Nested logit

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

Outline

Simple choice model

Mode choice

- Two alternatives: car and bus.
- There are red buses and blue buses.
- Car and bus travel times are equal: T.
- Only travel time is considered in the utility function.

Red bus/Blue bus paradox

Model 1

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

Choice probability

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

Red bus/Blue bus paradox

Model 2

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

Choice probability

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

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

Red bus/Blue bus paradox

Conclusion

If you paint the buses of a city red and blue, the mode share for public transportation increases from 50% to 66%.

Explaining the paradox

Model specification

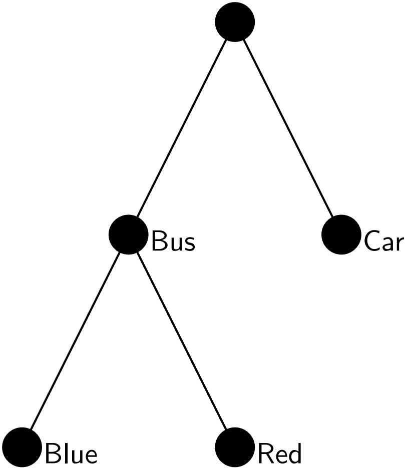
- Only travel time appears in the utility function.
- Other attributes are captured by the error term.
- Some of them are shared by $\varepsilon_{\text{blue bus}}$ and $\varepsilon_{\text{red bus}}$
  - fare
  - headway
  - comfort
  - convenience
  - etc.

Logit model

- Assumes that $\varepsilon_{\text{blue bus}}$ and $\varepsilon_{\text{red bus}}$ are independent.
- Inappropriate assumption in this case.

Outline

Capturing the correlation



Capturing the correlation

If bus is chosen then

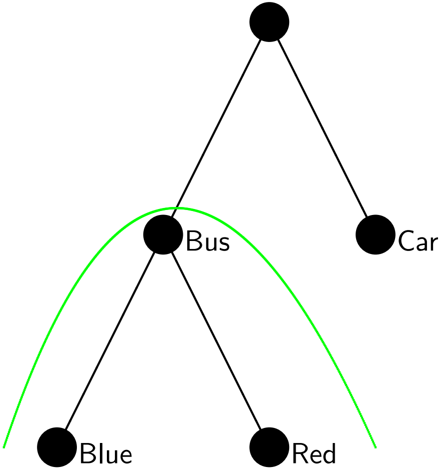
$$\begin{aligned} U_{\text{blue bus}} &= V_{\text{blue bus}} + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= V_{\text{red bus}} + \varepsilon_{\text{red bus}} \end{aligned}$$

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

Choice probability

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

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

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

Idea

- Use a logit model at the higher level.
- Define V_{bus} as the expected maximum utility of red bus and blue bus

Expected maximum utility

Definition

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

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

and

$$U_{\mathcal{C}} \equiv V_{\mathcal{C}} + \varepsilon_{\mathcal{C}}$$

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

For logit

$$U_{\mathcal{C}} = V_{\mathcal{C}} + \varepsilon_{\mathcal{C}}$$

$$V_{\mathcal{C}} = \frac{1}{\mu} \ln \sum_{i \in \mathcal{C}} e^{\mu V_i}$$

$$E[\varepsilon_{\mathcal{C}}] = \frac{\gamma}{\mu}$$

Expected maximum utility

Back to the blue/red buses

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

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

Outline

Nested Logit Model

Probability model: car

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

Extreme cases

- If $\mu = \mu_b$, then P(car) $= \frac{1}{3}$ (Model 2, logit with 3 alternatives)
- If $\mu_b \to \infty$, then $\frac{\mu}{\mu_b} \to 0$, and P(car) $\to \frac{1}{2}$ (Model 1, logit with 2 alternatives)

Nested Logit Model

Probability model: bus

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

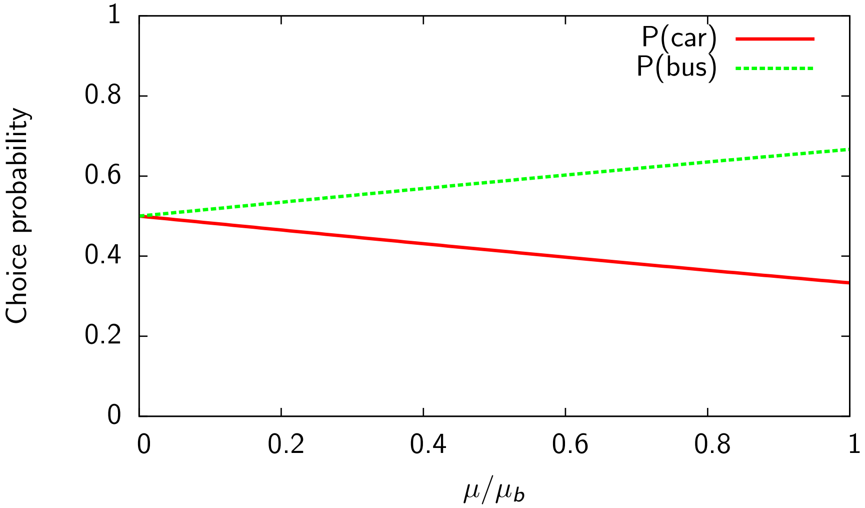
Extreme cases

- If $\mu = \mu_b$, then P(bus) $= \frac{2}{3}$ (Model 2)
- If $\mu_b \to \infty$, then $\frac{\mu}{\mu_b} \to 0$, then P(bus) $\to \frac{1}{2}$ (Model 1)

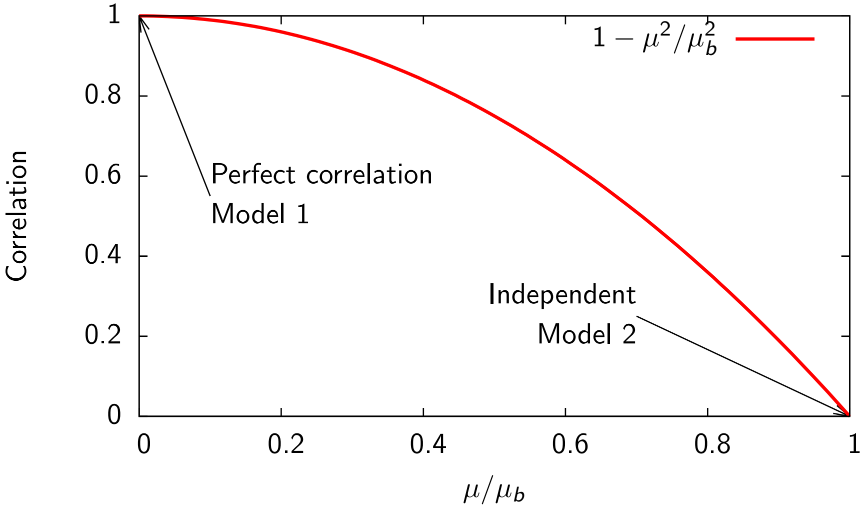
Utility of bus when $\mu_b \to \infty$

$$\lim_{\mu_b \to \infty} \beta T + \frac{1}{\mu_b} \ln 2 = \beta T$$

Nested Logit Model



Nested Logit Model



Solving the paradox

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

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

Nested logit model

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)
- Going down the tree, μ's must increase, variance must decrease

Outline

Airline itinerary case study: Logit model

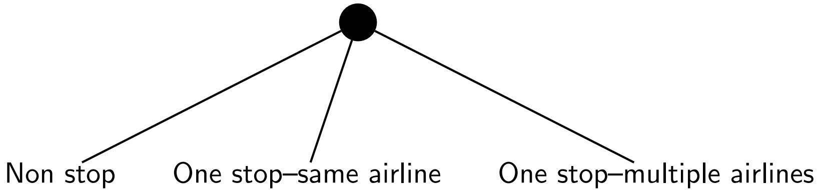
Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	One stop–same airline dummy	-0.898	0.218	-4.13	0.00
2	One stop–multiple airlines dummy	-1.24	0.223	-5.58	0.00
3	Round trip fare ($100)	-1.81	0.152	-11.90	0.00
4	Elapsed time (0 - 2 hours)	-0.856	0.226	-3.79	0.00
5	Elapsed time (2 - 8 hours)	-0.241	0.0820	-2.93	0.00
6	Elapsed time ($>$ 8 hours)	-0.936	0.314	-2.99	0.00
7	Leg room (inches), if male (non stop)	0.0972	0.0330	2.94	0.00
8	Leg room (inches), if female (non stop)	0.193	0.0315	6.15	0.00
9	Leg room (inches), if male (one stop)	0.128	0.0290	4.42	0.00
10	Leg room (inches), if female (one stop)	0.0845	0.0259	3.26	0.00
11	Being early (hours)	-0.150	0.0190	-7.89	0.00
12	Being late (hours)	-0.0993	0.0167	-5.94	0.00
13	More than 2 air trips per year (one stop–same airline)	-0.279	0.141	-1.98	0.05
14	More than 2 air trips per year (one stop–multiple airlines)	-0.0670	0.157	-0.43	0.67
15	Round trip fare / income ($100/$1000)	-23.0	8.11	-2.83	0.00

Summary statistics

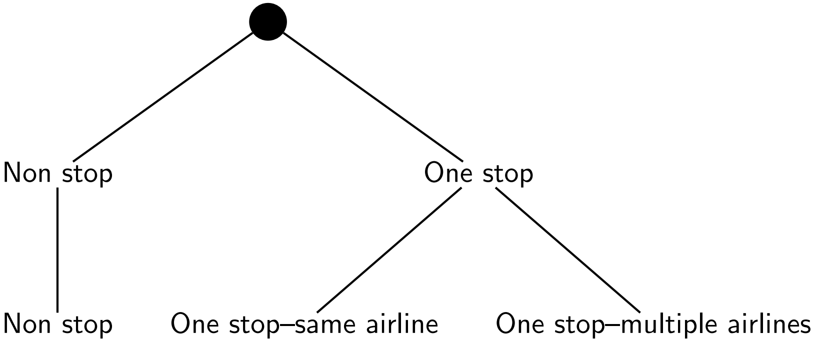
Number of observations = 2544

$$\mathcal{L}(0) = -2794.870$$
$$\mathcal{L}(c) = -2203.160$$
$$\mathcal{L}(\widehat{\beta}) = -1635.068$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\widehat{\beta})] = 2319.603$$
$$\rho^2 = 0.415$$
$$\bar{\rho}^2 = 0.410$$

Logit model



Nested logit model



Nested logit

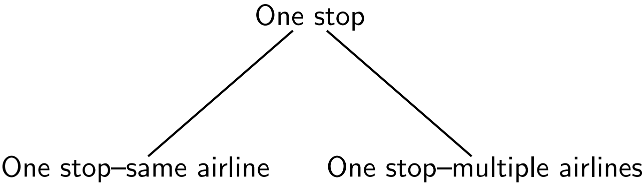
Marginal and conditional probabilities

$$
\begin{aligned}
\Pr(\text{NS}) &= \Pr(\text{NS}|\text{Non stop}) \Pr(\text{Non stop}|\{\text{Non stop}, \text{One stop}\}), \\
\Pr(\text{SAME}) &= \Pr(\text{SAME}|\text{One stop}) \Pr(\text{One stop}|\{\text{Non stop}, \text{One stop}\}), \\
\Pr(\text{MULT}) &= \Pr(\text{MULT}|\text{One stop}) \Pr(\text{One stop}|\{\text{Non stop}, \text{One stop}\}).
\end{aligned}
$$

Note

$$\Pr(\text{NS}|\text{Non stop}) = 1$$

Nest "one stop"



Nest "one stop"

Binary choice

- SAME: "One stop–same airline"
- MULT: "One stop–multiple airlines"

Specification of the utility functions

- Same as logit model.
- Up to normalization.
  - MULT constant normalized to zero
  - "More than two air trips per year (MULT)" normalized to 0
  - "Elapsed time (0–2 hours)" cannot be identified due to absence of data.

Nest "one stop"

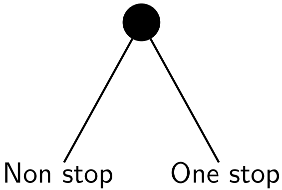
Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	One stop, same airline dummy	0.469	0.188	2.50	0.01
3	Round trip fare ($100)	-2.87	0.624	-4.61	0.00
5	Elapsed time (2–8 hours)	-0.387	0.136	-2.84	0.00
6	Elapsed time ($>$ 8 hours)	-2.33	0.759	-3.07	0.00
9	Leg room (inches), if male (one stop)	0.170	0.0464	3.65	0.00
10	Leg room (inches), if female (one stop)	0.104	0.0421	2.46	0.01
11	Being early (hours)	-0.250	0.0422	-5.91	0.00
12	Being late (hours)	-0.0942	0.0286	-3.29	0.00
13	More than two air trips per year (one stop, same airline)	-0.220	0.218	-1.01	0.31
15	Round trip fare / income ($100/$1000)	-37.8	40.8	-0.93	0.35

Summary statistics

Number of observations = 846

$$
\begin{aligned}
\mathcal{L}(0) &= -586.403 \\
\mathcal{L}(c) &= -585.258 \\
\mathcal{L}(\hat{\beta}) &= -318.994 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 534.816 \\
\rho^2 &= 0.456 \\
\bar{\rho}^2 &= 0.439
\end{aligned}
$$

Upper level



Upper level

Binary choice

- Non stop
- One stop

Non stop: specification of the utility functions

- Same as logit.
- Up to normalization.
  - Alternative specific constant
  - "More than 2 air trips per year"
  - Coefficients 0 and 12' are replacing 2 and 14 in logit.
- Coefficients already estimated at the lower level are not re-estimated.
- 5 coefficients must be estimated.

Upper level

One stop: specification

$$\tilde{V}_{\text{One stop}} = E[\max(U_{\text{SAME}}, U_{\text{MULT}})]$$

$$= \frac{1}{\mu_{\text{One stop}}} \log(e^{\mu_{\text{One stop}} V_{\text{SAME}}} + e^{\mu_{\text{One stop}} V_{\text{MULT}}}).$$

Normalization

$$\mu_{\text{One stop}} = 1$$

$$\tilde{V}_{\text{One stop}} = \log(e^{V_{\text{SAME}}} + e^{V_{\text{MULT}}}).$$

Upper level

Logit

$$\Pr(\text{One stop}|\{\text{Non stop}, \text{One stop}\}) = \frac{e^{\mu \tilde{V}_{\text{One stop}}}}{e^{\mu V_{\text{NS}}} + e^{\mu \tilde{V}_{\text{One stop}}}}.$$

Comment

As $\mu_{\text{One stop}}$ has been normalized, μ is identified.

Upper level

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
0	Non stop dummy	4.90	0.817	6.00	0.00
4	Elapsed time (0–2 hours)	-1.60	0.405	-3.95	0.00
7	Leg room (inches), if male (non stop)	0.170	0.0584	2.91	0.00
8	Leg room (inches), if female (non stop)	0.338	0.0565	5.98	0.00
12'	More than 2 air trips per year (non stop)	0.219	0.215	1.02	0.31
16	μ	0.526	0.0307	-15.42[1]	0.00

Summary statistics

Number of observations = 2544

$$\mathcal{L}(0) = -1763.366$$
$$\mathcal{L}(c) = -1617.902$$
$$\mathcal{L}(\hat{\beta}) = -1298.135$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 930.463$$
$$\rho^2 = 0.264$$
$$\bar{\rho}^2 = 0.260$$

[1] t-test against 1

Full model (sequential estimation)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
0	Non stop dummy	4.90	0.817	6.00	0.00
1	One stop, same airline dummy	0.469	0.188	2.50	0.01
3	Round trip fare ($100)	-2.87	0.624	-4.61	0.00
4	Elapsed time (0–2 hours)	-1.60	0.405	-3.95	0.00
5	Elapsed time (2–8 hours)	-0.387	0.136	-2.84	0.00
6	Elapsed time ($>$ 8 hours)	-2.33	0.759	-3.07	0.00
7	Leg room (inches), if male (non stop)	0.170	0.0584	2.91	0.00
8	Leg room (inches), if female (non stop)	0.338	0.0565	5.98	0.00
9	Leg room (inches), if male (one stop)	0.170	0.0464	3.65	0.00
10	Leg room (inches), if female (one stop)	0.104	0.0421	2.46	0.01
11	Being early (hours)	-0.250	0.0422	-5.91	0.00
12	Being late (hours)	-0.0942	0.0286	-3.29	0.00
12'	More than 2 air trips per year (non stop)	0.219	0.215	1.02	0.31
13	More than two air trips per year (one stop, same airline)	-0.220	0.218	-1.01	0.31
15	Round trip fare / income ($100/$1000)	-37.8	40.8	-0.93	0.35
16	μ	0.526	0.0307	-15.42[1]	0.00

Summary statistics

Number of observations = 2544

$$\mathcal{L}(0) = -2349.769$$
$$\mathcal{L}(c) = -2203.160$$
$$\mathcal{L}(\hat{\beta}) = -1617.129 (= -1298.135 - 318.994)$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 1465.28$$
$$\rho^2 = 0.312$$
$$\bar{\rho}^2 = 0.305$$

Logit vs. Nested logit

Scale

- What is being estimated: $\mu\beta$
- Logit: $\mu = 1$ (normalized)
- Nested logit: $\mu = 0.526$
- Make sure to compare $\mu\beta$ across models.

Normalization of constants

Parameter number	Description	Logit	Nested logit	Nested logit (scaled)	Nested logit (scaled & shifted)
0	Non stop	0.0	4.90	2.58	0.0
1	One stop–same airline dummy	-0.898	0.469	0.247	-2.33
2	One stop–multiple airlines dummy	-1.24	0.0	0.0	-2.58
12'	$>$ 2 trips/y (non stop)	0.0	0.219	0.115	0.0
13	$>$ 2 trips/y (one stop–same airline)	-0.279	-0.220	-0.115	-0.231
14	$>$ 2 trips/y (one stop–multiple airlines)	-0.0670	0.0	0.0	-0.115

Logit vs Nested logit

t-test

Logit: $\mu = 1$

$$\frac{0.526 - 1}{0.0307} = -15.42$$

Reject $H_0 : \mu = 1$

Likelihood ratio test

- Logit: -1635.068
- Nested logit: -1617.129
- Likelihood ratio test: -2(-1635.07+1617.13)=35.378
- Threshold: $\chi^2_{1,0.95} = 3.84$
- Logit is rejected.

Estimation

Sequential estimation

- Estimate first the lower levels.
- Transfer the estimated utility function to estimate the upper level.
- Consistent estimator.
- Not efficient.

Full information maximum likelihood

- All parameters estimated together.
- Consistent.
- Efficient.

Full model (full information estimation)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
0	Non stop dummy	1.74	0.337	5.16	0.00
1	One stop, same airline dummy	0.437	0.183	2.39	0.02
3	Round trip fare ($100)	-2.81	0.315	-8.91	0.00
4	Elapsed time (0–2 hours)	-1.49	0.417	-3.57	0.00
5	Elapsed time (2–8 hours)	-0.348	0.112	-3.10	0.00
6	Elapsed time ($>$ 8 hours)	-1.62	0.506	-3.21	0.00
7	Leg room (inches), if male (non stop)	0.168	0.0587	2.86	0.00
8	Leg room (inches), if female (non stop)	0.330	0.0624	5.28	0.00
9	Leg room (inches), if male (one stop)	0.175	0.0396	4.41	0.00
10	Leg room (inches), if female (one stop)	0.112	0.0344	3.25	0.00
11	Being early (hours)	-0.234	0.0338	-6.92	0.00
12	Being late (hours)	-0.135	0.0241	-5.61	0.00
12'	More than two air trips per year (non stop)	0.199	0.243	0.82	0.41
13	More than two air trips per year (one stop, same airline)	-0.237	0.210	-1.13	0.26
15	Round trip fare / income ($100/$1000)	-36.4	14.3	-2.55	0.01
16	μ	0.546	0.0595	-7.62[1]	0.00

Summary statistics

Number of observations = 2544

$$\mathcal{L}(0) = -2794.870$$
$$\mathcal{L}(c) = -2203.160$$
$$\mathcal{L}(\hat{\beta}) = -1613.858$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 2362.022$$
$$\rho^2 = 0.423$$
$$\bar{\rho}^2 = 0.417$$

Normalize $\mu = 1$, estimate μ_m

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	One stop–same airline dummy	-0.710	0.169	-4.20	0.00
2	One stop–multiple airlines	-0.949	0.173	-5.47	0.00
3	Round trip fare ($100)	-1.54	0.149	-10.29	0.00
4	Elapsed time (0–2 hours)	-0.815	0.215	-3.80	0.00
5	Elapsed time (2–8 hours)	-0.190	0.0610	-3.12	0.00
6	Elapsed time ($>$ 8 hours)	-0.887	0.267	-3.32	0.00
7	Leg room (inches), if male (non stop)	0.0919	0.0310	2.96	0.00
8	Leg room (inches), if female (non stop)	0.180	0.0296	6.08	0.00
9	Leg room (inches), if male (one stop)	0.0954	0.0219	4.35	0.00
10	Leg room (inches), if female (one stop)	0.0610	0.0193	3.16	0.00
11	Being early (hours)	-0.128	0.0160	-7.97	0.00
12	Being late (hours)	-0.0739	0.0141	-5.23	0.00
13	More than two air trips per year (one stop–same airline)	-0.239	0.124	-1.93	0.05
14	More than two air trips per year (one stop–multiple airlines)	-0.109	0.132	-0.82	0.41
15	Round trip fare / income ($100/$1000)	-19.9	7.47	-2.66	0.01
16	μ_m	1.83	0.199	4.17[1]	0.00

Summary statistics

Number of observations = 2544

$$\mathcal{L}(0) = -2794.870$$
$$\mathcal{L}(c) = -2203.160$$
$$\mathcal{L}(\hat{\beta}) = -1613.858$$
$$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] = 2362.022$$
$$\rho^2 = 0.423$$
$$\bar{\rho}^2 = 0.417$$

Normalization of the nested logit model

Best practice

- Normalize μ to 1.
- Estimate μ_m for each nest.
- As $0 \leq \mu/\mu_m \leq 1$, then $\mu_m \geq 1$.

Large models

- Normalize all μ_m to 1.
- Estimate μ.
- Note that it is not the most general specification.
- Imposing the same scale parameter for each nest is a strong assumption.
- Motivated only by models with a very high number of nests.

Outline

Derivation from random utility

- Let $\mathcal{C}$ be the choice set.
- Let $\mathcal{C}_1, \ldots, C_M$ be a partition of $\mathcal{C}$.
- The model is derived as

$$P(i|\mathcal{C}) = \sum_{m=1}^{M} \Pr(i|m, \mathcal{C}) \Pr(m|\mathcal{C}).$$

- Each i belongs to exactly one nest m.

$$P(i|\mathcal{C}) = \Pr(i|m) \Pr(m|\mathcal{C}).$$

- Utility: error components

$$U_i = V_i + \varepsilon_i = V_i + \varepsilon_m + \varepsilon_{im}.$$

Derivation: $\Pr(i|m)$

$$\begin{aligned}\Pr(i|m) &= \Pr(U_i \geq U_j, j \in \mathcal{C}_m) \\ &= \Pr(V_i + \varepsilon_m + \varepsilon_{im} \geq V_j + \varepsilon_m + \varepsilon_{jm}, j \in \mathcal{C}_m) \\ &= \Pr(V_i + \varepsilon_{im} \geq V_j + \varepsilon_{jm}, j \in \mathcal{C}_m)\end{aligned}$$

Assumption: ε_{im} i.i.d. $\text{EV}(0, \mu_m)$

$$\Pr(i|m) = \frac{e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_j}}.$$

Derivation: $\Pr(m|\mathcal{C})$

$$\Pr(m|\mathcal{C}) = \Pr\left(\max_{i\in\mathcal{C}_m} U_i \geq \max_{j\in\mathcal{C}_\ell} U_j, \forall \ell \neq m\right)$$

$$= \Pr\left(\varepsilon_m + \max_{i\in\mathcal{C}_m}(V_i + \varepsilon_{im}) \geq \varepsilon_\ell + \max_{j\in\mathcal{C}_\ell}(V_j + \varepsilon_{j\ell}), \forall \ell \neq m\right),$$

As ε_{im} are i.i.d. $\text{EV}(0, \mu_m)$,

$$\max_{i\in\mathcal{C}_m}(V_i + \varepsilon_{im}) \sim \text{EV}(\tilde{V}_m, \mu_m),$$

where

$$\tilde{V}_m = \frac{1}{\mu_m} \ln \sum_{i\in\mathcal{C}_m} e^{\mu_m V_i}.$$

Derivation: $\Pr(m|\mathcal{C})$

Denote

$$\max_{i \in \mathcal{C}_m}(V_i + \varepsilon_{im}) = \tilde{V}_m + \varepsilon'_m,$$

to obtain

$$\Pr(m|\mathcal{C}) = \Pr(\tilde{V}_m + \varepsilon'_m + \varepsilon_m \geq \tilde{V}_\ell + \varepsilon'_\ell + \varepsilon_\ell, \forall \ell \neq m).$$

where

$$\varepsilon'_m \sim \text{EV}(0, \mu_m).$$

Define

$$\tilde{\varepsilon}_m = \varepsilon'_m + \varepsilon_m,$$

to obtain

$$\Pr(m|\mathcal{C}) = \Pr(\tilde{V}_m + \tilde{\varepsilon}_m \geq \tilde{V}_\ell + \tilde{\varepsilon}_\ell, \forall \ell \neq m).$$

Derivation: $\Pr(m|\mathcal{C})$

Assumption: $\tilde{\varepsilon}_m$ i.i.d. $\mathrm{EV}(0,\mu)$

$$\Pr(m|\mathcal{C}) = \Pr(\tilde{V}_m + \tilde{\varepsilon}_m \geq \tilde{V}_\ell + \tilde{\varepsilon}_\ell, \forall \ell \neq m)$$

$$= \frac{e^{\mu \tilde{V}_m}}{\sum_{p=1}^{M} e^{\mu \tilde{V}_p}}.$$

We obtain the nested logit model

$$P(i|\mathcal{C}) = \frac{e^{\mu_m V_i}}{\sum_{j\in\mathcal{C}_m} e^{\mu_m V_j}} \frac{e^{\mu \tilde{V}_m}}{\sum_{p=1}^{M} e^{\mu \tilde{V}_p}}$$

$$= \frac{e^{\mu_m V_i}}{\sum_{j\in\mathcal{C}_m} e^{\mu_m V_j}} \frac{\exp\left(\frac{\mu}{\mu_m} \ln \sum_{\ell\in\mathcal{C}_m} e^{\mu_m V_\ell}\right)}{\sum_{p=1}^{M} \exp\left(\frac{\mu}{\mu_p} \ln \sum_{\ell\in\mathcal{C}_p} e^{\mu_p V_{\ell p}}\right)}$$

Nested Logit Model

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

Correlation

Correlation matrix is block diagonal:

$$\mathrm{Corr}(U_i, U_j) = \begin{cases} 1 & \text{if } i = j, \\ 1 - \dfrac{\mu^2}{\mu_m^2} & \text{if } i \neq j,\ i \text{ and } j \text{ are in the same nest } m, \\ 0 & \text{otherwise.} \end{cases}$$

Variance-covariance matrix is block diagonal:

$$\mathrm{Cov}(U_i, U_j) = \begin{cases} \dfrac{\pi^2}{6\mu^2} & \text{if } i = j, \\ \dfrac{\pi^2}{6\mu^2} - \dfrac{\pi^2}{6\mu_m^2} & \text{if } i \neq j,\ i \text{ and } j \text{ are in the same nest } m, \\ 0 & \text{otherwise.} \end{cases}$$

Summary

- Independence assumption of logit may lead to erroneous forecasts
- Relaxing the assumption: nests
- Closed form model