Advanced discrete choice models with applications to transport demand

Stephane Hess
Imperial College London

The Expanding Sphere of Travel Behaviour Research

11th International Conference on Travel Behaviour Research
Kyoto, August 16th - 20th 2006

Topics

- Model specification
- Model estimation
- Interpretation of results
- Use of advanced models in practice

Topics

- Model specification
- Model estimation $\Rightarrow$ strongly related
- Interpretation of results
- Use of advanced models in practice

Model specification I: correlation structure

- Standard models assume covariance homogeneity
- Mixed covariance structures

$$P_n(i) = \int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\lambda}) \, f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda}$$

- Allowing for covariance heterogeneity
  - Improvements in model fit
  - Differences in forecasts
  - Differences in trade-offs

Mixed Covariance example

	MNL	NL	MCOV_U	MCOV_N
Null LL	-3517.56	-3517.56	-3517.56	-3517.56
Final LL	-2329.07	-2320.23	-2191.42	-2189.8
Par	14	15	16	16
Adj.ρ^2	0.3339	0.3361	0.3725	0.3729
VTTS_{car}	54.64	46.03	42.43	41.96
VTTS_{PT}	23.64	27.25	30.13	30.94

Changes in market shares after 20% inc. in base TT

	MNL	NL	MCOV_U	MCOV_N
base alt.	-13.33%	-18.54%	-17.96%	-18.51%
early dep.	30.72%	46.17%	51.01%	53.36%
late dep.	31.27%	47.51%	39.22%	40.39%
PT	26.27%	17.81%	15.62%	14.93%

Model specification II: taste heterogeneity

- Continuous *vs* discrete mixtures

$$P(i) = \int_{\beta} \left[P\left(i \mid \beta\right) f\left(\beta, \Omega\right) \right] \mathrm{d}\beta$$

  - specification issue: choice of distributions

$$P(i) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P\left(i \mid \beta = \langle \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle \right) \pi_1^{j_1} \ldots \pi_K^{j_K}$$

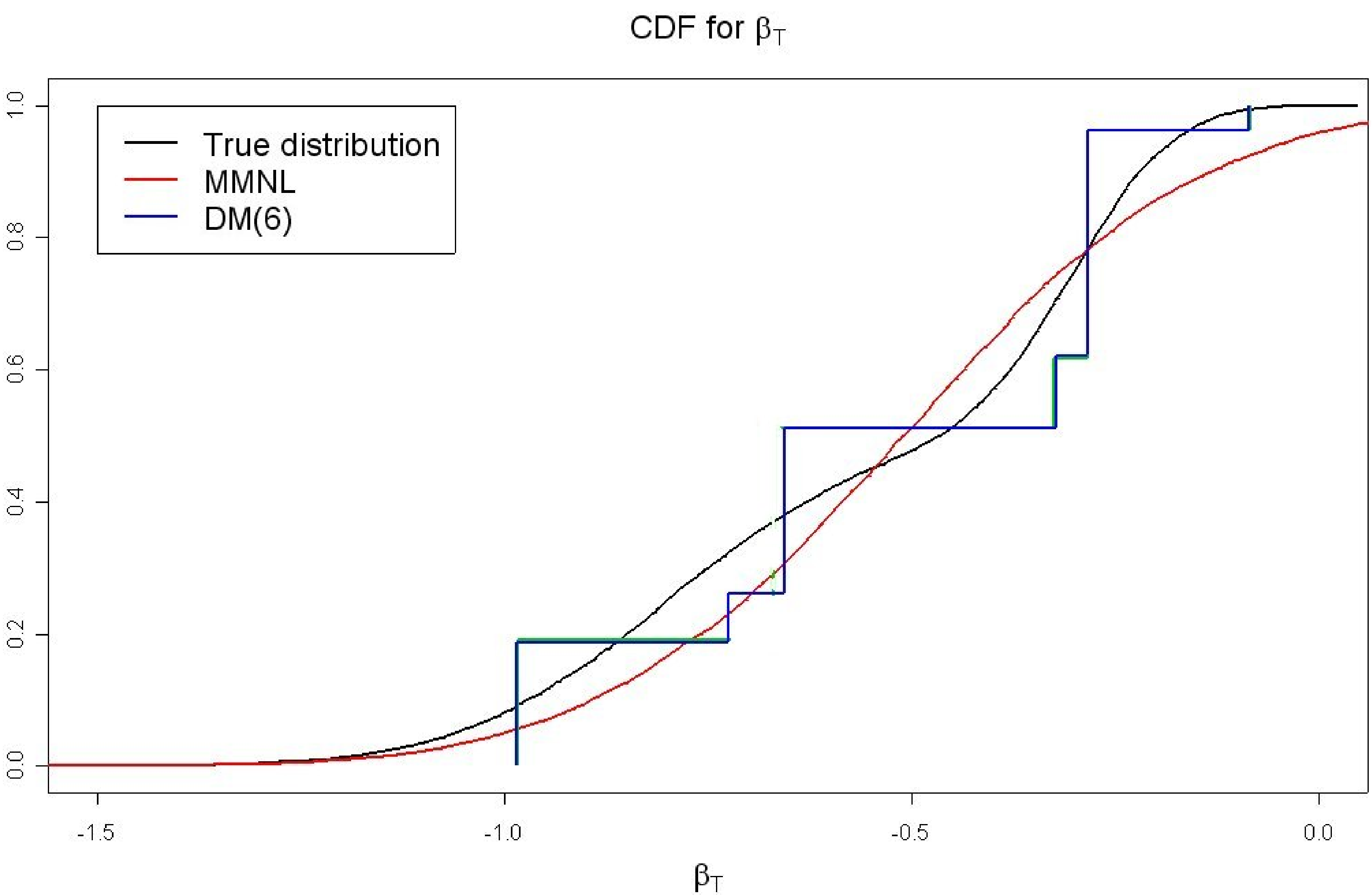
  - specification issue: selection of number of support points

Taste heterogeneity example I

Model	MNL	MMNL(N)	DM(2)	DM(3)
Respondents:	185	185	185	185
Observations:	1,421	1,421	1,421	1,421
Final LL:	-880.96	-849.65	-845.40	-844.60
adj. ρ^2:	0.1036	0.1343	0.1366	0.1354
Estimation time (s):	1	75	1	3
Mean VTTS (DKK/hour):	19.77	30.41	32.81	34.29
VTTS standard deviation	-	33.70	36.55	41.86

Imperial College London

Taste heterogeneity example II



Model specification III: joint structures

- Correlation & random taste heterogeneity
  - phenomena taking place in unobserved part of utility
- Both processes can act at the same time
  - Solution:
    use GEV mixture models, or joint RCL-ECL formulation
- Majority of applications account only for one of two phenomena
  - Big risk of confounding
    - wrongly retrieved taste heterogeneity: issues in CBA, etc
    - wrongly retrieved correlation: issues in forecasting

Confounding example

	True model	MNL	NL (rail-SM)	NL (rail-car)	RCL	NL mixture
Final LL		-1562.29	-1560.55	-1525.55	-1163.73	-1147.78
adj. $\rho^2(0)$		0.5239	0.5241	0.5347	0.6433	0.6478

	True model	MNL est.	MNL t-stat.	NL (rail-SM) est.	NL (rail-SM) t-stat.	NL (rail-car) est.	NL (rail-car) t-stat.	RCL est.	RCL t-stat.	NL mixture est.	NL mixture t-stat.
β_{TC} (μ)	-0.100	-0.035	-20.14	-0.035	-19.99	-0.028	-16.23	-0.152	-8.75	-0.093	-14.72
β_{TC} (σ)	0.035	-	-	-	-	-	-	0.061	8.64	0.037	14.38
β_{HW} (μ)	-0.020	-0.023	-14.16	-0.022	-10.98	-0.014	-9.91	-0.034	-10.05	-0.018	-9.74
β_{HW} (σ)	-	-	-	-	-	-	-	0.009	1.97	0.001	0.49
$\beta_{TT,car}$ (μ)	-0.030	-0.021	-14.15	-0.020	-13.51	-0.015	-11.83	-0.054	-7.78	-0.028	-10.05
$\beta_{TT,car}$ (σ)	-	-	-	-	-	-	-	0.016	4.10	0.002	0.73
$\beta_{TT,rail}$ (μ)	-0.040	-0.032	-20.79	-0.032	-15.72	-0.022	-13.63	-0.065	-12.13	-0.037	-12.85
$\beta_{TT,rail}$ (σ)	-	-	-	-	-	-	-	0.002	2.14	0.001	1.05
$\beta_{TT,SM}$ (μ)	-0.035	-0.022	-10.13	-0.022	-10.50	-0.013	-6.28	-0.059	-9.83	-0.034	-11.61
$\beta_{TT,SM}$ (σ)	-	-	-	-	-	-	-	0.003	1.17	0.001	0.56
$\lambda_{rail,SM}$	0.50	1.00	-	0.89	1.22	1.00	-	1.00	-	0.48	4.59
$\lambda_{rail,car}$	1.00	1.00	-	1.00	-	0.49	5.92	1.00	-	1.00	-

Imperial College London

Model estimation I: type of draws

- Simulation-based estimation of continuous mixture models
- Various alternatives to standard PMC draws
- Halton draws:
  - Problems even with only $4 - 5$ dimensions
  - Also problems with various adapted versions
- MLHS
  - outperforms Halton sequences, but high variance
- Advanced approaches
  - difficult to implement, lack of available code
  - sometimes no guarantee of better performance

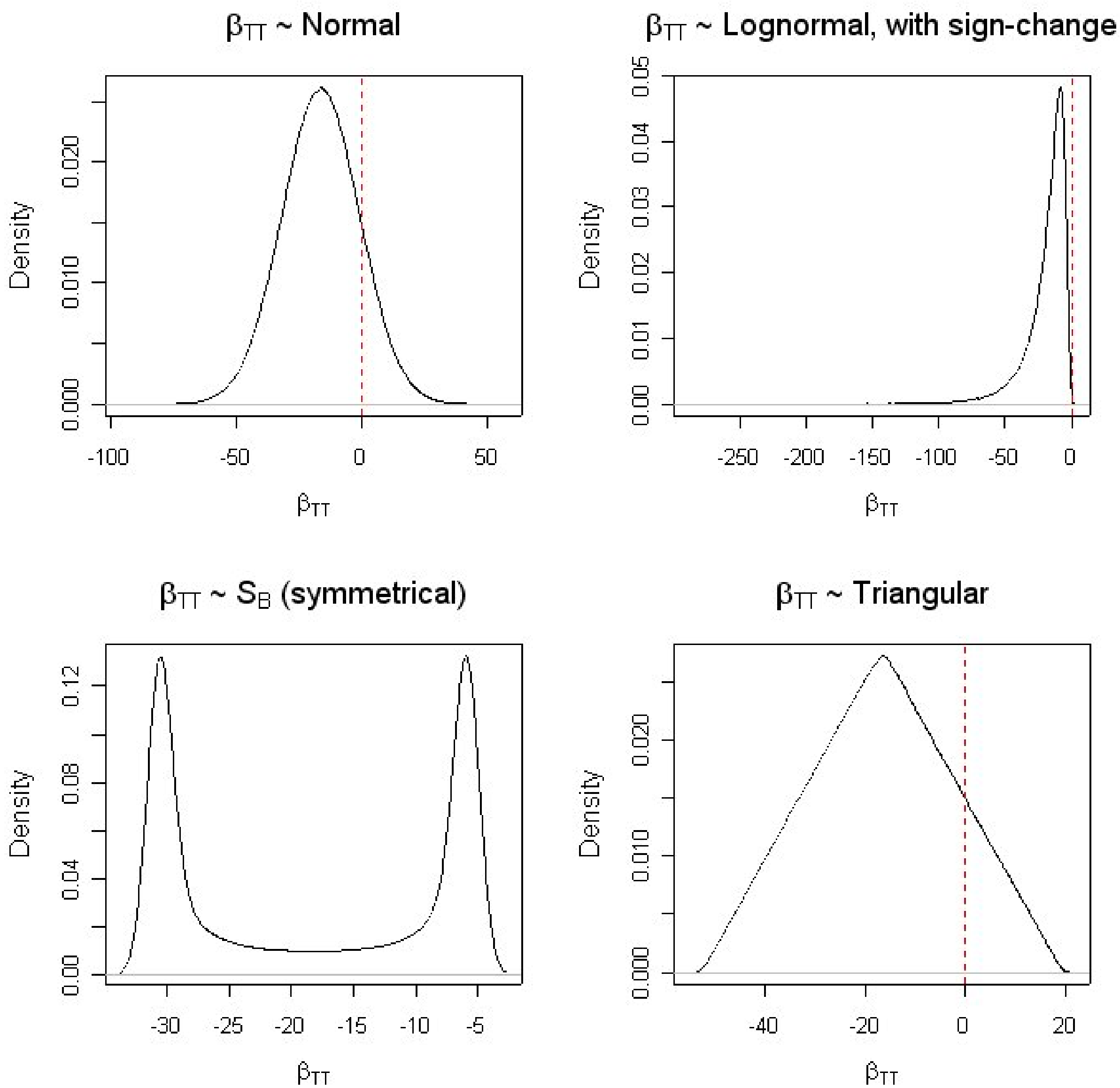
Model estimation II: number of draws

- Number of draws has very significant impacts on results
- Guidelines in literature
  - Worrying number of applications using 125 Halton draws, in up to 8 dimensions!
- Other approaches for reducing estimation cost
  - MNL starting values
  - Pre-estimation with lower number of draws
- Advisable to use multiple runs with different draws and starting values

Model interpretation

- Model specification has effects on model results
  - important to recognise in model interpretation
- Example: choice of distributions in MMNL
  - shape assumptions impact model results
- Application: Danish VOT data
  - compare models with different distributional assumptions

Model interpretation example



Imperial College London

Model interpretation example (continued)

- Significant differences in implied distributions

	Normal	Lognormal	S_B	Triangular
adj. ρ^2	0.1484	0.1518	0.1512	0.1486
% positive	14.53%	0%	0%	14.6%

- Differences in model fit very small

- Guidance: use flexible distributions

Theoretical rules *vs* practical considerations

- Applied part...
- Modelling air travel choice behaviour
- Joint choice of airport, airline and access mode
  - departure from state-of-the-art
  - major gains in model performance
- Three datasets
  - RP data (SF-bay area & Greater London)
  - SP data (US, internet-based)

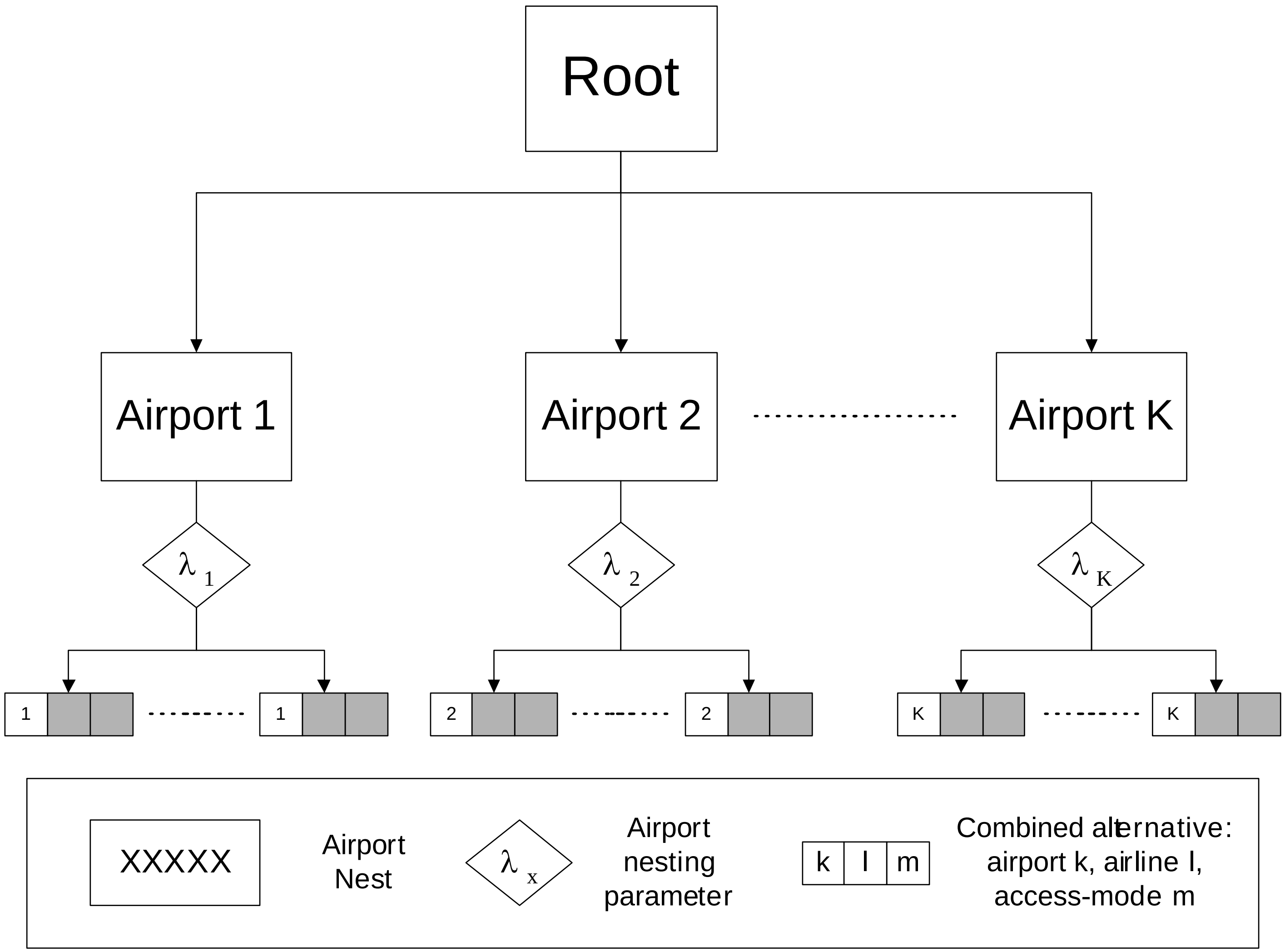
Random taste heterogeneity

- Some gains in model fit, and insights into behaviour
- Huge increases in estimation cost
- Use of Normal almost unavoidable
- Use of high number of draws almost impossible
- London data ($> 10,000$ observations, 324 alternatives)
  - basic MMNL models take in excess of 1 month to estimate
- Flexible interactions and non-linearities massively reduce scope for retrieving random taste heterogeneity

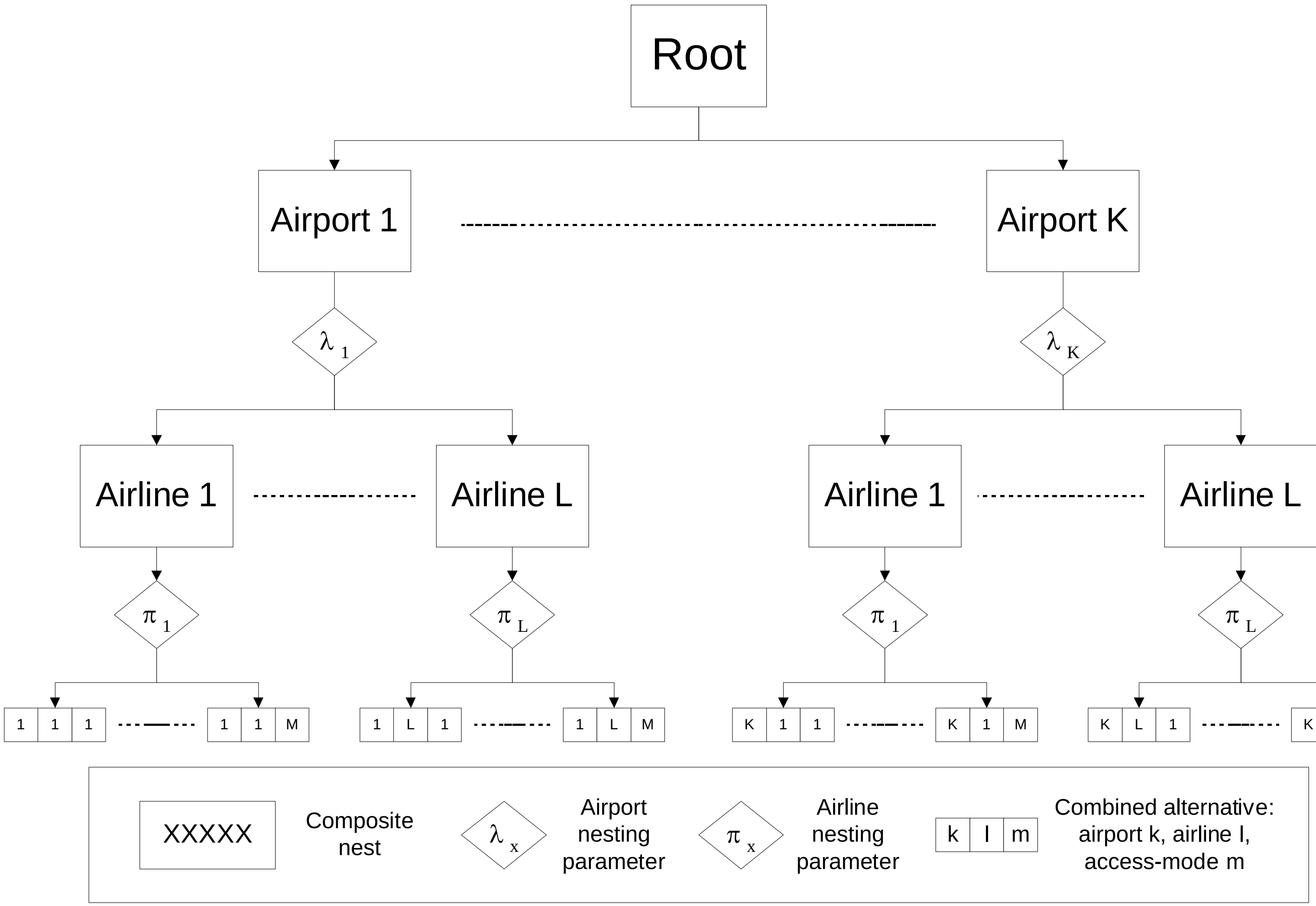
Correlation structure

- Shortcomings of NL model for multi-dimensional choice processes

2-level NL

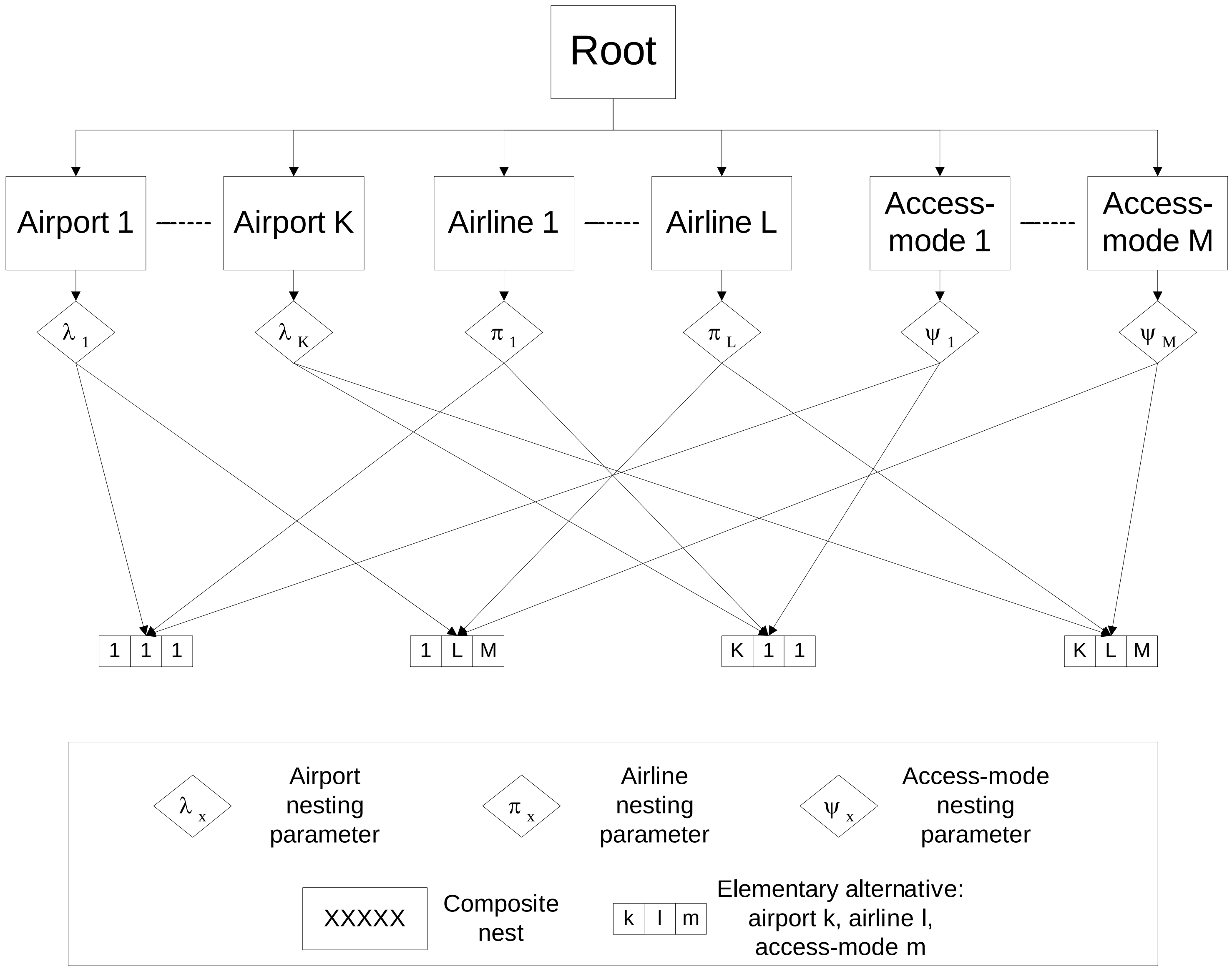

Root
Airport 1
Airport 2
Airport K
λ 1
λ 2
λ K
1
1
2
2
K
K
XXXXX
Airport Nest
λ x
Airport nesting parameter
k
l
m
Combined alternative: airport k, airline l, access-mode m

3-level NL



Imperial College London

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 1
1 L M
K 1 1
K L 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

Correlation structure

- Shortcomings of NL model for multi-dimensional choice processes
- Need cross-nesting structure

	Par	*Adj.* ρ^2	est.time
MNL	55	0.3445	minutes
NL airport	59	0.3465	hours
NL airline	74	0.3469	hours
NL access	60	0.3499	hours
CNL	91	0.3578	weeks

Advanced structures

- GEV mixtures
- Covariance heterogeneity
- Essentially impossible to use
  - Potential confounding and other mis-specification effects

Conclusions

- Major issues in specification, estimation and interpretation
- Certain guidelines
- Often need to be violated in practice

Discussion

- State-of-the-art has moved forward at great speed
- State-of-practice is trailing behind
- Advanced models often almost inapplicable
  - Estimation and application cost
  - Data requirements
- Often rather small gains in performance
  - log-likelihood is not everything
- Mixture models over-hyped
- Need to educate, and sell our models better

Thank you ...

Imperial College London