Mixed GEV Models

Stephane Hess
Centre for Transport Studies
Imperial College London
Structure

• Introduction
• Modelling issue
• Error components Logit
• Mixed GEV
• Comparison of performance
• Discussion
Introduction: Logit-based model structures

- **MNL**
- **NL**
- **CNL**
- **RNEV**
- **NGEV**

**Nested forms**:
- IIA within nests
- No nesting

**Simulation**:
- No IIA based
- Based

**Models**:
- EC MMNL
- Mixed GEV
- RC MMNL
- MNL

Centre for Transport Studies

Imperial College
London
Application

- Choice of departure airport in a multi-airport region
- Excludes arriving and connecting passengers
- Passengers on direct flights only
- Ignores unchosen transport modes

Passengers have already made choice of going by air
Application 2: Study area

[Map showing locations of SFO, OAK, and SJC]
Selection of destinations
Data Description

- Air-passenger survey data (August & October 1995)
- >21,000 individual passenger records
- 60 data entries per passenger
- Historic air-travel level-of-service information
- Detailed ground-access level-of-service information
- After data cleaning and selection of destinations
  ➔ 9,924 observations
- Some 3,474 passengers: no other airport possible
  ➔ Final sample: 6,450 passengers
Fare data and availability of flights

• No information on actual fares paid and on flight availability at unchosen airports at time of booking
  ➔ Need to use average fare information

• Two major assumptions
  ➔ Flights available from all 3 airports at time of booking
  ➔ Tickets sell at similar speed at the individual airports (e.g. availability of cheapest tickets)
Layout of study

• Two stages
• Stage 1: test for presence of taste heterogeneity
  ➔ Use aggregate information across airlines
  ➔ Use access journey characteristics for chosen mode
• Stage 2: elementary choice level
  ➔ Explicit modelling of 3 choices: airport, airline, access-mode
  ➔ Aim: Determine optimal model structure (substitution patterns)
Model specification

- Division into residents and visitors, and into business and leisure travellers
- Natural log-transform used for frequency
- Coefficients identified:
  - Fare
  - Frequency
  - Access-time
- Taste heterogeneity
  - Access-time (lognormal)
  - ASC SFO (Normal)
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Resident business</th>
<th>Resident leisure</th>
<th>Visitor business</th>
<th>Visitor leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>t-test</td>
<td>( \beta )</td>
<td>t-test</td>
</tr>
<tr>
<td>Fare (common)</td>
<td>-0.0475</td>
<td>-3.8</td>
<td>-0.0477</td>
<td>-3.7</td>
</tr>
<tr>
<td>Fare (income &lt; $21,000)</td>
<td>-0.0430</td>
<td>-2.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency (common)</td>
<td>1.9469</td>
<td>5.6</td>
<td>1.8881</td>
<td>7.7</td>
</tr>
<tr>
<td>Frequency (income &lt; $44,000)</td>
<td></td>
<td></td>
<td></td>
<td>1.9701</td>
</tr>
<tr>
<td>Frequency (income &gt; $44,000)</td>
<td></td>
<td></td>
<td></td>
<td>3.0328</td>
</tr>
<tr>
<td>Access time ( c )</td>
<td>-1.8571</td>
<td>-15.5</td>
<td>-1.8916</td>
<td>-17.1</td>
</tr>
<tr>
<td>Access time ( s )</td>
<td>0.6742</td>
<td>4.3</td>
<td>0.5102</td>
<td>3.6</td>
</tr>
<tr>
<td>Access time ( \mu )</td>
<td>-0.1960</td>
<td>N/A</td>
<td>-0.1718</td>
<td>N/A</td>
</tr>
<tr>
<td>Access time ( \sigma )</td>
<td>0.1487</td>
<td>N/A</td>
<td>0.0937</td>
<td>N/A</td>
</tr>
<tr>
<td>ASC SFO mean</td>
<td>1.1563</td>
<td>4.2</td>
<td>0.9289</td>
<td>3.9</td>
</tr>
<tr>
<td>ASC SFO std.dev</td>
<td>2.0260</td>
<td>3.6</td>
<td>1.3650</td>
<td>2.7</td>
</tr>
<tr>
<td>ASC SJC</td>
<td>-0.1045</td>
<td>-0.5</td>
<td>-0.1515</td>
<td>-0.8</td>
</tr>
<tr>
<td>LL</td>
<td>-604.03</td>
<td></td>
<td>-659.67</td>
<td></td>
</tr>
<tr>
<td>LL (MNL)</td>
<td>-615.53</td>
<td></td>
<td>-666.22</td>
<td></td>
</tr>
</tbody>
</table>
Results 1

• Fare:
  ➔ Significant only for leisure travellers and resident business travellers
  ➔ Poor data, but could also indicate indifference to cost

• Frequency:
  ➔ Income effect only for visiting leisure travellers

• MMNL model leads to modest gains in model fit, but important gains in explanatory power
<table>
<thead>
<tr>
<th></th>
<th>Resident business</th>
<th>Resident leisure</th>
<th>Visitor business</th>
<th>Visitor leisure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of access-time ($/min)</td>
<td>4.27 [2.69] b</td>
<td>3.47 [1.65]</td>
<td>N/A</td>
<td>3.48 [2.26]</td>
</tr>
<tr>
<td>Willingness to pay for frequency increases ($) a</td>
<td>45.26K b</td>
<td>38.63K</td>
<td>N/A</td>
<td>41.31K c</td>
</tr>
<tr>
<td>Mean willingness to accept access-time increases for one additional flight at a base frequency of 5 flights (min)</td>
<td>2.90</td>
<td>2.56</td>
<td>3.90</td>
<td>3.32 c</td>
</tr>
<tr>
<td>Willingness to pay for one additional flight at a base frequency of 5 flights ($)</td>
<td>8.25</td>
<td>7.04</td>
<td>N/A</td>
<td>7.53 11.59</td>
</tr>
</tbody>
</table>

\[ K = \ln(f+1) - \ln(f); \]  
\[ b \text{ low-income travellers only; } \]  
\[ c \text{ low-income and medium-income travellers only; } \]  
\[ d \text{ high-income travellers only } \]
Results 2

• Access-time:
  ➔ Higher VOT for business travellers
  ➔ Greater variation for visitors than for residents
  ➔ VOT very high
    ➔ Poor data
    ➔ Perception of risk

• Frequency:
  ➔ Visitors value increases more than residents
  ➔ High-income travellers more sensitive to changes
  ➔ On average: business travellers more sensitive
Model performance

• Probability of correct prediction of choices:
  - Resident business: 64.29%
  - Resident leisure: 67.95%
  - Visiting business: 66.46%
  - Visiting leisure: 65.85%

• Performance on validation sample (660 travellers)
  - Resident business: 67.61%
  - Resident leisure: 66.09%
  - Visiting business: 67.03%
  - Visiting leisure: 68.25%

• Also: high accuracy in recovering true market shares
Summary & Conclusions: Stage 1

- Prevalence of taste heterogeneity in population of air-travellers, both deterministic and random
- Higher sensitivity to fare for low-earners and leisure travellers
- Higher values of access time and flight frequency for business travellers
- Non-linear specification of flight frequency offers great benefits
- Similar results in ongoing study at elementary choice level
MNL/NL modelling

• Stage 2:
  ➔ Explicit modelling of choice of airport (3), airline (8) and access-mode (6)

• 144 elementary alternatives (airport-airline-access-mode)

• 2 stages:
  ➔ MNL: search for optimal utility specification
  ➔ NL: search for optimal nesting approach
Model specification

• 6 separate models:
  ➔ Separate models for residents and visitors
  ➔ Segmentation by purpose (business, holiday, VFR)
• Explanatory variables:
  ➔ Access cost, in-vehicle time, walk-time, wait-time
  ➔ Flight fare, frequency, flight time, turboprop dummy
  ➔ On-time performance (never significant)
  ➔ Past experience (always hugely significant)
  ➔ Log-transform used for frequency and experience
Results (summary…)

- No significant effect of fare for business travellers and visiting holiday travellers
- Negative effect of turboprop for resident business and holiday travellers
- Positive effect of increases in frequency
- Negative effect of increases in in-vehicle time
- Higher cost-sensitivity for low-income groups
- Higher values of time for high-income groups
Results of NL models

- Substantive results similar to MNL results
- Not possible to fit multi-level NL models
- Have to use nesting along 1 dimension only
- Best fit offered by nesting by access-mode
  - Reflection of low elasticity for changing mode
- Important differences across purposes and residents/visitors in correlation structures
Nesting by airport

- Structural parameter for SFO always needs to be set to 1
  ➔ No heightened correlation between SFO alternatives compared to non-SFO alternatives

<table>
<thead>
<tr>
<th></th>
<th>Business</th>
<th></th>
<th>Holiday</th>
<th></th>
<th>VFR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Resident</td>
<td>Visitor</td>
<td>Resident</td>
<td>Visitor</td>
<td>Resident</td>
</tr>
<tr>
<td>SFO</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SJC</td>
<td>0.7829</td>
<td>0.5259</td>
<td>0.7627</td>
<td>0.4399</td>
<td>0.6708</td>
</tr>
<tr>
<td>OAK</td>
<td>0.8925</td>
<td>0.7178</td>
<td>0.7258</td>
<td>0.7373</td>
<td>0.7828</td>
</tr>
</tbody>
</table>
Nesting by airline

- Airlines 1, 3 and 7 had very poor punctuality record
- Airlines 5 and 8 are low-cost carriers

| Airline | Business | | | Holiday | | | | VFR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Resident | Visitor | | | Resident | Visitor | | | Resident | Visitor | | | |
| 1 | 0.9499 | 0.9617 | | | 0.9237 | 0.6989 | | | 1.00 | 1.00 | | | |
| 2 | 0.6108 | 0.9822 | | | 0.7841 | 0.6249 | | | 0.8663 | 0.8606 | | | |
| 3 | 1.00 | 0.8895 | | | 1.00 | 0.7697 | | | 0.8617 | 0.8549 | | | |
| 4 | 1.00 | 0.6538 | | | 1.00 | 0.7237 | | | 1.00 | 0.6762 | | | |
| 5 | 0.7433 | 0.6317 | | | 0.7379 | 0.3917 | | | 0.6344 | 1.00 | | | |
| 6 | 1.00 | 1.00 | | | 0.9967 | 0.6761 | | | 1.00 | 0.7935 | | | |
| 7 | 1.00 | 1.00 | | | 1.00 | 1.00 | | | 1.00 | 1.00 | | | |
| 8 | 0.8389 | 0.7921 | | | 0.7240 | 0.5298 | | | 0.6664 | 0.8399 | | | |
Nesting by access-mode

- Low structural parameters for car and taxi, and limo (where identifiable)
- High variability especially for scheduled airport services

<table>
<thead>
<tr>
<th></th>
<th>Business</th>
<th>Holiday</th>
<th>VFR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RES</td>
<td>VIS</td>
<td>RES</td>
</tr>
<tr>
<td>CAR</td>
<td>0.1793</td>
<td>0.4531</td>
<td>0.1252</td>
</tr>
<tr>
<td>SCHEDULED</td>
<td>0.1919</td>
<td>0.6378</td>
<td>0.1763</td>
</tr>
<tr>
<td>PT</td>
<td>0.3118</td>
<td>0.2473</td>
<td>0.3023</td>
</tr>
<tr>
<td>D2D</td>
<td>0.2929</td>
<td>0.4988</td>
<td>0.1796</td>
</tr>
<tr>
<td>TAXI</td>
<td>0.1283</td>
<td>0.3805</td>
<td>0.0901</td>
</tr>
<tr>
<td>LIMO</td>
<td>1.00</td>
<td>0.3636</td>
<td>0.2211</td>
</tr>
</tbody>
</table>
Summary

- Important differences across purposes and between residents and visitors
- Nesting only leads to minor improvements in model fit
  - helps interpretation
  - makes model behaviour more realistic
- Multi-level nesting structures do not converge
- Solution: use CNL
Every alternative belongs to one nest in each group
Possible to add upper level of nesting, to have heightened correlation between lower-level nests, say between bus and train
Modelling requirements

- Presence of taste heterogeneity
  - Failure to include this can lead to wrong trade-offs (e.g. VOT)
- Presence of complex substitution patterns
  - Failure to include this can lead to wrongly specified substitution patterns
- Issue: simultaneous modelling of 2 phenomena
- Two possibilities:
  - Error-components Logit
  - Mixed GEV
ECL 1

- Mixed Logit integrates MNL probabilities over assumed distribution of unobserved part of utility
- Random coefficients:
  - Taste coefficient 1 has mean $b_1$ in population
  - $b_1 \times x_{1,i}$ part of observed utility for alternative $i$
  - Agent $n$ has taste $\beta_{n1}$, with $\beta_{n1} = b_1 - s_{n1}$, $s_{n1}$ positive or negative
  - $s_{n1} \times x_{1,i}$ part of observed utility for alternative $i$ for agent $n$
- Parameter $\beta_1$ distributed with mean $b_1$ and standard deviation $s_1$
ECL 2

• GEV models:
  ➔ substitution patterns result of correlation in unobserved part of utility
• Can similarly induce correlation in Mixed Logit
  ➔ Error-components Logit (ECL) formulation
• Principle:
  ➔ Additional vector of explanatory variables, $z_n$
  ➔ take values of 0 or 1, depending on alternative
  ➔ Normally distributed, with mean of zero
  ➔ Only enter unobserved part of utility
  ➔ Creates correlation in unobserved part of utility
• $\mu_n \sim N(0, W)$
• $W$ generally diagonal
  ➔ error components are independent
  ➔ no correlation between $(z_{n,j} \ast \mu_{nj})$ and $(z_{n,k} \ast \mu_{nk})$ for $k \neq j$
  ➔ utility still correlated between alternatives sharing common $z$s
• Correlation between alternatives 1 and 2 calculated as:

$$\text{Cov}(\mu'_n z_{n1} + \varepsilon_{n1}, \mu'_n z_{n2} + \varepsilon_{n2}) = z'_{n1} W z_{n2}$$
ECL example

- 6 alternatives, 3 nests (A&B), (C&D), (E&F):
  - $z_{nA} = z_{nB} = (1,0,0)'$
  - $z_{nC} = z_{nD} = (0,1,0)'$
  - $z_{nE} = z_{nF} = (0,0,1)'$

- Diagonal $W$:
  $$W = \begin{bmatrix}
  \sigma_1 & 0 & 0 \\
  0 & \sigma_2 & 0 \\
  0 & 0 & \sigma_3 \\
  \end{bmatrix}$$

- Covariance between A & B is $\sigma_1$
  - Variance for each alternative equal to $\sigma_1 + \pi^2/6$
  - Correlation equal to $\sigma_1/ (\sigma_1 + \pi^2/6)$
Identification 1

• Principle:
  ➔ One error component per nest
• But: certain conditions need to be satisfied
  ➔ Order condition (necessary)
    ➔ A maximum of $J(J-1)/2 – 1$ alternative-specific parameters in the covariance matrix can be identified
  ➔ Rank condition (sufficient)
    ➔ $R = \text{rank of Jacobian of column vector of unique elements in covariance matrix of utility differences}$
    ➔ can estimate $R-1$ parameters
Identification 2

- If not all parameters identifiable, need normalisation
- 2 conditions:
  - Covariance matrix of normalised model and non-normalized model must be equal (set of equations)
  - Normalised covariance matrix must be positive semi-definite
- Still: often more than 1 acceptable normalisations
Identification: examples

• Heteroscedastic Logit:
  ➔ Need to constrain one variance term to zero
• Nested Logit with 2 nests:
  ➔ Only one structural parameter identifiable
  ➔ Waker (2001): three normalisation approaches ($\sigma_1=0, \sigma_2=0, \sigma_1=\sigma_2$) equivalent
• Cross-nested Logit model:
  ➔ Generally all parameters identifiable
  ➔ But: alternative belonging to highest number of nests initially forced to have highest variance
Advantages & Disadvantages of ECL model

• Advantages:
  ➔ Jointly accommodates taste heterogeneity and variable correlation patterns
  ➔ Accommodates heteroscedasticity
  ➔ Uses integration of “easy” MNL form

• Disadvantages:
  ➔ Identification
  ➔ Estimation: one extra dimension of integration per EC
    ➔ Inapplicable for some problems (e.g. housing units)
Mixed GEV models

• Random coefficients MMNL:
  ➔ Integration of MNL choice probabilities over assumed distribution of taste coefficients
  ➔ Conditional on $\beta$, have MNL model
• Model with random taste heterogeneity and correlation between alternatives
  ➔ Conditional on $\beta$, have a GEV model
  ➔ Random taste variation accommodated by integration over $\beta$
  ➔ Mixed GEV model
    ➔ can use any type of GEV model inside MGEV framework
Advantages & Disadvantages of MGEV model

• Advantages
  ➔ Number of random parameters limited to number of random taste coefficients
    ➔ computational savings
  ➔ No additional issues with identification; same set of rules applies as for GEV models

• Disadvantages
  ➔ Based on more complicated integrand than ECL
    ➔ run-time advantage only kicks in at a certain dimensionality
  ➔ Issue of finding optimal nesting structure
Framework for testing 1

- Simulated dataset with 10,000 observations
- 6 alternatives
- 2 nests (A,B,C); (D,E,F)
- Same structural parameters for two nests
- Two attributes, generated from $N(0,3)$
- True model:
  - Mixed GEV
  - Two taste coefficients follow Uniform Distribution
Framework for testing 2

- 4 models estimated
  - Mixed GEV
  - 3 ECL models ($\sigma_1=0$, $\sigma_2=0$, $\sigma_1=\sigma_2$)
- Criteria:
  - Model fit
  - Recovery of true values of two taste coefficients
  - Recovery of market shares and individual choice probabilities
  - Substitution patterns
Results 1

- Model fit: LL & $Rho^2$

<table>
<thead>
<tr>
<th></th>
<th>LL</th>
<th>$Rho^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed GEV</td>
<td>-2536.88</td>
<td>0.8419</td>
</tr>
<tr>
<td>ECL_A</td>
<td>-2555.38</td>
<td>0.8408</td>
</tr>
<tr>
<td>ECL_B</td>
<td>-2562.26</td>
<td>0.8404</td>
</tr>
<tr>
<td>ECL_C</td>
<td>-2549.92</td>
<td>0.8411</td>
</tr>
</tbody>
</table>
Results 2

• True coefficients:
  \[ \beta_1 \sim U[1.4, 3.4] \]
  \[ \beta_2 \sim U[1.0, 3.0] \]
• Mixed GEV: \[ \beta_1 \sim U[1.5, 3.6] ; \beta_2 \sim U[1.1, 3.2] \]
• ECL_A: \[ \beta_1 \sim U[3.0, 8.1] ; \beta_2 \sim U[2.6, 7.6] \]
• ECL_B: \[ \beta_1 \sim U[3.1, 7.5] ; \beta_2 \sim U[2.5, 7.6] \]
• ECL_C: \[ \beta_1 \sim U[4.1, 10.7] ; \beta_2 \sim U[3.3, 9.0] \]

\[ \Rightarrow \text{Mixed GEV performs best, scale difference for others} \]
Results 3

- Recovery of market shares (estimate run)
  ➔ Very close, thanks to ASCs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>16.49%</td>
<td>17.63%</td>
<td>15.25%</td>
<td>17.85%</td>
<td>17.20%</td>
<td>15.58%</td>
</tr>
<tr>
<td>Mixed NL</td>
<td>16.38%</td>
<td>17.92%</td>
<td>15.31%</td>
<td>17.75%</td>
<td>17.08%</td>
<td>15.56%</td>
</tr>
<tr>
<td>ECL_A</td>
<td>16.33%</td>
<td>17.88%</td>
<td>15.30%</td>
<td>17.76%</td>
<td>17.12%</td>
<td>15.61%</td>
</tr>
<tr>
<td>ECL_B</td>
<td>16.41%</td>
<td>17.89%</td>
<td>15.34%</td>
<td>17.72%</td>
<td>17.11%</td>
<td>15.53%</td>
</tr>
<tr>
<td>ECL_C</td>
<td>16.40%</td>
<td>17.92%</td>
<td>15.36%</td>
<td>17.71%</td>
<td>17.10%</td>
<td>15.51%</td>
</tr>
</tbody>
</table>
Results 4

• Look at individual observations
• Calculate 6 choice probabilities for each observation
• Compare values to those produced by original model, use average RMSE over observations
  ➔ Mixed GEV: 0.00341
  ➔ ECL_A: 0.00442
  ➔ ECL_B: 0.00473
  ➔ ECL_C: 0.00376
Results 5

• Assume change in attribute 1 for first alternative by -20%
• Apply different models
• Compare results to those for original model (RMSE)
  => Mixed GEV: 0.00356
  => ECL_A: 0.00454
  => ECL_B: 0.00479
  => ECL_C: 0.00388
Results 6

- Theoretically, different ECL models should produce same results
- Look in detail at some specific observations
### Results 7

<table>
<thead>
<tr>
<th></th>
<th>Alt_1</th>
<th>Alt_2</th>
<th>Alt_3</th>
<th>Alt_4</th>
<th>Alt_5</th>
<th>Alt_6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Before</strong></td>
<td>67.22%</td>
<td>21.83%</td>
<td>N/A</td>
<td>10.94%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td>57.99%</td>
<td>27.79%</td>
<td>N/A</td>
<td>14.16%</td>
<td>0.05%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Change</strong></td>
<td><strong>-13.72%</strong></td>
<td><strong>27.30%</strong></td>
<td>N/A</td>
<td><strong>29.47%</strong></td>
<td>274.70%</td>
<td>399.20%</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Before</strong></td>
<td>66.88%</td>
<td>22.70%</td>
<td>N/A</td>
<td>10.37%</td>
<td>0.05%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td>58.55%</td>
<td>27.51%</td>
<td>N/A</td>
<td>13.75%</td>
<td>0.20%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Change</strong></td>
<td><strong>-12.46%</strong></td>
<td><strong>21.16%</strong></td>
<td>N/A</td>
<td><strong>32.63%</strong></td>
<td>280.61%</td>
<td>277.54%</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Before</strong></td>
<td>68.26%</td>
<td>21.45%</td>
<td>N/A</td>
<td>10.28%</td>
<td>0.02%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td>55.81%</td>
<td>28.17%</td>
<td>N/A</td>
<td>15.93%</td>
<td>0.10%</td>
<td>0.00%</td>
</tr>
<tr>
<td><strong>Change</strong></td>
<td><strong>-18.24%</strong></td>
<td><strong>31.30%</strong></td>
<td>N/A</td>
<td><strong>54.99%</strong></td>
<td>534.17%</td>
<td>518.32%</td>
</tr>
</tbody>
</table>
## Results 8

<table>
<thead>
<tr>
<th></th>
<th>Alt_1</th>
<th>Alt_2</th>
<th>Alt_3</th>
<th>Alt_4</th>
<th>Alt_5</th>
<th>Alt_6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>Before</td>
<td>13.34%</td>
<td>18.27%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>1.04%</td>
<td>23.94%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>-92.21%</td>
<td>31.03%</td>
<td>64.15%</td>
<td>1.61%</td>
<td>11.34%</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>Before</td>
<td>14.25%</td>
<td>19.64%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>1.13%</td>
<td>27.21%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>-92.10%</td>
<td>38.54%</td>
<td>54.34%</td>
<td>0.78%</td>
<td>3.06%</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>Before</td>
<td>15.21%</td>
<td>18.67%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>1.16%</td>
<td>26.83%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>-92.35%</td>
<td>43.68%</td>
<td>71.45%</td>
<td>0.12%</td>
<td>1.74%</td>
</tr>
</tbody>
</table>
Summary & Conclusions 1

• Many transportation problems
  ➔ Prevalence of random taste heterogeneity
  ➔ Existence of complex substitution patterns

• Two possible types of models: MGEV & ECL

• ECL has minor runtime advantage in case of low number of nests

• MGEV has very significant runtime advantage in case of high number of nests
Summary & Conclusions 2

• Comparison of ECL and MGEV
  ➔ Slightly better fit for MGEV
  ➔ MGEV better able to represent changes in market shares after changes in explanatory variables

• Problems with ECL
  ➔ Estimation
  ➔ Identification
    ➔ multiple possible normalisations can lead to different results
Outlook

- Application:
  - Different substitution patterns in different groups of passengers
  - Can similarly expect differences within groups
- One possibility: parameterise structural parameters
- But: some variation may be random
  - Mixed GEV with mixing over structural parameters
Questions ?