Nested logit models

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Red bus/Blue bus paradox

- Mode choice example
- Two alternatives: car and bus
- There are red buses and blue buses
- Car and bus travel times are equal: T

Red bus/Blue bus paradox

Model 1

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{bus}} &= \beta T + \varepsilon_{\text{bus}} \end{aligned}$$

Therefore,

$$P(\text{car}|\{\text{car}, \text{bus}\}) = P(\text{bus}|\{\text{car}, \text{bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T}} = \frac{1}{2}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Red bus/Blue bus paradox

Model 2

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{blue bus}} &= \beta T + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= \beta T + \varepsilon_{\text{red bus}} \end{aligned}$$

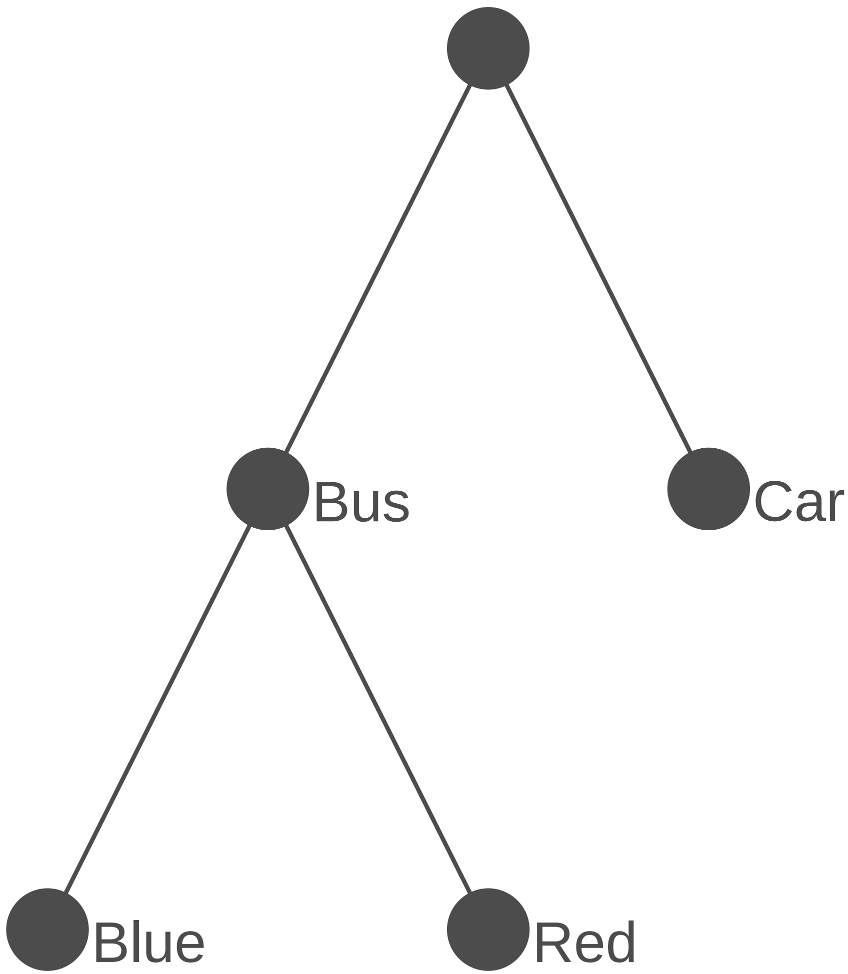
$$P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T} + e^{\beta T}} = \frac{1}{3}$$

$$\left.\begin{array}{r} P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{blue bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{red bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \end{array}\right\} = \frac{1}{3}.$$

Red bus/Blue bus paradox

- Assumption of logit: ε i.i.d
- $\varepsilon_{\text{blue bus}}$ and $\varepsilon_{\text{red bus}}$ contain common unobserved attributes:
  - fare
  - headway
  - comfort
  - convenience
  - etc.

Capturing the correlation



Capturing the correlation

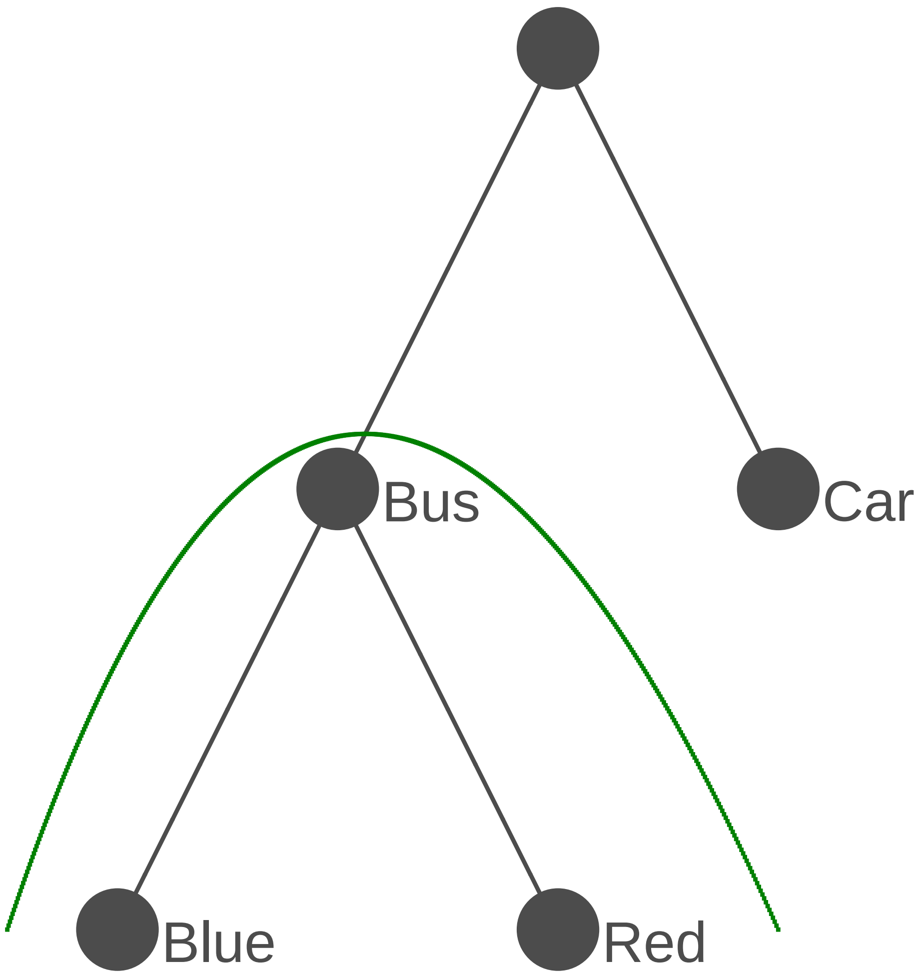
If bus is chosen then

$$\begin{aligned} U_{\text{blue bus}} &= V_{\text{blue bus}} + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= V_{\text{red bus}} + \varepsilon_{\text{red bus}} \end{aligned}$$

where $V_{\text{blue bus}} = V_{\text{red bus}} = \beta T$

$$P(\text{blue bus}|\{\text{blue bus}, \text{red bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T}} = \frac{1}{2}$$

Capturing the correlation



Capturing the correlation

What about the choice between bus and car?

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{bus}} &= V_{\text{bus}} + \varepsilon_{\text{bus}} \end{aligned}$$

with

$$\begin{aligned} V_{\text{bus}} &= V_{\text{bus}}(V_{\text{blue bus}}, V_{\text{red bus}}) \\ \varepsilon_{\text{bus}} &= ? \end{aligned}$$

Define V_{bus} as the expected maximum utility of red bus and blue bus

Expected maximum utility

For a set of alternative $\mathcal{C}$, define

$$U_{\mathcal{C}} = \max_{i \in \mathcal{C}} U_i = \max_{i \in \mathcal{C}} (V_i + \varepsilon_i)$$

and

$$V_{\mathcal{C}} = E[U_{\mathcal{C}}]$$

For logit

$$E[\max_{i \in \mathcal{C}} U_i] = \frac{1}{\mu} \ln \sum_{i \in \mathcal{C}} e^{\mu V_i} + \frac{\gamma}{\mu}$$

Expected maximum utility

$$\begin{aligned} V_{\text{bus}} &= \frac{1}{\mu_b} \ln(e^{\mu_b V_{\text{blue bus}}} + e^{\mu_b V_{\text{red bus}}}) \\ &= \frac{1}{\mu_b} \ln(e^{\mu_b \beta T} + e^{\mu_b \beta T}) \\ &= \beta T + \frac{1}{\mu_b} \ln 2 \end{aligned}$$

where μ_b is the scale parameter for the logit model associated with the choice between red bus and blue bus

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

Probability model:

$$P(\text{car}) = \frac{e^{\mu V_{\text{car}}}}{e^{\mu V_{\text{car}}} + e^{\mu V_{\text{bus}}}} = \frac{e^{\mu \beta T}}{e^{\mu \beta T} + e^{\mu \beta T + \frac{\mu}{\mu_b} \ln 2}} = \frac{1}{1 + 2^{\frac{\mu}{\mu_b}}}$$

If $\mu = \mu_b$, then P(car) = $\frac{1}{3}$ (Model 2)

If $\mu_b \to \infty$, then $\frac{\mu}{\mu_b} \to 0$, and P(car) $\to \frac{1}{2}$ (Model 1)

Note for $\mu_b \to \infty$

$$e^{\mu V_{\text{bus}}} = \frac{1}{2} e^{\mu V_{\text{red bus}}} + \frac{1}{2} e^{\mu V_{\text{blue bus}}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

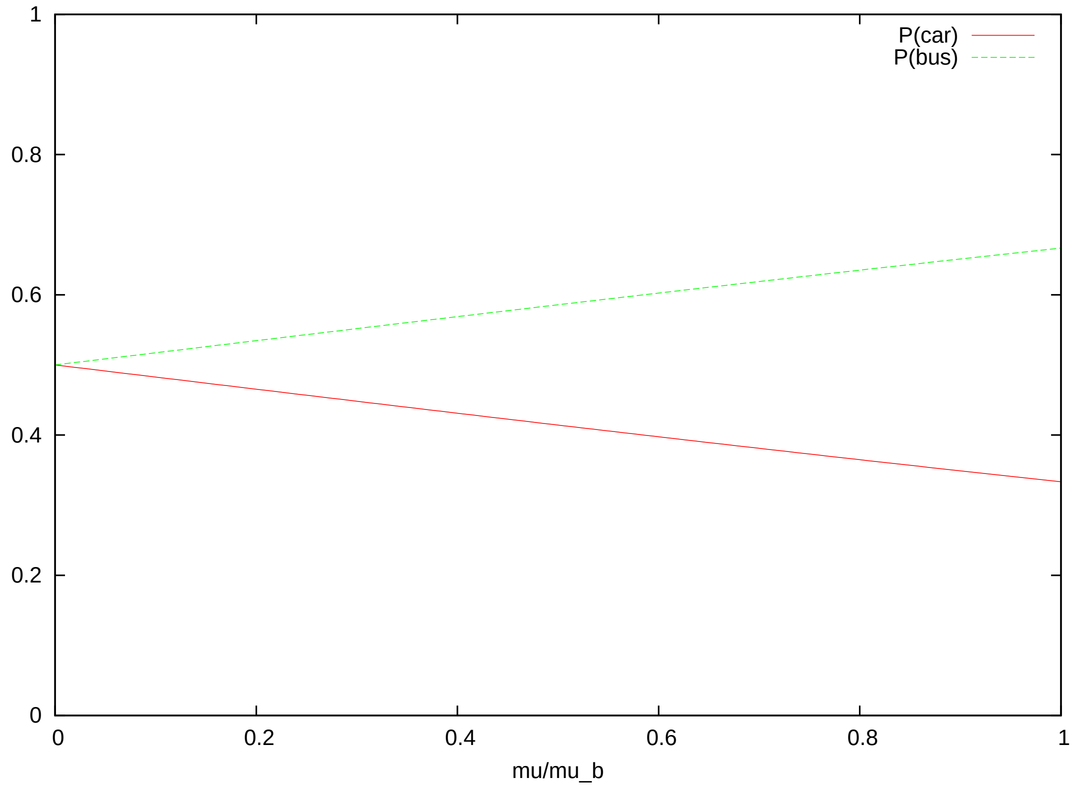
Probability model:

$$P(\text{bus}) = \frac{e^{\mu V_{\text{bus}}}}{e^{\mu V_{\text{car}}} + e^{\mu V_{\text{bus}}}} = \frac{e^{\mu\beta T + \frac{\mu}{\mu_b}\ln 2}}{e^{\mu\beta T} + e^{\mu\beta T + \frac{\mu}{\mu_b}\ln 2}} = \frac{1}{1 + 2^{-\frac{\mu}{\mu_b}}}$$

If $\mu = \mu_b$, then P(bus) = $\frac{2}{3}$ (Model 2)

If $\frac{\mu}{\mu_b} \to 0$, then P(bus) $\to \frac{1}{2}$ (Model 1)

Nested Logit Model



$$\frac{\mu}{\mu_b}$$

Solving the paradox

If $\frac{\mu}{\mu_b} \to 0$, we have

$$
\begin{aligned}
P(\text{car}) &= & & 1/2 \\
P(\text{bus}) &= & & 1/2 \\
P(\text{red bus}|\text{bus}) &= & & 1/2 \\
P(\text{blue bus}|\text{bus}) &= & & 1/2 \\
P(\text{red bus}) &= P(\text{red bus}|\text{bus})P(\text{bus}) &=& 1/4 \\
P(\text{blue bus}) &= P(\text{blue bus}|\text{bus})P(\text{bus}) &=& 1/4
\end{aligned}
$$

Comments

- A group of similar alternatives is called a nest
- Each alternative belongs to exactly one nest
- The model is named Nested Logit
- The ratio μ/μ_b must be estimated from the data
- $0 < \mu/\mu_b \leq 1$ (between models 1 and 2)

A case study

- Choice of a residential telephone service
- Household survey conducted in Pennsylvania, USA, 1984
- Revealed preferences
- 434 observations

A case study

Availability of telephone service by residential area:

	Metro area	Adjacent to metro area	Other non-metro areas
Budget Measured	yes	yes	yes
Standard Measured	yes	yes	yes
Local Flat	yes	yes	yes
Extended Area Flat	no	yes	no
Metro Area Flat	yes	yes	no

Logit Model

$$\mathcal{C} = \{\text{BM}, \text{SM}, \text{LF}, \text{EF}, \text{MF}\}$$

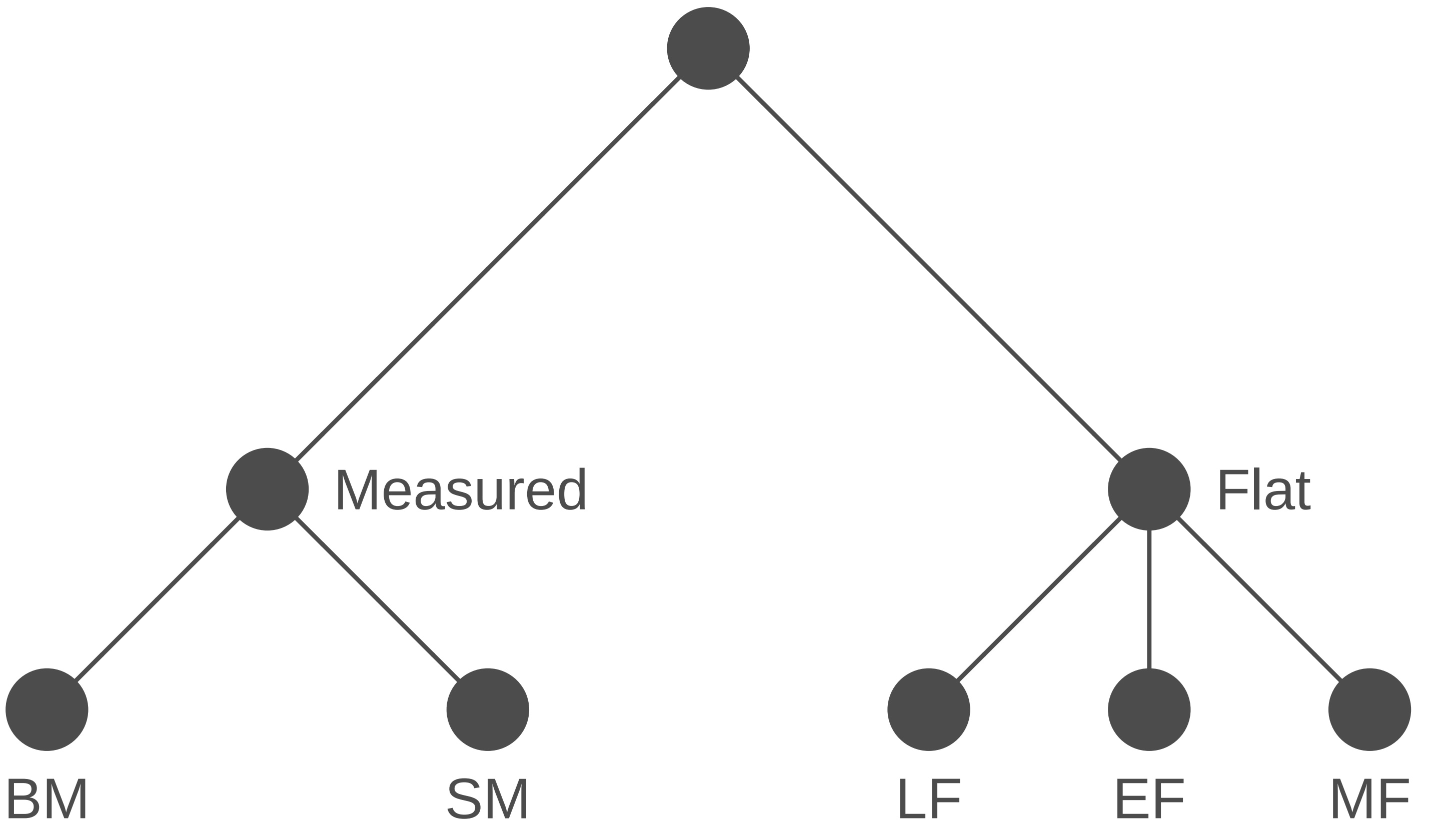
$$\begin{aligned}
V_{\text{BM}} &= \beta_{\text{BM}} + \beta_c \ln(\text{cost}_{\text{BM}}) \\
V_{\text{SM}} &= \beta_{\text{SM}} + \beta_c \ln(\text{cost}_{\text{SM}}) \\
V_{\text{LF}} &= \beta_{\text{LF}} + \beta_c \ln(\text{cost}_{\text{LF}}) \\
V_{\text{EF}} &= \beta_{\text{EF}} + \beta_c \ln(\text{cost}_{\text{EF}}) \\
V_{\text{MF}} &= \beta_c \ln(\text{cost}_{\text{MF}})
\end{aligned}$$

$$P(i|\mathcal{C}) = \frac{e^{V_i}}{\sum_{j \in \mathcal{C}} e^{V_j}}$$

Logit Model

Parameter	Logit	
	Value	(t-stat)
β_{BM}	-2.46	(-7.84)
β_{SM}	-1.74	(-6.28)
β_{LF}	-0.54	(-2.57)
β_{EF}	-0.74	(-1.02)
β_c	-2.03	(-9.47)
$\mathcal{L}_0$	-560.25	
$\mathcal{L}$	-477.56	
# Obs	434	

Nested Logit Model



Nested Logit Model

Model of the choice among “measured” alternatives

$$P(i|M) = \frac{e^{V_i}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}}} \quad i = \text{BM}, \text{SM}$$

We estimate the model with the 196 observations choosing either BM or SM, and calculate the inclusive value

$$I_M = \ln(e^{V_{\text{BM}}} + e^{V_{\text{SM}}})$$

for all observations (scale normalized to 1)

Nested Logit Model

Parameter	Logit		Measured	
	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)		
β_{SM}	-1.74	(-6.28)	0.76	(4.53)
β_{LF}	-0.54	(-2.57)		
β_{EF}	-0.74	(-1.02)		
β_c	-2.03	(-9.47)	-3.12	(-4.76)
$\mathcal{L}_0$	-560.3		-135.9	
$\mathcal{L}$	-477.6		-116.8	
# Obs	434		196	

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

Model of the choice among “flat” alternatives

$$P(i|M) = \frac{e^{V_i}}{e^{V_{\text{LF}}} + e^{V_{\text{EF}}} + e^{V_{\text{MF}}}} \quad i = \text{LF}, \text{EF}, \text{MF}$$

We estimate the model with the 238 observations choosing LF, EF or MF and calculate the inclusive value

$$I_F = \ln(e^{V_{\text{LF}}} + e^{V_{\text{EF}}} + e^{V_{\text{MF}}})$$

for all observations (scale normalized to 1)

Nested Logit Model

Parameter	Logit		Measured		Flat	
	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)				
β_{SM}	-1.74	(-6.28)	0.76	(4.53)		
β_{LF}	-0.54	(-2.57)			-1.21	(-3.17)
β_{EF}	-0.74	(-1.02)			-1.42	(-1.55)
β_c	-2.03	(-9.47)	-3.12	(-4.76)	-3.73	(-6.22)
$\mathcal{L}_0$	-560.3		-135.9		-129.5	
$\mathcal{L}$	-477.6		-116.8		-79.4	
# Obs	434		196		238	

Nested Logit Model

Model of the choice of type of service

$$P(M) = \frac{e^{\mu(\tilde{\beta}_M + I_M)}}{e^{\mu(\tilde{\beta}_M + I_M)} + e^{\mu I_F}} = \frac{e^{\beta_M + \mu I_M}}{e^{\beta_M + \mu I_M} + e^{\mu I_F}}$$

$$P(F) = \frac{e^{\mu I_F}}{e^{\mu(\tilde{\beta}_M + I_M)} + e^{\mu I_F}} = \frac{e^{\mu I_F}}{e^{\beta_M + \mu I_M} + e^{\mu I_F}}$$

- I_M and I_F are attributes of *measured* and *flat*, resp.
- $\beta_M = \mu\tilde{\beta}_M$ and μ are unknown parameters, to be estimated.
- $0 < \mu \leq 1$

Nested Logit Model

Parameter	Logit		Measured		Flat		Nested Logit	
	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)						
β_{SM}	-1.74	(-6.28)	0.76	(4.53)				
β_{LF}	-0.54	(-2.57)			-1.21	(-3.17)		
β_{EF}	-0.74	(-1.02)			-1.42	(-1.55)		
β_c	-2.03	(-9.47)	-3.12	(-4.76)	-3.73	(-6.22)		
β_M							-2.32	(-5.67)
μ							0.43	(5.49)
$\mathcal{L}_0$	-560.3		-135.9		-129.5		-300.8	
$\mathcal{L}$	-477.6		-116.8		-79.4		-280.4	
# Obs	434		196		238		434	

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

How to interpret the log-likelihood?

Assume that individual n has chosen alt. i in nest M.

$$P_n(i) = P_n(i|M)P_n(M)$$

Consider now all individuals choosing an alt. i in nest M

$$\sum_n \ln P_n(i) = \sum_n \ln P_n(i|M) + \sum_n \ln P_n(M) = \mathcal{L}_{\mathsf{M}} + \sum_n \ln P_n(M)$$

For individuals choosing an alternative j in nest F, we have

$$\sum_n \ln P_n(j) = \sum_n \ln P_n(i|F) + \sum_n \ln P_n(F) = \mathcal{L}_{\mathsf{F}} + \sum_n \ln P_n(F)$$

Nested Logit Model

Therefore, we obtain that

$$\mathcal{L} = \mathcal{L}_{\mathsf{M}} + \mathcal{L}_{\mathsf{F}} + \mathcal{L}_{\mathsf{NL}}$$

Nested Logit Model

Parameter	Logit		Measured		Flat		Nested Logit	
	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)						
β_{SM}	-1.74	(-6.28)	0.76	(4.53)				
β_{LF}	-0.54	(-2.57)			-1.21	(-3.17)		
β_{EF}	-0.74	(-1.02)			-1.42	(-1.55)		
β_c	-2.03	(-9.47)	-3.12	(-4.76)	-3.73	(-6.22)		
β_M							-2.32	(-5.67)
μ							0.43	(5.49)
$\mathcal{L}_0$	-560.3		-135.9		-129.5		-300.8	[-566.2]
$\mathcal{L}$	-477.6		-116.8		-79.4		-280.4	[-476.6]
# Obs	434		196		238		434	

Nested Logit Model

Which value of β_c should we use?

Measured: -3.12 (-4.76) or Flat: -3.73 (-6.22)

Equal β_c's:

- Jointly estimate *measured* and *flat* models and constrain β_C to be equal
- Declare “Measured” alternatives unavailable when a “Flat” alternative is chosen, and vice versa.

Nested Logit Model

Parameter	Logit		Nested Logit	
	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)		
β_{SM}	-1.74	(-6.28)	0.79	(4.80)
β_{LF}	-0.54	(-2.57)	-1.07	(-3.49)
β_{EF}	-0.74	(-1.02)	-1.28	(-1.46)
β_c	-2.03	(-9.47)	-3.47	(-8.01)
β_M			-1.66	(-5.92)
μ			0.42	(5.85)
$\mathcal{L}_0$	-560.3		-566.2	
$\mathcal{L}$	-477.6		-473.6	
# Obs	434		434	

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

Logit:

$$P(\text{BM}) = \frac{e^{V_{\text{BM}}}}{\sum_{j \in \mathcal{C}} e^{V}_{j}}$$

Nested Logit:

$$
\begin{aligned}
P(\text{BM}) &= P(\text{BM}|M)P(M) \\
&= \frac{e^{V_{\text{BM}}}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}}} \frac{e^{\beta_M + \mu I_M}}{e^{\beta_M + \mu I_M} + e^{\mu I_F}} \\
&= \frac{e^{V_{\text{BM}}}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}}} \frac{e^{\beta_M + \mu \ln(e^{V_{\text{BM}}} + e^{V_{\text{SM}}})}}{e^{\beta_M + \mu \ln(e^{V_{\text{BM}}} + e^{V_{\text{SM}}})} + e^{\mu \ln(e^{V_{\text{LF}}} + e^{V_{\text{EF}}} + e^{V_{\text{MF}}})}}
\end{aligned}
$$

Nested Logit Model

Let $\mu = 1$

$$P(\text{BM})$$

$$= \frac{e^{V_{\text{BM}}}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}}} \frac{e^{\beta_M + \ln(e^{V_{\text{BM}}} + e^{V_{\text{SM}}})}}{e^{\beta_M + \ln(e^{V_{\text{BM}}} + e^{V_{\text{SM}}})} + e^{\ln(e^{V_{\text{LF}}} + e^{V_{\text{EF}}} + e^{V_{\text{MF}}})}}$$

$$= \frac{e^{V_{\text{BM}}}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}}} \frac{e^{\beta_M}\left(e^{V_{\text{BM}}} + e^{V_{\text{SM}}}\right)}{e^{\beta_M}\left(e^{V_{\text{BM}}} + e^{V_{\text{SM}}}\right) + e^{V_{\text{LF}}} + e^{V_{\text{EF}}} + e^{V_{\text{MF}}}}$$

$$= \frac{e^{V_{\text{BM}}}}{e^{V_{\text{BM}}} + e^{V_{\text{SM}}} + e^{V_{\text{LF}} - \beta_M} + e^{V_{\text{EF}} - \beta_M} + e^{V_{\text{MF}} - \beta_M}}$$

Nested Logit Model

In general, if $\mathcal{C} = \bigcup_{m=1,\ldots M} \mathcal{C}_m$,

$$P(i|\mathcal{C}_m) = \frac{e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_i}} \text{ and } P(\mathcal{C}_m|\mathcal{C}) = \frac{e^{\mu V'_m}}{\sum_{k=1,\ldots,m} e^{\mu V'_k}}$$

where

$$V'_m = \frac{1}{\mu_m} \ln \sum_{i \in \mathcal{C}_m} (e^{\mu_m V_i})$$

When $\frac{\mu}{\mu_m} = 1$, for all m, NL becomes logit

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Simultaneous estimation

$$P(i|\mathcal{C}) = P(i|\mathcal{C}_m)P(\mathcal{C}_m|\mathcal{C})$$

Note that each i belongs to exactly one nest m i.e. the $\mathcal{C}_m$'s do not overlap

The log-likelihood for observation n is

$$\ln P(i_n|\mathcal{C}_n) = \ln P(i_n|\mathcal{C}_{mn}) + \ln P(\mathcal{C}_{mn}|\mathcal{C}_n)$$

where i_n is the chosen alternative.

Simultaneous estimation

Sequential estimation:

- Estimation of NL decomposed into two estimations of logit
- Estimator is consistent but not efficient

Simultaneous estimation:

- Log-likelihood function is generally non concave
- No guarantee of global maximum
- Estimator asymptotically efficient

Simultaneous estimation

Parameter	Logit		Seq. Nested Logit		Sim. Nested Logit	
	Value	(t-stat)	Value	(t-stat)	Value	(t-stat)
β_{BM}	-2.46	(-7.84)			-3.79	(-6.28)
β_{SM}	-1.74	(-6.28)	0.79	(4.80)	-3.00	(-5.32)
β_{LF}	-0.54	(-2.57)	-1.07	(-3.49)	-1.09	(-3.57)
β_{EF}	-0.74	(-1.02)	-1.28	(-1.46)	-1.19	(-1.41)
β_c	-2.03	(-9.47)	-3.47	(-8.01)	-3.25	(-6.99)
β_M			-1.66	(-5.92)		
μ			0.42	(5.85)	0.46	(4.17)
$\mathcal{L}_0$	-560.3		-566.2		-560.3	
$\mathcal{L}$	-477.6		-473.6		-473.3	
# Obs	434		434		434	

Compare $\beta_M = -1.66$ and $\mu\beta_{\text{BM}} = -1.74$

Compare $\beta_{\text{SM}} - \beta_{\text{BM}} = 0.79$ for Seq. and Sim.