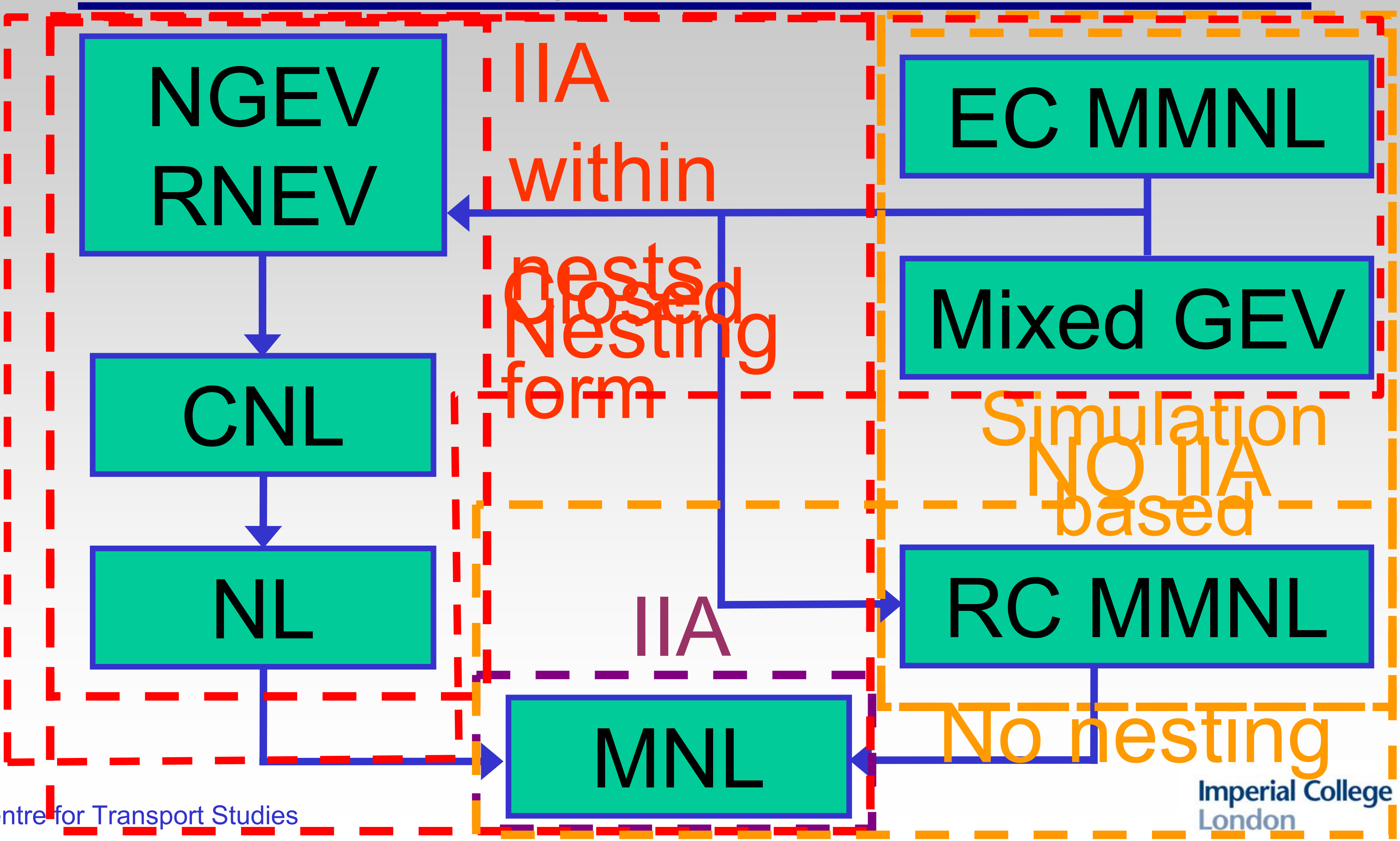
Mixed GEV Models

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



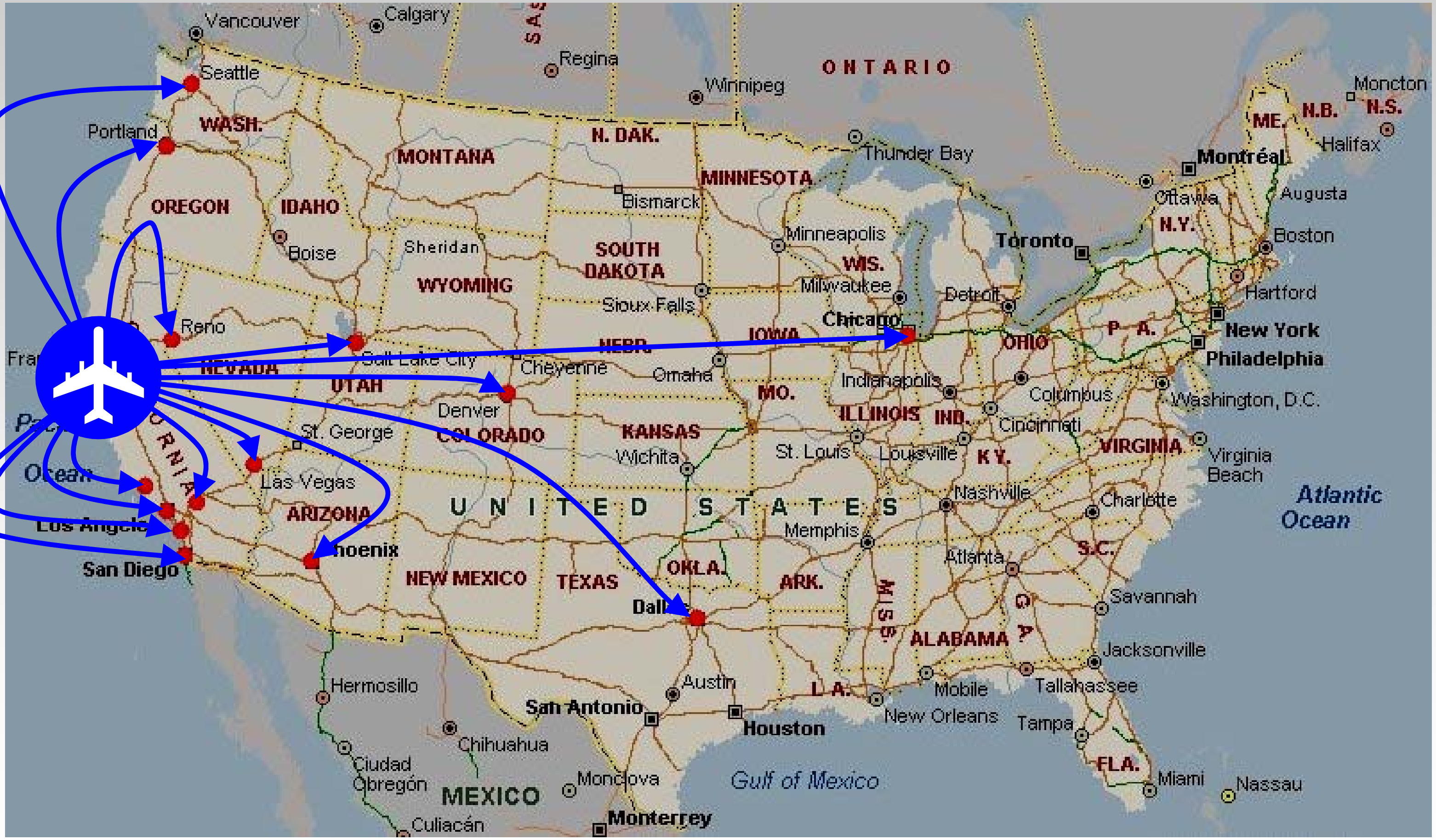
Application

- 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

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)

	Resident business		*Resident leisure*		*Visitor business*		*Visitor leisure*	
Parameter	**β**	**t-test**	**β**	**t-test**	**β**	**t-test**	**β**	**t-test**
Fare (common)			-0.0475	-3.8			-0.0477	-3.7
Fare (income < $21,000)	-0.0430	-2.55						
Frequency (common)	1.9469	5.6	1.8333	5.7	1.8881	7.7		
Frequency (income < $44,000)							1.9701	5.2
Frequency (income > $44,000)							3.0328	5.2
Access time **c**	-1.8571	-15.5	-1.8916	-17.1	-1.9706	-20.6	-1.9669	-13.0
Access time **s**	0.6742	4.3	0.5102	3.6	0.9373	5.4	0.6934	5.5
Access time μ	-0.1960	N/A	-0.1718	N/A	-0.2163	N/A	-0.1779	N/A
Access time σ	0.1487	N/A	0.0937	N/A	0.2566	N/A	0.1398	N/A
ASC SFO mean	1.1563	4.2	0.9289	3.9	0.3632	2.5	0.5028	1.9
ASC SFO std.dev	2.0260	3.6	1.3650	2.7			1.6019	2.2
ASC SJC	-0.1045	-0.5	-0.1515	-0.8	-0.7767	-3.7	0.7784	2.8
LL	-604.03		-659.67		-573.67		-514.62	
LL (MNL)	-615.53		-666.22		-592.05		-519.92	

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

	Resident business	*Resident leisure*	*Visitor business*	*Visitor leisure*
Value of access-time ($/min)	4.27 [2.69] [b]	3.47 [1.65]	N/A	3.48 [2.26]
Trade-off between frequency increases and access-time increases (min/flight) [a]	15.88$\boldsymbol{K}$ [11.75]	14.04$\boldsymbol{K}$ [7.49]	21.40$\boldsymbol{K}$ [24.63]	18.20$\boldsymbol{K}$ [13.97] [c] 28.02$\boldsymbol{K}$ [21.51] [d]
Willingness to pay for frequency increases ($) [a]	45.26$\boldsymbol{K}$ [b]	38.63$\boldsymbol{K}$	N/A	41.31$\boldsymbol{K}$ [c] 63.59$\boldsymbol{K}$ [d]
Mean willingness to accept access-time increases for one additional flight at a base frequency of 5 flights (min)	2.90	2.56	3.90	3.32 [c] 5.11 [d]
Willingness to pay for one additional flight at a base frequency of 5 flights ($)	8.25	7.04	N/A	7.53 11.59

[a] $\boldsymbol{K}=ln(f+1)-ln(f)$; [b] low-income travellers only;
[c] low-income and medium-income travellers only , [d] 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

	Business		Holiday		VFR	
	Resident	Visitor	Resident	Visitor	Resident	Visitor
SFO	1.00	1.00	1.00	1.00	1.00	1.00
SJC	0.7829	0.5259	0.7627	0.4399	0.6708	0.9333
OAK	0.8925	0.7178	0.7258	0.7373	0.7828	1.00

Nesting by airline

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

	Business		Holiday		VFR	
Airline	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

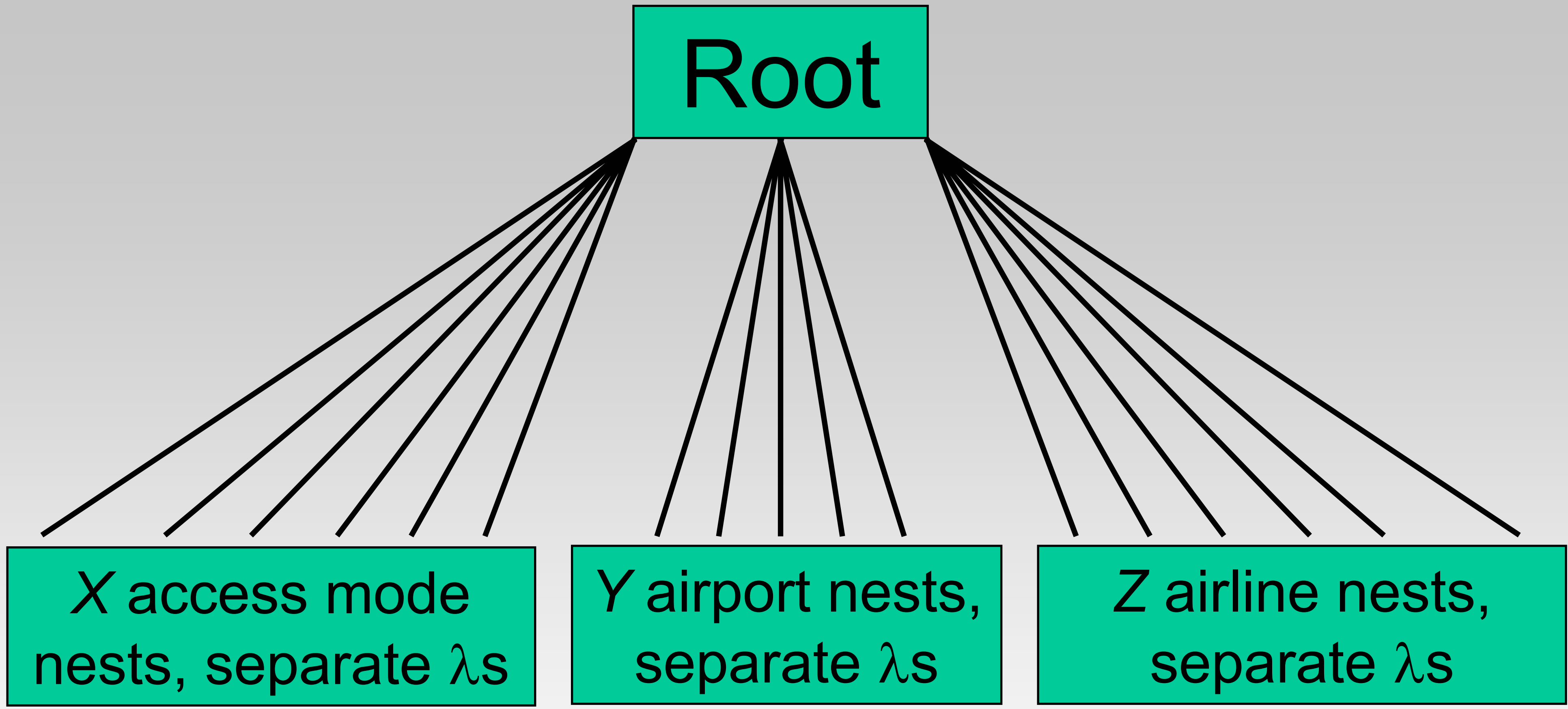
Nesting by access-mode

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

	Business		Holiday		VFR	
	RES	VIS	RES	VIS	RES	VIS
CAR	0.1793	0.4531	0.1252	0.1632	0.1325	0.0871
SCHEDULED	0.1919	0.6378	0.1763	0.1455	0.0455	0.7961
PT	0.3118	0.2473	0.3023	0.3299	1.00	0.0180
D2D	0.2929	0.4988	0.1796	0.1632	0.1792	0.1192
TAXI	0.1283	0.3805	0.0901	0.1632	0.1731	0.0543
LIMO	1.00	0.3636	0.2211	0.2475	0.3094	1.00

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
    - ➔ $b_1 * x_{1,i}$ part of observed utility for alternative *i*
  - ➔ Agent *n* has taste β_{n1}, with $\beta_{n1} = b_1 - s_{n1}$, s_{n1} positive or negative
    - ➔ $s_{n1} * x_{1,i}$ part of observed utility for alternative *i* for agent *n*
- Parameter β_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

ECL 3

- $\mu_n \sim N(0, W)$
- *W* generally diagonal
  - ➔ error components are independent
  - ➔ no correlation between ($z_{n \cdot j} * \mu_{nj}$) and ($z_{n \cdot k} * \mu_{nk}$) for $k \neq j$
  - ➔ utility still correlated between alternatives sharing common *zs*
- Correlation between alternatives 1 and 2 calculated as:

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

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 σ_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 = 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 β, have MNL model
- Model with random taste heterogeneity and correlation between alternatives
  - ➔ Conditional on β, have a GEV model
  - ➔ Random taste variation accommodated by integration over β
  - ➔ 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 (σ_1=0, σ_2=0, σ_1= σ_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

	LL	Rho^2
Mixed GEV	-2536.88	0.8419
ECL_A	-2555.38	0.8408
ECL_B	-2562.26	0.8404
ECL_C	-2549.92	0.8411

Results 2

- True coefficients:
  - ➔ β_1~U[1.4 , 3.4]
  - ➔ β_2~U[1.0 , 3.0]
- Mixed GEV: β_1~U[1.5 , 3.6] ; β_2~U[1.1 , 3.2]
- ECL_A: β_1~U[3.0 , 8.1] ; β_2~U[2.6 , 7.6]
- ECL_B: β_1~U[3.1 , 7.5] ; β_2~U[2.5 , 7.6]
- ECL_C: β_1~U[4.1 , 10.7] ; β_2~U[3.3, 9.0]

➔ Mixed GEV performs best, scale difference for others

Results 3

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

	P(Alt1)	P(Alt2)	P(Alt3)	P(Alt4)	P(Alt5)	P(Alt6)
Original	16.49%	17.63%	15.25%	17.85%	17.20%	15.58%
Mixed NL	16.38%	17.92%	15.31%	17.75%	17.08%	15.56%
ECL_A	16.33%	17.88%	15.30%	17.76%	17.12%	15.61%
ECL_B	16.41%	17.89%	15.34%	17.72%	17.11%	15.53%
ECL_C	16.40%	17.92%	15.36%	17.71%	17.10%	15.51%

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

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

Results 8

		Alt_1	Alt_2	Alt_3	Alt_4	Alt_5	Alt_6
A	Before	13.34%	18.27%	0.00%	0.01%	0.00%	68.38%
	After	1.04%	23.94%	0.00%	0.01%	0.00%	75.01%
	Change	-92.21%	31.03%	64.15%	1.61%	11.34%	9.69%
B	Before	14.25%	19.64%	0.00%	0.01%	0.00%	66.10%
	After	1.13%	27.21%	0.00%	0.01%	0.00%	71.65%
	Change	-92.10%	38.54%	54.34%	0.78%	3.06%	8.40%
C	Before	15.21%	18.67%	0.00%	0.01%	0.00%	66.11%
	After	1.16%	26.83%	0.00%	0.01%	0.00%	72.00%
	Change	-92.35%	43.68%	71.45%	0.12%	1.74%	8.91%

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 ?