case, we get:

$$\begin{aligned} P_{nj} &= P[y_n = j|i, j] = \mathrm{P}[U_{nj} - U_{nj} > 0, \forall\, i \neq j] \\ &= P[\varepsilon_{ni} - \varepsilon_{nj} < V_{nj} - V_{ni}, \forall\, i \neq j] \\ &= \int I\left[\xi_n < V_{nj} - V_{ni}, \forall\, i \neq j\,\right] f(\xi_n)\, d\xi_n \end{aligned}$$

• Specify *i.i.d.* Gumbel distribution for $f(\varepsilon_n)$ ⇒ Logit Model.

- independence across utility functions
- identical variances (means absorbed in constants)

MNL Model - Identification

• Logit Model for choice j:

$$P_{nj} = P[y_n = j|\, \boldsymbol{x}_n] = \frac{exp(x_n{'}\beta_j)}{\sum_{l=1}^{J} exp(x_n'\beta_l)}$$

• We call P_{nj} a conditional MNL model ($\boldsymbol{x}_n$: different across n and j)

• Normalization. If we add a constant to a parameter ($\beta_i + c$), given the Logistic distribution, exp(c) will cancel out. Cannot distinguish between ($\beta_i + c$) & β_i. A normalization is needed:

⇒ pick a category, say i, and set coefficients equal to 0 –i.e., $\beta_i = \mathbf{0}$.

$$P_{ni} = P[y_n = i|\, \boldsymbol{x}_n] = \frac{1}{1 + \sum_{l=1}^{J} exp(x_n'\beta_l)} \qquad \text{(Typically, } i = J.\text{)}$$

$$P_{nj} = P[y_n = j|\, \boldsymbol{x}_n] = \frac{exp(x_n{'}\beta_j)}{1 + \sum_{l\neq i}^{J} exp(x_n'\beta_l)}$$

MNL Model – Interpretation & Effects

• The interpretation of parameters is based on partial effects:

– Derivative (marginal effect)

$$\frac{\partial P[y_n = j|\, x_n]}{\partial x_{nk}} = P_{nj} * (1 - P_{nj}) * \beta_k$$

– Elasticity (proportional changes)

$$\frac{\partial log P[y_n = j|\, x_n]}{\partial log x_{nk}} = \frac{x_{nk}}{P_{nj}} * P_{nj} * (1 - P_{nj}) * \beta_k$$

$$= \boldsymbol{x}_{nk} * (1 - P_{nj}) * \beta_k$$

Note: The elasticity is the same for all choices "j." A change in the cost of air travel has the same effect on all other forms of travel. (This result is called **independence from irrelevant alternatives** (**IIA**). Not a realistic property. Many experiments reject it.)

MNL Model – Interpretation & Effects

• Interpretation of parameters

– Probability-ratio

$$\frac{P[y_n=j|\, \boldsymbol{x}_n]}{P[y_n=i|\, \boldsymbol{x}_n]} = \frac{exp(\boldsymbol{x}_{nj}'\boldsymbol{\beta})}{exp(\boldsymbol{x}_{ni}'\boldsymbol{\beta})}$$

$$\ln\left\{\frac{P[y_n=j|\, \boldsymbol{x}_n]}{P[y_n=i|\, \boldsymbol{x}_n]}\right\} = (\boldsymbol{x}_{nj} - \boldsymbol{x}_{ni})'\boldsymbol{\beta}$$

– Does not depend on the other alternatives! Again, we have IIA. Implication of MNL models pointed out by Luce (1959).

Note: The log-odds ratio of each response follow a linear model. A regression can be used for the comparison of two choices at a time.

MNL Model – Estimation

- Estimation
  - ML estimation

$$L(\beta) = \prod_n \prod_j P(y_n = j \mid x_n)^{D_{nj}}$$

$$LogL(\beta) = \sum_n \sum_j D_{nj} \ln P(y_n = j \mid x_n)$$

$$LogL(\beta) = \sum_n \sum_j D_{nj} \ln\left(\frac{\exp(x_{nj}'\beta)}{\sum_k \exp(x_{nj}'\beta)}\right)$$

$$= \sum_n \sum_j D_{nj} (\ln(\exp(x_{nj}'\beta) - \ln(\sum_k \exp(x_{nj}'\beta)))$$

$$= \sum_n (\sum_j D_{nj} (x_{nj}'\beta)) - \ln(\sum_k \exp(x_{nj}'\beta))$$

where $D_{nj} = 1$ if j is selected, 0 otherwise)

MNL Model – Estimation

• Estimation

- ML (continuation):

  A lot of *f.o.c.* equations, with a lot of unknowns (parameters).

  Each covariate has $J-1$ coefficients.

  We use numerical procedures, G-N or N-R often work well.

- Alternative estimation procedures

  Simulation-assisted estimation (Train, Ch.10)

  Bayesian estimation (Train, Ch.12)

MNL Model – Application - PIM

- **Example** (from Bucklin and Gupta (1992)):

$$P_t^h(i|inc) = \frac{exp(U_{it}^h)}{\sum_{l=1}^{J} exp(U_{lt}^h)}$$

$$U_{it}^h = u_i + \beta_1 BL_i^h + \beta_2 LBP_{it}^h + \beta_3 SL_i^h + \beta_4 LSP_{it}^h + \beta_5 Price_{it} + \beta_6 Pro_{it}$$

- u_i= constant for brand-size i
  - BL_i^h = loyalty of household h to brand of brandsize i
  - LBP_{it}^h = 1 if i was last brand purchased, 0 otherwise (at time t)
  - SL_i^h= loyalty of household h to size of brandsize i
  - LSP_{it}^h = 1 if i was last size purchased, 0 otherwise (at time t)
  - $Price_{it}$ = actual shelf price of brand-size i at time t
  - Pro_{it} = promotional status of brand-size i at time t

MNL Model – Application - PIM

- Data
  - A.C.Nielsen scanner panel data
  - 117 weeks: 65 for initialization, 52 for estimation
  - 565 households: 300 selected randomly for estimation, remaining hh = holdout sample for validation
  - Data set for estimation: 30,966 shopping trips, 2,275 purchases in the category (liquid laundry detergent)
  - Estimation limited to the 7 top-selling brands (80% of category purchases), representing 28 brand-size combinations (= level of analysis for the choice model)

MNL Model – Application I – PIM

• Goodness-of-Fit

Model	# param.	Log Likelihood	U^2 (pseudo R^2)	BIC
Null model	27	-5957.3	-	6061.6
Full model	33	-3786.9	.364	3914.3

• Estimation Results

Parameter	Coefficients (t-statistic)
BL β_1	3.499 (22.74)
LBP β_2	.548 (6.50)
SL β_3	2.043 (13.67)
LSP β_4	.512 (7.06)
$Price$ β_5	-.696 (-13.66)
Pro β_6	2.016 (21.33)

MNL Model – Application II – Travel Mode

• Data: 4 Travel Modes: Air, Bus, Train, Car. N=210

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                 Choice
Log likelihood function      -256.76133
Estimation based on N =    210, K =   7
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              2.51201      527.52265
Fin.Smpl.AIC     2.51465      528.07711
Bayes IC         2.62358      550.95240
Hannan Quinn     2.55712      536.99443
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .0951 .0850
Chi-squared[ 4]          =     53.99489
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      GC|     .03711**        .01484        2.500   .0124   (General Cost)
    INVC|    -.05480***       .01668       -3.285   .0010   (in-vehicle cost)
    INVT|    -.00896***       .00215       -4.162   .0000   (in-vehicle time)
   HINCA|     .02922***       .00931        3.138   .0017   (household income annual)
   A_AIR|   -1.88740***       .69281       -2.724   .0064   (air FE)
 A_TRAIN|     .69364***       .25010        2.773   .0055   (train FE)
   A_BUS|    -.20307          .24817        -.818   .4132   (bus FE)
--------+--------------------------------------------------
```

MNL Model – Application II – Travel Mode

CLOGIT Fit Measure:

• Based on the log likelihood

```
Log likelihood function      -256.76133
Constants only   -283.7588  .0951 .0850
Chi-squared[ 4]          =     53.99489
```

Values in parentheses below show the number of correct predictions by a model with only choice specific constants.

• Based on the model predictions

```
+-----------------------------------------------------+
| Cross tabulation of actual vs. predicted choices.    |
| Row indicator is actual, column is predicted.        |
| Predicted total is F(k,j,i)=Sum(i=1,...,N) P(k,j,i). |
| Column totals may be subject to rounding error.      |
+-----------------------------------------------------+
Matrix Crosstab has  5 rows and  5 columns.
         AIR          TRAIN         BUS          CAR          Total
        +----------------------------------------------------------------------
AIR     |    35.0000 (16)   7.0000        4.0000       13.0000        58.0000
TRAIN   |     7.0000       41.0000 (19)   4.0000       11.0000        63.0000
BUS     |     5.0000        4.0000       16.0000 (4)    4.0000        30.0000
CAR     |    11.0000       11.0000        6.0000       31.0000 (17)   59.0000
Total   |    58.0000       63.0000       30.0000       59.0000       210.0000
```

MNL Model – Scaling

• Scale parameter

• Variance of the extreme value distribution Var[ε] = $\pi^2/6$
- If true utility is $U^*_{nj} = \boldsymbol{x}_{nj}'\beta + \varepsilon^*_{nj}$ with Var(ε^*_{nj}) = σ^2 ($\pi^2/6$), the estimated representative utility $V_{nj} = \boldsymbol{x}_{nj}'\beta$ involves a rescaling of β^*
$\Rightarrow \beta = \beta^* / \sigma$

• β^* and σ can not be estimated separately.

$\Rightarrow$ Take into account that the estimated coefficients indicate the variable's effect *relative to* the variance of unobserved factors

$\Rightarrow$ Include scale parameters if subsamples in a pooled estimation (may) have different error variances

MNL Model – Scaling

• Scale parameter in the case of pooled estimation of subsamples with different error variance.

• For each subsamples, multiply utility by μ_s, which is estimated simultaneously with β.

• Normalization: set μ_s equal to 1 for 1 subs.

• Values of μ_s reflect diff's in error variation
  - $\mu_s > 1$: error variance is smaller in s than in the reference subsample
  - $\mu_s < 1$: error variance is larger in s than in the reference subsample

MNL Model – Application

Example (from Breugelmans et al. (2005), based on Andrews and Currim (2002); Swait and Louvière (1993)):

- Data from online experiment, 2 product categories
- Three different assortments, assigned to different respondent groups
  - Assortment 1: small assortment
  - Assortment 2 = ass.1 extended with additional brands
  - Assortment 3 = ass.1 extended with add types
- Explanatory variables are the same (hh char's, MM), with exception of the constants
- A scale factor is introduced for assortment 2 and 3 (assortment 1 is reference with scale factor =1)

MNL Model – Application

Table 1: Descriptive stats for each assortment (margarine and cereals)

	MARGARINE		
Attribute	**Assortment 1 (limited)**	**Assortment 2 (add new flavors of existing brands)**	**Assortment 3 (add new brands of existing flavors)**
Brand	Common [a]	Common	Common Add new brands
Flavor	Common	Common Add new flavors	Common
# alternatives	11	19	17
# respondents	105	116	100
# purchase occasions	275	279	278
# screens needed	< 1	> 1	> 1
	CEREALS		
Attribute	**Assortment 1 (limited)**	**Assortment 2 (add new flavors of existing brands)**	**Assortment 3 (add new brands of existing flavors)**
Brand	Common	Common	Common Add new brands
Flavor	Common	Common Add new flavors	Common
# alternatives	21	32	46
# respondents	81	97	87
# purchase occasions	271	261	281
# screens needed	> 1	> 1	> 1

MNL Model – Application

- MNL-model – Pooled estimation

$$P^h_{it|a} = \frac{\mu_a(u^h_{it|a})}{\sum^J_{l \epsilon C^h_{t|a}} \mu_a(u^h_{lt|a})}$$

Notation:

- $P^h_{it|a}$ = Probability that household h chooses item i at time t, facing assortment a
- $u^h_{it|a}$ = Choice utility of item i for household h facing assortment a
  = f(household variables, MM-variables)
- $C^h_{t|a}$= Set of category items available to household h within assortment a
- μ_a = Gumbel scale factor

MNL Model – Application

Estimation results

- Goodness-of-Fit
  – (average) LL: -0.045 (M), -0.040 (C)
  – BIC: 2929 (M), 4763(C)
  – CAIC: 2871 (M), 4699 (C)

- Scale factors:
  – M: 1.2498 (ass2), 1.2627 (ass3)
  – C: 1.0562 (ass2), 0.7573 (ass3)

MNL Model – Application

Margarine				Cereals			
Variable	Assortment 1	Assortment 2	Assortment 3	Variable	Assortment 1	Assortment 2	Assortment 3
Scale factor	[1.00][b]	1.2498***	1.2627***	Scale factor	[1.00][b]	1.0562***	0.7573***
Mean	2.0675***	[2.5840***][c]	[2.6106***][c]	**Mean**	0.6441***	[0.6803***][c]	[0.4888***][c]
Last purchase	2.8310***	[3.5382***][c]	[3.5747***][c]	Last purchase	5.2011***	[5.4934***][c]	[3.9109***][c]
Item preference	0.2805	0.4228**	0.5400*	Item preference	0.0077	0.6130	0.0969
Brand asymmetry	-0.0841	-0.0880	0.0169	Brand asymmetry	-0.0260	0.2938**	-0.1596
Size asymmetry	- [d]	0.3672**	-0.1190	Taste asymmetry	0.3119	-0.0614	0.3816**
Sequence	0.8332	1.0303***	0.6235	Type asymmetry	-0.3311	-0.0695	0.6190***
Proximity				Sequence	2.0041***	0.7214	4.1140***
				Proximity			

(Excluding brand/size constants)

MNL Model – Limitations

- Limitations of the MNL model:
  - Independence of Irrelevant Alternatives or IIA (proportional substitution pattern): the relative odds between any two outcomes are independent of the number and nature of other outcomes being simultaneously considered.
  - Order (where relevant) is not taken into account
  - Systematic taste variation can be represented, not random taste variation
  - No correlation between error terms (*i.i.d.* errors)

MNL Model – IIA

• This is the big weakness of the model. The choice between any two alternatives does not depend upon a third one -i.e., the ratio of choice probabilities for alternatives i and j does not depend on characteristics of other alternatives, say, x_{i3}.

$$\frac{P[y_n=j|\, x_n]}{P[y_n=i|\, x_n]} = \frac{exp(x_{nj}'\boldsymbol{\beta}_j)}{exp(x_{ni}'\boldsymbol{\beta}_i)}$$

• Implications: Proportional substitution patterns (or unrealistic substitution patterns!). It is possible to ignore third alternatives in estimation.

• But, it clashes with data

MNL Model – IIA

Example (McFadden (1974)): Blue Bus – Red Bus:

Suppose we have three equally distributed transportation categories:

- T1: Blue bus (P=33%), Car (P=33%), Red bus (P=33%)

Now, we paint the red busses blue. Then, we have two choices. Assuming IIA, we have: Blue bus (P=50%), Car (P=50%).

But, a more likely distribution: Blue bus (P=66%), Car (P=33%).

Note: Debreu (1960) has a similar example with Beethoven/Debussy.

• MNL model assumes that none of the categories can serve as substitutes (no correlation). If they can serve as substitutes, then the results of MNL may not be very realistic.

MNL Model – IIA - Testing

• We want to test IIA. We use an adaptation of the Hausman test: Hausman-McFadden specification test (*Econometrica,* 1983)

<u>Idea</u>: If a subset of the choice set is truly irrelevant, omitting it should not significantly affect the estimates. Two estimators: one efficient, one inefficient ⇒ Hausman test.

• Steps for Wald test:
- Estimate logit model twice:
(a) on full set of alternatives (with k "irrelevant" variables) ⇒ $\mathbf{b}_a$
(b) on subset of alternatives (& subsamples with choices from this set) ⇒ $\mathbf{b}_b$
- Compute the Wald test
- Under H_0 (IIA is true): $H = (\mathbf{b}_a - \mathbf{b}_b)'[\mathbf{V}_a - \mathbf{V}_b]^{-1}(\mathbf{b}_a - \mathbf{b}_b) \xrightarrow{a} \chi^2_k$

MNL Model – IIA - Testing

• Steps for LR test:
- Estimate logit model twice:
  a. on full set of alternatives ⇒ get LogL_a
  b. on subset of alternatives ⇒ get LogL_b
  (and subsample with choices from this set)

- Compute LR test as usual. Asymptotically, goodness of fit should be similar to Wald test.

Model – IIA: Alternative Models

• In the MNL model we assumed independent ε_{nj} with extreme value distributions. This essentially created the IIA property.

• This is not completely correct, because other distributions for the unobserved, say with normal errors, we would not get IIA exactly, but something close to it.

• The solution to the IIA problem is to relax the independence between the unobserved components of the latent utility, ε_i.

• There are a number of ways to go.

Model – IIA: Alternative Models

• Solutions to IIA

– **Nested Logit Model**, allowing correlation between some choices.

– Models that allow for correlation among the error terms, such as **Multinomail Probit Models**

– **Mixed or random coefficients Logit**, where the marginal utilities associated with choice characteristics are allowed to vary between individuals.

All of these originate in some form or another in McFadden's work (1981, 1982, and 1984).

Nested Logit Model

• Idea: We have J choices. We call a group of similar choices (alternatives) a **nest**. We allow correlations between the choices through nesting them. Each choice belongs to exactly one nest.

• Details:

- Grouping: We group together or cluster sets of choices into S sets: B_1, B_2, ..., B_S, with nest characteristics $\boldsymbol{Z}_s$.

- Correlations: Choices are correlated inside the nest (B_1 (Bus nest) = red bus, blue bus). But, we force independence between the nests.

- Preferences as before: Individuals choosing the option with the highest utility, where the utility of choice j in set B_s for individual n is

$$U_{nj} = \boldsymbol{x}_{nj}{}' \beta + \boldsymbol{Z}_s{}'\alpha + \varepsilon_{nj}, \qquad \varepsilon_{nj} \sim \text{GEV}$$

Nested Logit Model

• ε_{nj} has a the joint cumulative distribution function of error terms is

$$F(\varepsilon_{n1}, \varepsilon_{n2}, ..., \varepsilon_{nJ}) = \exp\left(-\sum_{s=1}^{S} \left(\sum_{j \in B_s} e^{-\varepsilon_{nj}/\lambda_s} \right)^{\lambda_s} \right)$$

• Within the sets the correlation coefficient for the ε_{nj} is approximately equal to $1-\lambda$. Between the sets choices i & j are independent.

Example: Green investment, with 4 choices: ETF, Passive Mutual Fund, Active Mutual Fund & Roboinvestment. We allow correlation among ETF & Passive MF (ε_{ETF}, ε_{P-MF} correlated) and among Active MF & Roboinvestment (ε_{A-MF}, ε_{Robo}).

Nested Logit Model – Probability

• In the MNL, we assume independence of errors. In the Nested Logit model framework, we assume the errors ε_{ETF} & ε_{P-MF} share common unobserved attributes.

• The key in this model is how we form the nests. Different nest structures will produce different result.

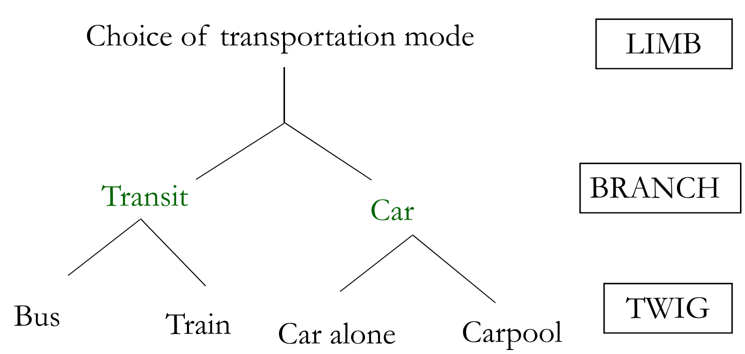
• We choose the choices that are potentially close, with the data being used to estimate the amount of correlation.

Nested Logit Model – Example

Example: Transportation mode choice.
Choices: Bus, Train, Car (alone) & Carpool $\Rightarrow J = 4$.

• Nests: - Transit: Bus and Train
- Car: Car (alone) and Carpool

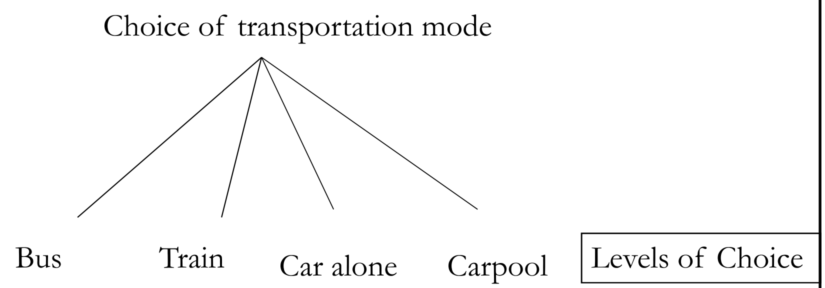
Levels of Choice



Nested Logit Model – Example

Example: Transportation mode choice.
Choices: Bus, Train, Car (alone) & Carpool $\Rightarrow J = 4$.

• No Nesting



Note: We are not assuming that individuals choose sequentially. The diagrams simply represents nesting patterns and structure of system of logit models.

Nested Logit Model – Structure

• We cluster similar choices in nests or branches.

• The RUM as usual: $U_{nj} = V_{nj} + \varepsilon_{nj}$.

• But, it gets complicated. Now, we compound utility:

$$U(\text{Choice}) = U(\text{Choice}|\text{Branch}) + U(\text{Branch})$$

- U(Branch) = function of some variables Z (characteristics of the branch, say comfort of ride, speed, price, etc).

- U(Choice|Branch) = function of some variables X (age, education, income). These variables vary across choices.

Nested Logit Model – Structure

• Within a branch
  – Identical variances (IIA applies)
  – Covariance (all same) = variance at higher level

• Branches have different variances (scale factors)

• Nested logit probabilities: Generalized Extreme Value

• Prob[Choice, Branch] = Prob(Branch) * Prob(Choice|Branch)

⇒ We need two models:
 1) Model of branch selection
 2) Model of Choice, given branch selection

Nested Logit Model – Probability

• Recall that Z_S be branch/set-specific characteristics. (It may be empty, an indicator variable for set S, etc.). This set influences your choice of branch.

• Let the conditional probability of choice j given that your choice is in the set B_S, or $y_n \in B_S$ ("twig level probability") be equal to:

$$P[y_n = j|\, \boldsymbol{x}_n, y_n \in\ B_S] = \frac{exp(V_{nj}/\lambda_S)}{\sum_{l=1}^{J} exp(V_{nl}/\lambda_S)}$$

for $j \in B_S$, and 0 otherwise.

This probability describes the *lower level model* –describes choice within the nest or at twig level, given a branch.

• Unusual notation: Correlation inside the nest = 1 – λ_S, $\lambda_S \in$ [0, 1].

Nested Logit Model – Probability

• Suppose the marginal probability of each choice in the set B_s:

$$P_{nBs} = P[Y_n \in B_s \mid X_n] \frac{\exp(Z_s'\alpha)\left(\sum_{k \in B_S} \exp(V_{nk}/\lambda_S)\right)^{\lambda_s}}{\sum_{l=1}^{S} \exp(Z_l'\alpha)\left(\sum_{k \in B_l} \exp(V_{nk}/\lambda_l)\right)^{\lambda_l}}$$

This is the *upper level model* –describes choices between nests (probability of a branch).

• If $\lambda_s = 1$ for all s –i.e., no correlation within the nest-, then

$$P_{nj} = \frac{exp(V_{nj} + \mathbf{Z}_s'\alpha)}{\sum_{s=1}^{S} \sum_{l \in B_s} exp(V_{ns} + \mathbf{Z}_s'\alpha)}$$

• We are back to the conditional logit model.

Nested Logit Model – Summary

• The nested logit probability can be decomposed into 2 logit models:

P_j = Prob[nest containing j] * Prob[j, given nest containing j]

$$P_{nj} = P_{nj} * P_{n,B_s}$$

where

$$P_{nj|B_s} = \frac{exp(V_{nj}/\lambda_s)}{\sum_{l \in B_s}^{J} exp(V_{nl}/\lambda_s)} \qquad \text{(1) Lower level model}$$

$$P_{n,B_s} = \frac{exp(\mathbf{Z}_{ns}'\alpha + \lambda_s\, IV_{ns})}{\sum_{s \in B_s}^{S} exp(\mathbf{Z}_{ns}'\alpha + \lambda_s\, IV_{ns})} \qquad \text{(2) Upper level model}$$

$$IV_{ns} = \ln(\textstyle\sum_{s \in B_s}^{S}(V_{ns}/\lambda_s)) \qquad \text{(3) Inclusive value}$$

• There is a link between $P_{nj|B_s}$ * $P_{n|B_s}$ (upper and lower level): the *inclusive value* IV_{nk} --the log of the denominator of *lower level model*.

Nested Logit Model – Summary

• IV_{B_s} is also called the log-sum for nest B_s. It represents the expected utility for the choice of alternatives within nest B_s.

$$IV_{B_s} = E[\max_{j \in B_s} (U_j)] = E[\max_{j \in B_s} (V_j + \varepsilon_j)]$$

• For consistency with RUM, λ_k must be in the [0, 1] interval (*sufficient condition*) –see McFadden (1981). The value of λ_k can serve as a check on the nested logit model.

• IIA within, not across nests:

$$\frac{P_{ni}}{P_{nj}} = \frac{\exp(V_{ni}/\lambda_k)\left(\sum_{m \in B_k} \exp(V_{nm}/\lambda_k)\right)^{\lambda_k - 1}}{\exp(V_{nj}/\lambda_l)\left(\sum_{m \in B_l} \exp(V_{nm}/\lambda_l)\right)^{\lambda_l - 1}}$$

• When $\lambda_k = 1 \Rightarrow$ no correlation within nests: $\frac{P_{ni}}{P_{nj}} = \frac{exp(V_{ni})}{exp(V_{nj})}$

Nested Logit Model – Summary

Example: Transportation mode with 4 choices(Bus, Train, Car (alone) & Carpool) and 2 nests (Transit: Bus and Train; Car: Car (alone) and Carpool).

- *Lower level model.* It gives conditional probability of transit choices – conditional on choosing transit mode. For exampe, conditional probability of choosing Bus, conditional on choosing the Transit nest:

$$P[Y_n = Bus \mid X_n, Y_n \in B_{Transit}] = \frac{\exp(x_{nBus}'\beta/\lambda_{Transit})}{\exp(x_{nBus}'\beta/\lambda_{Transit}) + \exp(x_{nTrain}'\beta/\lambda_{Transit})}$$

Similarly, conditional probability of choosing Carpool, conditional on choosing the Car nest:

$$P[Y_n = Carpool \mid X_n, Y_n \in B_{Car}] = \frac{\exp(x_{nCarpool}'\beta/\lambda_{Car})}{\exp(x_{nCarpool}'\beta/\lambda_{Car}) + \exp(x_{nCar-alone}'\beta/\lambda_{Car})}$$

Nested Logit Model – Summary

Example (continuation).

Note: β enters into both equations ⇒ simultaneous estimation

- IIA holds within nests:

$$\frac{P[Y_n = Bus \mid X_n, Y_n \in B_{Transit}]}{P[Y_n = Train \mid X_n, Y_n \in B_{Transit}]} = \frac{\exp(x_{nBus}'\beta/\lambda_{Transit})}{\exp(x_{nTrain}'\beta/\lambda_{Transit})}$$

it depends on $\boldsymbol{x}_{bus}$ and $\boldsymbol{x}_{Train}$ only.

- *Inclusive value*: Expected utility from choice given branch choice

$$IV_{Transit} = \ln\left[\exp(x_{nBus}'\beta/\lambda_{Transit}) + \exp(x_{nTrain}'\beta/\lambda_{Transit})\right]$$
$$IV_{Car} = \ln\left[\exp(x_{nCarpool}'\beta/\lambda_{Car}) + \exp(x_{nCar-alone}'\beta/\lambda_{Car})\right]$$

Nested Logit Model – Summary

Example (continuation).

- *Upper level model*. It gives the probability of choosing a nest/branch. For example, the probability of choosing Transit:

$$P[Y_n \in B_{Transit} \mid X_n] = \frac{\exp(Z_{nTransit}'\alpha + \lambda_{Transit} IV_{nTransit})}{\sum_{l=Transit,Bus} \exp(Z_l'\alpha + \lambda_l IV_{nl})}$$

Nested Logit Model – Estimation

• Estimation

- ML joint estimation.

Complicated, especially since the log likelihood function is not concave, but it is not impossible. Convergence is not guaranteed.

- Sequential estimation using nesting structure:

(1) Estimate lower model: Within the nest we have a conditional MNL with coefficients β/λ_s. (Easy to estimate, log likelihood is concave.)

(2) Compute *inclusive value*, $\ln(\sum_{l \in B_s}^{J} exp(V_{nl}/\lambda_s))$, using the estimates of β/λ_s.

(3) Estimate upper model with inclusive value as explanatory variable: Plug the estimates of β/λ_s in P_{ni}. Another conditional MNL model.

Nested Logit Model – Estimation

• Disadvantages sequential estimation

- The sequential (two-step) estimators are not efficient.
- The covariance has to be computed separately –McFadden (1981).
- Parameters that enter both levels are not constrained to be equal.
- It does not insure consistency with utility maximization.

Note: We can use the parameter estimates from the sequential estimation as starting values for joint ML estimation.

• Different nests can produce very different results. Partition choice set into mutually exclusive subsets within which

(a) unobserved factors are correlated, and

(b) relative odds are independent of other alternatives.

Nested Logit Model – Example 1

Example (from Disdier and Mayer (2004)): Location choices by French firms in Eastern and Western Europe

- We want to model the factors involving the selection of location j: $P_j = \text{P}(\pi_j > \pi_k \ \forall k \neq j\ )$

- Location choices are likely to have a nested structure (non-IIA)
  - First, select region (East or Western Europe)
  - Next, select country within region

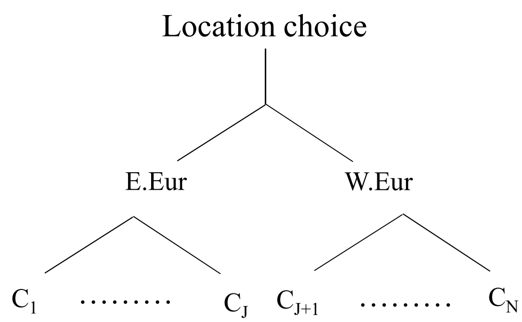
- Data
  - 1843 location decisions in Europe from 1980 to 1999
  - 19 host countries (13 West Europe, 6 East Europe)

Nested Logit Model – Example 1

- Location choices: Data
  - NF French firms already located in the country
  - GDP GDP
  - GPP/CAP GDP per capita
  - DIST Distance France – host country
  - W Average wage per capita (manufacturing)
  - UNEMPL unemployment rate
  - EXCHR Exchange rate volatility
  - FREE Free country
  - PNFREE Partly free and not free country
  - PR1 Country with political rights rated 1
  - PR2 Country with political rights rated 2
  - PR345 Country with political rights rated 3,4,5
  - PR67 Country with political rights rated 6,7
  - LI Annual liberalization index
  - CLI Cumulative liberalization index
  - ASSOC =1 if an association agreement is signed

Nested Logit Model – Example 1

- Location choices by French firms in Eastern and Western Europe



Nested Logit Model – Example 1

Table 1
Independent variables: data sources and expected sign

Variable	Definition	Source	Expected sign
NF	French firms already located in the country	DREE	?
GDP	GDP	CHELEM	+
GDP/CAP	GDP per capita	CHELEM	+
DIST	Weighted sub-national distance between France and the host country	REGIO (for the regional population)	–
W	Average wage per capita in manufacturing	OECD	–
UNEMPL	Unemployment rate	World Bank	?
EXCHR	Exchange rate volatility	IMF	?
FREE	Free country	Freedom House	+
PNFREE	Partly Free and Not Free country	Freedom House	var.
PR1	Country with political rights rated 1	Freedom House	+
PR2	Country with political rights rated 2	Freedom House	+
PR345	Country with political rights rated 3, 4, or 5	Freedom House	+
PR67	Country with political rights rated 6 or 7	Freedom House	var.
LI	Annual liberalization index	de Melo et al. (1997)	+
CLI	Cumulative liberalization index	de Melo et al. (1997)	+
ASSOC	= 1 if an association agreement is signed		+

Nested Logit Model – Example 1

Table 3
Location choice of French firms in Europe: the conditional logit model

Model	(1)	(2)	(3)	(4)	(5)	(6)
Ln *NF*	0.46^{***}	0.45^{***}	0.46^{***}	0.45^{***}	0.49^{***}	0.46^{***}
	(0.04)	(0.04)	(0.07)	(0.07)	(0.05)	(0.05)
Ln *GDP*	0.35^{***}	0.35^{***}	0.35^{***}	0.35^{***}	0.36^{***}	0.38^{***}
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Ln *GDP/CAP*	-0.34^{**}	-0.44^{**}	-0.70^{***}	-0.77^{***}	0.17	−0.08
	(0.15)	(0.18)	(0.22)	(0.24)	(0.26)	(0.29)
Ln *DIST*	-0.88^{***}	-0.89^{***}	-0.84^{***}	-0.83^{***}	-0.74^{***}	-0.78^{***}
	(0.09)	(0.10)	(0.14)	(0.14)	(0.14)	(0.14)
Ln *W*	-0.33^{***}	-0.36^{***}	−0.05	−0.09	-0.71^{***}	-0.62^{***}
	(0.11)	(0.13)	(0.17)	(0.18)	(0.20)	(0.20)
Ln *UNEMPL*	0.37^{***}	0.30^{***}	0.60^{***}	0.56^{***}	−0.01	−0.05
	(0.07)	(0.07)	(0.10)	(0.11)	(0.11)	(0.11)
EXCHR	-2.18^{**}	-2.50^{**}	2.65	−2.03	-2.28^{**}	-2.14^{*}
	(0.98)	(1.08)	(4.12)	(4.67)	(1.10)	(1.11)
FREE	1.83^{***}		2.09^{***}		0.84^{***}	
	(0.24)		(0.44)		(0.28)	
PNFREE	ref. var.		ref. var.		ref. var.	
PR1		4.09^{***}		3.36^{***}		1.16^{***}
		(0.73)		(0.76)		(0.31)
PR2		3.55^{***}		2.81^{***}		0.76^{***}
		(0.73)		(0.75)		(0.28)
PR345		2.61^{***}		1.89^{**}		
		(0.75)		(0.84)		
PR67		var.		var.		
Observations	1843	1843	825	825	1018	1018
Pseudo R^2	0.152	0.156	0.221	0.223	0.108	0.109

Notes. The dependent variable is location choice. Columns (1) and (2) contain the coefficients for the entire sample from 1980 to 1999. Columns (3) and (4) consider the initial time period from 1980 to 1990, while columns (5) and (6) report the results for the later period from 1991 to 1999. Standard errors in parentheses.

* Significant at the 10% level.
** Idem., 5%.
*** Idem., 1%.

Nested Logit Model – Example 1

292 A.-C. Disdier, T. Mayer / Journal of Comparative Economics 32 (2004) 280–296

Table 4
Location choice of French firms in Europe: the nested logit model

Model	(1)	(2)	(3)	(4)	(5)
Ln *NF*	0.68^{***}	0.71^{***}	0.62^{***}	0.70^{***}	0.83^{***}
	(0.06)	(0.06)	(0.10)	(0.13)	(0.10)
Ln *GDP*	0.30^{***}	0.29^{***}	0.42^{***}	0.24^{***}	0.20^{***}
	(0.04)	(0.04)	(0.07)	(0.09)	(0.08)
Ln *GDP/CAP*	−0.10	−0.13	−0.65	1.05	−0.18
	(0.29)	(0.29)	(0.51)	(0.79)	(0.43)
Ln *DIST*	-0.70^{***}	-0.60^{***}	-0.89^{***}	-0.69^{**}	−0.13
	(0.17)	(0.16)	(0.24)	(0.34)	(0.30)
Ln *W*	-0.68^{***}	-0.51^{**}	−0.57	−0.88	−0.33
	(0.25)	(0.24)	(0.40)	(0.67)	(0.36)
Ln *UNEMPL*	−0.06	−0.06	−0.16	0.35	−0.17
	(0.11)	(0.11)	(0.17)	(0.29)	(0.19)
EXCHR	−1.81	-4.79^{**}	-6.00^{***}	−0.81	-13.68^{*}
	(2.08)	(2.00)	(2.29)	(16.66)	(8.22)
FREE	1.03^{***}				
	(0.32)				
PNFREE	ref. var.				
Inclusive value	0.91^{***}	0.77^{***}	0.47^{***}	0.51^{***}	0.92^{***}
	(0.08)	(0.06)	(0.13)	(0.07)	(0.12)
Observations	1008	1008	430	223	355
Pseudo R^2	0.141	0.139	0.151	0.147	0.137

Notes. The dependent variable is location choice. Columns (1) and (2) contain the coefficients for the years of the sample ranging from 1991 to 1999. Column (3) considers the first time period from 1991 to 1993, column (4) reports the coefficients for the period from 1994 to 1995, while column (5) presents the results for the later period from 1996 to 1999. Standard errors in parentheses.

* Significant at the 10% level.
** Idem., 5%.
*** Idem., 1%.

Nested Logit Model – Example 2 (Greene)

Example: Transportation mode (Air, train, bus, caaaaa)

STATA comands: NLOGIT

; Lhs=mode

; Rhs=gc,ttme,invt,invc

; Rh2=one,hinc

; Choices=air,train,bus,car

; Tree=Travel[Private(Air,Car), Public(Train,Bus)]

; Show tree

; Effects: invc(*)

; Describe

; RU1 \$ (RU1: Random utility Model 1 – the one presented). This option selects branch normalization

Nested Logit Model – Example 2 (Greene)

- **Tree Structure Specified for the Nested Logit Model**

Sample proportions are marginal, not conditional.
Choices marked with * are excluded for the IIA test.

Trunk	(prop.)	Limb	(prop.)	Branch	(prop.)	Choice	(prop.)	Weight	IIA
Trunk{1}	1.00000	TRAVEL	1.00000	PRIVATE	.55714	AIR	.27619	1.000	
						CAR	.28095	1.000	
				PUBLIC	.44286	TRAIN	.30000	1.000	
						BUS	.14286	1.000	

Model Specification: Table entry is the attribute that multiplies the indicated parameter.

Choice	******	Parameter				
	Row 1	GC	TTME	INVT	INVC	A_AIR
	Row 2	AIR_HIN1	A_TRAIN	TRA_HIN3	A_BUS	BUS_HIN4
AIR	1	GC	TTME	INVT	INVC	Constant
	2	HINC	none	none	none	none
CAR	1	GC	TTME	INVT	INVC	none
	2	none	none	none	none	none
TRAIN	1	GC	TTME	INVT	INVC	none
	2	none	Constant	HINC	none	none
BUS	1	GC	TTME	INVT	INVC	none
	2	none	none	none	Constant	HINC

Nested Logit Model – Example 2 (Greene)

- **STARTING VALUES**

```
-----------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function      -172.94366
Estimation based on N =    210, K =  10
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only   -283.7588  .3905 .3787
Chi-squared[ 7]          =    221.63022
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=    210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]
--------+--------------------------------------------------
      GC|    .07578***       .01833        4.134    .0000
    TTME|   -.10289***       .01109       -9.280    .0000
    INVT|   -.01399***       .00267       -5.240    .0000
    INVC|   -.08044***       .01995       -4.032    .0001
   A_AIR|   4.37035***      1.05734        4.133    .0000
AIR_HIN1|    .00428          .01306         .327    .7434
 A_TRAIN|   5.91407***       .68993        8.572    .0000
TRA_HIN3|   -.05907***       .01471       -4.016    .0001
   A_BUS|   4.46269***       .72333        6.170    .0000
BUS_HIN4|   -.02295          .01592       -1.442    .1493
--------+--------------------------------------------------
```

Nested Logit Model – Example 2 (Greene)

- FIML Nested Multinomial Logit Model

```
Dependent variable                 MODE
Log likelihood function      -166.64835
The model has 2 levels.
Random Utility Form 1: IVparms = LMDAb|l
Number of obs.=    210, skipped    0 obs
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]
--------+--------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    .06579***       .01878        3.504    .0005
    TTME|   -.07738***       .01217       -6.358    .0000
    INVT|   -.01335***       .00270       -4.948    .0000
    INVC|   -.07046***       .02052       -3.433    .0006
   A_AIR|   2.49364**       1.01084        2.467    .0136
AIR_HIN1|    .00357          .01057         .337    .7358
 A_TRAIN|   3.49867***       .80634        4.339    .0000
TRA_HIN3|   -.03581***       .01379       -2.597    .0094
   A_BUS|   2.30142***       .81284        2.831    .0046
BUS_HIN4|   -.01128          .01459        -.773    .4395
        |IV parameters, lambda(b|l),gamma(l)
 PRIVATE|   2.16095***       .47193        4.579    .0000
  PUBLIC|   1.56295***       .34500        4.530    .0000
        |Underlying standard deviation = pi/(IVparm*sqr(6)
 PRIVATE|    .59351***       .12962        4.579    .0000
  PUBLIC|    .82060***       .18114        4.530    .0000
```

Nested Logit Model – Example 2 (Greene)

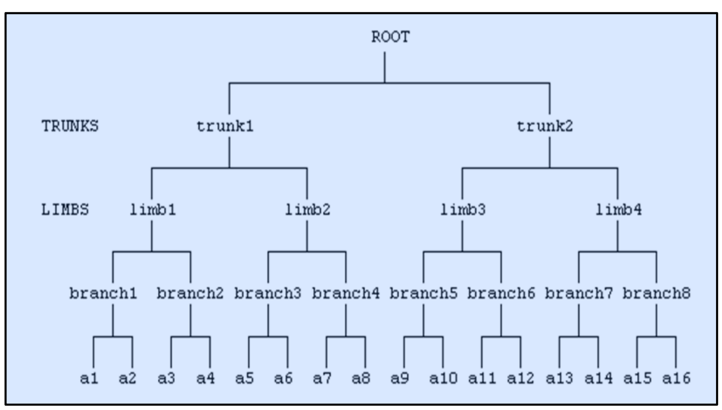
Elasticity averaged over observations.

Attribute is INVC		Decomposition of Effect if Nest				Total Effect	
		Trunk	Limb	Branch	Choice	Mean	St.Dev
in choice AIR							
Branch=PRIVATE							
*	Choice=AIR	.000	.000	-2.456	-3.091	-5.547	3.525
	Choice=CAR	.000	.000	-2.456	2.916	.460	3.178
Branch=PUBLIC							
	Choice=TRAIN	.000	.000	3.846	.000	3.846	4.865
	Choice=BUS	.000	.000	3.846	.000	3.846	4.865
in choice CAR							
Branch=PRIVATE							
	Choice=AIR	.000	.000	-.757	.650	-.107	.589
*	Choice=CAR	.000	.000	-.757	-.830	-1.587	1.292
Branch=PUBLIC							
	Choice=TRAIN	.000	.000	.647	.000	.647	.605
	Choice=BUS	.000	.000	.647	.000	.647	.605
in choice TRAIN							
Branch=PRIVATE							
	Choice=AIR	.000	.000	1.340	.000	1.340	1.475
	Choice=CAR	.000	.000	1.340	.000	1.340	1.475
Branch=PUBLIC							
*	Choice=TRAIN	.000	.000	-1.986	-1.490	-3.475	2.539
	Choice=BUS	.000	.000	-1.986	2.128	.142	1.321
in choice BUS							
Branch=PRIVATE							
	Choice=AIR	.000	.000	.547	.000	.547	.871
	Choice=CAR	.000	.000	.547	.000	.547	.871
Branch=PUBLIC							
	Choice=TRAIN	.000	.000	-.841	.888	.047	.678
*	Choice=BUS	.000	.000	-.841	-1.469	-2.310	1.119

Effects on probabilities of all choices in the model:
* indicates direct Elasticity effect of the attribute.

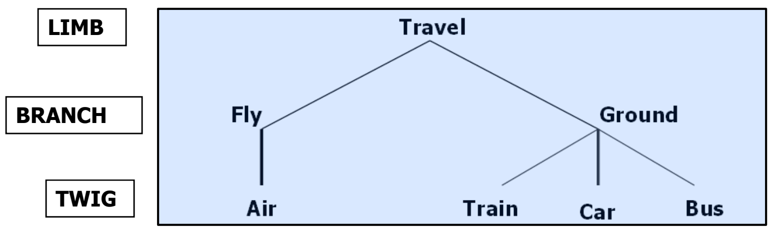
NL Model: Higher Level Trees (Greene)

• We can do higher order nesting. For example, housing choices can be divided by Location (Neighborhood); Housing Type (Rent, Buy, House, Apt); and Housing (# Bedrooms).



NL Model: Degenerate Branches (Greene)

- The branches do not have to have twigs. We can *degenerate* trees.



NL Model: Degenerate Branch (Greene)

- FIML Nested Multinomial Logit Model

```
Dependent variable                 MODE
Log likelihood function      -148.63860
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|    .44230***       .11318        3.908    .0001
    TTME|   -.10199***       .01598       -6.382    .0000
    INVT|   -.07469***       .01666       -4.483    .0000
    INVC|   -.44283***       .11437       -3.872    .0001
   A_AIR|   3.97654***      1.13637        3.499    .0005
AIR_HIN1|    .02163          .01326        1.631    .1028
 A_TRAIN|   6.50129***      1.01147        6.428    .0000
TRA_HIN2|   -.06427***       .01768       -3.635    .0003
   A_BUS|   4.52963***       .99877        4.535    .0000
BUS_HIN3|   -.01596          .02000        -.798    .4248
        |IV parameters, lambda(b|l),gamma(l)
     FLY|    .86489***       .18345        4.715    .0000
  GROUND|    .24364***       .05338        4.564    .0000
        |Underlying standard deviation = pi/(IVparm*sqr(6)
     FLY|   1.48291***       .31454        4.715    .0000
  GROUND|   5.26413***      1.15331        4.564    .0000
--------+--------------------------------------------------
```

NL Model: Degenerate Branch (Greene)

• STATA commands:

```
NLOGIT    ; lhs=mode
          ; rhs=gc,ttme,invt,invc
          ; rh2=one,hinc
          ; choices=air,train,bus,car
          ; tree=Travel[Fly(Air),
                      Ground(Train,Car,Bus)]
          ; show tree
          ; effects:gc(*)
          ; Describe
          ; ru2  $
```

(This is RANDOM UTILITY FORM 2. The different normalization shows the effect of the degenerate branch.)

NL Model: Degenerate Branch (Greene)

• Estimation of RU2 Form of Nested Logit Model

```
FIML Nested Multinomial Logit Model
Dependent variable                 MODE
Log likelihood function      -168.81283  (-148.63860 with RU1)
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
        |Attributes in the Utility Functions (beta)
      GC|     .06527***       .01787        3.652   .0003
    TTME|    -.06114***       .01119       -5.466   .0000
    INVT|    -.01231***       .00283       -4.354   .0000
    INVC|    -.07018***       .01951       -3.597   .0003
   A_AIR|    1.22545          .87245        1.405   .1601
AIR_HIN1|     .01501          .01226        1.225   .2206
 A_TRAIN|    3.44408***       .68388        5.036   .0000
TRA_HIN2|    -.02823***       .00852       -3.311   .0009
   A_BUS|    2.58400***       .63247        4.086   .0000
BUS_HIN3|    -.00726          .01075        -.676   .4993
        |IV parameters, RU2 form = mu(b|l),gamma(l)
     FLY|    1.00000         ......(Fixed Parameter)......
  GROUND|     .47778***       .10508        4.547   .0000
        |Underlying standard deviation = pi/(IVparm*sqr(6)
     FLY|    1.28255         ......(Fixed Parameter)......
  GROUND|    2.68438***       .59041        4.547   .0000
```

NL Model: An Error Components Model (Greene)

• We can allow for some heterogeneity in the utility within the branches.

Random terms in utility functions share random components

$$
\begin{aligned}
U(Air,i) &= \alpha_{AIR} + \beta_1 INVC_{i,AIR} + \ldots + \varepsilon_{i,AIR} + \boxed{w_{i,1}} \\
U(Train,i) &= \alpha_{TRAIN} + \beta_1 INVC_{i,TRAIN} + \ldots + \varepsilon_{i,TRAIN} + \boxed{w_{i,1}} \\
U(Bus,i) &= \alpha_{BUS} + \beta_1 INVC_{i,BUS} + \ldots + \varepsilon_{i,BUS} + \boxed{w_{i,2}} \\
U(Car,i) &= \beta_1 INVC_{i,CAR} + \ldots + \varepsilon_{i,CAR} + \boxed{w_{i,2}}
\end{aligned}
$$

$$
Cov\begin{pmatrix} Air \\ Train \\ Bus \\ Car \end{pmatrix} = \begin{bmatrix} \sigma_\varepsilon^2+\theta_1^2 & \theta_1^2 & 0 & 0 \\ \theta_1^2 & \sigma_\varepsilon^2+\theta_1^2 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2+\theta_2^2 & \theta_2^2 \\ 0 & 0 & \theta_2^2 & \sigma_\varepsilon^2+\theta_2^2 \end{bmatrix}
$$

This model is estimated by maximum simulated likelihood.

NL Model: An Error Components Model (Greene)

```
-----------------------------------------------------------
Error Components (Random Effects) model
Dependent variable                         MODE
Log likelihood function          -182.27368
Response data are given as ind. choices
Replications for simulated probs. =   25
Halton sequences used for simulations
ECM model with panel has         70 groups  <-----------
Fixed number of obsrvs./group=           3
Hessian is not PD. Using BHHH estimator
Number of obs.=    210, skipped     0 obs
--------+--------------------------------------------------
Variable| Coefficient     Standard Error  b/St.Er.  P[|Z|>z]
--------+--------------------------------------------------
        |Nonrandom parameters in utility functions
      GC|      .07293***        .01978        3.687    .0002
    TTME|     -.10597***        .01116       -9.499    .0000
    INVT|     -.01402***        .00293       -4.787    .0000
    INVC|     -.08825***        .02206       -4.000    .0001
   A_AIR|     5.31987***        .90145        5.901    .0000
 A_TRAIN|     4.46048***        .59820        7.457    .0000
   A_BUS|     3.86918***        .67674        5.717    .0000
        |Standard deviations of latent random effects
SigmaE01|     -.27336          3.25167        -.084    .9330
SigmaE02|     1.21988           .94292        1.294    .1958
--------+--------------------------------------------------
```

Testing the NL Model vs. the MNL (Greene)

• Log likelihood for the NL model

• Constrain IV parameters to equal 1 with
  ; IVSET(list of branches)=[1]

• Use LR test

• For the example:

- LogL = -166.68435
- LogL (MNL) = -172.94366

Chi-squared with 2 d.f. = 2 * (-166.68435 - (-172.94366))
= 12.51862

The critical value is 5.99 (95%) $\Rightarrow$ MNL model is rejected

• Check IV coefficients: A *sufficient condition* for consistency with RUM: they should be between (0, 1).