Understanding air travel choice behaviour

Stephane Hess

Outline

- Introduction
- San Francisco Bay area study
- Greater London study
- SP study
- Joint RP/SP application
- Conclusions

INTRODUCTION

Introduction

- Unprecedented increase in air traffic
- Capacity problems
  - ➔ Need for increases in capacity
  - ➔ Economic & environmental considerations
- Multi-airport regions
  - ➔ Distribution of passengers after expansion
  - ➔ Need to produce forecasts
    - ➔ Need understanding of passengers' choices

Existing research

- High number of existing studies
  - ➔ Focus on US and UK
- Most studies very simplistic
  - ➔ Only choice of airport
  - ➔ Highly aggregate data
  - ➔ Basic model structures preferred

Topic and scope of analysis

- Aims:
  - → Formulate multi-dimensional choice process
  - → Explore benefits of advanced model structures
  - → Use high level of disaggregation
- Excludes arriving and connecting passengers
- Ignores unchosen transport modes (car, coach, …)

Methodology 1

- Discrete choice models
- Choice between finite set of discrete alternatives
- Alternatives have an associated “utility”
- Principle: alternative with highest utility is chosen
- Taste coefficients “calibrated” on RP or SP data

Methodology 2

- Problem: not all of the utility is observed
  - ➔ assume random distribution for remainder
- Choice becomes probabilistic
  - ➔ random utility model (RUM)
- Different model forms
  - ➔ correlation between unobserved utility terms
  - ➔ random variations in taste coefficients

SAN FRANCISCO BAY AREA STUDY

Study area


SFO
OAK
SJC
San Francisco
San Francisco Bay
Oakland
San Jose
Pacifica
ONTARIO
MINNESOTA
IOWA
Chicago
New York
Philadelphia
Atlantic Ocean
Gulf of Mexico
Houston

Destinations

Data Description

- Air passenger survey data
- Collected in August and October 1995
- In addition: (historic) air travel and ground access level-of-service information
- Present study: resident business travellers
  ➔ 1,212 observations
- Multiplicative weights used to correct effects of sampling

Fare data and availability of flights

- No information on:
  - actual fares paid
  - availability of unchosen flights
- Need to use average fare information
- Two major assumptions
  - ➔ All options available at time of booking
  - ➔ Tickets sell at similar speed at the individual airports (e.g. availability of cheapest tickets)

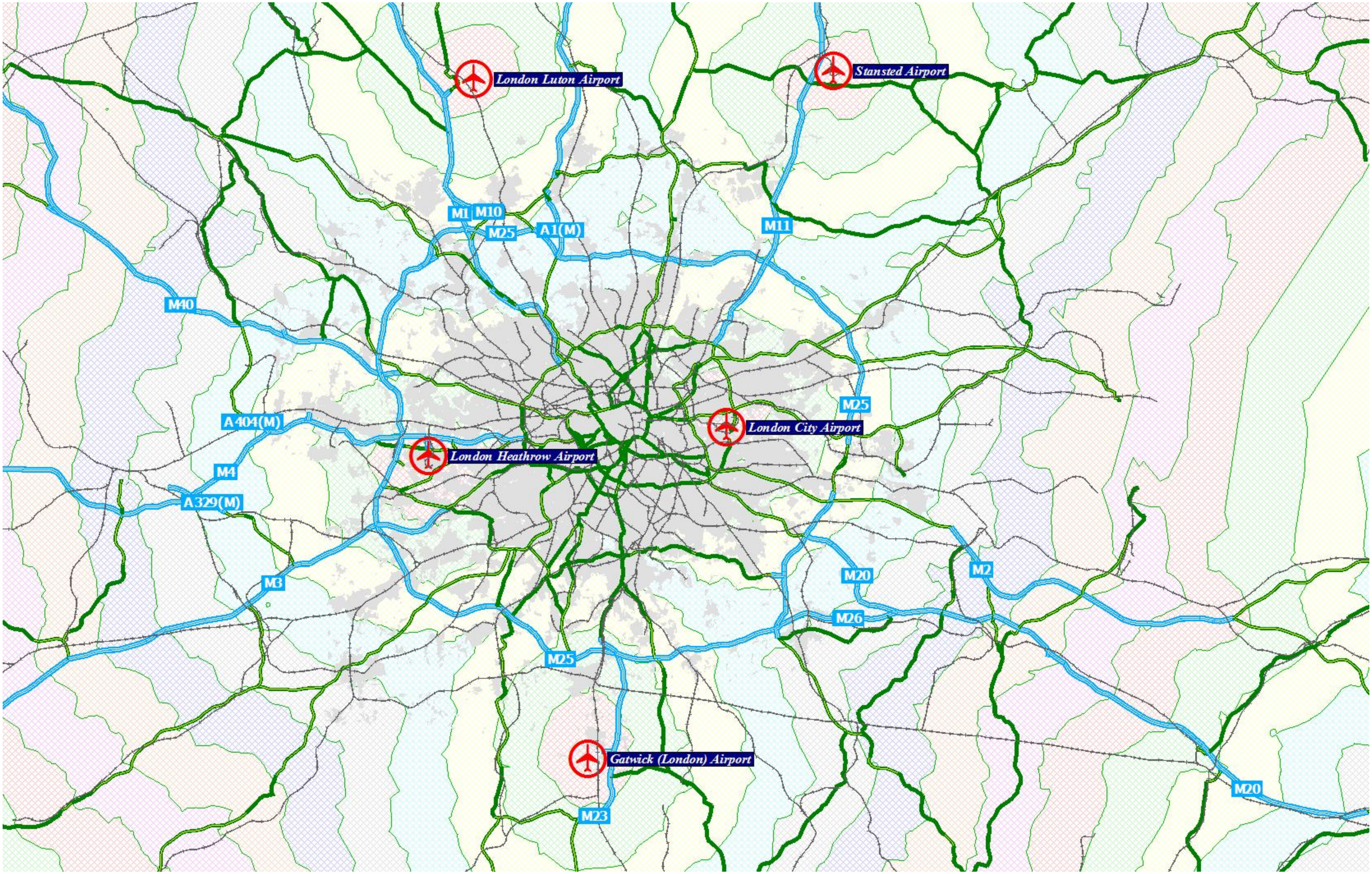
Model specification

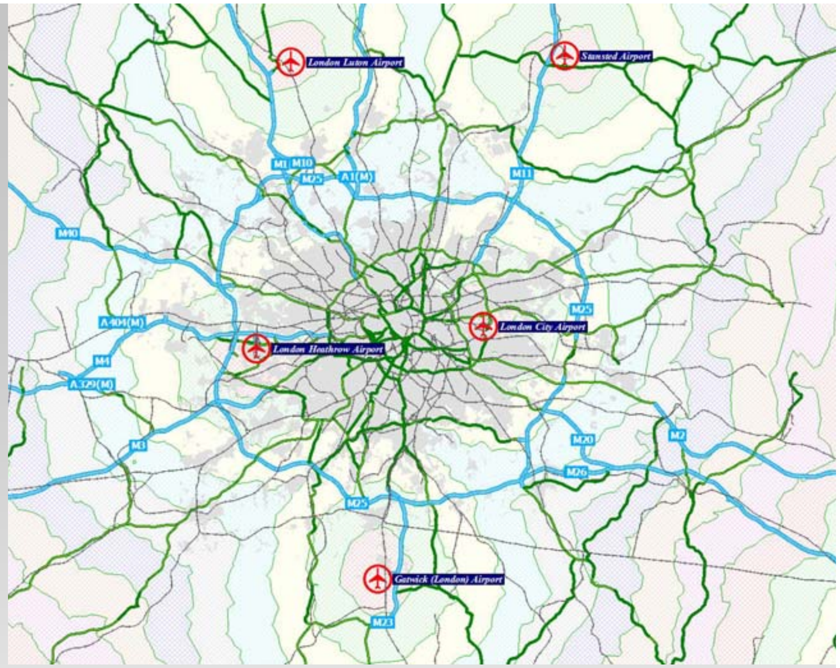
- Structure
  - ➔ 3 airports, 8 airlines, 6 access modes
  - ➔ 144 elementary alternatives (airport/airline/access mode)
- Explanatory variables:
  - ➔ Access cost, in-vehicle time, walk time, wait time
  - ➔ Fare, frequency, flight time, turboprop dummy
  - ➔ On time performance
  - ➔ Past experience

Results

- Fare: no significant effect
- Turboprop: negative effect
- Frequency: positive effect
- In-vehicle time: negative effect
- Walk time: negative effect
- Wait time: no effect
- On time: no effect
- Flight time: no effect
- Past experience: positive effect

GREATER LONDON STUDY


London Luton Airport
Stansted Airport
M1
M10
M25
A1(M)
M11
M40
A404(M)
M4
A329(M)
M25
London City Airport
London Heathrow Airport
M3
M20
M2
M26
M25
Gatwick (London) Airport
M23
M20



- LHR: 67.34 mppa (52.24%)
- LGW: 31.45 mppa (24.40%)
- STN: 20.91 mppa (16.22%)
- LTN: 7.54 mppa (5.85%)
- LCY: 1.68 mppa (1.30%)
- Total: 128.92 mppa
- LHR & LGW at capacity
- Market-share of STN has doubled in 5 years …
- Unconstrained growth predictions: 300 mppa by 2030

Topic and scope of analysis

- Combined choice of airport, airline and access mode
- 5 departure airports
- 31 destination airports
- 6 access modes
- 37 airlines
- 54 actual airport-airline combinations

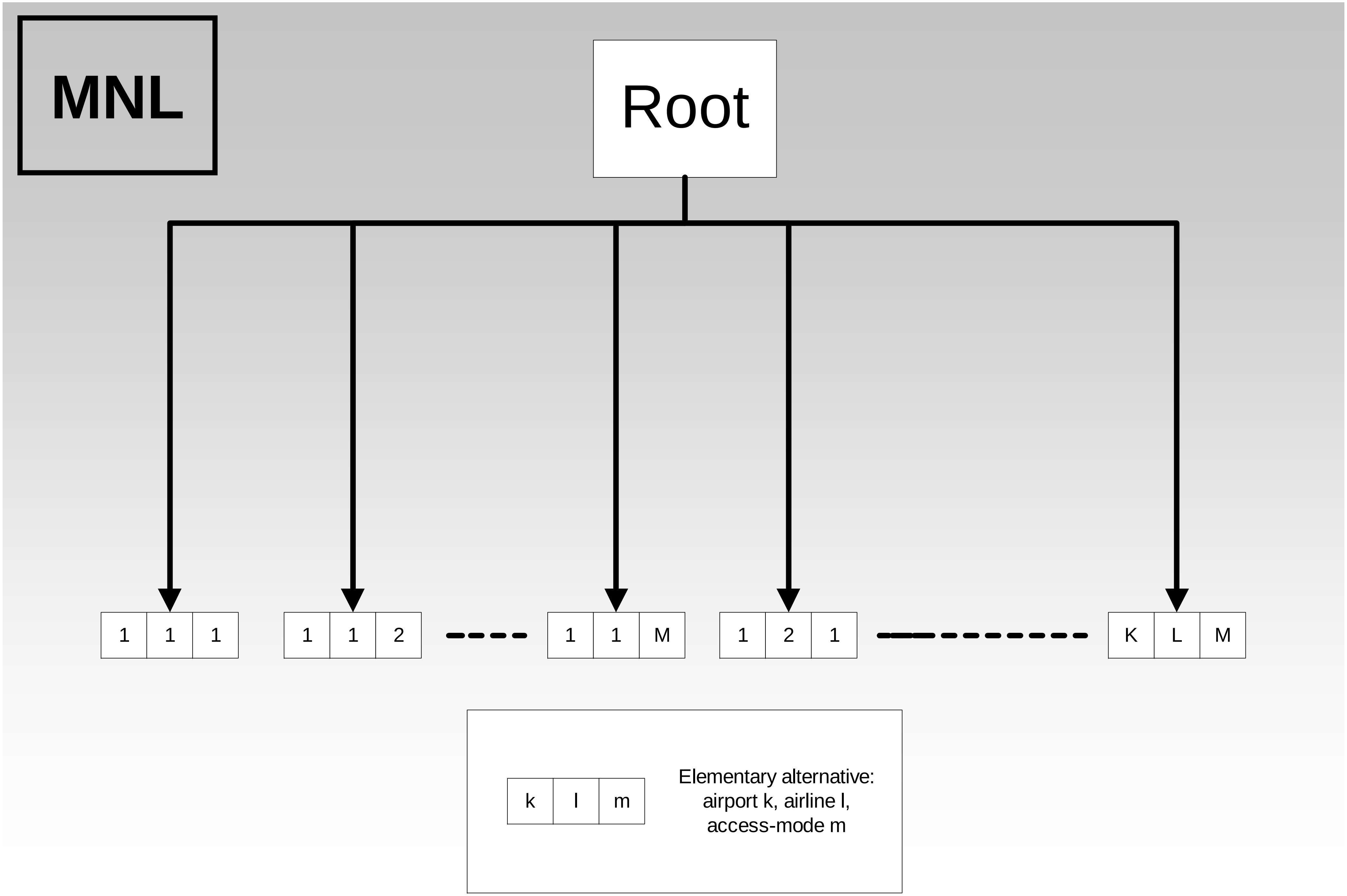
  → 324 alternatives
- Survey: 33,612 observations (1996)

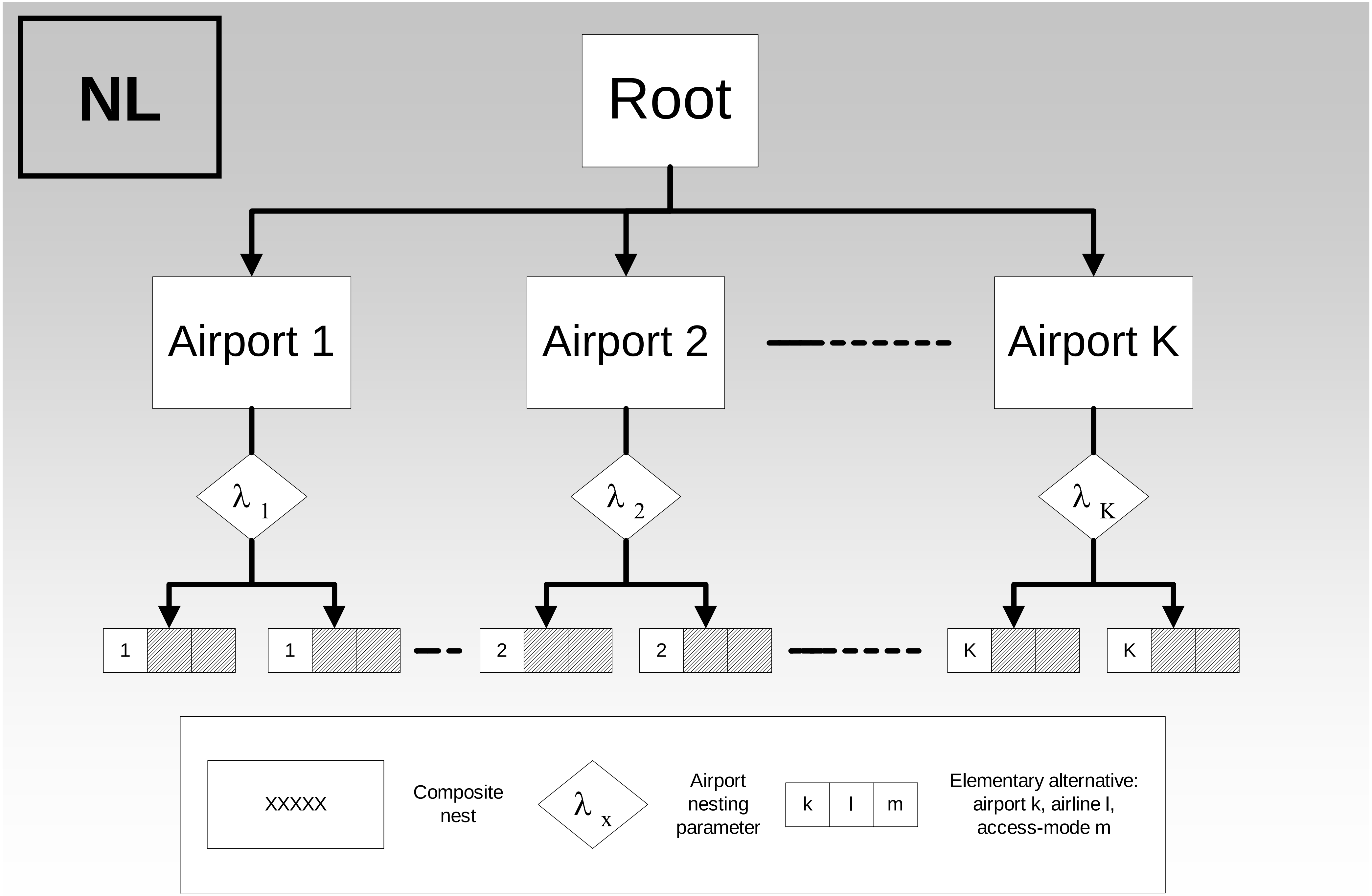
Model results

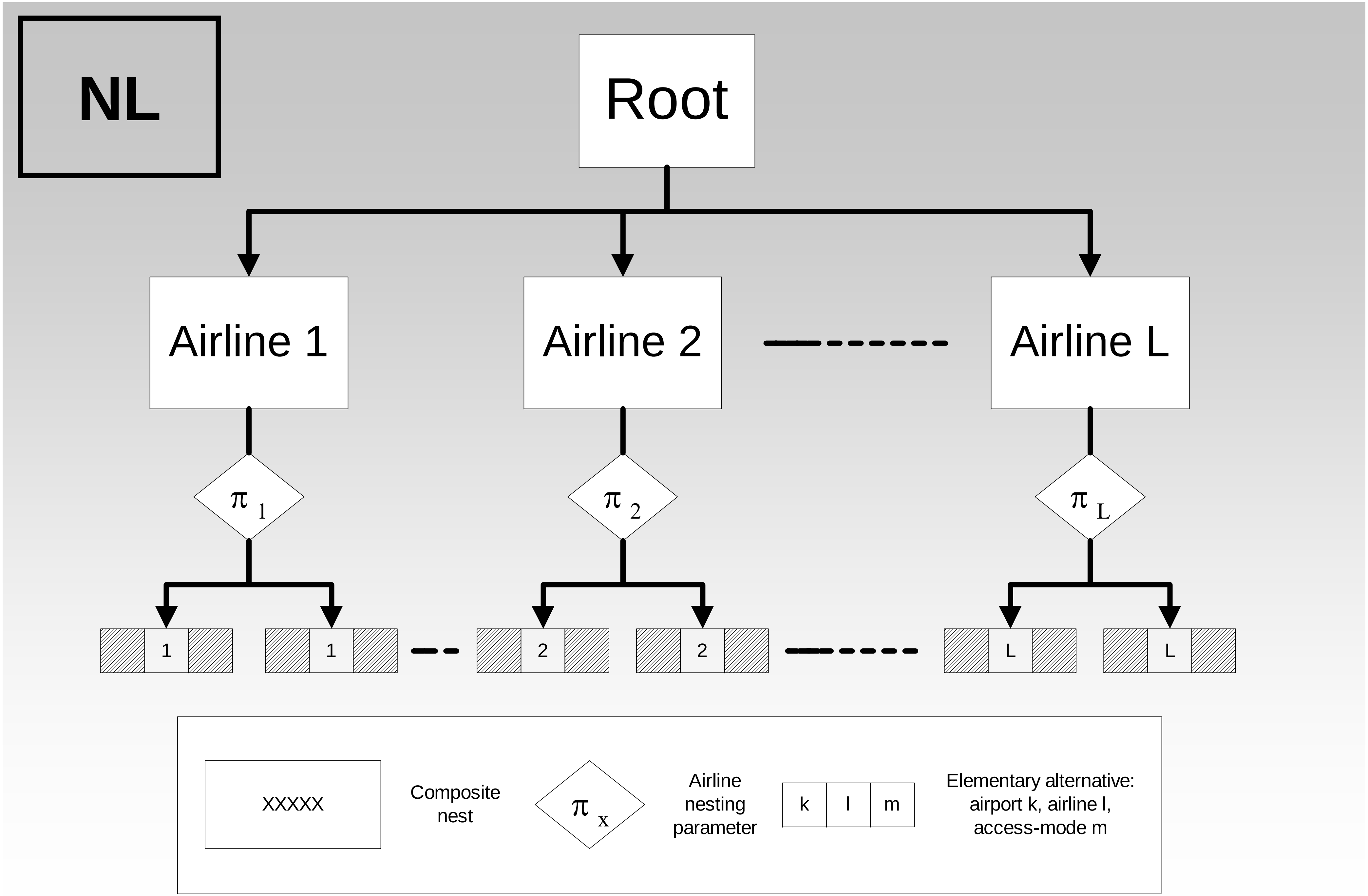
- Common effects:
  → access time (negative, log-linear)
  → access cost (negative, log-linear)
  → flight frequency (positive, log-linear)
  → flight time (negative, log-linear)
- Visiting business:
  → non-UK airline (positive)
- Visiting leisure:
  → fare (negative, linear)

Model structure

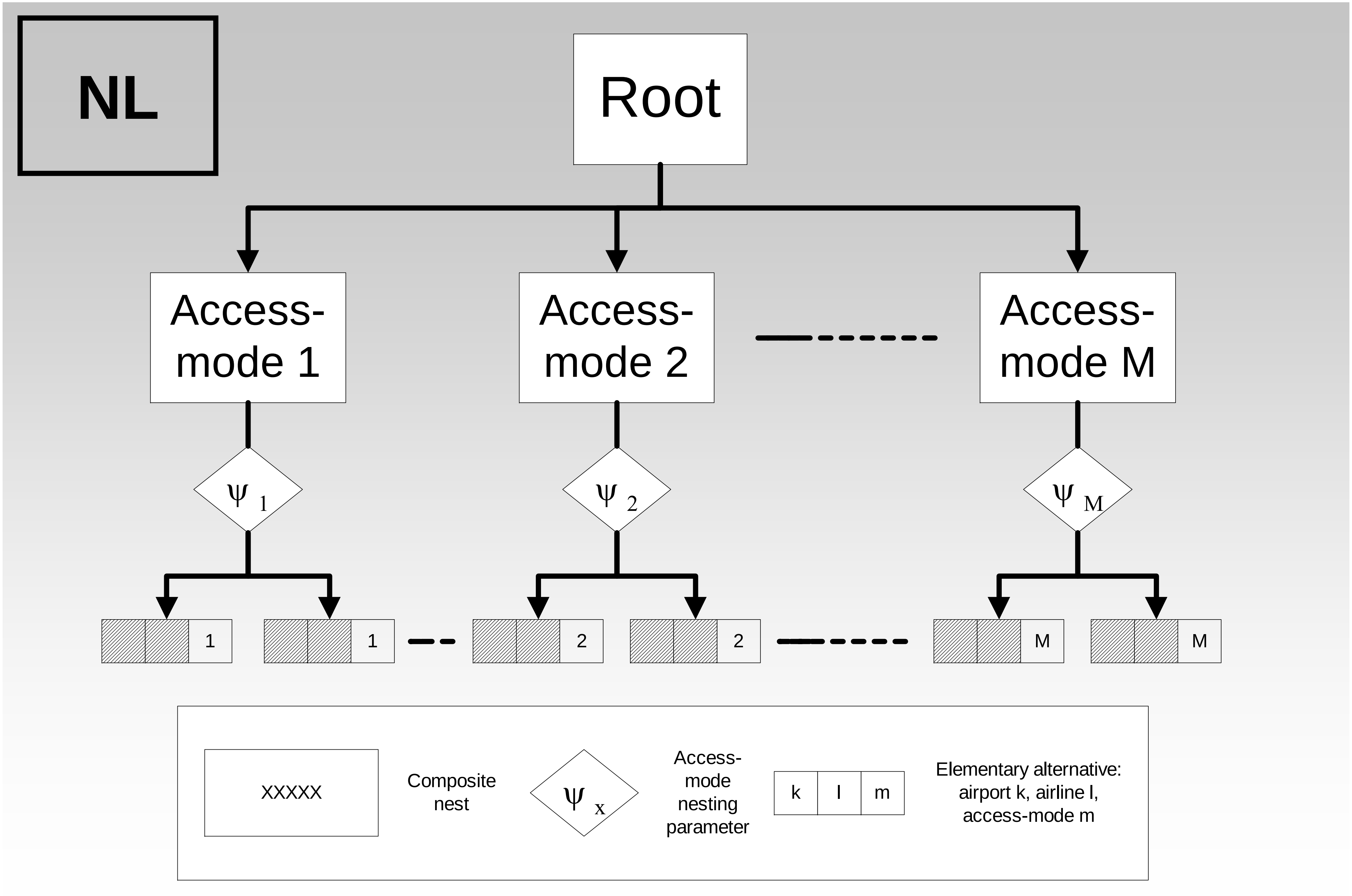
- Not all similarities between alternatives captured
- Allow for correlation in unobserved utility
- Investigate different possible nesting structures

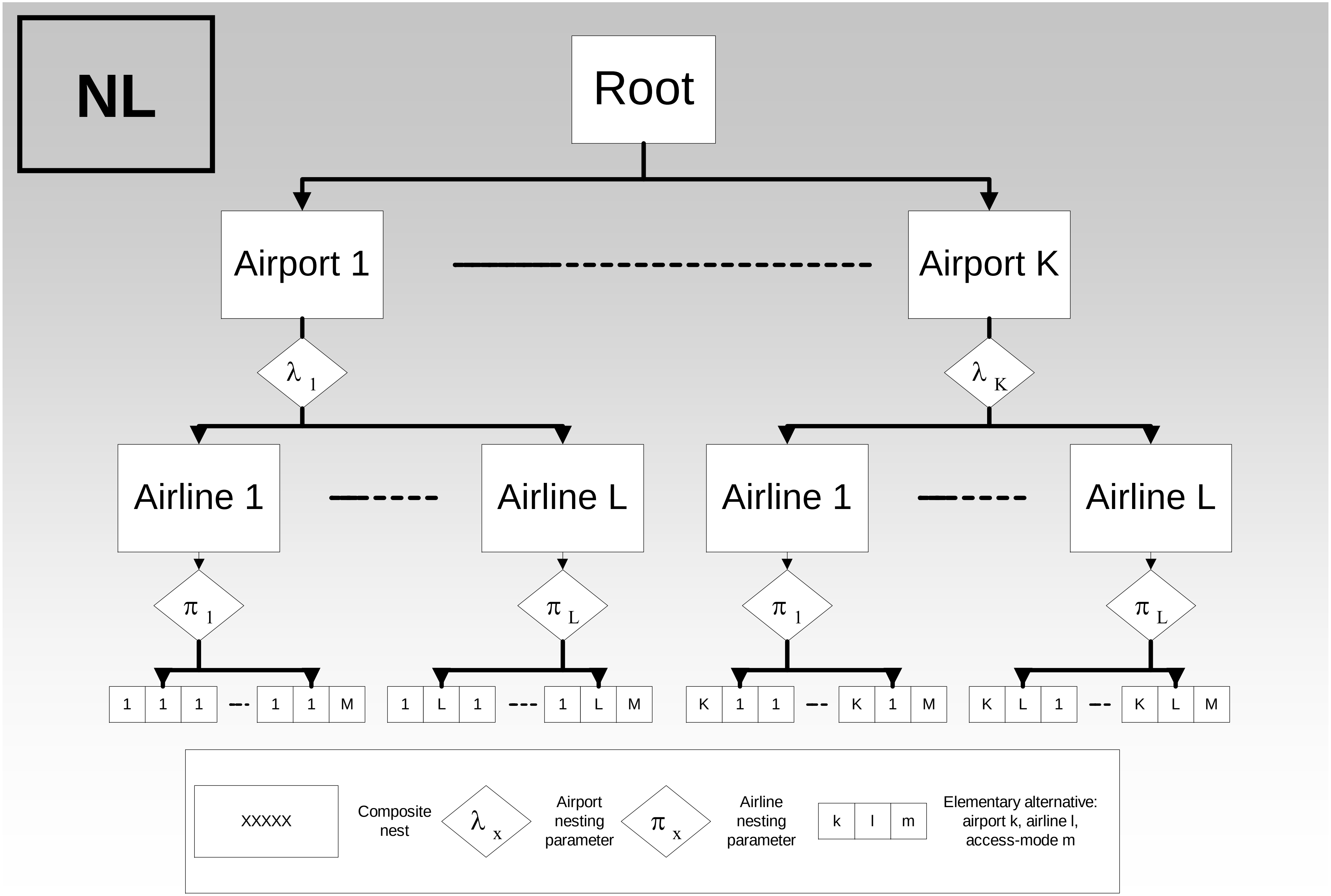

MNL
Root
1 1 1
1 1 2
1 1 M
1 2 1
K L M
k l m
Elementary alternative:
airport k, airline l,
access-mode m

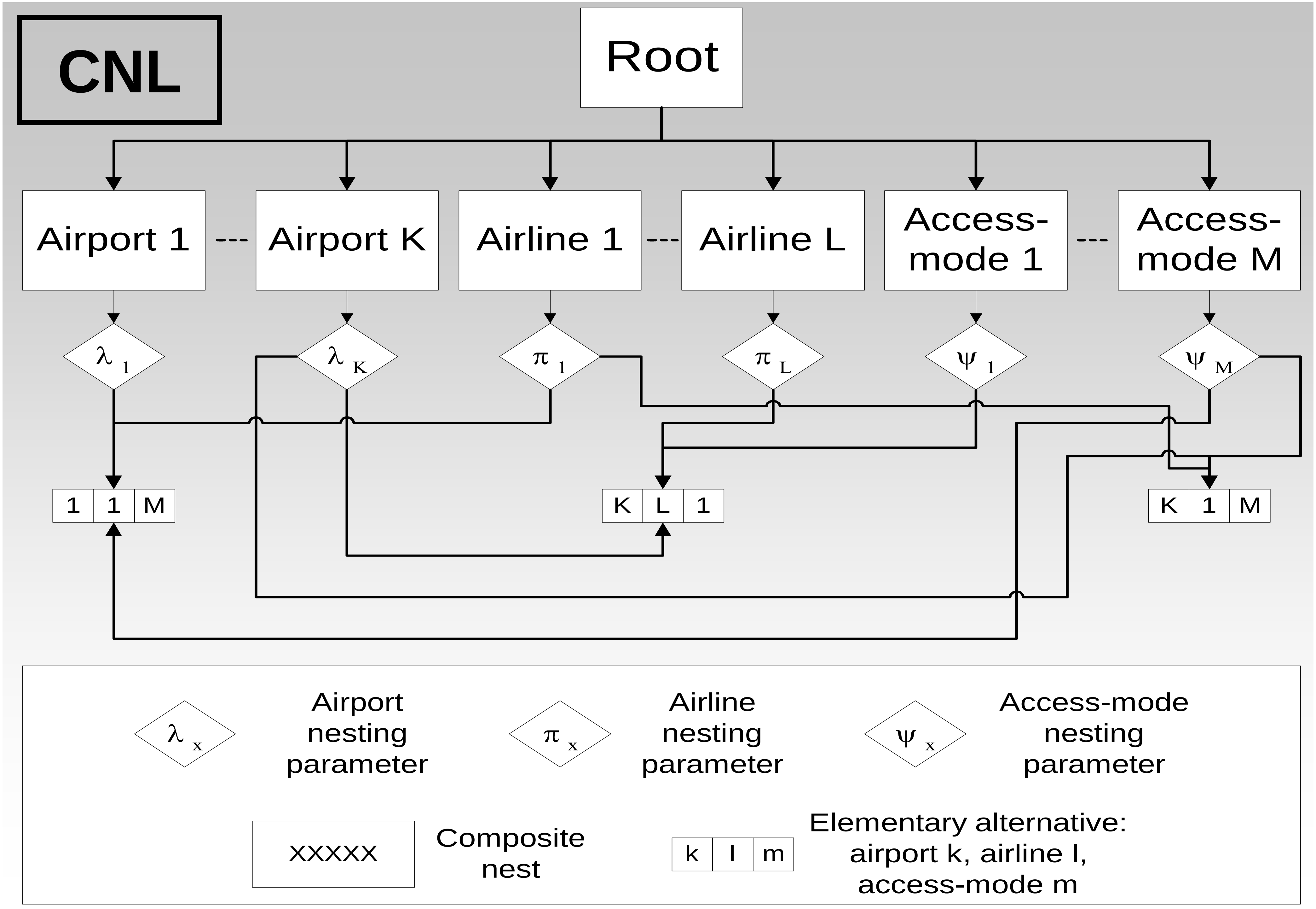

NL
Root
Airport 1
Airport 2
Airport K
λ 1
λ 2
λ K
1
1
2
2
K
K
XXXXX
Composite nest
λ x
Airport nesting parameter
k
l
m
Elementary alternative: airport k, airline l, access-mode m



NL
Root
Airline 1
Airline 2
Airline L
π 1
π 2
π L
1
1
2
2
L
L
XXXXX
Composite nest
π x
Airline nesting parameter
k
l
m
Elementary alternative: airport k, airline l, access-mode m


NL
Root
Access-mode 1
Access-mode 2
Access-mode M
ψ 1
ψ 2
ψ M
1
1
2
2
M
M
XXXXX
Composite nest
ψ x
Access-mode nesting parameter
k
l
m
Elementary alternative: airport k, airline l, access-mode m


NL
Root
Airport 1
Airport K
λ 1
λ K
Airline 1
Airline L
Airline 1
Airline L
π 1
π L
π 1
π L
1 1 1
1 1 M
1 L 1
1 L M
K 1 1
K 1 M
K L 1
K L M
XXXXX
Composite nest
λ x
Airport nesting parameter
π x
Airline nesting parameter
k l m
Elementary alternative: airport k, airline l, access-mode m


CNL
Root
Airport 1
Airport K
Airline 1
Airline L
Access-mode 1
Access-mode M
λ 1
λ K
π 1
π L
ψ 1
ψ M
1 1 M
K L 1
K 1 M
λ x
Airport nesting parameter
π x
Airline nesting parameter
ψ x
Access-mode nesting parameter
XXXXX
Composite nest
k l m
Elementary alternative: airport k, airline l, access-mode m

Model comparison

	LL at convergence	Parameters	Adjusted ρ^2	Estimation time
MNL	-14945.3	55	0.3445	< 10 mins
NL airport	-14896.1	59	0.3465	5 hours
NL airline	-14870.7	74	0.3469	5 hours
NL access	-14816.7	60	0.3499	3 hours
CNL	**-14603.9**	**91**	**0.3578**	**> 1 week**

- But: differences in trade-offs and substitution patterns!

SP CASE STUDY

Limitations of RP survey data

- Collected by airport or airline authorities
  - priorities different from those of modellers
- Problems with auxiliary data:
  - poor quality of fare data
  - no information on availabilities
  - no frequent flier information
  - not possible to measure scheduling effects
- Big assumptions necessary, poor results …
  - generally no effect of air fares …

Alternatives to RP survey data

- Self-collected RP survey data:
  - expensive & limited sample size
  - problems in getting clearances
  - problems with availabilities remain
- RP bookings data:
  - information on availabilities
  - generally no information on socio-demographics etc
- SP data:
  - doubts about reliability
  - but: information on actually faced alternatives
  - information on fares, frequent flier, scheduling,…

SP choice data

- Airport & airline choice in US
- Internet-based
- 600 travellers, 10 observations per traveller
- Binomial choice
- 3 groups: Business, holiday, VFR

Air Travel Study2002

Which would you choose for a trip to Jacksonville International, Jacksonville?

	Your Current Flight	Alternate Flight
CARRIER	American Airlines	Northwest
ON-TIME PERFORMANCE	This flight was on time	80% of these flights are on-time
SCHEDULED IN-THE-AIR TRAVEL TIME	5 hrs. 45 mins.	5 hrs. 45 mins.
ARRIVAL TIME	5:45 PM	7:45 PM
NUMBER OF CONNECTIONS	1	None
AIRCRAFT TYPE	Regional Jet and Standard Jet	Standard Jet
FARE	$250	$188
DEPARTURE AIRPORT	Manchester Airport, Manchester NH	Rutland State Airport, Rutland VT

○ I prefer my current trip ○ I prefer the alternate trip

Question 3 of 10

Progress 80%

NEXT

Results 1

- Effects of:
  - air fare (log-linear)
    - with income interaction & distance interaction
  - access time (log-linear)
    - with distance interaction
  - flight time (log-linear)
  - aircraft type
  - on time performance
    - with distance interaction
  - schedule delay (SDL log-linear for business)
    - with distance interaction
  - airline allegiance (FF & *preferred* airline)
  - airport allegiance

Results 2

- MMNL models:
  - Business: LL = -338.35 vs -395.14 (6 df)
  - Holiday: LL = -468.97 vs -532.42 (7 df)
  - VFR: LL = -757.57 vs -829.732 (7 df)
- Random variation greatest for access time and air fare
- Normal distribution used …

WTP 1 ($)

<table>
<tr><th></th><th>Business</th><th>Holiday</th><th>VFR</th></tr>
<tr><td>Reduction in access time (1 hour)</td><td>75.40</td><td>35.80</td><td>35.48</td></tr>
<tr><td>Reduction in SDE (1 hour)</td><td>13.27</td><td rowspan="2">2.61</td><td>3.68</td></tr>
<tr><td>Reduction in SDL (1 hour)</td><td>11.08</td><td>2.25</td></tr>
<tr><td>On time (+10%)</td><td>10.39</td><td>7.02</td><td>5.57</td></tr>
</table>

WTP 2 ($)

<table>
<tr><th></th><th>Business</th><th>Holiday</th><th>VFR</th></tr>
<tr><td>FF elite or elite-plus vs none</td><td>125.24</td><td rowspan="2">11.44</td><td>-</td></tr>
<tr><td>FF standard vs none</td><td>49.12</td><td>-</td></tr>
<tr><td>Top airport vs worst</td><td>83.22</td><td>53.97</td><td>55.73</td></tr>
<tr><td>2nd airport vs worst</td><td>30.56</td><td>41.42</td><td>54.63</td></tr>
<tr><td>3rd airport vs worst</td><td>-</td><td>18.54</td><td>25.89</td></tr>
<tr><td>Airport closes to home</td><td>-</td><td>-</td><td>28.02</td></tr>
<tr><td>Top airline vs worst</td><td>-</td><td>25.07</td><td>21.06</td></tr>
<tr><td>2nd airline vs worst</td><td>-</td><td>18.16</td><td>15.27</td></tr>
<tr><td>3rd airline vs worst</td><td>-</td><td>20.09</td><td>4.77</td></tr>
</table>

WTP 3 ($)

	Business	Holiday	VFR
No connection vs one connection	44.15	19.60	18.98
No connection vs two connections		62.21	

RP/SP MODELLING APPLICATION

Combined RP/SP model I

- SP data has advantages in terms of data quality
- RP data has advantages in terms of *response* quality
- Combined RP/SP models
- Need to account for differences in variance of error term (scale)

Combined RP/SP model II

- Two completely unrelated data sets
  - Issues with compatibility
- Also: only some of the coefficients shared
  - Use additional rescaling parameter
- Shared coefficients:
  - Access cost & access time
  - Flight time & air fare

Combined RP/SP model III

- β_{RP} = marginal utility coefficients limited to RP alternatives
- β_{SP} = marginal utility coefficients limited to SP alternatives
- $\beta_{RP/SP}$ = shared coefficients
- α = additional rescaling parameter
- $U_{i(RP)} = \sum_{k=1,\ldots,K} \beta_{RP,k} \cdot x_k + \sum_{l=1,\ldots,L} \beta_{RP/SP,l} \cdot x_l$
- $U_{i(SP)} = \sum_{m=1,\ldots,M} \beta_{SP,k} \cdot x_m + \sum_{l=1,\ldots,L} \beta_{RP/SP,l} \cdot \boldsymbol{\alpha} \cdot x_l$

Combined RP/SP model IV

- All effects significant in combined models
- Adjusted ρ^2

	Resident business	Resident leisure
RP only	0.3413	0.3529
SP only	0.5185	0.5669
RP/SP	0.3427	0.3610

FORECASTING APPLICATION

Forecasting application I

- Changes in market shares following:
  - Decrease in air fares at STN by 20%
  - Decrease in PT access time at STN by 20%

STN fares down by 20%

RESIDENT BUSINESS

	LCY	LGW	LHR	LTN	STN
Original	0.0129	0.2723	0.5965	0.0176	0.0958
New	0.0123	0.2655	0.5860	0.0169	0.1143
Change	**-4.10%**	**-2.51%**	**-1.77%**	**-3.83%**	**+19.40%**

RESIDENT LEISURE

	LCY	LGW	LHR	LTN	STN
Original	0.0057	0.5303	0.3663	0.0149	0.0778
New	0.0044	0.5133	0.3557	0.0138	0.1076
Change	**-21.84%**	**-3.20%**	**-2.88%**	**-7.05%**	**+38.27%**

STN PT access time down by 20% (A)

RESIDENT BUSINESS

	LCY	LGW	LHR	LTN	STN
Original	0.0129	0.2723	0.5965	0.0176	0.0958
New	0.0128	0.2717	0.5956	0.0176	0.0973
Change	**-0.31%**	**-0.20%**	**-0.16%**	**-0.22%**	**+1.64%**

RESIDENT LEISURE

	LCY	LGW	LHR	LTN	STN
Original	0.0057	0.5303	0.3663	0.0149	0.0778
New	0.0056	0.5297	0.3659	0.0148	0.0788
Change	**-0.62%**	**-0.12%**	**-0.10%**	**-0.14%**	**+1.34%**

STN PT access time down by 20% (B)

RESIDENT BUSINESS AT STN

	CAR	RENTAL	LDC	MINICAB	PT	TAXI
Original	0.0623	0.0033	0.0017	0.0095	0.0077	0.0113
New	0.0618	0.0033	0.0017	0.0094	0.0100	0.0111
Change	**-0.89%**	**-0.66%**	**-0.77%**	**-0.97%**	**+30.65%**	**-0.97%**

RESIDENT LEISURE AT STN

	LCY	LGW	LHR	LTN	STN	TAXI
Original	0.0591	0.0002	0.0017	0.0034	0.0092	0.0042
New	0.0584	0.0002	0.0017	0.0034	0.0110	0.0042
Change	**-1.13%**	**-0.97%**	**-1.03%**	**-1.22%**	**+19.88%**	**-1.22%**

CONCLUSIONS

Conclusions

- Problems with fare and availability data in RP studies
- SP data poses fewer problems
  - questions about reliability
- Solution: use combined RP/SP
  - issues with compatibility remain

Interpretation of results

- High sensitivity to access time
  - → Explains success of centrally located airports, especially for business travellers (LCY)
  - → New airports need good ground level access
- Outlying airports
  - → Attract passengers only with good access
  - → Without good access: only in return for low fares (STN)

Competition in London

- High competition between London airports
- Helped by product differentiation in conjunction with route overlap
- Government decision:
  - expand Stansted
  - option to build short runway at Heathrow
- Question: will Stansted become *new* Heathrow
  - could reduce product differentiation (costs of STN development could push up fares)
  - lower competition

Questions