ADVANCED DISCRETE CHOICE MODELS WITH APPLICATIONS TO TRANSPORT DEMAND

Stephane Hess

September 2005

A thesis submitted as fulfilment
of the requirements for the degree of
Doctor of Philosophy
of the University of London and for the
Diploma of Membership of Imperial College

Centre for Transport Studies
Imperial College London

Abstract

The area of discrete choice modelling has, over recent years, witnessed the development of ever more flexible model structures that allow for an increasingly realistic representation of travel behaviour.
With these developments have also come important issues of specification, estimation and interpretation, some of which are addressed in this thesis, mainly in the context of models allowing for random taste heterogeneity across respondents. As such, it is shown that severe risks of misinterpretation arise when relying on the commonly used Normal distribution, and the advantages of several alternative distributions are illustrated, while also discussing the benefits of a discrete mixture approach. The thesis also highlights risks of confounding between different components of the error structure, and discusses the development of approaches that can lead to computational savings in simulation-based model estimation and application. Finally, a framework is developed for the representation of random variations in a model's covariance structure.
With the pace of theoretical developments, the gap between theory and practice in the use of discrete choice models has widened. The applied part of the thesis aims to partly bridge this gap in one area of travel-behaviour research, looking at the modelling of choices made by air-passengers departing from multi-airport regions, with applications to Greater London and the San Francisco Bay area. The case-studies show the benefits of using advanced model approaches, in this case cross-nesting and random coefficients structures. At the same time however, the work shows that the appeal of such models in large-scale analyses is reduced by heightened data requirements, and the significant rise in estimation cost. Finally, the case-studies show that, while the issues discussed in the theoretical part of the thesis need to be taken into account in interpretation, for practical purposes, the guidelines in terms of specification often need to be violated.

Acknowledgements

The number of people who deserve to be mentioned here is so large that I am bound to forget some. Any omissions are not deliberate, and their support, whatever form it took, will always be gratefully remembered ...

Of those people I have remembered to thank here, the first will have to be my supervisor, John Polak, for all the advice he's given me over the course of my PhD. John has been a constant source of inspiration and guidance, from the very first week all the way to the final writing-up. I couldn't possibly have wished for a better and more dedicated supervisor.

I would also like than my two examiners, Staffan Algers and Benjamin Heydecker, for valuable discussions which have provided me with a number of ideas for how the work presented in this thesis can be developed further.

During my research I have had the almost unique opportunity of working with some of the leading players in the area of discrete choice modelling. I am immensely grateful for the contributions of Kay Axhausen, Michel Bierlaire, Denis Bolduc, Andrew Daly, and, last but not least, Kenneth Train. Andrew and Michel deserve some additional recognition for their endless support in helping me use ALogit and BIOGEME for my research.

I am similarly grateful to Fabian Bastin, Bryn Battersby, Gregory Coldren, Nigel Dennis, Laurie Garrow, and Hugh Gunn for helpful discussions, and useful feedback on my research.

My research has also benefited from feedback I received after seminars; here, I am grateful for the invitations from Richard Batley, Moshe Ben-Akiva, Mogens Fosgerau, Rodrigo Garrido, Sergio Jara-Diaz, Frank Koppelman, Marcela Munizaga, Juan de Dios Ortúzar, and Riccardo Scarpa, which allowed me to present my work to very different sets of audiences; heterogeneity is good!

I would also like to thank my sponsors, the Rees Jeffreys Road Fund, for their financial support. Similarly, I would like to express my thanks to RAND Europe, for providing me with interesting and challenging work, giving me a taste of *real-world* transport modelling.

A number of people have provided invaluable support in helping in the procurement of the data used in this thesis, especially so, but not limited to, in the context of the air-travel work. Here, I would like to thank Georg Abay, Stan Abrahams, Tom Adler, Mark Judd, Josh Lovegrove, Phil Osler, Chuck Purvis, Trevor Smedley, John Weber, and most of all, Washington Ochieng.

I have also benefited from discussions with other members of CTS, most notably

Elisabetta Cherchi, Charles Lindveld, and Bob Noland. Here, I would also like to thank Moazzam Ishaque, for being the perfect office colleague, never once complaining about my loud outbursts of frustration when the CNL models just wouldn't converge. Thanks also to Steve Robinson, for looking after my model estimations for me while I was off on my travels. I have also enjoyed the company of my fellow PhD researchers; Robin North, Mohammed Quddus, and all the others. Additionally, I would like to thank Jackie Sime for all her help; much more than just a section secretary, she's the one who runs the place ...

Finally, the biggest thanks of all will have to go to my parents. I owe it all to them; they gave me a hunger for learning, encouraged me to embark on this mad journey that's called a PhD, and gave me invaluable moral support throughout. I also want to thank my two brothers, Cedric and Philippe, and especially my sister, Sarah, for their encouragements throughout.

Stephane Hess
August 2005

Contents

Abstract 1

Acknowledgements 2

Contents 4

List of Tables 9

List of Figures 11

1 Introduction 12
- 1.1 Background . . . 12
- 1.2 Aims . . . 13
- 1.3 Outline of the thesis . . . 15
- 1.4 Contributions . . . 16

2 Existing model structures 17
- 2.1 Introduction . . . 17
- 2.2 Basic concepts . . . 17
- 2.3 Logit and Probit models . . . 20
  - 2.3.1 The Multinomial Logit model . . . 20
  - 2.3.2 The Multinomial Probit model . . . 21
- 2.4 The GEV family and the Nested Logit model . . . 22
  - 2.4.1 GEV choice probabilities . . . 23
  - 2.4.2 The Nested Logit model . . . 24
- 2.5 Models allowing for cross-nesting . . . 26
  - 2.5.1 The Generalised Nested Logit model . . . 27
  - 2.5.2 The Cross-Nested Logit model . . . 28
  - 2.5.3 The Link-Nested Logit model . . . 30
  - 2.5.4 The Paired Combinatorial Logit model . . . 30
  - 2.5.5 The Paired GNL model . . . 31
  - 2.5.6 The Ordered Generalised Extreme Value model . . . 31
  - 2.5.7 The Principles of Differentiation model . . . 31
- 2.6 Multi-level GEV models . . . 32
- 2.7 Recursive GEV models . . . 33
  - 2.7.1 Recursive NL models . . . 33
  - 2.7.2 General recursive GEV models . . . 34
- 2.8 Some alternative model forms . . . 35

2.8.1 Models with heteroscedastic error-terms 35
2.8.2 The Universal Logit family 36
2.8.3 Choice set generation models 37
2.9 Models using mixture distributions 38
2.9.1 The Mixed Multinomial Logit model 38
2.9.2 GEV mixture models 40
2.10 Discussion 41

3 Efficiency in simulation-based estimation and application 43
3.1 Introduction and context 43
3.2 Simulation processes in estimation and application 44
3.3 Alternatives to standard Monte Carlo integration 48
3.4 Halton sequences 49
3.5 Adapted versions of Halton sequences 56
3.5.1 Scrambled Halton sequences 56
3.5.2 Shuffled Halton sequences 57
3.6 Modified Latin Hypercube sampling 59
3.7 Empirical comparisons 63
3.7.1 Application A 64
3.7.2 Application B 68
3.8 Alternative QMC approaches 70
3.9 Summary and Conclusions 71

4 Specification and interpretation of random coefficients models 74
4.1 Introduction and context 74
4.2 Choice of distribution: the state of practice 75
4.3 VTTS case-study 76
4.3.1 Introduction 76
4.3.2 Data and model specification 79
4.3.3 Estimation results 81
4.3.4 Summary of findings 87
4.4 Interpretation of counter-intuitively signed coefficients 89
4.5 Summary and Conclusions 92

5 Discrete mixture models 95
5.1 Introduction and context 95
5.2 Methodology 98
5.3 Testing the validity of the discrete mixture structure 101
5.4 VTTS case-study 102
5.5 Summary and Conclusions 107

6 Confounding between substitution patterns and random taste heterogeneity 109
6.1 Introduction and context 109
6.2 Case-studies 111
6.2.1 True model: NL with two nests 112
6.2.2 True model: CNL with two nests 114
6.2.3 True model: RCL with single random coefficient 116

6.2.4 True model: RCL with single random coefficient shared by all three alternatives . . . 117
6.2.5 True model: RCL with two random coefficients . . . 118
6.2.6 True model: NL mixture . . . 120
6.3 Impact of confounding on model forecasts . . . 121
6.4 Summary and Conclusions . . . 126

7 Mixed Covariance models **128**
7.1 Introduction and context . . . 128
7.2 Methodology . . . 129
7.2.1 General model form . . . 129
7.2.2 Model with purely random covariance heterogeneity . . . 132
7.2.3 Model with deterministic and random covariance heterogeneity 133
7.3 Application . . . 135
7.4 Summary and Conclusions . . . 142

8 Modelling air-travel choice-behaviour **144**
8.1 Introduction . . . 144
8.2 Air-travel choice-behaviour . . . 146
8.2.1 General framework . . . 146
8.2.2 Dimensions of choice in air-travel . . . 146
8.2.3 Choice processes in air-travel . . . 149
8.2.4 Discussion . . . 150
8.3 Literature review . . . 151
8.3.1 Introduction . . . 151
8.3.2 Air as a modal alternative . . . 151
8.3.3 Airport choice . . . 152
8.3.4 Airline and fare-class choice . . . 154
8.3.5 Access-mode choice . . . 155
8.3.6 SP data . . . 155
8.3.7 Other air-travel applications . . . 157
8.3.8 Summary . . . 157
8.4 Overview of research . . . 158
8.4.1 Scope of applications . . . 158
8.4.2 Aims of research . . . 161
8.4.3 Model structure . . . 163
8.4.4 Issues . . . 168
8.5 Summary . . . 170

9 San Francisco Bay area case-study **172**
9.1 Introduction and context . . . 172
9.2 Description of data . . . 175
9.2.1 Air-passenger survey data . . . 175
9.2.2 Air-travel level-of-service data . . . 177
9.2.3 Ground-access level-of-service data . . . 177
9.3 Choice of airport . . . 178
9.3.1 Model specification and estimation . . . 178
9.3.2 Model prediction performance . . . 180

9.3.3 Summary of findings for airport choice study . . . 181
9.4 Multi-dimensional choice process . . . 181
9.4.1 Structure of choice set . . . 182
9.4.2 Base specification of utility function . . . 182
9.4.3 MNL models . . . 184
9.4.4 Nesting structures . . . 187
9.4.5 Mixture structures . . . 192
9.4.6 Comparison of substantive results . . . 196
9.4.7 Model validation . . . 199
9.5 Summary and Conclusions . . . 201

10 Greater London case-study 203
10.1 Introduction and context . . . 203
10.2 Description of data . . . 208
10.2.1 Air-passenger survey data . . . 208
10.2.2 Air-travel level-of-service data . . . 209
10.2.3 Ground-access level-of-service data . . . 210
10.3 Model structure and specification . . . 211
10.3.1 Choice set . . . 211
10.3.2 Re-weighting of survey data . . . 212
10.3.3 Specification of constants . . . 212
10.3.4 Non-linearities . . . 213
10.4 Model estimation . . . 213
10.4.1 Preliminary data analysis . . . 213
10.4.2 Utility specification . . . 215
10.4.3 MNL models . . . 216
10.4.4 NL models . . . 216
10.4.5 CNL models . . . 221
10.4.6 Comparison of substantive results . . . 223
10.5 Model validation . . . 230
10.6 Summary and Conclusions . . . 232

11 Stated preference case-study of airport and airline choice 234
11.1 Introduction and context . . . 234
11.2 Description of data . . . 235
11.3 Model specification . . . 237
11.3.1 Explanatory variables used in specification search . . . 237
11.3.2 Non-linearities . . . 240
11.3.3 Continuous interactions . . . 240
11.4 MNL models . . . 242
11.4.1 MNL model for business travellers . . . 242
11.4.2 MNL model for holiday travellers . . . 244
11.4.3 MNL model for VFR travellers . . . 245
11.4.4 Comparison of results across population segments . . . 247
11.5 Extension to MMNL models . . . 253
11.5.1 MMNL model for business travellers . . . 254
11.5.2 MMNL model for holiday travellers . . . 255

11.5.3 MMNL model for VFR travellers 257
11.5.4 Conclusions for MMNL analysis 258
11.6 Summary and Conclusions 259

12 Summary, conclusions and directions for future research 261
12.1 Summary 261
12.1.1 Theoretical part 261
12.1.2 Applied part 263
12.2 Directions for future research 266
12.2.1 Discrete choice theory 266
12.2.2 Air-travel behaviour research 266
12.3 Conclusions 267

References 269

A Confounding between substitution patterns and random taste heterogeneity: ECL results 290
A.1 Introduction 290
A.2 Estimation results 290
A.3 Forecasting analysis 292

B Covariance heterogeneity: Development of ECL approach 295
B.1 General ECL formulation 295
B.2 Deterministic covariance heterogeneity in ECL models 296
B.3 Purely random covariance heterogeneity in ECL models 297
B.4 Deterministic and random covariance heterogeneity in ECL models . 297
B.5 Discussion 298

C Frequently Used Acronyms 299

D Publications 301

List of Tables

3.1 Estimation results on vehicle-type choice data 65
3.2 Estimation performance on vehicle-type choice data 67
3.3 Estimation performance on *Mobidrive* data 69

4.1 Estimation results on Danish VOT data: part I 82
4.2 Estimation results on Danish VOT data: part II 83

5.1 Generation of synthetic data for discrete mixture experiments 101
5.2 Discrete mixture results: first synthetic data experiment 102
5.3 Discrete mixture results: second synthetic data experiment 103
5.4 VTTS case-study: MNL estimation results 103
5.5 VTTS case-study: results with constrained discrete mixture model . . 104
5.6 VTTS case-study: results with unconstrained discrete mixture model 106

6.1 Estimation results: first confounding case-study 113
6.2 Estimation results: second confounding case-study 115
6.3 Estimation results: third confounding case-study 116
6.4 Estimation results: fourth confounding case-study 118
6.5 Estimation results: fifth confounding case-study 119
6.6 Estimation results: sixth confounding case-study 120
6.7 Forecasting exercise: first confounding case-study 122
6.8 Forecasting exercise: fourth confounding case-study 123
6.9 Forecasting exercise: sixth confounding case-study 125

7.1 Estimation results on mixed covariance data 139
7.2 Forecasting on mixed covariance data, part I 140
7.3 Forecasting on mixed covariance data, part II 141

9.1 Annual passenger counts for three main SF-bay area airports 174
9.2 Forecasts of passenger levels in the SF-bay area 174
9.3 Summary of choice data for SF-bay area case-study 176
9.4 Estimation results for airport choice models in the SF-bay area 180
9.5 Prediction performance of MMNL models for airport choice 181
9.6 MNL results for travellers resident in the SF-bay area 184
9.7 MNL results for travellers not resident in the SF-bay area 185
9.8 NL results using nesting by airport on SF-bay data 188
9.9 NL results using nesting by airline on SF-bay data 189
9.10 NL results using nesting by access-mode on SF-bay data 190
9.11 MMNL results for travellers resident in the SF-bay area 193

9.12 MMNL results for travellers not resident in the SF-bay area 194
9.13 Trade-offs between flight frequency and access time 198
9.14 Prediction performance of NL models on SF-bay area validation data 200

10.1 Annual passenger counts at London's five main airports 205
10.2 Passenger counts in survey data used in London study 210
10.3 Stated main reason for choice of airport in London survey data 214
10.4 MNL results for London data 216
10.5 NL results for London data, using nesting by airport 217
10.6 NL results for London data, using nesting by airline 219
10.7 NL results for London data, using nesting by access-mode 220
10.8 Structural constraints required in NL and CNL models 221
10.9 CNL estimation results for London data, part I 222
10.10 CNL estimation results for London data, part II 224
10.11 Trade-off between in-vehicle access time and access cost 225
10.12 Trade-off between flight time and in-vehicle access time 226
10.13 Trade-off between frequency and in-vehicle access time 226
10.14 Prediction performance of London models on validation samples . . . 231
10.15 Prediction performance of London models on estimation samples . . . 232

11.1 SP case-study: MNL results for business travellers 242
11.2 SP case-study: MNL results for holiday travellers 244
11.3 SP case-study: MNL results for VFR travellers 246
11.4 SP case-study: MNL trade-offs, part I 250
11.5 SP case-study: MNL trade-offs, part II 251
11.6 SP case-study: effects of continuous interactions 252
11.7 SP case-study: MMNL results for business travellers 255
11.8 SP case-study: MMNL results for holiday travellers 256
11.9 SP case-study: MMNL results for VFR travellers 257

A.1 ECL estimation results: confounding case-studies 291
A.2 ECL forecasting exercise: first confounding case-study 292
A.3 ECL forecasting exercise: fourth confounding case-study 293
A.4 ECL forecasting exercise: sixth confounding case-study 294

List of Figures

3.1 Illustration of two-dimensional filling-in effect of Halton sequences . . 52
3.2 Two-dimensional Halton sequences for different prime numbers 53
3.3 Two-dimensional pseudo-random number sequences 54
3.4 Correlation in two-dimensional Halton sequences 55
3.5 Correlation in two-dimensional scrambled Halton sequences 57
3.6 Two-dimensional shuffled Halton sequences 59

4.1 MMNL results based on Normal and Lognormal distributions 84
4.2 MMNL results based on Johnson S_B distribution 85
4.3 MMNL results based on Johnson S_U and Gamma distributions 86
4.4 MMNL results based on Triangular and Exponential distributions . . 87
4.5 MMNL results based on Logistic and Weibull distributions 88

8.1 Two-level NL model: nesting along airport dimension 164
8.2 Three-level NL model: nesting along airport and airline dimensions . 165
8.3 CNL model for airport, airline and access-mode choice 166

9.1 Map of San Francisco Bay area 173

10.1 Map of Greater London 204

11.1 Example screen-shot for SP survey: third choice-situation 237
11.2 Example screen-shot for SP survey: ninth choice-situation 238
11.3 Example screen-shot for SP survey: tenth choice-situation 239

Chapter 1

Introduction

1.1 Background

Discrete choice models belonging to the family of Random Utility Models (RUM) have been used extensively in the area of travel behaviour research for over thirty years[1]. For many years after the initial methodological developments, the high cost of estimating advanced models meant that most applications, even in an *academic* context, were limited to the use of the most basic model structures, such as Multinomial and Nested Logit. Over the past ten years however, gains in computing power as well as improvements in estimation techniques have led to the increased use of advanced nesting structures, and more recently, models based on mixture distributions, such as Mixed Logit (MMNL). Aside from allowing for the increased exploitation of such structures that had previously been confined mostly to theoretical discussions[2], these gains in estimation capability have also spurred new developments, for example in the form of advanced mixture models.

On the basis of this evolution in the state-of-the-art, modellers now have tools at their disposal that allow for the representation of complex inter-alternative substitution patterns, deterministic and random variations in tastes across respondents, correlation across observations, and differences in the error-terms across alternatives as well as respondents. However, with these gains in flexibility have also come a number of issues relating to model specification and interpretation. While also applying to some extent in the case of advanced nesting structures, these issues arise especially in the case of mixture models, such as MMNL. Indeed, as noted by Hensher & Greene (2003, page 133), *"the learning curve is steep and the unwary are likely to fall into a chasm if not careful. These chasms are very deep indeed given the complexity of the mixed logit model."* As such, it can be seen that, especially in the context of the representation of random taste heterogeneity, the assumptions made during model specification have a direct influence on model results, and an inappropriate choice of mixture distribution for a given taste coefficient can lead to problems in interpretation and potentially misguided policy-decisions. Additionally, important issues of confounding may arise in the case where assumptions made

[1]See McFadden (2000) for a history of the evolution of the state-of-the-art in the area of random utility models.

[2]These model structures had been *known* for many years, but were simply inapplicable, especially in large applications.

with regards to the model's error-structure are not reflected in the data. Finally, it should be noted that, despite the gains in estimation efficiency and computing power, the cost of estimating advanced model structures remains high, and while this again applies mainly in the case of mixture models, the numerical issues faced in the estimation of models based on complex nesting structures are not trivial either.

While the above discussion is an illustration of the dramatic evolution of the state-of-the-art in discrete choice modelling, it should be noted that there have been no such fundamental changes in the state-of-practice. Indeed, although the occasional analysis makes use of a cross-nesting structure or a Mixed Logit specification, the vast majority of large-scale *real-world* applications still rely mainly on the use of Multinomial and Nested Logit. Several reasons can be identified for this. The first is that the cost of estimating and applying advanced model structures remains so high that, while acceptable in the small-scale analyses typically conducted in research studies, the size of applications in actual policy analysis, in terms of choice set and sample size, limits modellers to more basic structures. In this context, another factor is the higher cost of such models in terms of data requirements. However, it must also be noted that there is a general lack of appreciation amongst practitioners as to the potential benefits of using advanced model structures in large-scale applications, in terms of providing further insights into choice-behaviour, but also in terms of avoiding sources of bias that are specific to the more assumption-bound models. Here, researchers have an important responsibility in promoting the use of advanced model structures while at the same time also devising ways that facilitate their use in large-scale applications.

The points raised in the preceding two paragraphs, namely the issues of specification, estimation and interpretation of advanced models, and the bridging of the gap between the state-of-the-art and the state-of-practice, are the two central themes of this thesis. As such, the thesis is divided into two separate, yet interrelated parts, one theoretical, and one applied.

The first part of the thesis is concerned with the state-of-the-art, and presents several theoretical developments, aimed mainly at addressing the prevailing issues of specification, estimation and interpretation in advanced discrete choice models.

The second part of the thesis looks at the state-of-practice in one area of travel behaviour research, namely air-travel[3]. Three case-studies are conducted in this part of the thesis, looking at the joint choice of airport, airline, and access-mode for passengers departing from multi-airport regions, with Revealed Preference applications to Greater London and the San Francisco Bay area, in addition to a Stated Preference study of airport and airline choice in the United States.

1.2 Aims

As mentioned at the end of Section 1.1, the work presented in this thesis is divided into two separate parts. As such, it is useful to also present the aims of the two

[3] This choice is partly motivated by its appeal from a methodological point of view, given the complex nature of the choice processes undertaken by air-travellers. However, the choice is also based on the reasoning that, in the face of important policy decisions (cf. Chapter 8), there is a need for reliable forecasts of passenger behaviour, calling for the use of flexible modelling approaches.

parts separately, looking first at the theoretical part, and then at the applied part.

The main aims of the theoretical part of the research were:

- An analysis of the issue of the specification and interpretation of random taste heterogeneity
- Research into ways of reducing the cost of estimation of mixture models
- A study of the issue of confounding between individual components in the error-structure of the discrete choice models
- The development of a model structure allowing for random variations in inter-alternative substitution patterns across respondents

As hinted at in the discussion in Section 1.1, the aims in the theoretical part of the thesis relate to the specification, estimation, and interpretation of advanced discrete choice models. The main exception to this is the final item in the above list; here, a specific gap in the existing state-of-the-art from the point of view of model structure is addressed.

A number of aims can also be identified for the applied part of the thesis, which are described in more detail in Chapter 8. In summary, the main aims for this part of the research were:

- The development of a framework for the joint modelling of the choice of airport, airline and access-mode
- The representation of random taste heterogeneity in the context of air-travel choice behaviour
- The simultaneous analysis of correlation along multiple dimensions of choice in air-travel behaviour research, using cross-nesting structures
- The study of air-travel choice behaviour in London

While the first three aims listed above relate directly to the earlier point about bridging the gap between the state-of-the-art and the state-of-practice in air-travel behaviour research, the fourth aim is of a topical rather than methodological nature. The motivation here is partly one of addressing a geographical imbalance[4], but is also motivated by the notion that levels of competition between airports in the London area are potentially much higher than in many other multi-airport regions (cf. Chapter 10).

The research also had a number of other, more *local* aims, and these are highlighted as part of the discussions presented in the appropriate chapters.

[4]There has been a relative lack of research into air-travel choice behaviour in the UK over recent years, with a focus on other markets, such as the US.

1.3 Outline of the thesis

To give an overview of the structure of the remainder of this thesis, we will now look briefly at the contents of the individual chapters:

- **Chapter 2** presents a review of existing work in the area of discrete choice modelling, focussing on model structure.
- **Chapter 3** looks at the use of alternatives to pseudo-random draws in the simulation of the integrals representing the choice probabilities of mixture models. After a review of existing work in this area, the chapter discusses the development of the Modified Latin Hypercube Sampling (MLHS) approach.
- **Chapter 4** looks at the issues of specification and interpretation of random taste heterogeneity on the basis of continuous distributions, with an emphasis on the case of the distribution of the value of travel time savings (VTTS). The chapter includes an analysis making use of a large number of different continuous distributions, including some that have not previously received widespread exposure in the context of random coefficients modelling.
- **Chapter 5** discusses an alternative to the use of continuous distributions in random coefficients models, with the random variations accommodated with the help of discrete mixtures. The chapter also discusses the risk of biased estimates in the presence of individuals with zero valuations of changes in explanatory variables, and shows how the use of discrete mixture models can reduce the risk of such bias.
- **Chapter 6** highlights an important issue in the context of advanced discrete choice models, namely the risk of biased results due to confounding in the case of inappropriate assumptions with regards to the underlying error-structure, and shows how the use of sufficiently flexible models can help reduce the risk of such bias.
- **Chapter 7** presents the development of a model that allows for random variations in the inter-alternative substitution patterns across individuals.
- **Chapter 8** acts as the introduction to the applied part of the thesis. The chapter presents a review of existing work in the area of air-travel behaviour research, sets out the scope and aims for the three case-studies, and discusses some generic issues that need to be faced.
- **Chapter 9** presents the findings of a case-study of the combined choice of airport, airline and access-mode in the San Francisco Bay area, making use of Multinomial, Nested, and Mixed Logit structures.
- **Chapter 10** presents the findings of a case-study of the combined choice of airport, airline and access-mode in Greater London, making use of Multinomial, Nested and Cross-Nested Logit structures.

- **Chapter 11** presents the findings of a case-study of the combined choice of airport and airline using SP data[5], making use of Multinomial and Mixed Logit structures.

- **Chapter 12** provides a summary of the work discussed in this thesis, presents the conclusions, and describes some avenues for future research.

1.4 Contributions

Several contributions are made in this thesis. In addition to the in-depth discussions of important issues, such as that of the distributional assumptions in random coefficient models, and the problem of confounding between individual components of the error-structure of the *true* underlying model, these are (in order of appearance in the thesis):

- The development of the MLHS approach in Chapter 3.

- The use of discrete mixture models to allow for the presence of respondents with zero VTTS[6] in Chapter 5.

- The development of the Mixed Covariance model in Chapter 7.

- The use of a cross-nesting structure for the simultaneous analysis of the correlation along multiple dimensions of choice in air-travel in Chapter 10.

Several other, more small-scale contributions are also made, and these are discussed at appropriate stages in the remaining part of the thesis.

[5] Collected via an internet-based survey in the US.

[6] Although discrete mixture models have been used before, as described in Section 5.1, their exposure has been limited, and it seems that they have not previously been used to address the issue of zero valuations of changes in explanatory variables.

Chapter 2

Existing model structures

2.1 Introduction

The review presented in this chapter looks at existing work on model structures in the area of discrete choice analysis, while the discussion of the existing literature in relation to topics addressed in more detail in this thesis is presented in the appropriate chapters, in conjunction with the new contributions, with the aim of improving readability. As such, existing work addressing the issue of the specification of random taste heterogeneity in mixture models is not discussed in this chapter, but is presented in Chapter 4. The same applies for existing work on the use of discrete choice models in air-travel research, where an extensive review is presented in Chapter 8. Finally, the estimation of closed form models is not discussed in detail here, with the focus of the thesis being primarily on the use of mixture models, where a detailed discussion of existing work is presented in Chapter 3.

The discussion in this chapter is structured as follows. After a brief review of the main concepts of discrete choice models in Section 2.2, we look at the Multinomial Logit and Multinomial Probit models in Section 2.3. This is followed in Section 2.4 by a discussion of the GEV family and its best-known member, the Nested Logit model. We then in turn look at cross-nesting structures in Section 2.5, multi-level nesting structures in Section 2.6, and recursively defined model structures in Section 2.7. After a brief presentation of some alternative model forms in Section 2.8, we turn our attention to models based on mixture distributions in Section 2.9. The chapter closes with a summary in Section 2.10.

2.2 Basic concepts

In this section, we briefly look at some of the concepts that the theory of discrete choice modelling is based on, along the line introducing some notation that is used during the remainder of this thesis.

In a discrete choice experiment, a *decision-maker* n chooses a single alternative from a choice set C_n, made up of a finite[1] number of mutually exclusive alternatives,

[1]Problems with the requirement of having a finite number of alternatives arise especially in the case of continuous dependent variables, where some form of aggregation into a set of discrete alternatives needs to be used, as discussed for example by Train et al. (1987). The decision

where the choice set is exhaustive, and the ordering of alternatives has no effect on the choice process undertaken by the decision-maker. Each alternative $i = 1, \ldots, I$ in the choice set is characterised by a utility $U_{i,n}$, which is specific to decision-maker n, due to variations in attributes of the individuals, as well as in the attributes of the alternative, as faced by different decision-makers[2]. The use of the concept of utility, along with the need for a decision-rule, leads to the single most important assumption in the field of discrete choice modelling, namely that of *utility maximising behaviour* by respondents. As such, respondent n will choose alternative i if and only if $U_{i,n} > U_{j,n}\ \forall j \neq i$, with $i, j \in C_n$, where this notation excludes the possibility of equality between the utilities of two alternatives.

In an actual modelling analysis, the aim is to express the utility of an alternative as a function of the attributes of the alternative and the tastes and socio-demographic attributes of the decision-maker. Here, the limitations in terms of data and the randomness involved in choice-behaviour mean that, in practice, modellers will only be able to observe part of the utility. As such, we have:

$$U_{i,n} = V_{i,n} + \epsilon_{i,n}, \tag{2.1}$$

with $V_{i,n}$ and $\epsilon_{i,n}$ giving the *observed* and *unobserved* parts of utility respectively. Here, $V_{i,n}$ is defined as $f(\beta_n, x_{i,n})$, where $x_{i,n}$ represents a vector of measurable (to the researcher) attributes of alternative i[3] as faced by decision-maker n[4], and β_n is a vector of parameters representing the tastes of decision-maker n, which is to be estimated from the data. The function $f(\beta_n, x_{i,n})$ is free from any a priori assumptions, and although, in many cases, the use of a non-linear formulation has clear advantages[5], the majority of work in the area of discrete choice modelling relies on a linear formulation, such that $V_{i,n} = \beta_n' x_{i,n}$[6]. The inclusion of the unobserved utility term, $\epsilon_{i,n}$, means that the deterministic choice process now becomes probabilistic, leading to a random utility model (RUM), with the alternative with the highest observed utility having the highest probability of being chosen[7].

on what approach to use in the transformation of a continuous variable into a discrete one can have significant effects on the performance of the model and the results produced, as for example observed recently by Daly et al. (2005) in the context of departure time modelling. Here, the use of the term *aggregation* must not be confused with its use in the context of calculating aggregate choice probabilities, i.e. using aggregation over decision-makers. This topic is not discussed in this thesis; a thorough comparison of the various approaches that can be used is presented by Koppelman (1975).

[2] Here, a purely cross-sectional notation is used, with one observation per individual.

[3] Working on the assumption that, unlike in a Universal Logit type model, the utility of alternative i is not affected by the attributes of other alternatives $j \neq i$.

[4] The vector $x_{i,n}$ potentially also includes interactions with socio-demographic attributes of respondent n.

[5] As shown in some of the applications in this thesis, and discussed for example by Gaudry & Wills (1978*a*), and Mandel et al. (1997).

[6] In this context, the recent discussions by Rajaonarison et al. (2005) are of interest, looking at replacing the exponential function in GEV-based models by a deformed counterpart, where the estimates for the deformation parameter allow modellers to determine whether the linear-in-attributes specification is appropriate.

[7] With perfect knowledge of utility (leading to the *rational* model), the choice probabilities can be expressed graphically as a step function, while, in a RUM, the choice probability is a sigmoid curve, which approaches the step function as the variance of the *unobserved* part of utility decreases.

It can be seen that the probability of decision-maker n choosing alternative i is now given by:

$$P_n(i) = P\left(\epsilon_{j,n} - \epsilon_{i,n} < V_{i,n} - V_{j,n} \;\; \forall j \neq i\right). \tag{2.2}$$

With the *unobserved* part of utility varying randomly across respondents, the mean of this term can be added to the observed part of utility, in the form of an alternative-specific constant (ASC). The vector $\epsilon_n = \{\epsilon_{1,n}, \ldots, \epsilon_{I,n}\}$ is now defined to be a random vector with joint density $f(\epsilon_n)$, zero mean and covariance matrix Ω, and by noting that the probability of alternative i in equation (2.2) is the cumulative distribution of the random term $\epsilon_{j,n} - \epsilon_{i,n}$, we can write:

$$P_n(i) = \int_{\epsilon_n} I\left(\epsilon_{j,n} - \epsilon_{i,n} < V_{i,n} - V_{j,n} \;\; \forall j \neq i\right) f\left(\epsilon_n\right) \,\mathrm{d}\epsilon_n, \tag{2.3}$$

where $I(\cdot)$ is the indicator function which equals 1 if the term inside brackets is true and 0 otherwise. The probability is now given by a multi-dimensional integral which only takes a closed form for certain choices of distribution for ϵ_n, where the choice of $f(\epsilon_n)$ has a crucial impact on the behaviour of the choice-model, as described in the remainder of this chapter.

Equation (2.2) shows that the choice is invariant to the addition of the same constant to all utilities, or the multiplication of all utilities by the same constant, leading to the conclusion that *only differences in utility matter*. Aside from allowing us to reduce the dimensionality of the integral in equation (2.3) from I to $I-1$, this statement also indicates that the absolute levels of the ASCs are of no importance. As such, given the infinite number of possible values leading to the same differences in the ASCs, it becomes necessary to normalise the constants to avoid overspecification of the log-likelihood function. While the use of a constraint setting one ASC to zero has become common practice, Bierlaire et al. (1997) show that there are an infinite number of approaches that can be used to avoid over-specification, with, despite yielding the same results, substantial differences in estimation performance between them. Here, it should be noted that, even after an appropriate normalisation, problems with identification can occur in the case of very large choice sets (and an accordingly high number of ASCs), such that a parameterisation of the ASCs may in some cases be preferable, even though this leads to a violation of the zero-mean assumption for the error-terms (cf. Hess, Polak & Bierlaire 2005).

Just as we normalise the absolute level of utility through taking differences of the errors and normalising ASCs, we must also normalise the scale of utility, given that an infinite number of different scalings of utility will lead to the same model. The normalisation process depends on the defined relationship between the errors in the utility function, and becomes more complicated in models with more complex interrelationships between the errors. A detailed discussion of these issues, in the case of independent and identically distributed (*iid*), heteroscedastic, as well correlated errors is given by Train (2003). In the present discussion, except where otherwise stated, we work on the assumption of a homogeneous sample, with the scale of the underlying extreme-value term (in the Logit-based models) normalised to a value of 1.

2.3 Logit and Probit models

2.3.1 The Multinomial Logit model

The Multinomial Logit (MNL) model is the most basic member of the family of GEV models discussed in Section 2.4, and remains one of the most-widely used model forms. The *Logit formula* was first derived by Luce (1959); Marschak (1960) later showed that the model is consistent with utility maximisation and McFadden (1974) showed that the form of the Logit formula necessarily implies the use of the type I extreme value (Gumbel) distribution for the unobserved part of the utility.

The MNL model, which was at first also referred to as the *conditional Logit* model, is the extension of the binary Logit model to the multinomial case. There are numerous approaches leading to the derivation of the Logit choice probabilities. Ben-Akiva & Lerman (1985) base their derivation on the one given by Domencich & McFadden (1975), using the properties of the Gumbel distribution. Train (2003) on the other hand uses explicit integration of the multivariate cumulative distribution of the different $\epsilon_{j,n}$ for $j \neq i$ over the distribution of ϵ_i. A very similar approach was taken by McFadden (1974), using integration of the first derivative (with respect to its i^{th} term) of the multivariate cumulative distribution of the $\epsilon_{j,n}\ \forall j$ over the range of possible values of $\epsilon_{i,n}$. The MNL choice probability for alternative i and decision-maker n is given by:

$$P_n(i) = \frac{e^{V_{i,n}}}{\sum_{j=1}^{I} e^{V_{j,n}}}, \tag{2.4}$$

where the fact that the choice probabilities no longer involve the error terms ϵ_n means that the model can be estimated and applied without the use of simulation.

The assumption of *iid* errors is essential to the derivation of the Logit model; the model belongs to a class that Manski (1977) calls the *"independent and identically distributed random utility"* models (IIDRU). The nature of the error-terms prevents a treatment of correlation in the errors across alternatives or observations for the same respondent, while the assumption of identically distributed errors prevents a treatment of random variations in tastes across respondents.

It is well-documented that the behaviour of the MNL model is strictly governed by the *independence from irrelevant alternatives* (*IIA*) assumption; the ratio of the MNL choice probabilities for two different alternatives i and j is independent of the attributes or even existence of other alternatives, leading to the conclusion that any changes in the probability of a given alternative draw equally from the probabilities of all the other alternatives in the choice set (i.e. equal cross-elasticities). According to Hausman & Wise (1978), L.J. Savage was the first author to fully recognise and also criticise the effects of the *IIA* property. Tversky (1972) attributes the first critique of Luce's use of the *IIA* assumption to Debreu (1960). Hausman & Wise (1978) argue with some justification that the *IIA* property should rather be called *independence of relevant alternatives* property or *independence among alternatives* property. Another way of describing the *IIA* property (e.g. Vovsha 1997) is that, in a model exhibiting *IIA* over alternatives, all alternatives are equally independent from each other. McFadden (1974) describes the *IIA* property by noting that in the case of an additional alternative, the proportional decrease in the selection probabilities of the existing alternatives is equal to the selection probability of the new alternative (with an analogous situation when an alternative is removed).

While there are many examples where the *IIA* property is not acceptable (cf. Ben-Akiva & Lerman 1985), such that the MNL model should not be used, there are however also cases where *IIA* is a valid assumption, namely in those cases where the single alternatives are virtually unrelated, or when the relationship (closeness) between any two alternatives is the same for all pairs of alternatives.

One common mistake made in the existing MNL literature, and which McFadden et al. (1978, page 40) call *unwarranted criticism of the MNL model*, is to state that the *IIA* property is exhibited by the MNL model not only for individual choice probabilities and homogeneous population market shares, but also in the market shares observed in heterogeneous populations, whereas, due to the non-linearities in aggregation, this is clearly not the case. Another effect of the *IIA* property is that the coefficients estimated on a subsample of the original sample of alternatives are the same as those estimated on the original sample; this, in the case where the *IIA* assumptions is valid, is an important asset for modelling work in the presence of a very large or unbounded choice set, as for example in destination choice modelling. The approach involved with the use of a subset of alternatives does not require any *a priori* analysis of the alternatives (as would be the case with other aggregation methods); the alternatives are simply sampled randomly from the choice set.

It is important to test the validity of the *IIA* assumption for a given choice set, as a MNL model fitted to a choice set that violates *IIA* can lead to very misleading results. For a detailed discussion, see McFadden et al. (1978), who give a number of different tests and look into the possible sources of violation of the *IIA* assumption. Here, an important observation is that the validity of the *IIA* assumption does not necessarily depend on the nature of the choice set, but can also depend on the specification of the observed utility function, in such that a poor specification of the observed part of utility can lead to correlation in the unobserved part of utility between alternatives, which can result in a violation of the *IIA* assumption. Suggestions for tests are also made by Hausman & Wise (1978), who advocate comparisons between Logit and Probit models.

2.3.2 The Multinomial Probit model

Although not used in this thesis[8], it is worth briefly looking at the main alternative to GEV-based model structures in discrete choice analysis, namely the Multinomial Probit (MNP) model. The underlying assumption of Probit models is that the error terms follow a joint Normal distribution with zero mean and covariance matrix Σ, with no a priori restrictions on the correlation structure in the distribution. This means that the Probit model allows for any degree of correlation between the single error terms and can hence represent very different substitution patterns. The absence of the assumption of identically distributed error terms means that taste variation can also be incorporated in Probit models. Additionally, repeated choices, such as those observed with panel data, can be fully incorporated in the existing

[8]The focus in the present thesis on GEV and GEV mixture models reflects current trends, but is also due to the fact that, for the air-travel behaviour analyses conducted in the applied part of the thesis, the cost of using Probit models would be prohibitively high.

framework of the MNP model, by allowing for correlation in the *unobserved* part of utility both between alternatives and between choice situations.

The Probit model has its roots in psychology (Thurstone 1927); the work by Marschak (1960) led to its use as a utility maximising model. Hausman & Wise (1978) and Daganzo (1979) further discussed the Probit model, and especially its ability to represent very diverse substitution patterns, as well as random taste variation across the population and between choice situations.

The MNP choice probabilities are given by:

$$\begin{aligned} P_n(i) &= P\left(\epsilon_{j,n} - \epsilon_{i,n} < V_{i,n} - V_{j,n}\ \forall j \neq i\right) \\ &= \int_{\epsilon_n} I\left(\epsilon_{j,n} - \epsilon_{i,n} < V_{i,n} - V_{j,n}\ \forall j \neq i\right) \phi\left(\epsilon_n\right) \mathrm{d}\epsilon_n, \end{aligned} \tag{2.5}$$

where $I(\cdot)$ is the indicator function and

$$\phi\left(\epsilon_n\right) = \frac{1}{(2\pi)^{\frac{I}{2}} \left|\Sigma_n\right|^{\frac{1}{2}}} e^{-\frac{1}{2}\epsilon_n' \Sigma_n^{-1} \epsilon_n}. \tag{2.6}$$

From equation (2.5), it can be seen that the calculation of the Multinomial Probit (MNP) choice probabilities requires the solution of an I-dimensional integral[9] that does not have a closed form expression. This thus needs to be approximated through simulation or other numerical approaches (cf. Train 2003, pp. 118-137). Initially, the high cost of estimation meant that the use of Probit models was restricted to choice situations with a very low number of alternatives. While recent gains in computer speed and estimation efficiency have somewhat lessened the problem, the use of the MNP model still imposes a heavy computational burden, as noted for example by Weeks (1997). Another discussion of the estimation of the MNP model is given by Horowitz (1991), who also discusses the issues involved with forecasting. Another issue that needs to be faced with the use of Probit models is that of identification, an issue discussed amongst others by Train (2003), and which is strongly related to that of a choice of an appropriate covariance structure (cf. Hausman & Wise 1978).

Finally, a major disadvantage of the Probit model, especially in the context of this thesis, is the requirement to use a Normal distribution for representing random taste heterogeneity, leading to significant losses in terms of flexibility, and issues of interpretation in the case of counter-intuitive results suggesting large shares of *wrongly*-signed coefficient values. At least from the point of view of random taste heterogeneity, the continued development of GEV mixture models, which do not impose any restrictions on the choice of distribution, has further decreased the appeal of the MNP model.

2.4 The GEV family and the Nested Logit model

The Generalised Extreme Value (GEV) family of models, introduced by McFadden (1978), is a set of closed form discrete choice models that are all based on the use of the extreme-value distribution, and which allow for various levels of correlation

[9] Respectively $I - 1$ after taking differences.

among the unobserved part of utility across alternatives. This is done through dividing the choice set into nests of alternatives, with increased correlation, and thus higher cross-elasticities, between alternatives sharing a nest. As such, alternatives sharing a nest are more likely substitutes for each other. The use of such a nesting structure means that GEV models are most easily understood in the form of trees, with the root at the top, elementary alternatives at the bottom, and composite alternatives, or nests, in between.

The MNL model is the most basic member of this GEV family, using a single nest of alternatives, resulting in equal cross-elasticities across all alternatives. While in the MNL model, the error terms are distributed *iid* extreme value, in the general GEV formulation, the error terms follow a joint generalised extreme value distribution; the individual error terms follow a univariate extreme value distribution, but the error terms associated with alternatives sharing a nest are correlated with each other. This structure leads us away from the diagonal variance-covariance matrix of the MNL model.

2.4.1 GEV choice probabilities

The GEV theory is based on the use of a generating function, usually defined as G, which takes the vector $\langle e^{V_1}, \ldots, e^{V_I} \rangle$ as its argument[10]. A general notation is to set $Y_1 = e^{V_1}$. The arguments of G are then $\langle Y_1, \ldots, Y_I \rangle$, which are clearly strictly positive. McFadden (1978) defines the function G through four conditions:

GEV$_1$: $G(Y_1, \ldots, Y_I)$ is non-negative

GEV$_2$: $G(Y_1, \ldots, Y_I)$ is homogeneous of degree 1, meaning that $G(\alpha Y_1, \ldots, \alpha Y_I) = \alpha G(Y_1, \ldots, Y_I)$.

GEV$_3$: $\lim_{Y_i \to +\infty} G(Y_1, \ldots, Y_I) = +\infty, \;\; \forall i \in (1, \ldots, I)$

GEV$_4$: For any distinct $(i_1, \ldots, i_k)$ from $(1, \ldots, I)$, the kth order partial derivative of G, $\frac{\partial^k G(Y_1,\ldots,Y_I)}{\partial Y_{i_1},\ldots,\partial Y_{i_k}}$ is ≥ 0 if k is odd and is ≤ 0 if k is even.

McFadden (1981) adds a fifth condition:

GEV$_5$: Assume that $G(Y_1, \ldots, Y_I)$ has parameters x and β, such that the Y_i are functions of these parameters (e.g. $Y_i = e^{\beta_i' x_i}$), and that two choice sets $B = (a_1, ..., a_I)$ and $B' = (b_1, ..., b_I, ..., b_{I+m})$ have alternatives with parameter vectors given by $x_B = \{x_{a_j} \; \forall a_j \in B\}$ and $x_{B'} = \{x_{b_j} \; \forall b_j \in B'\}$. If these parameters satisfy the condition that $x_{a_j} = x_{b_j}$ for $j = 1, ..., I$, then $G((Y_{a_1}, ..., Y_{a_I}), x_B, \beta) = G((Y_{b_1}, ..., Y_{b_I}, 0, ..., 0), x_{B'}, \beta)$.

GEV$_5$ essentially states that the addition of any number of alternatives with zero utility does not affect the choice process. Daly (2001*b*) notes that this might be necessary in order to ensure the symmetry of the partial derivatives. This condition was not included in the original definition of the GEV family (McFadden 1978) and was also ignored in most well-known texts (e.g. Train 2003); conditions GEV$_1$-GEV$_4$ thus seem to be sufficient to define a GEV model.

[10]The association with a given respondent n is dropped for now.

Ben-Akiva & Francois (1983) generalise condition GEV_2 such that $G(...)$ needs to be homogeneous of degree μ with $\mu > 0$ rather than $\mu = 1$, as given by McFadden (1978). This means that $G(\alpha Y_1, \ldots, \alpha Y_I) = \alpha^\mu G(Y_1, \ldots, Y_I)$, with $\mu > 0$.

The GEV choice probability for alternative i is now given by:

$$P(i) = \frac{Y_i G_i(Y_1, ..., Y_I)}{G(Y_1, ..., Y_I)}, \tag{2.7}$$

where $G_i(Y_1, ..., Y_I) = \frac{\partial G(Y_1,...,Y_I)}{\partial Y_i}$.

As mentioned above, the MNL model is the most basic member of the GEV family, using a single nest grouping together all alternatives. Its generating function is simply given by:

$$G\left(Y_1, \ldots, Y_I\right) = \sum_{i=1}^{I} Y_i, \tag{2.8}$$

where the use of a single level of nesting (i.e. all alternatives nested below the root) is reflected in the use of single level of a summation.

2.4.2 The Nested Logit model

The most basic example of a GEV model with an actual nesting structure is the two-level Nested Logit (NL) model, which is based on the use of mutually exclusive subsets S_m $(m = 1, ..., M)$ of the choice set C, such that $C = \bigcup_{m=1}^{M} S_m$. This can be extended to multi-level NL models, in which the nests on any given level are again mutually exclusive, with the lower-level nests (grouping of elementary alternatives) themselves grouped into nests at higher levels of nesting. A general formulation, with multi-level choice probabilities, is for example given by Koppelman & Sethi (2000).

A first version of the NL model was proposed by Domencich & McFadden (1975), but as McFadden (1978) notes, this model was using an 'unsatisfactory definition of inclusive value'. The correct form for the inclusive value term was developed by Ben-Akiva (1973, 1974*b*), the consistency of the NL model with utility maximisation was later proved independently by Williams (1977), and Daly & Zachary (1978), with McFadden (1978) developing the GEV form. A more detailed history of the development of the NL model is given by Ortúzar (2001).

The NL choice probabilities are represented through a product of successive choice probabilities that represent a chain from the root of the tree to the elementary alternative for which the probability is calculated. The utility of a composite alternative (nest) is determined by the utilities of those nodes that have the composite alternative as their direct ancestor, this way, the utility of elementary alternatives is propagated up through the tree. This is done through the use of a *logsum* term (also referred to as inclusive value or accessibility), where the logsum of nest m is given by the denominator of the *conditional* choice probability in nest m. A structural (logsum) parameter is associated with each node, determining the correlation between error-terms of the alternatives within the different nests.

The generating function for a two-level NL model with M nests is given by:

$$G(Y_1, \ldots, Y_I) = \sum_{m=1}^{M} \left(\sum_{j \in S_m} Y_j^{\frac{1}{\lambda_m}} \right)^{\lambda_m}, \tag{2.9}$$

where λ_m is the logsum parameter associated with nest m. The notation used here, which is equivalent to that used by Koppelman & Sethi (2000) and Train (2003), corresponds to the interpretation of the logsum parameter as the *independence parameter*, with higher λ_m meaning lower correlation. Two alternative notations exist; with λ_m giving the structural parameter as used in equation (2.9), Ben-Akiva & Bierlaire (1999) (amongst others) define the structural parameter as $\frac{1}{\lambda_m}$, while McFadden (1978) (amongst others) defines the structural parameter as $1-\lambda_m$. The different notations yield the same results, but lead to different requirements in terms of the acceptable range for λ_m and the interpretation of the estimates of λ_m. In the present notation, we have used *normalisation from the top*, such that the structural parameter at the root, λ_0, is equal to 1[11].

The value of the different structural parameters is generally constrained to lie between 0 and 1, where values below 0 are inconsistent with utility maximisation while values above 1 are only consistent with utility maximisation for special ranges of the explanatory variables. In a two-level NL model, the correlation in nest m with structural parameter λ_m is given by $1 - \lambda_m^2$. As described by Koppelman & Sethi (2000), the estimated parameters at each node are equal to the ratio between the actual structural parameters for that node and its immediate predecessor, and the range conditions apply to this ratio. At the highest level of nesting, this means that, in the case of *normalisation from the top*, the actual structural parameter needs to be lower than 1, while, for the remainder of the tree, the range requirements for the ratios of parameters prescribe decreasing structural parameters as we move down the tree[12].

It can easily be seen from equation (2.9) that the NL model reduces to the MNL model (cf. equation (2.8)) if all structural parameters are equal to 1. On the other hand, with all nesting parameters λ_m tending to zero, the choice among the alternatives in a nest becomes a deterministic choice such that the alternative with the maximum utility is chosen. Train (2003) notes that this effectively means that, as λ_m tends to zero $\forall m$ (while being positive), we move towards the elimination by aspects (EBA) model (cf. Tversky 1972).

In a NL model, the direct-elasticities for single-nest alternatives are the same as in the MNL model, as are the cross-elasticities for two alternatives not sharing a nest. However, for an alternative sharing a nest with other alternatives, the direct-elasticity is higher than in the MNL model, as is the cross-elasticity for two

[11]An alternative normalisation is to set one of the structural parameters at the lowest intermediary level equal to 1 (*normalisation from the bottom*), and, although McFadden (1978) argues that this approach may lead to a simpler formulation, it seems to have become common practice to *normalise from the top*, i.e. to set $\lambda_0 = 1$.

[12]I.e. with λ_u and λ_l giving the structural parameter in the upper and lower intermediary level in a three-level NL model, the ratios $\frac{\lambda_u}{\lambda_0}$ and $\frac{\lambda_l}{\lambda_u}$ need to be constrained between 0 and 1, such that we get the condition that $0 < \lambda_l \leq \lambda_u \leq 1$, where, with equality between the two parameters, the lower level of nesting becomes obsolete.

alternatives sharing a nest, where this stays constant across alternatives sharing a nest, such that *IIA* still holds within nests. Both direct and cross-elasticities increase with decreasing λ_m. Finally, in models with multiple levels of nesting, the correlation between two alternatives depends on the structural parameter associated with the lowest nest shared by the two alternatives, meaning that the cross-elasticities are higher when the lowest common nest is situated lower down the tree.

A final point needs to be addressed before we proceed to more complicated nesting structures. Indeed, for years following the introduction of the NL model, there existed two different forms of the model, a fact that was largely ignored, with the choice of approach often determined solely on the basis of the estimation software used. The problem was first discussed in detail by Koppelman & Wen (1998). The difference between the structure presented here, which can also be referred to as the Utility Maximising Nested Logit (UMNL) model, and the alternative form, which Koppelman & Wen (1998) call the Non-normalised Nested Logit (NNNL) model, is that, in the latter, the exponent of Y_J in equation (2.9) is equal to 1, rather than $\frac{1}{\lambda_m}$. Koppelman & Wen (1998) note that the two models are equivalent only if all structural parameters are the same across nests, when the only difference between the models lies in the scale of utility. They state that, when this condition is not satisfied, the NNNL model is not consistent with utility maximisation as the addition of a constant to the utility of each elementary alternative is, in the absence of normalisation, not equivalent to the addition of a constant at the level of the nest. Although Daly (2001*a*) discusses cases where this reasoning may not apply, such as in the case where no coefficients are shared across nests[13], the *normalised* version has now essentially become established as the *correct* form. The issue described here has also been addressed by Hensher & Greene (2002), Carrasco & Ortúzar (2002), and Heiss (2002), who also focusses on implementation.

2.5 Models allowing for cross-nesting

The number of different GEV models introduced since the initial development of the GEV framework (McFadden 1978), and especially over the last few years, has shown the flexibility allowed by this structure. With the increases in complexity and flexibility however also comes an increased cost of estimation, along with increased risks of misspecification.

The characteristic that differentiates advanced GEV models from the basic NL model is that they allow for cross-nesting (multi-nest-membership). A cross-nesting structure can be used in the case where there is heightened substitution between alternatives A and B, as well as between alternatives A and C, without heightened substitution between alternatives B and C. There are many problems in which the extra flexibility of the CNL model has the potential to offer considerable improvements, and a number of recent applications (e.g. Bierlaire et al. 2001) have confirmed these theoretical advantages, even with a relatively low number of nests or alternatives.

As well as allowing for multi-nest membership, the majority of cross-nesting models allow for differences in the degree of allocation of alternatives to nests, hence

[13] See also Koppelman & Wen (2001).

reflecting the different degrees of similarity between them. The more two alternatives belong to the same nests, the higher will be the cross-elasticity between their probabilities. The cross-elasticity will be at its maximum if all inclusion factors are equal for the two alternatives; Vovsha & Bekhor (1998) note that this is analogous to the common nest situation in a NL model.

The degree of cross-nesting allowed by the different models, and hence their flexibility, varies across models. While some of the existing GEV models offer major improvements in flexibility, others are simply minor extensions of existing models, often constructed for a very specific modelling application. The relationship between the different model forms is best illustrated by starting with a very general model form, and by showing how restrictions of this form yield the more specific model forms. This is the topic of the following few sections, where in each case, only the generating function of the models is reproduced; a more detailed description of the main models, including the actual choice probabilities, is again given by Koppelman & Sethi (2000).

2.5.1 The Generalised Nested Logit model

A very general version of a two-level GEV model is given by the Generalised Nested Logit (GNL) model of Wen & Koppelman (2001), which allows alternatives to belong to different nests, with different degrees of allocation for different nests, and differential levels of correlation in different nests. In the GNL model (and other cross-nested models), the choice probabilities (and hence also the elasticities) of an alternative are represented by a sum over nests, taking into account the differences in the membership (allocation) to different nests. Only nests for which an alternative has a strictly positive allocation parameter are included in this sum, while for cross-elasticities, strictly positive allocation parameters are necessary for both alternatives considered.

The generating function of the GNL model is given by:

$$G\left(Y_1,\ldots,Y_I\right)=\sum_{m=1}^{M}\left(\sum_{j\in S_m}\left(\alpha_{j,m}Y_j\right)^{\frac{1}{\lambda_m}}\right)^{\lambda_m}, \tag{2.10}$$

where $\alpha_{j,m}$ is the allocation parameter for alternative j and nest m, and where, primarily for reasons of interpretation, we set $0\leq\alpha_{j,m}\leq 1\,\forall j,m$ and $\sum_{m=1}^{M}\alpha_{j,m}=1\,\forall j$.

By imposing certain restrictions on the structural parameters and allocation parameters, the GNL model can be adapted to represent all existing two-level GEV models. The MNL model is a special case of the GNL model, using a single nest, with structural parameter equal to 1 and allocation parameters equal to 1 for all alternatives. In a two-level NL model, each alternative belongs to one nest only; the allocation parameter of a given alternative is thus equal to 1 in this nest, and 0 in all other nests.

The GNL model is restricted to a two-level structure, by using only two levels of summation in equation (2.10). This can be seen as a serious restriction in cases where a multi-level structure is required. Wen & Koppelman (2001) discuss how the GNL model can be used to approximate a multi-level NL structure by a two-level cross-nesting structure; the precision of this approximation however clearly depends on the application at hand.

Wen & Koppelman (2001) note that there exist two models that are very similar to the GNL model; these are the Choice Set Generation Logit Model (GenL or GenMNL, referred to as General Logit by Koppelman & Sethi 2000 and Wen & Koppelman 2001) and the Fuzzy Nested Logit (FNL) model. The GenL model was proposed by Swait (2001*a*) with the aim of simultaneously evaluating choice set generation and choice within the generated choice set, where, mathematically, the only difference to the GNL model comes in the form of strict allocation parameters. The FNL model addresses the GNL model's biggest shortcoming in that it allows for multi-level models. There is no published material on the FNL model; the only available material (Vovsha 1999) describes an application of the FNL model, defining the model as the "full cross-nested Logit" model, in the context of combined residential place and workplace choice. The generating function of this FNL model is given by a sum of two Cross-Nested Logit (CNL) generating functions, the resulting model structure essentially consists of two individual CNL models, which have the same elementary alternatives (one with residential choice at the upper level and one with workplace choice at the upper level). This leads to a three-level FNL model, with cross-nesting at the lower level. Extensions to higher levels of nesting are possible, but such a form is not given explicitly by Vovsha (1999).

2.5.2 The Cross-Nested Logit model

The first use of the notion "Cross-Nested Logit" is generally attributed to Vovsha (1997); however, the history of the CNL model reaches back to the initial developments of the GEV family. A predecessor of the common CNL model form was proposed by McFadden (1978), where each alternative belongs to every nest, and where the allocation parameters can vary across nests, but are required to stay constant across all alternatives in the same nest, leading to a generating function given by:

$$G(Y_1, \ldots, Y_I) = \sum_{m=1}^{M} \alpha_m \left(\sum_{j \in S_m} Y_j^{\frac{1}{\lambda_m}} \right)^{\lambda_m}, \tag{2.11}$$

with $\alpha_m > 0$.

The *non-normalised* CNL model

The CNL model proposed by Vovsha (1997) was developed for the modelling of mode choice. Unlike McFadden (1978), Vovsha (1997) allows the allocation parameters to vary over alternatives, but on the other side requires the structural parameters to be the same for all nests. The generating function for this model is given by:

$$G(Y_1, \ldots, Y_I) = \sum_{m=1}^{M} \left(\sum_{j \in S_m} \alpha_{j,m} Y_j \right)^{\lambda}, \tag{2.12}$$

with $\alpha_{j,m} > 0 \, \forall j$, and $\sum_{m=1}^{M} \alpha_{j,m}^{\lambda} = 1 \, \forall j$.

Aside from the use of a common structural parameter λ across all nests, and the special form for the constraint on the allocation parameters, the other main characteristic of this model is the absence of the $\frac{1}{\lambda}$ exponent inside the sum over alternatives sharing a common nest. While the absence of this exponent has been picked up in the existing literature, there has been little or no discussion as to the reasons for this form of the generating function. However, the absence of the exponent can be quite easily explained by the fact that the derivation of the CNL model given by Vovsha (1997) is based on the NNNL model rather than the standard NL model. The fact that this version of the CNL model lacks the $\frac{1}{\lambda}$ exponent means that it cannot be seen as a special case of the GNL model without some redefinition of the GNL parameters.

The *normalised* CNL model

A different version of the CNL model was developed by Vovsha & Bekhor (1998) for the modelling of route choice; this version includes the $\frac{1}{\lambda}$ exponent in the inner sum, and can thus be referred to as the *normalised* CNL model, essentially being a version of the UMNL model that allows for cross-nesting. The generating function of this model is given by:

$$G(Y_1, \ldots, Y_I) = \sum_{m=1}^{M} \left(\sum_{j \in S_m} (\alpha_{j,m} Y_j)^{\frac{1}{\lambda}} \right)^{\lambda}, \tag{2.13}$$

with the condition on the allocation parameters simplified to $\sum_{m=1}^{M} \alpha_{j,m} = 1 \; \forall \, j$. The CNL model given by Vovsha & Bekhor (1998) can be constrained to be equal to a NL model with equal structural parameters across nests, and can similarly be seen to be a special case of the GNL model, with all structural parameters taking on the same value.

Alternative CNL model forms

Various alternative forms of the *normalised* CNL model have been proposed. These include the model given by Papola (2004), which, unlike the two models given by Vovsha (1997) and Vovsha & Bekhor (1998), allows the structural parameters to be different in different nests, making the model more general and essentially equivalent to the GNL mode. Papola (2004) also discusses the flexibility of the CNL model in terms of correlation structure, stating that for any given homoscedastic variance-covariance matrix, there is a specific CNL model counterpart.

Ben-Akiva & Bierlaire (1999) (further expanded by Bierlaire 2005) give a formulation of the *general* CNL model that generalises some of the other existing formulations, by making fewer assumptions about the relationship of allocation parameters across nests. This version of the CNL model is characterised by the partial absence of the exponent in the inner sum of the generating function, where the exponent $\frac{1}{\lambda_m}$ is used only for Y_i and not for $\alpha_{i,m}$. This partial absence, which is essentially the only difference with the GNL model, hampers comparisons between the models, leading to a requirement for the redefinition of some of the variables and parameters in the GNL model to achieve equivalence. Bierlaire (2005) shows that conditions on the allocation parameters (i.e. sum to 1) used in previous derivations are not necessary for model validity. This revelation is similar to that made by Wen & Koppelman (2001) in the context of the GNL model; constraints on the sum of the allocation parameters are not strictly necessary but can facilitate interpretation and make the models more intuitive, aside from enabling parameter identification.

2.5.3 The Link-Nested Logit model

The Link-Nested Logit (LNL) model, proposed by Vovsha & Bekhor (1998), is a specialised model for route choice based on the *normalised* CNL model. The spatial overlap of routes is represented by a nesting structure that allows for different degrees of nest allocation, depending on the extent of overlapping. The links (i.e. sections of path) where the individual routes overlap become the nests, the allocation parameters reflect the proportion of the route length (or time) spent on each link. The LNL model uses a maximum degree of nesting, such that the structural parameters are close to zero. This means that two routes are treated as very similar on the section of the path where they overlap, with the obvious effect this has on the cross-elasticities.

2.5.4 The Paired Combinatorial Logit model

The Paired Combinatorial Logit (PCL) model (Chu 1989, Koppelman & Wen 2000) uses a two-level nesting structure with a single nest for each pair of alternatives, leading to $\frac{I!}{2!(I-2)!}$ different nests. Equal proportions of an alternative are assigned to each of the $I-1$ nests that the alternative belongs to; the PCL model is thus a restricted version of the GNL model, with equal allocation of an alternative to its $I-1$ nests[14]. The generating function of the PCL model is given by:

$$G(Y_1,\ldots,Y_I)=\sum_{i=1}^{I-1}\sum_{j=i+1}^{I}\left((\alpha Y_i)^{\frac{1}{\lambda_{ij}}}+(\alpha Y_j)^{\frac{1}{\lambda_{ij}}}\right)^{\lambda_{ij}}, \tag{2.14}$$

where λ_{ij} defines the structural parameter associated with the nest containing alternatives i and j.

The main characteristic of the PCL is that any two alternatives share exactly one nest, whereas in the other cross-nesting models, two alternatives can share multiple nests. Koppelman & Wen (2000) note that while both the NL model and the PCL model are generalisations of the MNL model, neither model is a restriction of the other model. They also observe that, while the PCL model achieves a more important relaxation of the *IIA* property than the NL model, the upper limit on the cross-elasticities in the PCL model is lower than in the NL model, given the equal allocation of an alternative to each nest in the PCL model.

[14]The original PCL model sets $\alpha_{i,m}=1$ for all nests m that alternative i belongs to, while the conditions on the allocation parameters in the GNL model lead to values of $\frac{1}{I-1}$. The choice probabilities in the two models are identical, as the use of equal allocation parameters for all nests means that they cancel out.

2.5.5 The Paired GNL model

Wen & Koppelman (2001) also propose the Paired GNL (PGNL) model, which imposes no restrictions on the allocation parameters. By allowing for different degrees of association between different alternatives, including the possibility of zero association between some alternatives, important gains in flexibility are obtained, which however need to be offset against heightened estimation cost.

2.5.6 The Ordered Generalised Extreme Value model

Another well-known member of the family of Cross-Nested Logit models is the Ordered Generalised Extreme Value (OGEV) model, even though this model was introduced some time before the first use of the term "Cross-Nested Logit". The OGEV model was proposed by Small (1987) in the context of departure time (or arrival time) choice modelling. The OGEV model is another example of a semi-normalised model, in that it uses the inverse of the structural parameter as an exponent for the arguments of the generating function, but not for the allocation parameters. The generating function of the OGEV model is given by:

$$G(Y_1, \dots, Y_I) = \sum_{m=1}^{M} \left[\sum_{j \in S_m} w_{m-j} Y_j^{\frac{1}{\lambda_m}} \right]^{\lambda_m}, \tag{2.15}$$

where the nests are defined such that they contain up to $L+1$ contiguous alternatives, with nest m defined as:

$$S_m = \{j \in \{1, ..., I\} \mid m - L \leq j \leq m\} \tag{2.16}$$

In the OGEV model, allocation parameters do not depend on the nest, but on the position of an alternative within a nest. As such, the number of different weight parameters (i.e. allocation parameters) is equal to $L+1$ (the $L+1$ different positions that an alternative can take), where Small (1987) imposes the conditions that all weights are non-negative and sum to 1.

The OGEV model nests adjacent time intervals, where the number of alternatives in a nest depends on the specification. The fact that neighbouring time intervals necessarily share more nests than non-neighbouring nests increases the correlation and cross-elasticities between such alternatives.

2.5.7 The Principles of Differentiation model

Another model that is sometimes grouped together with other models of the family of cross-nested models is the Principles of Differentiation (PD) model, which was developed by Bresnahan et al. (1997) to overcome the problems with the sensitivity of the NL model to the order of the levels of nesting.

The PD model works by clustering alternatives according to their attributes, known as principles of differentiation. A certain number of dimensions are used for nesting, where each dimension defines an *attribute* of the alternative, with the subnests in a nest representing the different *values* that the respective *attribute* can take. In the resulting model, each alternative belongs to exactly one subnest in each such group. To avoid the use of multi-level nesting, Bresnahan et al. (1997) then define the model as a weighted sum of individual two-level NL models, with the weights assigned to the different two-level NL models indicating the relative importance of the different attributes in the calculation of probabilities and cross-elasticities. This makes the model mathematically (though not conceptually) equivalent to a two-level cross-nested model, with the weight parameters equating to allocation parameters (see also Wen & Koppelman 2001).

The generating function for a PD model with M dimensions (principles of differentiation, i.e. number of two-level NL models used in weighted sum) is given by:

$$G(Y_1,\ldots,Y_I) = \sum_{m=1}^{M}\left[\alpha_m \sum_{k\in m}\left(\sum_{j\in(k,m)} Y_j^{\frac{1}{\lambda_m}}\right)^{\lambda_m}\right], \tag{2.17}$$

where $j \in (k,m)$ defines the group of alternatives that exhibit aspect k as the value of attribute m, while α_m represents the weight assigned to dimension m, with Bresnahan et al. (1997) imposing the constraint that $\sum_{m=1}^{M}\alpha_m = 1$.

The model can be rewritten as a restricted version of the GNL model on the basis that the structural parameter λ_m is the same for all "subnests" of nest m and that the weight parameters α_m are the same for all elementary alternatives that belong to (subnests of) nest m. This then yields a generating function of the required form, with two levels of summation:

$$G(Y_1,\ldots,Y_I) = \sum_{(m,k|1\leq m\leq M\&k\in m)}\left[\left(\sum_{j\in(k,m)} (\alpha_m Y_j)^{\frac{1}{\lambda_m}}\right)^{\lambda_m}\right], \tag{2.18}$$

The fact that the weight parameters are identical for all alternatives in a given nest makes the model similar to the McFadden (1978) model, while the fact that higher weights are assigned to more *important* attributes (thus giving them more discriminatory power) makes the model conceptually similar to the EBA model.

2.6 Multi-level GEV models

As Daly (2001*b*) notes, while the structure of multi-level NL models has frequently been discussed and while such models have occasionally been used in practice, there is a distinct lack of more advanced multi-level GEV models. Indeed, except for the applied version of the FNL model given by Vovsha (1999), all advanced GEV models discussed so far in this section use a classical two-level structure with choice of nest on top of choice of elementary alternative. The absence of such multi-level advanced GEV models may in part be seen as a reflection of the fact that the flexibility offered by cross-nested structures allows for many multi-level processes to be approximated closely by two-level structures.

Besides the "full cross-nested Logit" model, a notable exception to the two-level GEV structure is the MNL-OGEV model proposed by Bhat (1998*b*). This model is used for the simultaneous choice of mode and departure time for shopping trips.

The model uses a MNL model for mode choice at the upper level, with an OGEV model for departure time choice attached to each 'leaf' of the upper tree. Each mode is thus associated with every possible departure time. The generating function of this model is given by:

$$G\left(Y_{1,1},\ldots,Y_{1,J},\ldots,Y_{I,1},\ldots,Y_{I,J}\right)=\sum_{i=1}^{I}\left[\sum_{j=1}^{J+1}\left(\frac{1}{2}Y_{i,j-1}^{\frac{1}{d}}+\frac{1}{2}Y_{i,j}^{\frac{1}{d}}\right)^{\frac{\lambda_d}{\lambda_m}}\right]^{\lambda_m}, \quad (2.19)$$

where $Y_{i,j}$ defines the generating function variable associated with the i^{th} mode and j^{th} departure time, with the conditions that $Y_{i,0} = Y_{i,J+1} = 0$ (as in the OGEV model). The structural parameters are constrained to be the same for all nests on a given level, where, with λ_m used for the mode level and λ_d used for the departure time level, we have $0 < \lambda_d \leq \lambda_m \leq 1$. In the formulation in equation 2.19, the OGEV sub-model allows for correlation only between directly adjacent time-periods. Finally, it can be seen that the presence of three levels of nesting means that the MNL-OGEV generating function in equation (2.19) cannot be written as a special case of the GNL generating function.

The version of the FNL model given by Vovsha (1999) is very similar to the MNL-OGEV model, as it is essentially a CNL model grafted onto the leafs of a MNL model. As the OGEV model is a special version of a CNL model, the MNL-OGEV model can thus be seen as a restricted version of the FNL model, a fact that also illustrates the greater flexibility of the FNL model when compared to the GNL model.

2.7 Recursive GEV models

One of the major issues with GEV models is that of efficient estimation, a complication that arises especially in the case of multi-level or cross-nesting structures. In this context, an appealing approach comes in the form of a recursive formulation of the model structure[15].

2.7.1 Recursive NL models

The development of such an approach was first discussed by Daly (1987) in the context of the NL model, i.e. in the absence of cross-nesting. In the recursive formulation, a hierarchy (or tree) function t is defined such that $t(s)$ returns s', the tree element (node) located directly above s; s' is thus the nest containing s. Two sets of nodes are defined for each node, one containing its *ancestors* (up to but not including the root) and one containing its *siblings* (nodes with the same "immediate ancestor"). For any node in the tree (elementary or composite), the probability of choosing the alternative represented by this node is now equal to the probability of the alternative having higher utility than any of its siblings. The likelihood of an elementary alternative is given by the product of the probabilities of all nodes

[15] The presentation of the full details of the mathematical implementation of these methods is very lengthy, and as such, is not repeated here; a complete description of the models can be found in Daly (1987) and Daly & Bierlaire (2005).

included in the path going from the elementary node to the upper-most non-root element in the path. This form of the likelihood, and of the resulting log-likelihood, greatly facilitates estimation of the NL model. Daly (1987) develops the recursive approach in the context of the NNNL model, but Daly (2001*a*) shows that the approach is also applicable for the "normalised" NL (UMNL) model. In either case, the approach is free from any assumptions regarding the number of nesting levels.

2.7.2 General recursive GEV models

While the recursive approach discussed above has the potential to offer significant savings in the estimation of (multi-level) NL models, it does not allow for cross-nesting and can thus not represent any of the more general GEV models. The approach discussed by Daly & Bierlaire (2005), which is a combination of the work of Daly (2001*b*) on the Recursive Nested Extreme Value (RNEV) model, and Bierlaire (2002) on the Network GEV model, overcomes these shortcomings and is able to represent any existing GEV model. Defining D to be the set of all nodes in the tree (elementary and composite), this approach uses a non-circular single-root function $N(m)$, which, for a non-elementary node m ($m \in D \setminus E$, where E is the set of elementary alternatives), returns the elements (either elementary or composite) contained in the nest represented by node m. The use of a single root, together with a condition of non-circularity, implies that for every elementary alternative e, there exists at least one sequence leading from the root to the elementary alternative e. In the inverse direction, this sequence is equivalent to the ancestor sequence used by Daly (1987). However, in the new model, for any given node m, there is no requirement that the parent function $t(m)$ yields a single node from the upper level; the absence of this requirement means that the model allows for cross-nesting. The fact that the model also allows for multiple levels of nesting means that it generalises all existing GEV models, including the GNL model.

The choice probabilities in a recursive model allowing for cross-nesting are effectively sums of recursive NL probabilities as defined by Daly (1987), accounting for the different ways of moving between the root and an elementary alternative, due to the presence of cross-nesting. The individual sequences leading from the root to the different elementary alternatives can be seen as alternatives in an extended recursive NL model. The final choice probability of an alternative i is then simply the sum of the probabilities of all those alternatives (in the recursive NL model) that represent sequences to the elementary alternative i in the combined model. As the model is thus a sum of recursive NL models, it generalises the NL model, with no restriction on the number of levels of nesting. By allowing for cross-nesting, the resulting two-level model is identical to the GNL model, hence also generalising all the models that GNL generalises. The fact that the model is not limited to a two-level structure however means that it is more flexible than the GNL model, and is also able to represent the MNL-OGEV model. The model is thus very similar in flexibility to the FNL model, although it is inherently different from this model, due to its recursive approach.

2.8 Some alternative model forms

Several other model structures deserve to be mentioned briefly. In this section, we look specifically at three types; heteroscedastic extreme value models (Section 2.8.1), models related to the Universal Logit model (Section 2.8.2), and choice set generation models (Section 2.8.3).

Two additional models that could be included here are the Covariance Nested Logit (COVNL) model of Bhat (1997), and the Elimination by aspects (EBA) model of Tversky (1972). The COVNL model is not discussed here as a detailed description is presented in Chapter 7, which is dedicated entirely to the representation of covariance heterogeneity. For the EBA model, space considerations and the scope of the thesis led to the decision to avoid a detailed description of the model; however, the similarity between the EBA model and GEV structures is a topic of continuing interest (cf. Batley & Daly 2003).

2.8.1 Models with heteroscedastic error-terms

Although not central to this thesis, a topic of great importance is the treatment of heteroscedasticity in discrete choice models, either in terms of differences in error variances across alternatives or across individuals. Aside from the use of an error-components formulation (cf. Section 2.9.1), an alternative approach is based on a modified specification for the scale parameters of the error-terms.

Bhat (1995) presents such a Heteroscedastic Extreme Value (HEV) model based on a MNL structure, with different scale parameters for different alternatives. Here, the resulting model does not possess a closed form expression, leading to a more expensive estimation process, which can however be justified in the presence of significant levels of heteroscedasticity, which could lead to biased results in models based on a homogeneous error-structure. Hensher (1999) exploits the approach in the search for appropriate nesting structures; here, the insights obtained with regards to the variance of the error-terms for the different alternatives can provide some guidance for the specification of a nesting structure.

Another treatment of heteroscedasticity arises in the case of datasets from two separate sources, with the potential of different scale factors for the two groups. The most common example of such a scenario comes in the combined use of Stated Preference (SP) and Revealed Preference (RP) data. Bradley & Daly (1996) address this issue with the help of a NL structure based on single-alternative nests split into two groups, according to the data-source, where the nesting parameter stays constant across nests within the same group. This approach can clearly also be used to account for heteroscedasticity across different population groups. As suggested by Munizaga et al. (2000), the model can be described as the Single Element Nested Logit (SENL) model.

Accounting for heteroscedasticity across individual respondents, as opposed to groups of respondents, poses a more formidable task, given that it becomes essentially impossible to estimate individual-specific scale parameters. Here, one example of a possible approach is that of the Heterogeneous Conditional Logit (HCL) model of Steckel & Vanhonacker (1988), where the scale parameter of the underlying Logit model is distributed across individuals according to a Gamma distribution, where a closed form version of the model is obtained with the use of an Exponential distribution. An alternative approach in this context is to relate the scale parameter to individual-specific attributes, as discussed by Swait & Adamowicz (1996, 2001) in the development of their Heteroscedastic Multinomial Logit (HMNL) model.

A discussion of the representation of heteroscedasticity, across observations as well as across alternatives, is given by Munizaga et al. (2000), who also offer a comparison of different model structures (MNL, NL, SENL, HEV and MNP) on simulated data with heteroscedasticity between two groups of alternatives, or two groups of respondents. The results show a relative level of robustness of the MNL model in the case of heteroscedasticity across observations, which is however not the case when dealing with heteroscedasticity across alternatives, where, in addition to the MNP model, the NL model offers good performance. The performance of the MNL model in the presence of heteroscedasticity across observations should however be put into context by the use of just two segments of the population, and the use of common coefficients in the two groups in the generation of the data.

2.8.2 The Universal Logit family

In the models described thus far in this chapter, the utility of an alternative i depends only on attributes of that alternative. A number of models, described as Universal Logit (or Mother Logit) models, have been developed which relax this restriction.

The Universal Logit (UL) model was first introduced by McFadden (1975), based on an underlying MNL model, with the probability for alternative i given by:

$$P_n(i) = \frac{e^{g_{i,n}}}{\sum_{j=1}^{I} e^{g_{j,n}}}, \tag{2.20}$$

where $g_{j,n}\,\forall j$ can be a function of the attributes of all alternatives in the choice set. Although the UL model has been used in several applications (e.g. McFadden et al. 1978, Timmermans et al. 1991, Lafferiere & Gaudry 1993), the popularity of the model has been hampered by the doubts expressed by McFadden et al. (1977) in relation to the model's consistency with utility maximisation, and the difficulty of finding an appropriate specification of utility (cf. Ben-Akiva 1974*a*).

Another example of a model belonging to this wider family is the Dogit model (introduced by Gaudry & Degenais 1978), which allows some or all of the ratios of alternatives to violate the *IIA* property, thus *dodging* the dilemma of choosing a priori between a model that is fully constrained by the *IIA* assumption and a model which is entirely free of the assumption. The Dogit model is also popular because of an alternative interpretation, which divides the market share of each alternative into two parts; one part (captivity level) satisfying a compulsive need by the decision-makers, and another part that is in competition with other alternatives. The choice probability for alternative i is then given by:

$$P_n(i) = \frac{1}{1+\sum_{j=1}^{I}\theta_{j,n}}\frac{e^{V_{i,n}}}{\sum_{j=1}^{I} e^{V_{j,n}}} + \frac{\theta_{i,n}}{1+\sum_{j=1}^{I}\theta_{j,n}} \tag{2.21}$$

where $\frac{\theta_{i,n}}{1+\sum_{j=1}^{I}\theta_{j,n}}$ in equation (2.21) is a (positive) captivity term, which satisfies a compulsive need for alternative i, while the remaining share $\frac{1}{1+\sum_{j=1}^{I}\theta_{j,n}}$ is spent according to the *MNL* model. This not only means that the minimum market share for an alternative is no longer necessarily zero, but also allows for two alternatives with identical attributes to have different market shares. The Dogit model has rarely been used in practice, with one example of an application being Gaudry & Wills (1978*b*).

The Dogit model was generalised by Swait & Ben-Akiva (1987), in the form of the Parameterised Logit Captivity (PLC) model (originally developed by Ben-Akiva 1977), in which the captivity level parameters are allowed to vary over observations, as a function of the attributes of the decision-maker and the alternative in the current observation. The development of this model is linked to the issue of choice set generation, from the point of view that the model does not represent a choice process on a free choice set, due to the presence of captivity (cf. Swait & Ben-Akiva 1987). Finally, Fry & Harris (2002) have recently proposed the Dogit Ordered Extreme Value (DOGEV) model, which combines the DOGIT model with an OGEV model, thus accounting simultaneously for an ordering of the variables and the presence of captivity.

Another model belonging to the UL family is the C-Logit model developed by Cascetta et al. (1996) in the context of route choice modelling. The C-Logit model uses a "commonality factor" which is subtracted from the utility of a certain path linking the origin and the destination. The "commonality factor" measures the similarity of the path with other possible paths linking origin and destination. The higher the "commonality factor" of a path, the lower will be the (individual) utility of this path, and hence also its choice probability.

With I being the set of possible paths linking origin and destination, we can measure the probability of decision-maker n choosing path i as:

$$P_n(i) = \frac{e^{V_{i,n}-CF_i}}{\sum_{j=1}^{I} e^{V_{j,n}-CF_j}} \tag{2.22}$$

where CF_i is the "commonality factor" for path i, and where different specifications for CF_i lead to different forms of the C-Logit model and hence also different choice probabilities (Cascetta et al. 1996).

2.8.3 Choice set generation models

Choice set generation models form an interesting class of sub-models in the field of discrete choice modelling. Such models acknowledge the fact that not all alternatives will be considered by all respondents. This can be achieved either in the context of the discussions in Section 2.8.2 in terms of an individual being captive to a specific alternative, or on the basis of the two-stage choice paradigm of Manski (1977), with probabilistic choice set generation in the first step, followed by the choice of an alternative from this choice set in the second step.

Aside from the PCL model (Swait & Ben-Akiva 1987) mentioned in the previous section and the GenL model (Swait 2001*a*) discussed in Section 2.5.1, models accounting for differences in choice set formation are discussed for example by Swait &

Ben-Akiva (1985), Andrews & Srinivasan (1995), and Ben-Akiva & Boccara (1995). The study of Basar & Bhat (2004) is of special interest in the current thesis, given its application to airport choice modelling, and will be touched on in more detail in Section 8.3.

2.9 Models using mixture distributions

Over recent years, the Mixed Multinomial Logit (MMNL) model has become one of the most widely used tools in the area of demand modelling. In this section, we first look at the existing body of work on MMNL models, before proceeding to the more general family of GEV mixture models.

2.9.1 The Mixed Multinomial Logit model

The first applications of the MMNL model came in the work of EPRI (1997), Boyd & Mellman (1980) and Cardell & Dunbar (1980). However, the MMNL model has only become widely used over the past five to ten years, as the cost of estimation was previously prohibitively high.

The choice probabilities in the MMNL model are calculated as the integral of MNL choice probabilities over the assumed distribution of random terms. Formally, whereas in the MNL model, the utility that decision-maker n gets from choosing alternative i is given by

$$U_{i,n} = V_{i,n} + \epsilon_{i,n}, \tag{2.23}$$

the corresponding utility in the MMNL model is given by:

$$U_{i,n} = V_{i,n} + \eta_{i,n} + \epsilon_{i,n}. \tag{2.24}$$

In both cases, the ϵ terms are assumed to be distributed *iid* extreme-value over alternatives and decision-makers. But, whereas in the MNL model, this leads to a closed form expression for the choice probabilities (cf. Section 2.3.1), the presence of the additional vector of error terms η_n (grouping together $< \eta_{1,n}, \ldots, \eta_{I,n} >$) in the MMNL model results in general in an integral without a closed form expression. As such, the model needs to be estimated with the help of numerical processes, where typically, simulation is used. This issue is not discussed here, but is the topic of Chapter 3.

In the MMNL model, the mean of η_n is set to be zero, and no *a priori* constraints exist on the distribution of η_n; the researcher is free to make an appropriate and convenient choice. The resulting model form is very flexible, and free of any restrictive assumptions, such as *IIA*. For a given choice of distribution $f()$ for η_n, with parameter vector Ω, the MMNL choice probability is given by:

$$P_n(i) = \int_{\eta_n} P_n\left(i \mid \eta_n\right) f\left(\eta_n \mid \Omega\right) \mathrm{d}\eta_n, \tag{2.25}$$

where $P_n\left(i \mid \eta_n\right)$ is the MNL choice probability for alternative i and decision-maker n, conditional on η_n.

Two conceptually different, yet mathematically equivalent modelling approaches can arise from this notation; the Random Coefficients Logit (RCL) model, and the Error-Components Logit (ECL) model. The former exploits the error structure to allow for random taste heterogeneity, while the latter allows for inter-alternative correlation and heteroscedasticity. The two approaches can also be combined to allow jointly for random taste heterogeneity, inter-alternative correlation, and heteroscedasticity. We will now look at the two approaches in turn.

In the RCL formulation, some elements in the parameter vector β_n used in the calculation of the utility are assumed to be randomly distributed rather than fixed, such that the error term η_n represents the deviation from the mean observed utility V_i caused by the fact that β_n is no longer the same for all decision-makers. As such, we obtain:

$$P_n(i) = \int_{\beta_n} P_n\left(i \mid \beta_n\right) f\left(\beta_n \mid \Omega\right) \mathrm{d}\beta_n, \tag{2.26}$$

where $P_n\left(i \mid \beta_n\right)$ is now the MNL choice probability for alternative i and decision-maker n, conditional on the vector of taste-coefficients β_n, which is distributed across the population according to $f\left(\beta_n \mid \Omega\right)$. Generally, a continuous distribution will be used for $f\left(\beta_n \mid \Omega\right)$; the use of discrete distributions is discussed in detail in Chapter 5. This specification can also be adapted to allow for repeated choice (i.e. panel data), where the typical approach, which is addressed in several places in this thesis, is to assume that the tastes vary across respondents, but stay constant across the observations for a given respondent. Along the same line, it is possible to account for the effects of past experience on decision-makers' tastes, a subject that is however addressed less frequently in the existing literature. With the *random coefficients* specification, the crucial part of the model formulation is the choice of which coefficients are to be randomly distributed, and the choice of what random distributions should be used. This issue is the topic of extensive discussions in Chapter 4, which also includes a summary of existing work. Applications of the RCL formulation include Algers et al. (1998), Revelt & Train (1998), Bhat (2000*b*), and Han et al. (2001).

The second (and less frequent) interpretation of the MMNL model is that of the *error-components* specification. Here, the utility of an alternative i for decision-maker n is rewritten as:

$$U_{i,n} = V_{i,n} + \gamma_n' z_{i,n} + \epsilon_{i,n}, \tag{2.27}$$

where $\epsilon_{i,n}$ and $V_{i,n}$ are defined as in the context of the corresponding MNL model. The two additional terms in the utility function define the error-components structure; γ_n is a vector of random coefficients with a mean of zero and a covariance matrix Σ, which is generally specified to be diagonal, and where typically, a Normal distribution is used, while $z_{i,n}$ is a vector of 0 and 1 terms which determine what error-components enter the utility of alternative i. The choice probability for alternative i is then obtained by integration over the distribution of γ_n, with:

$$P_n(i) = \int_{\gamma_n} \frac{e^{V_{i,n}+\gamma_n' z_{i,n}}}{\sum_{j=1}^{I} e^{V_{j,n}+\gamma_n' z_{j,n}}} \phi\left(\gamma_n \mid 0, \Sigma\right) \mathrm{d}\gamma_n, \tag{2.28}$$

where $\phi(\gamma_n \mid 0, \Sigma)$ is the joint Normal density function of the elements in γ_n, with mean 0, and covariance matrix Σ.

If a given element is equal to 1 jointly in $z_{i,n}$ and $z_{j,n}$, correlation is introduced between the error-terms of the utilities for the two alternatives, with the extent of correlation depending on the variance of the associated error-component term in γ_n. The inclusion of different error-components in different utilities induces heteroscedasticity, where *controlled* heteroscedasticity can be introduced for single alternatives by ensuring that the associated dummy term in $z_{j,n}\,\forall j$ is equal to 1 only for the concerned alternative i[16]. It can be seen that, with $z_{j,n}$ containing only zero entries $\forall j$, the model reduces to MNL. In practice, such a treatment of correlation by an ECL structure can come at the cost of a high number of dimensions of integration, as well as important issues of identification, which are addressed in detail by Walker (2001), Walker et al. (2003), and also Bowman (2004). Two recent applications of the ECL formulation are given by de Jong et al. (2003), and Hess, Polak, Daly & Hyman (2004).

Finally, a number of applications have also looked at incorporating deterministic heterogeneity components into the distribution of the random terms, either in the mean or the standard deviation, hence allowing the modeller to relate the variation of random coefficients to individual-specific observed attributes. This can be useful in an RCL as well as ECL context. A recent example of such an approach is given by Greene et al. (2005).

2.9.2 GEV mixture models

As mentioned in Section 2.9.1, the RCL and ECL approaches can be combined straightforwardly, allowing for the joint modelling of random taste heterogeneity and inter-alternative correlation. This however comes at the cost of important issues in identification, and heightened cost of estimation and application when using error-components for the representation of correlation. Furthermore, although the MMNL model has the theoretical property of being able to approximate other random utility models arbitrarily closely[17], this may not always be as straightforward in practice (cf. Garrow 2004).

While integration over mixture distributions is necessary in the representation of continuous random taste heterogeneity, this is not strictly the case for inter-alternative correlation. Indeed, just as, conditional on a given value of the taste-coefficients, a RCL model allowing for random taste heterogeneity reduces to a MNL model, a model allowing for inter-alternative correlation in addition to random taste heterogeneity can in this case be seen to reduce to a given GEV model (assuming that an appropriate GEV model exists). As such, the correlation structure can be represented with the help of a GEV model, while the random taste heterogeneity is accommodated through integration over the assumed distribution of β. The use of the choice probability of a more complicated GEV model as the integrand in equation (2.26) leads to a more general type of a GEV mixture model, of which the

[16]The issue of *controlled* versus *uncontrolled* heteroscedasticity is discussed in more detail in the context of the ECL applications in Appendix A.

[17]As noted by Garrow (2004), although generally attributed to McFadden & Train (2000), these approximation qualities were already discussed earlier, by Dalal & Klein (1988).

RCL model is simply the most basic form. Applications of this approach include for example Chernew et al. (2002), Bhat & Guo (2004) and Hess, Bierlaire & Polak (2005*a*). In such a GEV mixture model, the number of random terms, and hence the number of dimensions of integration (and thus simulation) is limited to the number of random taste coefficients, whereas, in the ECL model, one additional random term is in principle needed for the representation of each separate nest. It should be noted that the potential runtime-advantage resulting from this difference in dimensions of integration only manifests itself beyond a certain number of nests, as the more complicated form of the integrand in GEV mixture models initially gives the ECL model a computational advantage. The use of GEV mixture models does however have another advantage over the use of the ECL model in that it avoids the issues of identification that are specific to this latter model form, although other identification issues, specific to the underlying GEV structure, may arise.

Finally, it should be noted that while the error-components method has historically only been used with a MNL model as the basis, the approach can theoretically also be used when the underlying model is of GEV form, for example in the case where some correlation is to be captured by the GEV structure, with a remaining amount of correlation (or indeed heteroscedasticity) to be explained by the error-components. This can be useful in the case where existing GEV structures are incapable of capturing the full array of correlation in the data, while the exclusive reliance on error-components would lead to excessive computational cost or issues of identification. The work presented in this thesis concentrates on the use of random coefficients GEV mixture models; the analysis of the potential of advanced error-components GEV models (not based on MNL) is an important area for future research.

2.10 Discussion

The review of existing work presented in this chapter has highlighted the rapid pace of theoretical developments in the field of discrete choice modelling, especially over the past few years. It has also shown that the work has pursued two main, and quite separate directions, namely the development of ever more flexible GEV structures, and the further investigation of models based on mixture distributions.

In the face of this rapid pace of developments, and the high number of different angles of research, the discussion presented in this chapter has focussed only on the main developments. Many interesting extensions are possible, and have been pursued, and these cannot all be discussed here. One such extension of the basic framework is the treatment of cutoffs, the inclusion of which moves us away from a purely compensatory model. Recent discussions of this topic, which can be related to choice set formation, include (Swait 2001*b*) and Cantillo & Ortúzar (2005), who allow the thresholds to vary across respondents. Other extensions of interest include an incorporation of learning effects into the models, or the modelling of group decision-making.

The theoretical progress in the area of discrete choice modelling is set to continue, for example with the development of mixture models based on more advanced GEV models, and the exploration of hybrid structures.

Despite the excitement about the development of new, more flexible model structures, a word of warning is required. Indeed, with the gains in flexibility also comes an increased risk of misspecification and misinterpretation. This applies especially in the context of mixture models, and with the surge in popularity of the MMNL model have also come a number of warnings as to the potential risks involved with the use of such a powerful model structure, where an uninformed specification can lead to misleading results. This extends particularly to the choice of mixture distributions in random coefficients modelling, an issue that is central to this thesis. Discussions of the risks involved are given for example by Walker (2002) and Hensher & Greene (2003), while Munizaga & Alvarez-Daziano (2002) also urge for caution and an informed choice of specification, especially in terms of covariance structure in the ECL formulation.

It must be stressed that the risks of misspecification apply not just with mixture models, but can similarly cause problems in the case of advanced GEV models, where important issues with identification can arise, as illustrated in the use of cross-nesting structures in Chapter 10. Furthermore, in both cases, the gains in flexibility need to be put into context by the often significant increases in the cost of estimation.

Another issue is that, with the pace of theoretical developments, models are often only tested in a handful of applications, primarily by the actual authors developing the structure. This leads to a lack of insight into the general performance of given model structures when faced with a variety of problems. The same issue extends to the testing of the robustness of the models in the face of misspecification. Additionally, it should be noted that the sheer number of different models, especially in the GEV context, can lead to problems in deciding which model to use for a given problem, as the different structures are often very similar to each other. Finally, it should also be noted that more work is required to determine whether advanced models *consistently* lead to different policy conclusions than more basic models, along the lines of the work of Viton (2004) in the context of the MMNL model.

Despite these issues, the new models have the potential of offering a more realistic representation of real-world behaviour, and as such should be preferred, if the issues of applicability can be overcome. This however leads us to the final point, which relates to the gap between the state-of-the-art and the state-of-practice. Indeed, it should be noted that, despite the rapid theoretical developments, only a handful of models, namely MNL, NL, and more recently MMNL have found widespread use in actual practical modelling work, where even CNL models are used only sparsely. As such, it can be seen that the pace of theoretical developments much exceeds the actual implementation and application of models in practice. Here, more effort should go into large-scale testing, education, and improved implementation of the advanced models, given their potential benefits when applied to real-world problems. As such, while technical endeavours are laudable, the raison-d'être of research should at least in part be seen to be the development of techniques that improve the realism of modelling processes used to assist policy-makers in their strategic decisions. Here, the advanced models clearly have a role to play, such that more should be done to encourage their widespread use. This reasoning forms part of the motivation for the choice of air-travel as the main area of application in this thesis.

Chapter 3

Efficiency in simulation-based estimation and application

3.1 Introduction and context

The requirement to use simulation[1] in the estimation of discrete choice models whose choice probabilities do not posses a closed form expression is well-documented in the existing literature (see Section 2.9 and Train 2003). Although often not mentioned, it should be stressed that this requirement to use simulation extends not just to the estimation of such models, but also to their application. While only a single run of the simulation-process is generally required in each model application, the aim is often to analyse a number of different policy scenarios, each requiring a separate application of the model, hence leading to substantial computational costs. As such, the more advanced model structures are at present not seen as tools that could readily be implemented in large-scale forecasting systems. This means that reductions in the computational cost of simulation processes are crucial not just from the point of view of estimation, but indeed also application.

This chapter discusses ways of improving the efficiency of simulation-processes for mixture models, focussing on the use of alternatives to classical Monte-Carlo integration. Such alternative approaches, broadly referred to as quasi-Monte Carlo approaches, can lead to a more accurate approximation of integrals, hence leading to a lower requirement in terms of the number of draws used in these processes, with consequent reductions in computational cost. In the context of discrete choice modelling, only one type of approach, namely the Halton sequence, has received widespread exposure, mainly due to its advantages in terms of implementation. However, the Halton sequence has severe limitations in high-dimensional applications, which have been discussed at length, for example by Bhat (2003).

The main aim of this chapter is to propose an alternative approach, referred to as Modified Latin Hypercube Sampling (MLHS), which is free of the problems exhibited by Halton sequences, while still maintaining the advantages in terms of easy implementation. Additionally, the chapter aims to further investigate the problems

[1] It should be noted that there are alternative approaches for solving integrals that do not have a closed form solution. These are however only slowly being introduced to the field of discrete choice modelling (e.g. Breffle et al. 2005), where simulation remains the standard approach. As such, alternative approaches are excluded from the discussion presented in this chapter.

with Halton sequences, and the various approaches proposed to address these problems. Here, the description of one such approach, referred to as the Shuffled Halton sequence, and introduced into the area of discrete choice modelling as part of the research described in this thesis, serves as the stepping stone in the development of the MLHS approach.

The remainder of this chapter is organised as follows. We first look at the need for simulation in the use of GEV mixture models (Section 3.2), and discuss the use of alternatives to standard Monte Carlo integration, in the form of quasi-Monte Carlo approaches (Section 3.3). Sections 3.4 and 3.5 look at the most commonly used type of quasi-Monte Carlo approach, the Halton sequence, in its original as well as adapted versions. The development of the MLHS approach is described in Section 3.6, and Section 3.7 presents two empirical applications. After a discussion of some more advanced quasi-Monte Carlo approaches in Section 3.8, the conclusions of the chapter are presented in Section 3.9. To a large extent, this chapter is based on material published in Hess et al. (2003) and Hess, Train & Polak (2005).

3.2 Simulation processes in estimation and application

Using a notation slightly adapted from that used in Section 2.9.1[2], we let $P_n\left(i \mid \beta, \eta\right)$ define the probability of an underlying GEV model, say MNL, conditional on η, which is a vector of random variables included in the utility functions in addition to the usual extreme-value terms. From this, we get:

$$P_n(i \mid \beta, \Omega) = \int_{\eta} P_n\left(i \mid \beta, \eta\right) f\left(\eta \mid \Omega\right) \mathrm{d}\eta, \tag{3.1}$$

to be the unconditional (on η) probability of alternative i for respondent n, obtained through integration over the distribution of η, where, in addition to β, the resulting probability is now conditional on a given value for Ω, the vector of parameters for the distribution of η.

In discrete choice models, the choice probabilities of the individual alternatives, with a given set of parameters, are a requirement in model estimation as well as application. In the absence of a closed form solution for the integral in equation (3.1), the value of these choice probabilities needs to be approximated with the help of numerical processes, where typically, simulation is used.

The process by which an integral is approximated through simulation is the same in estimation and in application, though certain external factors change, notably the number of simulation iterations. Here, an iteration is defined to be a separate evaluation of the integral in equation (3.1), with a different value for the vectors β and Ω.

The integral given in equation (3.1) is of the form:

$$\bar{g} = \int_{x} g\left(x\right) f\left(x\right) \mathrm{d}x, \tag{3.2}$$

[2]Here, we additionally show the dependency on β, which is to define a vector of fixed coefficients, constant across individuals.

where x is distributed according to $f(x)$. The calculation of $\bar{g}$ involves the averaging of $g(x)$ over the entirety of the continuous domain of $f(x)$. When it is not possible to solve the integral directly, its value can be approximated by a process referred to as Monte-Carlo integration (*MCI*). In this, the value of $g(x)$ is averaged over a finite set of draws from $f(x)$.

Going back to the example at hand, in *MCI*, the integral of $P_n(i \mid \beta, \eta)$ over the continuous domain of η is replaced by a summation over a finite set of realisations of the vector η, each carrying equal weight. Formally, with $\eta_r, r = 1, \ldots, R$ representing R independent draws from $f(\eta \mid \Omega)$, the integral representing the choice probability in equation (3.1) can be approximated as:

$$\widetilde{P_n}(i \mid \beta, \Omega) = \sum_{r=1}^{R} \left[\frac{1}{R} P_n(i \mid \beta, \eta_r) \right]. \tag{3.3}$$

Each independent draw from $f(\eta \mid \Omega)$ carries the same weight, $\frac{1}{R}$, such that this term can be placed outside of the summation. Typically, the draws from $f(\eta \mid \Omega)$ are based on appropriate transformations of uniform draws contained in the $0-1$ interval, where the required procedures are described in detail by Train (2003, Chapter 9). Essentially, in the case of univariate densities $f(\varepsilon)$ of a form where the corresponding cumulative distribution function $F(\varepsilon)$ is invertible, the inverse cumulative distribution function can be used, such that a draw from $f(\varepsilon)$ is given by $F^{-1}(\mu)$, where μ is a standard uniform variate. The situation becomes more complicated in the case of truncated distributions, where the standard uniform draw needs to be transformed so as to fall into the permissible range. Further complications arise in the case of multivariate distributions. While, in simple scenarios (e.g. multivariate Normal), a Choleski transformation can be used, there are cases in which more advanced approaches, such as accept-reject simulators, importance sampling, Gibbs sampling (Geman & Geman 1984), or even the Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970) are required. Such cases are not discussed in this work.

With the use of independent draws $\eta_r, r = 1, \ldots, R$, the estimator in equation (3.3) is unbiased by construction. The variance of the estimator is inversely proportional to the number of draws used, R, such that, as R increases, the simulation error decreases. The use of a finite number of draws invariably results in some simulation error; this is however not avoidable in the absence of a closed form solution.

The above discussion applies directly in the case of model application. Some further clarifications are required in the case of estimation. The purpose of estimation is the calibration of the model parameters on a given set of choice data, during which process the value of the model parameters is chosen so as to maximise the likelihood of the choices observed in the dataset.

For now, we will ignore the special case of repeated choices by the same respondent, to which we will return below. Let j_n define the choice observed for individual n, with $n = 1, \ldots, N$. In a fixed parameter model, let $P_n(j_n \mid \beta)$ give the probability of the choice observed for individual n, conditional on the vector of fixed model parameters β. The log-likelihood[3] of the model given the choices observed in the

[3]For practical (numerical) reasons, it is generally preferable to replace the likelihood function

data, conditional on β, is then given by:

$$\begin{aligned}\mathcal{LL}\left(\beta\right) &= \ln\left(\mathcal{L}\left(\beta\right)\right) \\ &= \sum_{n=1}^{N} \ln\left(P_n\left(j_n \mid \beta\right)\right),\end{aligned} \tag{3.4}$$

with $\mathcal{L}\left(\beta\right) = \prod_{n=1}^{N} P_n\left(j_n \mid \beta\right)$.

Typically, *Maximum Likelihood Estimation* (*ML*) is used in the estimation of discrete choice models, where, at the *Maximum Likelihood Estimator* (*MLE*) $\hat{\beta}$, we have:

$$\frac{\partial \mathcal{LL}\left(\hat{\beta}\right)}{\partial \hat{\beta}} = 0 \tag{3.5}$$

The extension to the case of models with random parameters is straightforward. In the presence of choice probabilities of the form given in equation (3.1), we use *MCI* simulation as in equation (3.3) to approximate the choice probability $P_n\left(j_n \mid \beta, \Omega\right)$ for each n.

The log-likelihood function in equation (3.4) is now replaced by the simulated log-likelihood, given by:

$$\mathcal{SLL}\left(\beta, \Omega\right) = \sum_{n=1}^{N} \ln\left(\widetilde{P_n}\left(j_n \mid \beta, \Omega\right)\right), \tag{3.6}$$

where, in addition to β, this is conditional on the vector Ω, giving the parameters of the distribution of the random terms η. For ease of notation, let Θ define a vector grouping together the fixed model parameters and the parameters of the distribution of the random parameters, such that $\Theta = \left(\beta, \Omega\right)$. The *MLE* is now replaced by the *Maximum Simulated Likelihood Estimator* (*MSLE*), where in this case, we have:

$$\frac{\partial \mathcal{SLL}\left(\hat{\Theta}\right)}{\partial \hat{\Theta}} = 0 \tag{3.7}$$

It has been shown (Lee 1995) that if R is increased faster than the square root of the number of observations ($\sqrt{N}$), *Maximum Simulated Likelihood* (*MSL*) estimation is asymptotically equivalent to *ML* estimation[4]. Just as was the case in model

by the log-likelihood. Because the logarithm is an increasing function, the maxima occur at the same parameter values.

[4] While the vast majority of applications using mixture models rely on MSL for model estimation, it is worth noting that there are two main alternative options that could be used instead of MSL, namely the Method of Simulated Moments (MSM) proposed by McFadden (1989), and the Method of Simulated Scores (MSS) proposed by Hajivassiliou & McFadden (1998). Both of these approaches have advantages and disadvantages when compared to MSL, as discussed for example by Train (2003), and it is not the aim of this thesis to discuss these alternative methods in great detail. However, one point is worth noting. By the nature of the models under investigation, simulation is required with the use of either of these three approaches. As such, it can be seen that the developments presented in this chapter are applicable to all three methods, and are not limited to MSL. Given that at least in some cases, MSM and MSS have advantages over MSL, their

application, the use of a fixed number of draws R induces simulation bias and variance. While in application, if uncontrolled, this can lead to biased forecasts, in estimation, it can lead to biased estimates of model parameters, which in turn will lead to problems in model application. It is thus crucial to minimise the potential effects of simulation error. This can be achieved by the use of a sufficiently high number of draws, R, and hence steps in the simulation process.

The problem is that, in model estimation, a very high number of iterations are potentially required, each involving the simulation of the probabilities of all observed choices. Increases in the cost of individual runs (iterations) of the simulation algorithm are thus multiplied by the number of iterations used by the optimisation algorithm. This makes the use of a very high number of draws impractical in many cases. The exploration of alternative approaches is the main topic of this chapter.

Before moving on to the issue of the efficiency of simulation processes, we will briefly look at the case of repeated choices by the same individuals. In this case, we no longer model independent choices, but sequences of choices for individuals. In the most basic treatment of repeated choice, we assume that the tastes vary across individuals, but stay constant across observations for the same individual. For this, we replace the individual choice probabilities in the log-likelihood function by the probabilities of the observed sequence of choices. Specifically, let $n_1, \ldots, n_{T_n}$ give the observed choices for individual n, across the T_n choice situations faced by that respondent. We can then write the probability of the observed sequence of choices for individual n as:

$$L_n(\beta, \Omega) = \prod_{j=n_1}^{n_{T_n}} P_n(j \mid \beta, \Omega), \tag{3.8}$$

with $P_n(j \mid \beta, \Omega)$ given by equation (3.1). The simulated analogue of this probability is given by:

$$SL_n(\beta, \Omega) = \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{j=n_1}^{n_{T_n}} P_n(j \mid \beta, \eta_r) \right], \tag{3.9}$$

where the order of summation and product is crucial. This term is then used to replace the basic simulated choice probabilities $\widetilde{P_n}(j_n \mid \beta, \Omega)$ in equation (3.6), giving:

$$\mathcal{SLL}(\beta, \Omega) = \sum_{n=1}^{N} \ln(SL_n(\beta, \Omega)) \tag{3.10}$$

It is worth mentioning that this specific treatment of repeated choice applies only in the case of random parameter models; in the absence of the summation over draws in equation (3.9) (i.e. R=1), the log-likelihood function is identical for the panel (repeated choice) and cross-sectional cases.

wide-scale use in the estimation of discrete choice models remains an important avenue for future research. Another topic of interest which is not discussed here is the use of Bayesian instead of classical approaches for estimation (cf. Train 2001). The reliance of such methods on random draws however again means that the use of alternatives to standard pseudo-random draws is possible.

The above specification makes the considerable assumption that the tastes of decision-maker n stay constant across the T_n choice-situations faced by this respondent. It can easily be argued that this is preferable to the case where the repeated choice information is not taken into account at all, equating to an assumption of equal inter-agent and intra-agent variations in tastes. Although the vast majority of applications do indeed rely on the above described specification, there are cases where its underlying assumption may not be fully satisfied, such as for example with RP data collected over a long period of time, where tastes can indeed be expected to vary, or SP surveys with a high number of choice situations, where factors such as fatigue, habit formation and learning need to be taken into account. To some extent, such effects can be represented with the help of dummy variables that do not lead to an increase in the dimensionality of the integral. However, there are limits to what can be achieved with such an approach, and in many cases, it may not be sufficient. One example of a more flexible approach is given by Train (2003, page 151), where the tastes for a given individual are serially correlated across observations. While this increases flexibility, it also leads to a requirement for additional levels of integration, such that the merit of such approaches needs to be evaluated on a case-by-case basis.

As a final point, it should be noted that, in order to avoid correlation in simulation error across individuals, separate sets of draws are generally used across individuals[5]. As such, it can be seen from equations (3.6) and (3.10) that, with K dimensions of integration, $\left(\sum_{n=1}^{N} T_n\right)$ K-dimensional sets of R draws are required when using a cross-sectional approach despite the presence of multiple choices per individual, while only N such sets are required in the panel data approach. This does however not guarantee a lower computational cost for the panel approach, given the more complicated form of the derivatives of the log-likelihood function.

3.3 Alternatives to standard Monte Carlo integration

As noted in Section 3.2, the use of a very high number of draws in MCI may not always be practical or even possible, due to the heightened computational cost it engenders. On the other hand, the use of a comparatively low number of pseudo-random (also known as Pseudo-Monte Carlo, or PMC) draws can lead to high simulation error, and biased results, which is similarly unacceptable. This is caused primarily by the fact that, especially in short sequences, the *randomness* of PMC draws will lead to an uneven distribution of draws across the area of integration, along individual dimensions, as well as the distribution in the unit hypercube. This phenomenon, which is also referred to as low uniformity or poor quality of coverage, translates directly into poor coverage for the draws obtained after transformation to the appropriate distribution. The uneven distribution of draws means that the simulation will assign unequal weight to different parts of the domain of the (density) functions used inside the integrals, resulting in a biased approximation. As such, it

[5] A discussion of the properties of the MSLE in the case where the same draws are used across all observations is given by Lee (1992)

is desirable to use alternatives to PMC draws that offer a better quality of coverage of the $0-1$ space, in single as well as multiple dimensions.

Another point worth considering is that of correlation between successive draws. It can be seen that, in the presence of negative correlation across draws, the variance of the simulator is lower[6]. As noted by Train (2003), the same concept also applies in the summation of separate simulators, as in a simulated log-likelihood function, where negative correlation in the draws across observations/individuals leads to lower variance for the simulated log-likelihood function. It can easily be seen that, when looking at individual simulated terms, the issues of covariance across draws and coverage of the $0-1$ area are inter-related, with lower variance equating to better coverage. This is most aptly described by noting that with negative correlation, a draw in one extreme of the domain will be compensated by a draw in the other extreme of the domain, a principle that is used directly in the case of antithetic draws, as described by Hammersley & Morton (1956). Clearly, by being uncorrelated, PMC draws are not able to exploit the advantages of negative correlation.

From the above discussion, it can be seen that the use of alternatives to PMC approaches in Monte Carlo integration can have great benefits in estimation and application. Although not strictly applicable in all cases, the general term Quasi-Monte Carlo (QMC) integration can be used to describe this set of approaches. QMC sequences are designed in a deterministic fashion, with the aim of providing more uniform coverage of the area of integration and negative correlation across draws. The use of such QMC sequences can lead to significant improvements in the precision of the simulated probabilities, and consequently MSL estimation, hence leading to lower requirements in the number of draws used, with corresponding reductions in the computational cost of model estimation and application.

In the following sections, we look at existing work and new developments in terms of QMC approaches that are applicable in the simulation-based estimation and application of mixture models, such as MMNL. The work centres mostly on quality of coverage offered by these sequences, as opposed to an in-depth discussion of negative correlation across draws. Here, we look solely at the case of a uniform distribution in the unit hypercube, which is the base for any transformations to other domains and distributions. It should be noted that this work does not aim to offer a definitive answer to the problem of simulation efficiency, or even define a near-optimal type of drawing algorithm. This belongs firmly to the domain of computational statistics; the work set out in this chapter merely aims to discuss approaches that offer gains in efficiency in the context of discrete choice modelling, while remaining easy to implement and widely applicable.

3.4 Halton sequences

The only widely used type of QMC series in the field of discrete choice modelling is the Halton sequence (Halton 1960), introduced to this area by Bhat (2001), with some examples of applications given by Revelt & Train (1999), Bhat (2000*a*) and Hensher (2001*a*). Halton sequences are generated according to a purely determin-

[6] An illustration of this is given by Train (2003, page 218).

istic approach based on the use of prime numbers. Specifically, a one-dimensional sequence based on prime p (≥ 2) fills the $0-1$ space by dividing this space into p segments, and by systematically filling in the empty spaces, using cycles of length p that place one draw in each segment. The use of prime numbers as the base of the Halton approach reduces the problems with collinearity, by avoiding the case where the cycle-length of one sequence is an integer multiple of the cycle-length of another sequence.

Formally, the i^{th} element in the Halton sequence based on prime p is obtained by taking the radical inverse of integer i in base p by reflection through the radical point. We have:

$$i = \sum_{l=0}^{L} b_l(i)\, p^l, \tag{3.11}$$

with $0 \leq b_l(i) \leq p-1$ and $p^l \leq i \leq p^{L+1}$, and use the values for $b_l(i)$ that solve equation (3.11), in writing the resulting Halton element in base p as:

$$\varphi_p(i) = 0.b_0(i)\, b_1(i) \ldots b_L(i)\,. \tag{3.12}$$

This can be rewritten in decimal form as:

$$\varphi_p(i) = \sum_{l=0}^{L} b_l(i)\, p^{-l-1}. \tag{3.13}$$

Aside from offering good quality of one-dimensional coverage, Halton sequences additionally have the advantage of negative correlation - later draws fill in the spaces left by existing draws. This is a direct result of the cyclical nature of the sequences. This also leads to negative correlation in the sets of draws used across individuals/observations, resulting in a correction effect, and a reduction in simulation error.

A problem caused by the deterministic nature of QMC sequences is that it is not possible to practically estimate the simulation error obtained when using these sequences. As Bhat (2003) notes, theoretical results can be used for this purpose, but they tend to lead to conservative bounds on the estimation error. As such, the use of randomisation approaches is preferable; here, the simulation error is calculated by measuring the deviation between the results obtained with differently randomised versions of the original sequence. The deterministic Halton sequence can be randomised in several ways. Tuffin (1996) and Bhat (2003) suggest random shifting, which is implemented by adding the same random draw to all elements in the Halton sequence, and by subtracting 1 from any of the elements whose value exceeds 1 as a result of this process. Another procedure, suggested by Wang & Hickernell (2002), is to eliminate the first G elements of the sequence, where G is chosen randomly.

Multi-dimensional Halton sequences are constructed through combination of one-dimensional sequences generated from different primes. The same principle applies in the case of randomised or otherwise transformed versions of Halton sequences.

It can be observed that individual Halton sequences are highly correlated, especially at the start of the sequences, manifesting itself most visibly in the form of collinearity between the two sequences. A cure commonly suggested in the existing literature is to remove the first few points (say L) in each sequence, hence changing the starting position in the individual cycles, where, according to Train (2003), the number of draws removed should be at least as large as the largest prime used in generating the sequences. While, depending on the value of L used, this does indeed remove the correlation at the start of the sequence, it is not true that the correlation does not reappear at later stages in the sequence. Indeed, the collinearity occurs in the case where, for a given draw from the multi-dimensional sequence, more than one of the one-dimensional sequences have just completed a cycle, and are about to start a new cycle.

Aside from offering high uniformity along individual dimensions, it can be observed that Halton sequences similarly offer good quality of multi-dimensional coverage when combining individual sequences, at least in the case of sequences based on low primes. It should be noted that this is in fact a direct result of the collinearity described above, a point that is often not recognised. Indeed, the way in which the individual sequences combine means that the unit hypercube is filled in with the help of diagonal lines (of collinear points) of varying length (which depends on the position in the individual one-dimensional cycles). An illustration of this is provided in Figure 3.1, which shows the two-dimensional Halton sequence generated from primes 7 and 11, for various sequence lengths. The plots clearly show how the multi-dimensional sequence is composed of individual diagonal lines of points, of varying length. While this leads to poor coverage when using a low number of draws, the uniformity of coverage increases as a higher number of draws is used. With the use of prime numbers whose ratio is not close to an integer, the average length of the multi-dimensional cycles is low relative to the length of cycles in the sequence based on the highest prime, a fact that leads to low correlation between sequences.

While the desirable properties of single and multi-dimensional Halton sequences based on low primes have received a lot of attention in the existing literature, the reason for this good behaviour (as described above) is not generally explained. As such, it is of little surprise that a link is not generally made between the good performance in low dimensions and the poor performance in high dimensions, at least in the discrete choice literature. Indeed, it is well-known that, when using high primes, the individual Halton sequences can be so highly correlated that, at least with the use of a moderate number of draws, the multi-dimensional sequences are formed by a set of diagonal lines and offer very poor quality of multi-dimensional coverage, which in turn can lead to poor simulation and estimation performance.

With the benefit of the above discussion on collinearity, it can be seen that the same principle of matching up of sequences that led to good quality of multi-dimensional coverage in low dimensions now leads to problems with correlation when using high primes. With the use of low prime numbers as the base, the relatively low number of perfectly collinear parts of the sequences means that the overall level of correlation between the sequences is at an acceptably low level. At the same time, the short individual cycle lengths means that the number of multi-dimensional cycles is sufficiently high to guarantee good quality of multi-dimensional coverage. With the use of high prime numbers, these advantages begin to unravel. Indeed, the

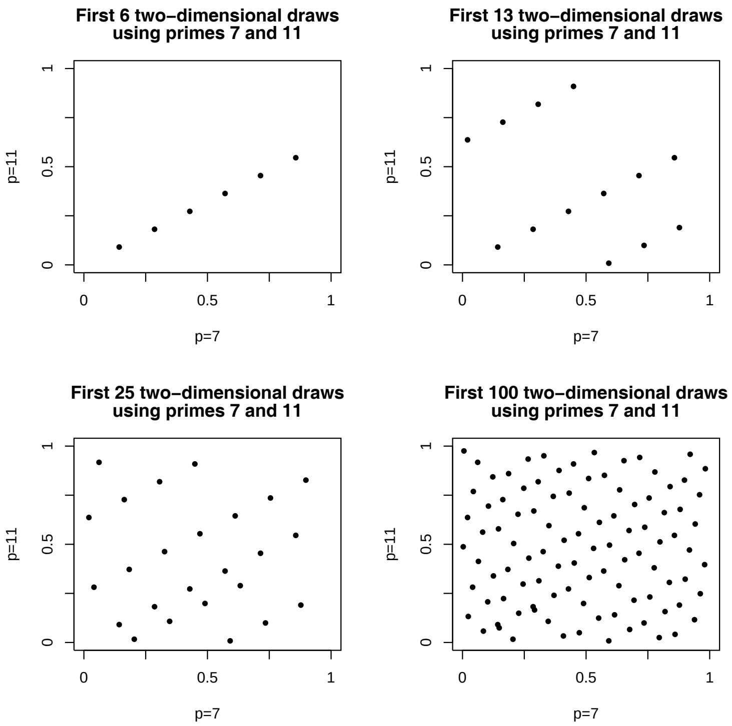


Figure 3.1: Illustration of two-dimensional filling-in effect using Halton sequences based on primes 7 and 11

length of the individual cycles now increases the scope for collinearity, leading to high correlation between individual sequences. The low number of multi-dimensional cycles in turn leads to a poor filling-in of the unit hypercube.

While most readily recognised with high primes, these problems actually arise to some degree whenever the ratio of any two prime numbers used is close to an integer value, as observed by Hess & Polak (2003*b*). Although Bhat (2003, page 841) similarly notes that "*the deterioration becomes clearly noticeable beyond five dimensions*", there still seems to be a wide-held belief that the problems with correlation occur only in high dimensions (e.g. well above 5), and untransformed Halton sequences continue to be used widely, something that is helped by the fact that most estimation packages for MMNL only offer a choice between PMC and standard Halton draws.

As an illustration of the correlation problems, Figure 3.2 shows four sets of two-dimensional Halton sequences, using different pairs of prime numbers, with three different sequence lengths (100, 200 and 500 draws), while Figure 3.3 shows three sets of corresponding PMC sequences using the various sequence lengths. It should be noted that, for the Halton sequences, the plots always show the start of the sequences; on the basis of the above discussion, it was decided not to use the approach of removing the initial part of the sequence. The conclusions in terms of correlation are not affected by this decision.

The plots clearly show the more uniform distribution in the Halton sequences based on the lowest possible pair of primes (2 and 3), when compared to the corre-

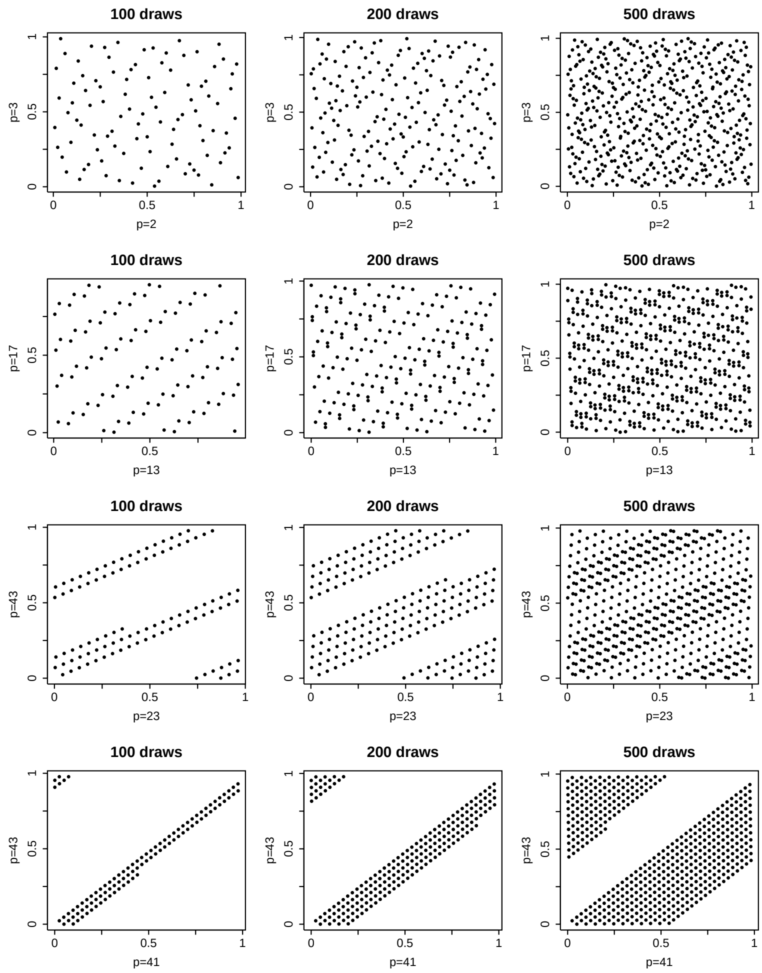


Figure 3.2: Two-dimensional Halton sequences for different sequence lengths and combinations of prime numbers

sponding PMC sequences. They also show that the differences are most visible in the case of a low number of draws; at higher values of R, the effects of the *random* distribution of the PMC draws are more or less cancelled out. This relates directly to the results of Train (1999) and Bhat (2001), who suggest that as little as $100 - 125$ Halton draws can offer better performance than 1000 PMC draws.

The plots in Figure 3.3 show that the use of a low number of PMC draws can result in clumping of draws in some areas of the multi-dimensional $0 - 1$ area, and consequently low coverage in other areas. The results also show the variability in performance of the PMC sequences over different runs. On the other hand, the plots for the Halton sequences based on higher primes clearly reveal how the high correlation between individual sequences can lead to very poor quality of coverage. In fact, it can be seen that in the presence of problems caused by heightened correlation in Halton sequences, the coverage offered by PMC sequences of corresponding length is visibly better. While the problems clearly increase in severity with higher primes, it can be seen that they are already substantive as of dimensions 6 and 7 (primes 13 and 17), making standard Halton sequences inapplicable in a high number of cases.

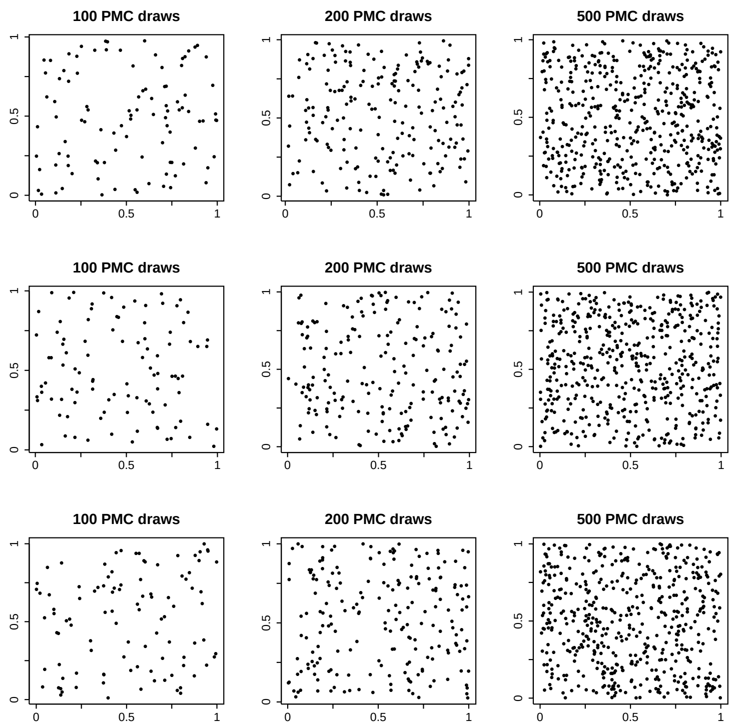


Figure 3.3: Two-dimensional pseudo-random number sequences for different sequence lengths

As a further illustration, Figure 3.4 shows two-dimensional plots for all combinations of primes between dimensions 5 and 15, with the commonly used sequence length of 100 draws. These plots reinforce the point that the problems with correlation are not limited to very high primes. Additionally, several cases clearly support the conclusions of Hess & Polak (2003*b*) with regards to primes whose ratio is close to an integer value (e.g. dimensions 7 and 8, 5 and 9, and 10 and 11). The problems with poor coverage are admittedly eased when using a higher number of draws, but even with relatively low primes, the required number of draws can be sufficiently high for the advantages of Halton sequences to disappear when compared to PMC sequences. Finally, it should be noted that, although the discussion here focusses on the two-dimensional case (for ease of illustration), problems also occur in the K-dimensional $0-1$ space, with $K > 2$.

Although, in the case of low primes, Halton sequences can lead to improvements in performance over PMC draws (and have been used successfully in different areas of research, including transport studies), the problems with correlation in higher dimensions limit their applicability. Furthermore, the above discussion has highlighted that the maximal allowable dimension for using Halton sequences may in fact be lower than suggested in the existing literature (where typically, authors have tended to discuss the problems above prime 41, i.e. dimension 13). Additionally, some existing results cast further doubt with regards to the performance of Halton draws. For example, Train (2003, pp.231-233) reports the results of a five-dimensional example using 100 and 125 Halton draws, where the first 5 primes were used, and

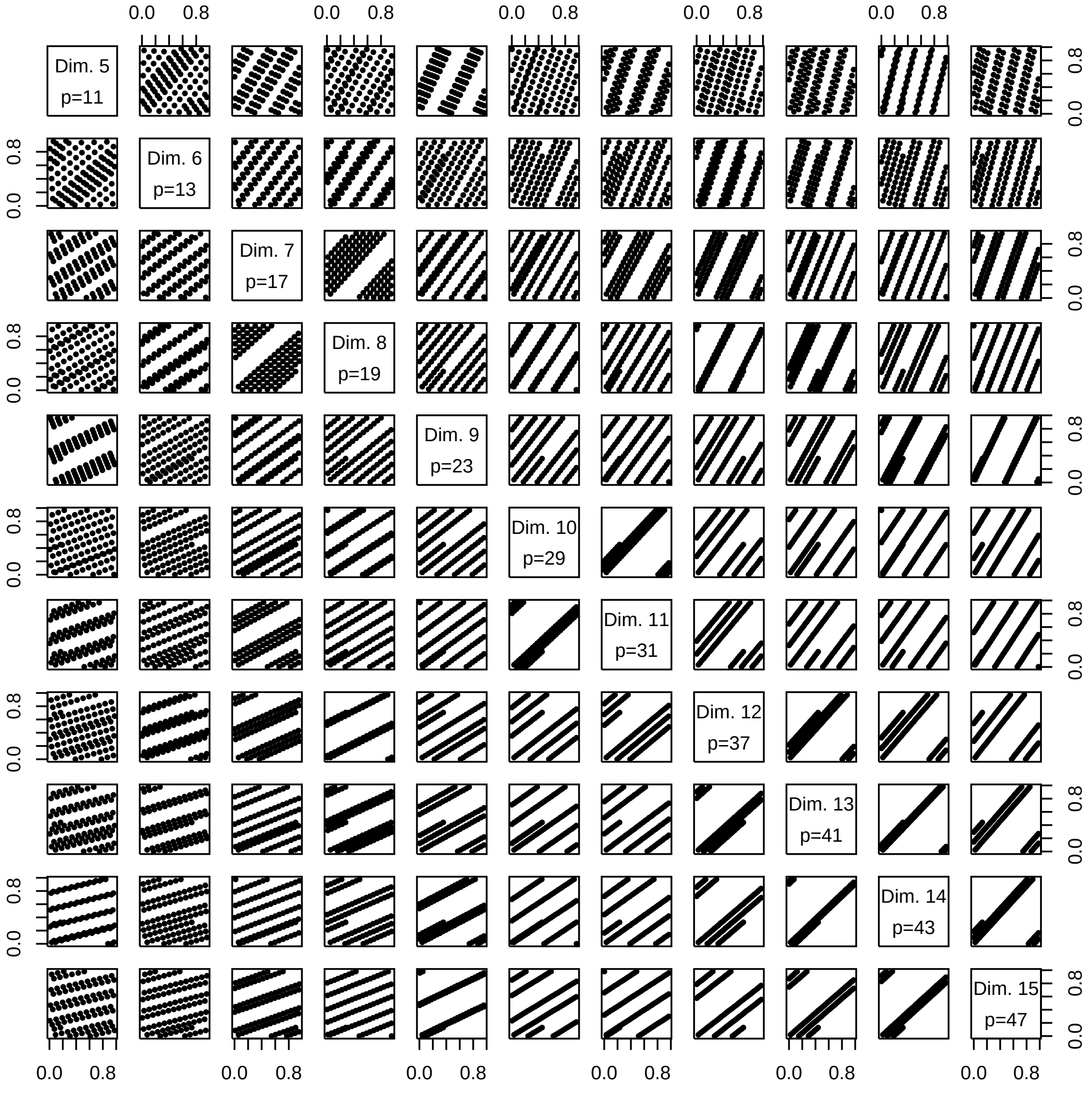


Figure 3.4: Correlation between individual one-dimensional Halton sequences for dimensions 5 to 15

where a different ordering of the primes was used in 5 separate runs. While the variation across runs was in both cases lower than that observed with $1,000$ PMC draws, there was greater variability when using 125 draws than when using 100 draws, with no apparent reason. This is contrary to the principle that a rise in the number of draws should lead to lower simulation error.

All these observations, in conjunction with the fact that Halton sequences are still gaining in popularity in the field of transport studies (where they remain the only wildly-used type of QMC approach), are a cause of concern. This is just furthered by the fact that seemingly an increasing number of studies rely on the use of a mere 100 draws or less per dimension and individual[7], which could be seen as a misguided interpretation of the results of Train (1999) and Bhat (2001) with regards to the superior performance with $100 - 125$ Halton draws when compared to 1000 PMC draws.

3.5 Adapted versions of Halton sequences

Two main methods for reducing the correlation in Halton sequences have been discussed in the discrete choice literature; scrambling and shuffling. We will now look at the two approaches in turn.

3.5.1 Scrambled Halton sequences

The scrambled Halton sequence for prime p is written as:

$$\varphi_{sp}(i) = \sum_{l=0}^{L} \sigma_p(b_l(i))\, p^{-l-1}, \tag{3.14}$$

where σ_p is the operator of permutations for the possible values of $b_l(i)$ in base p. Scrambling permutations aim at disrupting the cyclical nature of the Halton sequence while maintaining high uniformity of coverage. However, it has not proved possible to obtain optimal permutations, so that heuristic methods are used to derive *good* permutations, and these become increasingly onerous to derive as the prime p increases (cf. Hess & Polak 2003*a*). This can reduce the appeal of using scrambled Halton sequences in high-dimensional problems, if permutations for the appropriate number of dimensions are not readily available from previous research. Different methods for producing the permutations have been proposed in the literature, namely by Braaten & Weller (1979), Hellekalek (1984), Kocis & Whiten (1997) and Tuffin (1998). Scrambled Halton sequences were first discussed in the context of discrete choice modelling by Bhat (2003), using the scrambling approach proposed by Braaten & Weller (1979).

Bhat (2003) compares the performance of scrambled Halton sequences to that of standard Halton and PMC sequences using a 10-dimensional application. The results suggest that while both Halton-based approaches outperform the PMC se-

[7]There are examples in the transport literature of modellers relying on as little as $50-100$ Halton draws per individual and per dimension in applications with as many as 8 randomly distributed coefficients.

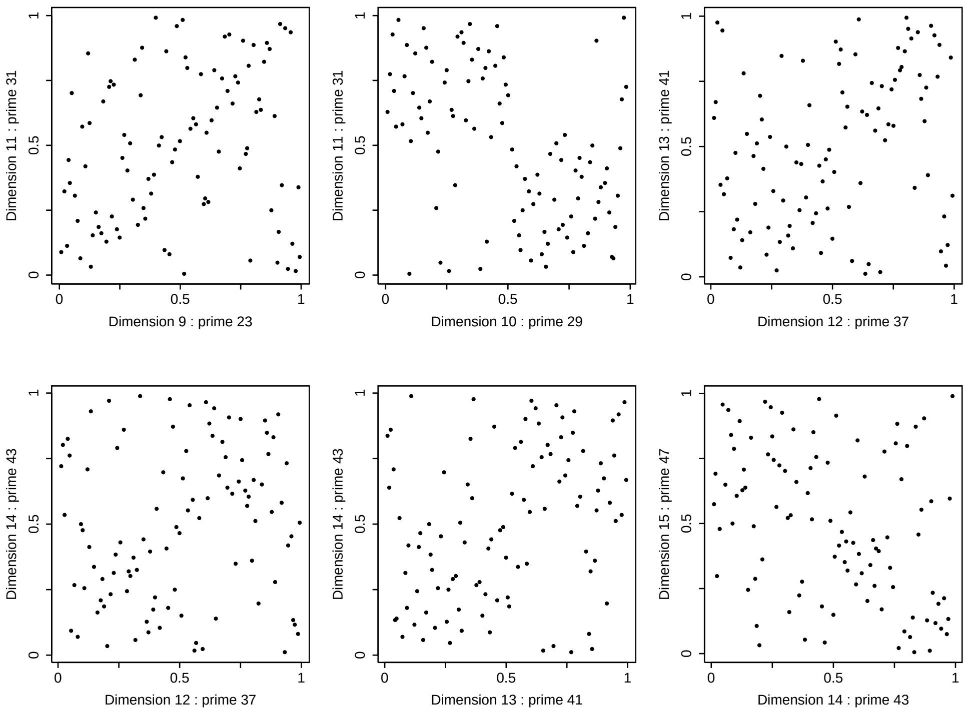


Figure 3.5: Correlation in two-dimensional scrambled Halton sequences

quences, the scrambled Halton sequences also offers lower bias in the estimated parameters than the standard Halton sequence, as well as lower simulation error (evaluated on the basis of multiple randomised versions of the sequences).

At the same time however, it has been observed that, while the scrambling approach generally achieves a significant reduction in the level of correlation, for some choices of prime numbers, the performance of the method is unsatisfactory, with high residual correlation and very uneven coverage (Hess & Polak 2003*a*). This is illustrated in Figure 3.5, which shows plots of six two-dimensional scrambled Halton sequences, generated from primes between 23 and 47. These plots show that, for some choices of primes, the use of the scrambling approach leads to a grouping of draws around either of the two diagonal lines, and subsequently poor coverage in two of the four corner areas. Although the problems again decrease with a higher number of draws, this remaining high level of correlation and poor coverage is still a cause for concern, and previous discussions have seemingly failed to address this.

3.5.2 Shuffled Halton sequences

The approach based on generating multi-dimensional sequences by combining randomly shuffled one-dimensional Halton sequences was first suggested by Morokoff & Caflisch (1994), and was introduced to the field of discrete choice modelling by Hess & Polak (2003*a*) in the context of the research described in this thesis.

The generation of a shuffled Halton sequence uses a one-dimensional standard

Halton sequence of length R, generated from prime p, as its input. Let:

$$H_p = \langle \varphi_p(1), \ldots, \varphi_p(R) \rangle, \tag{3.15}$$

where $\varphi_p(i)$ is defined as in equation (3.13). Differently shuffled versions of this sequence are then used in different runs of the shuffling algorithm, with the sequence used in the j^{th} run being given by:

$$H_{p,\sigma(j)} = \mathbf{H}(I_j, H_p), \tag{3.16}$$

where $\mathbf{H}$ is a function that creates a sequence with the elements of H_p arranged in the order given in the vector I_j, where I_j is a random permutation of the ordered index vector of length R, and where σ is a permutation operator that yields the permutation I_j of the index vector I in the j^{th} run. The use of different permutations of the index vector for different dimensions (for which different primes will still be used) disrupts the cyclical ordering in the different dimensions in different ways and hence manages to reduce correlation between the individual sequences.

There are no great difficulties involved in implementing the shuffling approach, and procedures to randomly permute the order of elements in a vector are in fact included in many numeric libraries. If no such procedure is readily available, the method can be implemented with just a few lines of computer code. The easiest and most efficient way of coding the shuffling of a vector v of length R seems to be to start with a vector of length R of uniformly distributed draws, and to create a vector w containing the ranking of the elements in this vector. The shuffled vector is then produced by drawing the elements from vector v according to the order given by vector w[8].

In theory, it is with this approach no longer necessary to base the generation of N K-dimensional shuffled Halton sequences of length R on the use of K one-dimensional Halton sequences of length NR generated from different primes. Indeed, the use of a different ordering for different individuals will lead to different sets of multi-dimensional draws, albeit still based on the same one-dimensional sequences. Even though this does entail significant computational savings, the use of different one-dimensional draws across individuals may be seen as preferable.

Although the performance of the shuffled Halton sequence is clearly dependent on the specific shuffling used in a given run, it has been shown that the method can offer stable performance over runs in the estimation of MMNL models (Hess et al. 2003). Compared to the scrambling approach, the shuffling approach disrupts the correlation between the one-dimensional sequences much further, to such an extent that the levels of correlation are comparable to those observed in multi-dimensional PMC sequences (cf. Hess & Polak 2003*b*). In the case where the cyclical nature of the individual sequences and its associated filling-in effect guarantees good multi-dimensional coverage in the standard sequences (i.e. low primes), the shuffling almost surely leads to a reduction in the quality of coverage, and poorer performance than with the standard (and possibly scrambled) sequences. On the other hand, in the case where the correlation has a detrimental effect (i.e. higher primes), the use of the shuffling approach can lead to better coverage than that obtained with the

[8] I am grateful to William Greene for this suggestion.

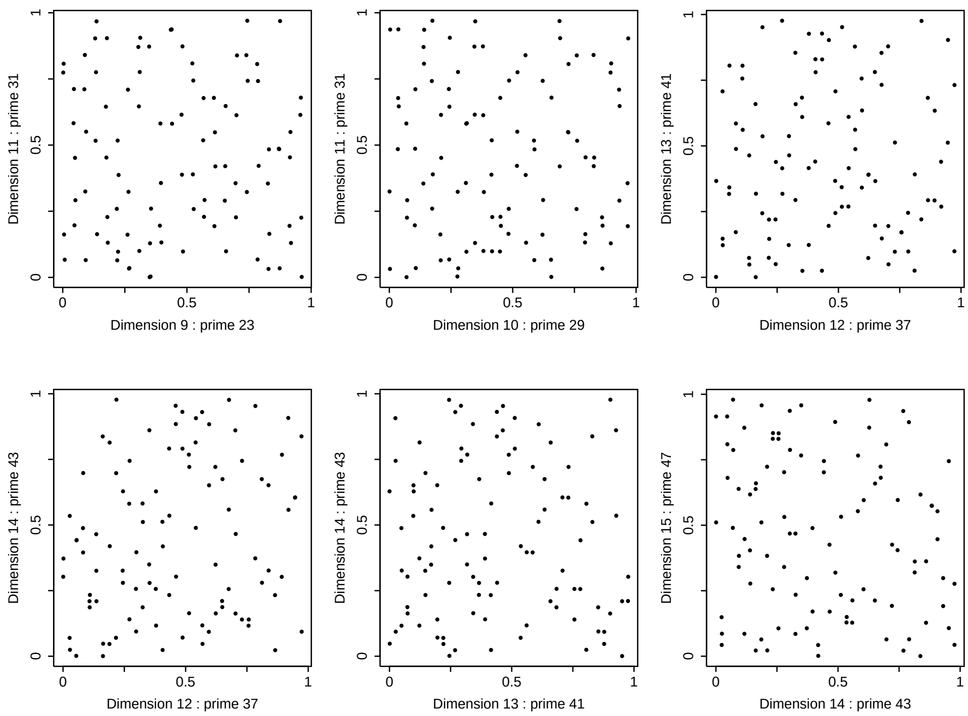


Figure 3.6: Examples of two-dimensional shuffled Halton sequences for combinations of primes 23 to 47

standard and scrambled sequences. This latter point is most readily illustrated in a graphical fashion by producing plots of shuffled Halton sequences on the basis of the same primes used for the scrambled sequences shown in Figure 3.5. The resulting plots (Figure 3.6) show that, for these choices of primes, the use of the shuffling approach leads to a lower level of correlation, and hence more uniform coverage, than the use of the scrambling approach[9].

As hinted at previously, the random shuffling removes desirable as well as undesirable effects of the cyclical nature of the one-dimensional sequences. As such, there is, with this approach, no guarantee of obtaining heightened multi-dimensional quality of coverage, and it should only be used in the presence of significant problems with correlation. The approach does however maintain one advantage over PMC sequences in these cases, namely its better one-dimensional coverage.

3.6 Modified Latin Hypercube sampling

The above discussion has highlighted the serious issues that can arise with the use of Halton sequences in the estimation of models based on integrals even of modest dimensionality. The discussion has also shown that the classic solution to dealing with the shortcomings of the standard Halton sequences, namely the

[9]It should be noted that, as the shuffling approach produces different results in different runs, the plots shown in Figure 3.6 are specific to the given runs.

scrambled Halton sequence, may not be appropriate in all cases. On the other hand, the shuffled Halton sequence is able to reliably deal with the problems caused by correlation between individual sequences, yet cannot offer any guarantees in terms of the quality of coverage in the unit hypercube.

These problems with Halton sequences cast some doubts as to their usefulness in the simulation-based estimation of discrete choice models, despite their appeal in terms of simplicity. On the other hand, more advanced methods are often difficult to implement in the framework of existing estimation packages, something that is reflected in the fact that such tools generally only offer a choice between PMC and Halton draws.

These arguments formed the main motivation for the work described here, leading to the development of the Modified Latin Hypercube Sampling (MLHS) approach[10].

Two main points need to be considered in explaining the reasoning underlying the MLHS approach. The first of these is the way in which Halton sequences are based on the principle of obtaining good coverage with the help of sub-sequences of evenly spaced points, with each additional sub-sequence filling in spaces left unfilled by previous sub-sequences. Halton sequences were developed in the context of series, which are defined as being sequences whose length can be extended without changing the original points, while still maintaining their statistical properties. This requirement allowed theoretical analysis of the asymptotic properties of the series, such as measuring the uniformity of coverage as the length of the series extends without bound. However, for any given length of the sequence, the points are not evenly spaced, and hence the one-dimensional coverage is not as uniform as possible. This is an effect of the sequential nature of these sequences, making their performance depend heavily on the exact number of draws used; this in turn can explain the fact that increases in sequence length do not necessarily lead to constant improvements in simulation performance, as for example observed by Hess et al. (2003), and, as described above, by Train (2003). With a given number of points, the most uniform coverage possible in single dimensions is necessarily attained by spacing those points evenly. Such a sequence of evenly-spaced points cannot be extended without changing the original points and therefore does not qualify as a series in the way that a Halton sequence does. This could be seen as a disadvantage, since the researcher cannot add points without recalculating the original points. However, for the purposes of numerical simulation rather than theoretical analysis of asymptotic properties, some preset number of points is always used[11], and with a given number of points, the best coverage is attained by spacing them evenly. This constitutes the first aspect of the MLHS approach.

The second aspect arises with respect to combining individual sequences to create multi-dimensional sequences. When combining one-dimensional Halton sequences into multi-dimensional vectors, the cyclical nature of the individual sequences leads to a varying level of correlation between points in the individual sequences. In the

[10]The MLHS work is published in Hess, Train & Polak (2005). Independently of the research presented here, the idea of using MLHS was also put forward in an e-mail by John Bowman, shortly after work on this analysis had started.

[11]This is not entirely true in the case of adaptive drawing algorithms, as briefly referred to in Section 3.9.

case of sequences based on low primes, this has desirable effects in that it leads to high uniformity in the multi-dimensional distribution of points. However, as of dimensions $5-8$, the collinearity has undesirable impacts in that it leads to poor multi-dimensional quality of coverage, as described in Section 3.4. The use of a scrambling approach still allows for a deterministic matching-up of individual sequences, hence maintaining the desirable effect of a multi-dimensional filling-in process (conditional on finding an appropriate scrambling heuristic). However, as shown in Section 3.5.1, the approach does not always lead to a satisfactory reduction in the level of correlation. While such reductions are offered by shuffled Halton sequences, there is, with this approach, no guarantee as to the quality of multi-dimensional coverage in the resulting shuffled sequences, given that points from individual sequences are now matched-up in a completely random fashion. As such, it can be argued that the best guarantee of obtaining good multi-dimensional coverage when combining randomly shuffled one-dimensional sequences is to base the approach on sequences with the highest possible one-dimensional uniformity. This constitutes the second aspect of the MLHS approach.

The two aspects discussed above can now be combined, and it can be seen that there is no reason to use Halton sequences as the input to the shuffling algorithm, but that rather, sequences with the highest possible one-dimensional quality of coverage should be used. This forms the basic theory behind the MLHS approach. The method combines randomly shuffled versions of sequences of evenly spaced points, where the shuffling eliminates the correlation between individual sequences. As the shuffling of the uniform vectors does not change the coverage in any of the one-dimensional sequences, the resulting sequence provides more uniform coverage in each dimension than any type of Halton sequence, or other type of QMC approach.

We now describe the sequence more formally. It can easily be seen that the most basic way of guaranteeing equal distances between a set of R draws is to set the j^{th} draw to be equal to $\frac{j}{R+1}$; the difference between any two adjacent draws is now equal to $\frac{1}{R+1}$, as is the difference between the extreme points and the border of the $0-1$ interval. However, this approach is clearly impractical as it prohibits the use of different draws in different dimensions, as well as in different runs. The problem thus lies in devising a method that leads to equal distances between adjacent draws without requiring the draws to be the same in different dimensions and runs.

We propose an approach that starts with a sequence of draws defined by:

$$\varphi(j) = \frac{j-1}{R}, \quad j = 1, \ldots, R. \tag{3.17}$$

A random number x is then drawn, such that $0 < x < \frac{1}{R}$; this can be obtained by drawing a pseudo-random number[12] contained in the interval $]0,1[$, thus using strict upper and lower limits, and dividing this number by R. The elements in the final sequence are then given by:

$$\psi(j) = \varphi(j) + x, \quad j = 1, \ldots, R. \tag{3.18}$$

[12]The spread of the values of x across sequences can be made more uniform by basing the draws of x themselves either on a quasi-random approach, or by using a pseudo-random number generator with proven properties.

In the resulting sequence, the distances between adjacent draws are all equal to $\frac{1}{R}$, satisfying the condition of equal spacing. Furthermore, the combined distance between the two extreme points and the respective borders of the $0-1$ interval is also equal to $\frac{1}{R}$; if the interval is regarded as being cyclical, all "adjacent" distances are equal to $\frac{1}{R}$. Multi-dimensional sequences are constructed by simple combination of randomly shuffled one-dimensional sequences, leading to low cross-dimensional correlation. The addition of the random variate x has the desired effect that the draws used are different across dimensions, as well as across runs, while still keeping the distance between adjacent draws at $\frac{1}{R}$, such that the quality of one-dimensional coverage remains unchanged.

An important difference arises between MLHS and Halton-based sequences. Indeed, in the latter approaches, to generate a set of R draws for N individuals, subsets of length R from a sequence of length NR are used, while, given that MLHS sequences are not created as series, it is necessary to generate the N sets of R draws independently, to guarantee equal quality of one-dimensional coverage across the N sequences. At this point, it should also be noted that the MLHS approach is not a quasi-Monte Carlo approach per se; rather, it should be described as an approach that attempts to pick random numbers in a stratified way.

The procedure described here is very similar to the Latin Hypercube Sampling (LHS) approach proposed by McKay et al. (1979) - hence the name, MLHS. LHS is the same as MLHS, except that for LHS, a different draw is taken for each element in each dimension, rather than using the same draw for all elements in each dimension. That is, for LHS, equation (3.18) is changed to:

$$\psi(j) = \varphi(j) + x_j, \quad j = 1, \ldots, R \tag{3.19}$$

where x_j is a separate draw from a Uniform distribution between 0 and $\frac{1}{R}$.

The use of these different drawing approaches leads to important differences between MLHS and LHS. The motivation for MLHS was to attain a sequence that has more uniform coverage in each dimension than Halton sequences. An evenly spaced set of points necessarily attains that goal and is the most uniform set possible. Using one common draw to shift all points retains this uniformity. In contrast, when using a separate draw for each point, as in LHS, the resulting set of points is less uniform than evenly spaced points and can potentially be less uniform than the Halton sequence. It should be noted however that the LHS method allows for the possibility of a type of cancelling out, by which a draw that moves one point to the higher end of its segment is counteracted by a draw that moves another point to the lower end of its segment. Inherently, however, the goal of uniformity operates against the potential for cancelling out in this way.

The MLHS approach has a disadvantage in that it is not easily possible to accurately define the asymptotic properties of the method, since it is not a series. This is a drawback in that it does not allow us to compare the method asymptotically with other approaches. However, some comparisons can be made. Firstly, the coverage offered by MLHS is necessarily better than that of the shuffled Halton sequence, given that both approaches rely on random shuffling of one-dimensional sequences in the construction of multi-dimensional sequences, where with our new approach, the coverage offered by the one-dimensional uniform sequences is superior to that of the one-dimensional Halton sequences. Secondly, the same reasoning can be used in the comparison with pseudo-random sequences. Indeed, multi-dimensional pseudo-random number sequences are in fact simply combinations of one-dimensional pseudo-random sequences, where the ordering of points in these one-dimensional sequences is similarly random, just as in the case of MLHS. The fact that the one-dimensional ordering of points is thus random in both cases means that, asymptotically, any differences between the two approaches in terms of the resulting multi-dimensional quality of coverage depend solely on the quality of coverage of the one-dimensional sequences, which is better in the case of MLHS.

3.7 Empirical comparisons

The previous sections have looked at the theoretical differences between the various types of Halton sequences and the MLHS approach. In this section, we turn our attention to the actual substantive differences between the different approaches when used in the simulation-based estimation of discrete choice models.

These issues have been discussed in great detail by Hess et al. (2003), Hess, Train & Polak (2005), Bastin et al. (2005), and other work referenced in these publications. In the present work, we limit the comparison to two separate case-studies. Aside from space constraints, this is motivated by two main principles.

- The first principle is based on the discussion in Section 3.6, which has described the clear advantages of MLHS over shuffled Halton, effectively making the latter approach redundant[13].
- The second principle is that, given that MLHS draws are constructed on the basis of independently shuffled one-dimensional sequences, their multi-dimensional quality of coverage is almost guaranteed to be lower than that obtained with standard (and potentially) scrambled Halton sequences generated from ***low*** primes.

As such, the two applications described in this section have the following rather separate aims:

- The first application compares the performance of the three types of Halton sequences, the MLHS approach, and PMC draws, in the estimation of a MMNL model with sixteen dimensions of integration. The high dimensionality of the integral allows us to gauge the relative performance of the different methods in high-dimensional problems.
- The second application compares the performance of MLHS draws and PMC draws in the estimation of a MMNL model with a low number of dimensions of integration (4). This allows us to validate the theoretical claims made with regards to the advantages of MLHS draws over PMC draws.

[13]Comparisons between the shuffled Halton sequence and standard and scrambled Halton sequences in the estimation of MMNL models are presented by Hess et al. (2003).

3.7.1 Application A

The first application uses a sixteen-dimensional MMNL model estimated on SP data. This application was conducted as part of work described by Hess, Train & Polak (2005).

The data used in this application were collected as part of a survey looking into potential customers' choices between three types of vehicles in California; gas internal combustion (ICV), electric (EV) and gas-electric hybrid (HV). The data are described in more detail by Train & Hudson (2000), and have previously been used by Sándor & Train (2004) and Train & Sonnier (2005). A total of 500 respondents were included in the dataset, and each respondent was presented with up to 15 different choice situations between three vehicles, where the individual choice sets do not necessarily contain one vehicle of each of the three types considered. The total number of observations in the dataset is $7,437$. Two different datasets were available for the original study; a basic dataset and an *enhanced* dataset. The present analysis uses the latter dataset, in which respondents were provided with prior information on EV vehicles and on air quality in California. This has been observed to have a significant positive effect on the attitude of respondents towards EV vehicles (Train & Hudson 2000).

Each alternative used in the dataset is described by a total of 6 attributes:

- Car type (ICV, EV or HV)
- Body type (10 different types, ranging from mini car to mini van)
- Purchase price ($\$1,000$'s)
- Operating cost ($/month)
- Performance (grouped into high, medium and low performance)
- Range (100's of miles between refueling/recharging)

Although the performance of the vehicles is simply divided into three levels, the respondents were actually provided with more detailed information on top speed and seconds needed to reach 60mph; these were however directly linked to the three levels of performance. In addition to this, the range attribute was set to constant values for ICV and HV vehicles; the reason for including this attribute was simply to gauge the effect on respondents' choices of increases in the range of EV vehicles.

For the empirical estimation, a MMNL model was used, acknowledging the repeated choice nature by assuming constant tastes across responses for the same individual. The dummy variables for type of vehicle take the role of alternative specific constants; for reasons of identification, no coefficient was associated with ICV vehicles, such that the coefficients associated with the dummy variables of the two other types of vehicle represent the net impact of unmeasured variables (including general attitude of respondents) on the utility of EV and HV vehicles relative to ICV vehicles. Similarly, some normalisation was performed for the other variables included in the model. For the body type, midsize car was taken as the base, whereas for the performance variables, medium performance was chosen as the base. Finally, as the range variable had been kept constant for ICV and HV vehicles, a

Parameter		Estimate	Std.err.	Parameter		Estimate	Std.err.
Price	c	-2.546	0.06	Sm. Car	μ	-1.326	0.17
	s	0.731	0.04		σ	1.120	0.29
Op.Cost	c	-3.540	0.11	Lrg. Car	μ	-0.463	0.17
	s	0.853	0.08		σ	1.183	0.27
Range	c	-0.586	0.25	Sm. SUV	μ	-0.796	0.16
	s	0.122	0.22		σ	0.758	0.28
Electric	μ	-1.979	0.21	Mid. SUV	μ	0.331	0.15
	σ	1.278	0.13		σ	0.778	0.33
Hybrid	μ	0.791	0.10	Lrg. SUV	μ	-0.160	0.24
	σ	1.140	0.10		σ	1.580	0.41
High Perf.	μ	0.184	0.06	Comp. PU	μ	-1.290	0.18
	σ	0.609	0.09		σ	1.038	0.28
Low Perf.	μ	-0.492	0.06	Full PU	μ	-0.771	0.19
	σ	0.551	0.10		σ	1.589	0.31
Mini Car	μ	-2.983	0.23	Minivan	μ	-0.479	0.19
	σ	1.936	0.31		σ	1.500	0.25

Table 3.1: Estimation results on vehicle-type choice data

coefficient associated with this attribute was only estimated for EV cars. This leads to a total of two estimated ASCs, three marginal utility coefficients, and eleven dummy coefficients.

All parameters used in the utility function were specified to vary randomly across respondents, where a Lognormal distribution was used for the three marginal utility coefficients (with an appropriate sign change for the price and cost attributes), while a Normal distribution was used for the various constants. The final model thus uses 16 randomly distributed coefficients, leading to 32 parameters to be estimated. It seems that this makes it the highest-dimensional MMNL analysis comparing different types of Halton sequences to date. Kenneth Train's Gauss code[14] was used for the estimation of the models.

In order to obtain *true* values of the parameters for use as reference points in the comparison of the different types of draws, ten runs based on $2,000$[15] PMC draws (per individual and per dimension) were used. From the results, mean values of the parameters were calculated over runs, in addition to standard errors of the *true* parameters (calculated as the square root of the average of the squared standard errors from individual runs).

The results from this estimation are summarised in Table 3.1, which gives the estimates of the different parameters, along with their standard errors. For the coefficients following a Normal distribution, these correspond to the mean (μ) and standard deviation (σ), while for the coefficients following a Lognormal distribution, they give the mean (c) and standard deviation (s) of the underlying Normal distribution. The corresponding mean and standard deviation for the actual Lognormal

[14]See http://elsa.berkeley.edu/∼train/

[15]A stability analysis showed this number to be sufficient for obtaining stable estimates, with no significant changes in likelihood, parameter estimates or standard errors when using a higher number of draws.

distribution can be obtained by

$$\mu = \exp\left(c + \frac{s^2}{2}\right) \tag{3.20}$$

and

$$\sigma = \mu\sqrt{\exp\left(s^2\right) - 1} \tag{3.21}$$

respectively. For the three coefficients in question, price, cost and range, this calculation leads to mean values of −0.102, −0.042 and 0.561 respectively, after a sign-change for the price and cost-coefficient, along with standard deviations of 0.086, 0.043 and 0.069 respectively.

The estimated values of the coefficients indicate the changes in utility following a unit change in the respective attribute, with all other attributes kept constant. The results produced in the present analysis are broadly consistent with those produced by Train & Hudson (2000), showing significant negative impacts of increases in price and operating cost, along with positive impacts of increases in range, with high variations in sensitivity levels across respondents especially for price and operating cost. With medium performance used as the base, the signs of the coefficients for high and low performance are as expected, indicating that high performance is preferred to medium performance while low performance leads to lower utility than medium performance. The results suggest that the (mean) effect of moving from low to medium performance is around 2.5 times as important as the effect of moving from medium to high performance. The estimate for the coefficient associated with the dummy variable for electric vehicles can be most easily interpreted by summing it with the range coefficient (which was set to zero for all non-electric vehicles); this shows that on average, ceteris paribus, an electric vehicle needs a range of 353 miles to be valued equally highly as an ICV vehicle. With inferior range, EV vehicles are valued lower than ICV vehicles, while HV vehicles are valued more highly than ICV vehicles. In terms of body type, only the mid-size SUV is on average preferred to the mid-size car (used as base). It should also be noted that the mean of the coefficient associated with large SUV's is not significantly different from zero, while its standard deviation is significant.

For the comparison of the different types of draws, four sequence lengths were used, namely 50, 100, 200 and 500. Ten runs were used for the non-deterministic methods (shuffled Halton draws, MLHS and PMC), while, for the Halton-based approaches, the first sixteen eligible primes (2 to 53) were used. With the aim of testing the ability of the draws to recover the *true* parameters, the root-mean-squared-error (RMSE) between the estimated value and the *true* value of each parameter was calculated across runs. For standard and scrambled Halton sequences, this equates to the absolute difference between the *true* and estimated parameter, given that only a single run was used. In order to account for the shape of the log-likelihood function, these error values were then expressed as a proportion of the standard error of the true parameter (cf. Sándor & Train 2004), before being averaged across the 32 parameters that were estimated. This approach is based on the understanding that standard errors are related to the shape of the log-likelihood function at the maximum. A high standard error for a parameter indicates a flat shape of the

Draws	Standard Halton	Scrambled Halton	Shuffled Halton	MLHS	PMC
50	0.7579	0.7682	0.7577	0.7377	0.7627
100	0.5807	0.5964	0.5521	0.5017	0.5664
200	0.3762	0.3878	0.3682	0.3630	0.3858
500	0.2278	0.2177	0.2280	0.2152	0.2448

Table 3.2: Estimation performance of different types of draws on vehicle-type choice data: average (across parameters) RMSE as proportion of standard error

log-likelihood function for this parameter at the maximum of the function; slight changes in the value of the parameter have little effect on the value of the log-likelihood function. The converse is the case for low standard errors. The inverse of these standard errors can thus be used as weights for the parameters in the calculation of estimation performance (when comparing different types of draws), such that a higher simulation error is tolerated for parameters that have a higher standard error (c.f Sándor & Train 2004).

The first observation that can be made from Table 3.2 is that the performance of the standard Halton draws is surprisingly good, despite the problems with high correlation that would be expected in such a high-dimensional application. Another interesting observation is that the use of scrambled Halton sequences actually leads to poorer performance than the use of standard Halton sequences for three out of the four sequence lengths (50, 100 and 200 draws). Except for the longest choice of sequence length (500 draws), the standard and scrambled Halton draws are outperformed by the shuffled Halton draws, suggesting that, especially with short sequences, the shuffling approach can offer improvements over the other two approaches in high dimensional problems. The MLHS method leads to the best performance for all four sequence lengths. Finally, it should be noted that the use of PMC draws leads to surprisingly good relative performance, given earlier results, for example by Train (1999) and Bhat (2001). Stated more directly, the alternative methods did not provide very much improvement over PMC draws.

A possible reason for the small improvement could be the high dimensionality of the application compared with those of earlier applications. As stated above, the Halton and MLHS procedures are not specifically designed in multiple dimensions; rather they are all designed as one-dimensional sequences that are combined to create multi-dimensional sequences. It might be the case that, as the number of dimensions rises, the importance of one-dimensional coverage diminishes. Or, stated alternatively, any unstructured combination of one-dimensional sequences may start to more closely resemble a purely random sequence as the number of dimensions rises. This conclusion suggests the need for comparisons with sequences like systematic sampling and (t, m, s)-nets, which, as discussed in Section 3.8, provide better uniformity over dimensions by relinquishing uniformity in each dimension.

The results of the application do suggest that the lowest error is obtained by the MLHS approach. However, aside from being based on a single application, it should be clear that the differences in performance between the single methods are probably too small to allow us to generalise the results. Systematic analyses using synthetic datasets would provide more insight, as would an extension of the parameter space used in comparison to the entire variance-covariance matrix. Nevertheless, the results in this study are interesting and intuitive. They indicate that if a multi-dimensional sequence is going to be created by randomly combining one-dimensional sequences, then performance is enhanced by obtaining more uniform coverage in each dimension, which is attained by MLHS relative to shuffled Halton sequences, and by MLHS and shuffled Halton sequences relative to pure pseudo-random sequences.

3.7.2 Application B

The second application uses a four-dimensional MMNL model estimated on RP mode choice data with 5 alternatives. This application was conducted as part of the work published in Bastin et al. (2005), although the runs using 1,000 MLHS draws were repeated due to some inconsistencies in the original results.

The Mobidrive dataset used here was collected in 1999 in two cities of Germany (Karlsruhe and Halle-Salle), from 160 households and 360 individuals, where each individual was observed during six continuous weeks (cf. Axhausen et al. 2002). In the present analysis, only the dataset for Karlsruhe is used; appropriate level-of-service data for the used and non-used alternatives were added separately. All trips are grouped into tours, and the population is divided into workers (commuters and education) and non-workers. For each worker, the daily chain is divided into morning, commute and evening patterns. For non-workers, we define the main activity as the longest out-of-home activity recorded, where the daily activity chains are represented in relation to this pivotal activity and organised into morning, principal and evening patterns. With this definition, a total of 5,795 tours were identified, performed by 136 individuals belonging to 66 households, with an average daily number of 1.72 tours per individual.

The final model contained 21 parameters, of which four were specified to be randomly distributed, namely the coefficients associated with time, cost, total travel time and time budget. For ease of estimation, a Normal distribution was used for these coefficients. This however led to significant probabilities of counter-intuitively signed coefficients, which should be seen as an artefact of the use of the Normal distribution (cf. Chapter 4). As such, the results produced by this model are not reliable from a policy-decision perspective, but the model can still be used for the present analysis, where the aim is of a purely mathematical nature.

The models in this application were estimated with AMLET (Bastin et al. 2003), where, in the present example, a basic trust region optimiser was used. The *true* parameters were generated by running the model ten times with 10,000 pseudo-random draws per individual. As in the first example, the square-root of the average of the squared errors was again calculated over runs. Ten independent runs were performed in order to produce the performance indicators. Aside from the RMSE, two additional measures were used to compare the results to those obtained with the *reference* model, namely the bias (difference between the mean estimate across runs and the *reference* estimate), and the standard deviation of estimates across runs. In each case, the measures were calculated as a proportion of the standard error of the associated parameter, before being averaged across parameters.

Given the above discussion about the quality of the estimates on the basis of

MLHS	**200 draws**	**500 draws**	**1000 draws**	**2000 draws**
Bias/std.err.	0.0269	0.00961	0.00558	0.00311
RMSE/std.err.	0.04516	0.02486	0.02414	0.02261
St.dev./std.err.	0.03678	0.02387	0.02449	0.02316
Est. time (s)	381	967	2,474	4,468

PMC	**1000 draws**	**2500 draws**	**5000 draws**
Bias/std.err.	0.01946	0.01076	0.00491
RMSE/std.err.	0.05128	0.04215	0.03615
St.dev./std.err.	0.04927	0.04244	0.03769
Est. time (s)	2,280	5,769	11,064

Table 3.3: Estimation performance of different types of draws on *Mobidrive* data

the distributional assumptions, the substantive model estimates are not reproduced here, but can be found in Bastin et al. (2005). Rather, we look only at the performance of the models using MLHS and PMC draws in terms of reproducing the estimates obtained with the *reference* model. These performance measures are reported in Table 3.3. For MLHS, the results are presented for sequences of length 200, 500, 1,000 and 2,000 draws, while for PMC, the results for 1,000, 2,500 and 5,000 draws are shown.

The first observation that can be made from the results is that when comparing the performance with the only common number of draws, 1,000, the MLHS draws provide better performance than the PMC draws for all three measures of performance. Except for the bias measure, the same applies when comparing the results obtained with 200 MLHS draws to those obtained with 1,000 PMC draws, while, with 500 MLHS draws, all three indicators show better performance than with 1,000 PMC draws. Additionally, the results show better performance with 500 MLHS draws than with 2,500 PMC draws, as well as for 2,000 MLHS draws when compared to 5,000 PMC draws, while, with 1,000 MLHS draws, the bias is only slightly higher than with 5,000 PMC draws, whereas RMSE and standard deviation are lower. The results for MLHS show decreasing bias and RMSE with increases in R, while the standard deviation, which is much lower than with PMC draws (showing greater stability), remains relatively stable beyond 500 draws (with a minor increase between 500 and 1,000, which is however almost surely an observation biased by the low number of runs). Although more runs are required to improve the reliability of the conclusions, the overall results are very encouraging, and the implied possibilities in terms of computational savings are quite impressive. Indeed, the ability to obtain better performance with 500 MLHS draws than with 2,500 PMC draws would mean savings by around 80% in runtime, while, with the results for 2,000 MLHS draws and 5,000 PMC draws, the savings are still around 60%.

3.8 Alternative QMC approaches

As described in the preceding sections, multi-dimensional Halton and MLHS sequences are created by combining one-dimensional sequences. Several QMC sequences have been proposed that are created directly in the multi-dimensional space. Generally, these sequences obtain more uniform spacing of points over the multiple dimensions while attaining less uniformity in individual dimensions.

One such method is the systematic sampling approach proposed by McGrath (1970). This approach produces a K-dimensional uniform grid of R points by randomly drawing $\frac{R}{M}$ points in one of the M K-dimensional parts of the grid and translating these points into the remaining $M-1$ grid areas. Alternatively, the draws in each sub-area are drawn randomly. The precision of either approach increases with M. The main difficulty with these grid methods is that of finding appropriate values of M (and R), such that $\frac{R}{M}$ and $\sqrt[K]{M}$ are integer values. Clearly, the best performance with this method is obtained in the case where a division can be used such that each sub-cube of the $0-1$ hypercube contains exactly one draw. This is only possible in the rare case where $\sqrt[K]{R}$ is an integer value. Alternatively, it is possible to use a division of the $0-1$ hypercube into hyperrectangles, rather than hypercubes; the use of different divisions along different dimensions however potentially reduces multi-dimensional uniformity.

It is of interest to briefly compare the systematic sampling method to the MLHS approach, hence highlighting the differences between a sequence designed in multiple dimensions and a sequence based on the combination of one-dimensional sequences. Indeed, it should be highlighted again that, with a method based on the combination of randomly shuffled one-dimensional sequences, there is no guarantee of good multi-dimensional coverage, even in the case of very good one-dimensional uniformity. It is clear that the one-dimensional quality of coverage of a grid method increases as M and hence $\sqrt[K]{M}$ (the number of intervals per dimension) increases[16]. This quality of coverage will however necessarily be inferior to that obtained in single dimensions by the MLHS method, as this effectively uses R equally-sized intervals along each dimension with one draw per interval. Only in the case where $K=1$ and $M=R$ will the systematic sampling approach lead to the same degree of one-dimensional uniformity of coverage. As the systematic sampling approach does not rely on any shuffling of the one-dimensional draws, but explicitly creates a grid of multi-dimensional points, the multi-dimensional quality of coverage should be superior to that of MLHS; this however presumes sufficiently high values for M (with correspondingly small cells and low numbers of draws per cell), which are often not possible. These difficulties, which hamper general implementation of the systematic sampling approach, have thus far prevented the deployment of this approach in the area of discrete choice modelling.

Another type of QMC sequences designed in multiple dimensions are (t,m,s)-nets, as discussed in the context of discrete choice modelling by Sándor & Train (2004). These nets constitute a general class that includes Sobol, Faure, Niederreiter, Niederreiter-Xing, and other sequences. The specification and construction of (t,m,s)-nets differ across different numbers of dimensions and points. They are restrictive in the number of points that can be used for any given number of dimen-

[16] A similar reasoning applies in the case of hyperrectangles, instead of hypercubes.

sions, and, like the grid procedures, generally attain more uniform multi-dimensional coverage at the expense of uniformity along each individual dimension. Sándor & Train (2004) compare four kinds of (t, m, s)-nets with Halton sequences in the estimation of MMNL models. They find that two of the (t, m, s)-nets performed better than Halton sequences, while the other two performed worse, suggesting that more complicated methods do not necessarily offer universal improvements in performance.

Moving back to sequences produced in single dimensions, another potentially interesting solution is the use of Sobol sequences, which have significant advantages over Halton sequences in terms of much lower degradation in performance in higher sequences. A recent application by Garrido (2003) confirms these advantages in the context of MMNL estimation, showing that Sobol sequences offer more stable performance than Halton draws. However, Garrido (2003) not only repeats findings by Train (2003) showing that increases in sequence length with Halton draws do not always lead to increases in performance, but worryingly, observes a similar phenomenon with Sobol sequences. While such problems should occasionally be expected in the case of approaches using random perturbations, purely deterministic approaches should be immune to such problems.

As Train (2003) observes, the number of different approaches for producing QMC numbers is so high that it is impossible to describe all of them in a single text. This, in any case, is not the aim of this chapter, as set out in Section 3.3. There are major differences between the various approaches, in terms of the ease of implementation, as well as in their performance. Examples of more advanced types of QMC number sequences are for example given by Bratley & Fox (1988), Bratley et al. (1992), Owen (1995) and Niederreiter & Xing (1998), while good reviews of the different available methods are given by Bratley et al. (1992) and Krommer & Ueberhuber (1994). Spanier & Li (1997) discuss the potential of combining QMC and PMC approaches, and Shaw (1988) discusses the use of QMC sequences in Bayesian statistics.

In concluding, it should be noted that the increased complexity of advanced methods often counterbalances their advantages in terms of performance, especially with the relatively simple problems faced in transport studies. This is reflected in the fact that in this area of research, Halton sequences have established themselves as the most popular choice of QMC sequence. As Train (2003, page 239) notes; *"It is important to remember, however, in the excitement of these methods, that accuracy can always be improved by simply using more draws. The researcher needs to decide whether learning and coding new methods of taking draws is more expedient, given her time constraints, than simply running her model with more draws."*

3.9 Summary and Conclusions

This chapter has discussed issues relating to the estimation and application of models based on integrals without a closed form solution, and specifically, the high cost incurred as a result of the numerical processes that are required to evaluate the choice probabilities of such models. The discussion has centred on simulation, and specifically, on the use of alternatives to PMC draws in classical Monte-Carlo integration. As such, the chapter has discussed the development of an alternative approach to the commonly used Halton sequence for use in simulating the choice probabilities of GEV mixture models. The approach, MLHS, has certain advantages over other approaches, in terms of simplicity of implementation, as well as reduced risk of problems with inter-dimensional correlation in high-dimensional problems.

The two brief applications presented in this section lead to mixed conclusions. On the one hand, the second application shows that, in low-dimensional problems, MLHS draws offer more stable performance than PMC draws, and have the potential to lead to impressive savings in computational costs. On the other hand, the first application has shown that, in high-dimensional problems, there is little to pick between the different QMC approaches attempted (although the MLHS approach seemingly offers the best performance), and that the performance of PMC draws is actually quite acceptable. This can be interpreted in two ways; either that the QMC sequences attempted are not adequate for the use in high-dimensional problems, or that the disadvantage of PMC draws decreases with increasing dimensionality.

More research remains to be done to provide an answer to this question, as well as in the general testing of the MLHS approach, in terms of the analysis across a wider range of modelling scenarios, and the comparison in performance across a wider set of QMC sequences. In this work, the comparison was limited to those approaches developed during the research presented in this thesis, along with the most commonly used types of sequences in discrete choice modelling. While the results from the high-dimensional application cast some doubt as to the usefulness of basic QMC approaches in high-dimensional MMNL models, it should be remembered that the majority of applications in choice modelling use only a very limited number of randomly distributed coefficients. Given the persisting doubts surrounding the commonly used Halton sequence, it seems that MLHS draws could provide an interesting alternative, which has by now been implemented in AMLET as well as BIOGEME, while the manual implementation in ALogit and a number of other estimation packages is straightforward.

In closing, one topic that deserves a brief mention is the use of adaptive drawing algorithms. Such approaches are based on the notion that a lower level of simulation accuracy is needed in the early stages of estimation, when the optimisation takes only rough steps in the general direction of the optimum. On the basis of this reasoning, a lower number of draws can be used in early iterations, leading to significant reductions in the cost of estimating MMNL models, as shown by Bastin et al. (2003). As such, while QMC approaches aim to reduce the average cost per iteration through the use of a lower number of more uniformly positioned draws, adaptive drawing approaches, which are generally based on PMC draws, reduce the average cost by accepting lower levels of precision in initial iterations. Two different versions of such an algorithm have been tested in the context of mixture models; the basic trust-region with dynamic accuracy (BTRDA[17]) approach of Bastin (2004), which is currently restricted to MMNL, but allows for flexible changes in the number of draws in each iteration, and the BIOMC[18] optimiser implemented in BIOGEME, which can be used with any type of GEV mixture model, but where the changes in the number of draws are determined prior to the start of the estimation work. The main issue with the use of adaptive drawing techniques is that of implementation. Here, it should be noted that the use of a manual approach, which precedes the

[17]Used for example by Bastin et al. (2005).

[18]Used for example by Hess, Bierlaire & Polak (2005*a*).

actual *full-scale* estimation by one using a lower number of draws with a less strict convergence criterion, can yield comparable savings, without a need for a special implementation, as illustrated by Hess (2005). Finally, a common sense approach, in which MNL estimates are used as starting values in MMNL estimation[19], at least for the fixed coefficients and the mean values of any normally distributed coefficients, can also lead to important reductions in overall estimation time, while yielding the same results, as observed by Hess (2005).

[19]Respectively GEV estimates in the case of GEV mixture models.

Chapter 4

Specification and interpretation of random coefficients models

4.1 Introduction and context

As described in Chapter 2, researchers and practitioners are increasingly using the MMNL model for a representation of random variations in sensitivities across respondents[1]. This not only offers great benefits in terms of providing insights into variations in tastes across respondents, but also potentially avoids bias in the estimated trade-offs in models based on the use of fixed taste coefficients (see for example Algers et al. 1998, Hess & Polak 2004*a*).

However, while the MMNL model can offer great gains in flexibility, the pitfalls are similarly significant, as stressed for example by Hensher & Greene (2003). Aside from the greater cost in terms of model estimation and application, as discussed in Chapter 3, two main issues arise with the use of the MMNL model; the choice of statistical distribution for randomly distributed coefficients, and the economic interpretation of such coefficients. The additional issue of deciding which parameters should be modelled as being randomly distributed across agents can be addressed relatively straightforwardly on the basis of statistical tests.

This chapter looks specifically at the issues of the choice of distribution and the interpretation of the resulting parameter values, and discusses how the two are strongly inter-related. The discussion centres on one specific application of MMNL models, namely the estimation of variations in the valuation of travel time savings (VTTS).

While the issue of the choice of distribution especially has been discussed repeatedly in the existing literature, with some examples being the work of Hensher & Greene (2003), Sørensen (2003) and Train & Sonnier (2005), it should be noted that the number of distributions tested in previous work has generally been rather limited. Additionally, the vast majority of actual modelling analyses still rely exclusively on the use of the Normal distribution, while only a handful of alternatives to the Normal distribution have received even modest exposure. This is despite the fact that there is a body of work that shows that misspecification of taste hetero-

[1] The issues described in this chapter apply to mixture models in general, and are not constrained to MMNL. For reasons of simplicity, the discussion here centres on the MMNL model, which is the only widely used mixture model.

geneity, as for example in the use of an inappropriate distribution, can undermine the reliability of widely used benefit measures such as the *rule of a half* or the *log-sum* (cf. Cherchi & Polak 2005). From this point of view, there is an urgent need for extensive research into the applicability of hitherto unused (or rarely used) distributions in the context of mixture models. As such, the aim of the applied part of this chapter is to conduct a case-study using a high number of different continuous distributions in a single application, hence highlighting the potential differences in performance as well as substantive results[2].

The other issue addressed in this chapter is that of the interpretation of results produced with the help of models allowing for a random distribution of tastes across respondents. This discussion is positioned in the context of recent results showing counter-intuitive findings in terms of a share of the population with negative VTTS, i.e. travellers who actually seek increases in travel time (e.g. Cirillo & Axhausen 2004). Here, the aim of this chapter is to establish whether such results should be seen as actual evidence of the existence of such individuals, or should rather be regarded as effects of model misspecification or data problems. The evidence from the case-study is used to support the more theoretical claims.

To a large extent, the material presented in this section has been published in Hess, Bierlaire & Polak (2005*c*), but the case-study presented here is far more comprehensive, where the original application made use of only three distributions (Normal, Lognormal and S_B). Additionally, the work is motivated by the findings of Hess & Axhausen (2005), who, in a comparison of the approximation power of different distributions to simulated datasets, highlight problems with the tail-behaviour of unbounded symmetrical distributions, such as the Normal.

The remainder of this chapter is organised as follows. The next section discusses the state-of-practice in terms of the choice of distribution in MMNL modelling. The importance of the distributional assumptions and the issues faced in the interpretation of MMNL results are illustrated in the VTTS case-study presented in Section 4.3. Section 4.4 looks at the interpretation of results showing a significant share of respondents with counter-intuitively signed travel time coefficients. Finally, Section 4.5 summarises the findings of the chapter, and offers some guidance for good practice.

4.2 Choice of distribution: the state of practice

As indicated in Section 4.1, the choice of distribution for a randomly distributed coefficient plays a crucial role in the specification of a MMNL model. In practice, only the *Normal* (Gaussian) and *Lognormal* distributions have found widespread application in MMNL modelling. Several authors have also advocated the use of the *Triangular* distribution (e.g. Hensher & Greene 2003), while recently, good results have also been obtained with *Johnson's* S_B distribution (cf. Train & Sonnier 2005).

While the use of the Normal distribution can be appropriate in the case of coefficients without a strict sign assumption, the fact that it is unbounded can cause

[2]The work presented in this chapter is based entirely on standard continuous distribution functions, and ignores the potential use of empirical distributions. Additionally, any estimation work is based purely on standard techniques, with options such as the conditioning on respondents' choices (cf. Sillano & Ortúzar 2004) not explored here.

severe problems with interpretation when used for coefficients where such an *a priori* assumption exists in principle (e.g. travel time coefficients). Additionally, problems can arise in the case of asymmetrical *true* distributions. While the use of the Lognormal distribution can at least partly address these issues, problems can be caused by the long tails (cf. Hess & Polak 2004*a*). Additionally, computational problems and slow convergence limit the applicability of the Lognormal distribution. The issues of the long tail can be avoided with the use of the Triangular distribution, which has the advantage of being bounded to either side. However, in its standard form, the distribution faces the same problem as the Normal in terms of a symmetrical shape, where the asymmetrical variant is difficult to implement, given the complications with estimating the location of the peak.

The main factor leading to the almost exclusive reliance in MMNL modelling on the above listed distributions is the relatively limited repertoire of distributions supported by existing MMNL estimation packages, where only a small subset of modellers make use of their own, purpose-written code. Within the set of available distributions, the choice is often influenced by issues of numerical problems when using the advanced distributions, but it must be said that the limited awareness of the potential benefits of using a broader range of distributions also plays a role in the preeminent position of the Normal distribution.

4.3 VTTS case-study

We now turn our attention to the VTTS case-study, which serves as an indication of the effects of distributional assumptions in the use of MMNL models.

4.3.1 Introduction

The computation of VTTS measures has been one of the main applications of random utility models, with some recent discussions of the topic including Algers et al. (1998), Hensher (2001*a*,*b*,*c*), Lapparent & de Palma (2002), Cirillo & Axhausen (2004) and Sillano & Ortúzar (2004). The VTTS is an important willingness-to-pay indicator, used for example for cost-benefit analysis in the context of planning new transport systems, or for pricing. In discrete choice models, the computation of VTTS measures is relatively straightforward, given by the ratio of the partial derivatives of the utility function with respect to travel time and travel cost (i.e. the marginal rate of substitution between travel time and travel cost, at constant utility). Although this is an intuitively plausible approach, it is important to appreciate that the justification for this approach to the valuation of travel time savings rests not on plausibility but rather on a substantial body of microeconomic theory that addresses the issue of how individuals allocate time amongst alternative activities, including travel. Indeed, the topic of time allocation and valuation has been the subject of intense study from a variety of different perspectives for several decades (see, among others, Becker 1965, Oort 1969, De Serpa 1971, Evans 1972, Truong & Hensher 1985, Bates 1987 and Jara-Diaz & Guevara 2003). The papers by Jara-Diaz (2000) and Mackie et al. (2001) provide excellent overviews of the development of this literature.

Under the strong but necessary assumption that all effects of these two attributes (travel time and travel cost) are captured in the observed part of utility, the VTTS measure is simply computed as:

$$\frac{\partial V/\partial TT}{\partial V/\partial TC}, \tag{4.1}$$

with V giving the observed part of utility, and TT and TC representing the travel time and travel cost attributes respectively. In the case of fixed taste coefficients, and with the commonly used linear-in-attributes utility function, this formula reduces to β_{TT}/β_{TC}, where β_{TT} and β_{TC} are the time and cost coefficients, giving the marginal utilities of increases by one unit in travel time and travel cost respectively. Estimates of these marginal utilities are produced by calibrating the model on the choice data used in the estimation[3]. Even with the use of non-linear transforms, such as the natural logarithm, the computation remains relatively straightforward, although the actual values of the attributes now enter into the computation of the trade-offs.

With the increased use of the MMNL model in the area of transport studies, researchers have begun to increasingly exploit the power of this model to represent a random variation in the marginal utility of travel time and travel cost across respondents (e.g. Algers et al. 1998, Cirillo & Axhausen 2004). However, the extension of the theoretical foundations of the calculation of the VTTS to the case where β_{TT} and/or β_{TC} are modelled as random parameters is not straightforward. Indeed, in the MMNL analysis of the distribution of the VTTS measures across a population of respondents, the distributional assumptions play a crucial role, and have a significant effect on model interpretation.

In this discussion, we focus on the marginal utility of travel time, but a similar principle applies in the case of the marginal utility of travel cost, or indeed in the case where the VTTS is modelled directly, as opposed to being based on the ratio of β_{TT} and β_{TC} (cf. Fosgerau 2004). It should be noted that this approach has potential advantages, especially in the case of random coefficients models, as it avoids the issues involved with calculating a distribution of the VTTS on the basis of a ratio of two randomly distributed coefficients. As such, the continued exploration of such approaches in random coefficients models is an important avenue for future research.

In models that are based on the use of fixed taste coefficients, researchers generally have an *a priori* expectation of obtaining a negative travel time coefficient, and models producing positive values will normally be rejected on the grounds of model misspecification or lack of explanatory power in the data. While the sign-issue is thus relatively straightforward in the case of fixed coefficients models, it becomes more complicated in the case of models allowing for random taste heterogeneity. Indeed, in such models, the use of an unbounded distribution can lead to a non-zero probability of positive as well as negative travel time parameters. In this case, it is however not clear a priori whether such estimates do in fact indicate the presence of respondents with negative VTTS in the population, or whether they are simply an artefact of the model specification or the poor quality of the data used in model

[3] Here, it should be noted that the coefficient values themselves are estimators which are asymptotically normally distributed. As such, the ratio of β_{TT} and β_{TC} is itself a random variable, as discussed by Armstrong et al. (2001).

estimation.

One potential source of model misspecification can come in the form of an inappropriate choice of mixture distribution for the travel time coefficient. Like for most other coefficients, the most common choice of distribution for the travel time coefficient is the Normal distribution. Here, the unbounded nature of the Normal distribution can lead to major complications. Indeed, in the case where the *true* distribution yields strictly negative values, but has a mean close to zero with a long tail into the negative space of numbers, the symmetrical nature of the Normal distribution can, in approximation, lead to a significant share of positive values, even though such values are not actually *revealed* by the data. On the other hand, in the case where the source for such estimates are some problems with the data, the Normal distribution has the potential to produce such values, hence allowing the modeller to identify the problem and motivate an investigation into its causes. Although, without an in-depth investigation, it is desirable not to explain a significant probability of a positive travel time coefficient by the notion that some agents have a negative VTTS, it is similarly bad practice to simply constrain the model to purely negative values for β_{TT}, hence ignoring the impact of data or model imperfections.

The issue with the Normal distribution is thus the problem of deciding whether a non-zero probability of a positive coefficient is *revealed* by the data or is simply an artefact of the symmetrical nature of the distribution. Here, it can be seen that the Triangular distribution, in its symmetrical form, leads to similar problems, although by being bounded, it at least avoids the long tails. As such, the aim should be to use distributions that can signal the presence of such effects with a minimal risk of the effects actually being caused by the distribution itself. Here, the above arguments in relation to the Normal and the symmetrical Triangular suggest that such a distribution should not make too strict an a priori shape assumption. Additionally, it should be clear that distributions with estimated bounds have an advantage in terms of not making an a priori assumption about the range of the distribution. Here, it should be noted that with any distribution bounded on one side, the estimation of an additional offset parameter eliminates the issue of *fixed* bounds alluded to by Hess, Bierlaire & Polak (2005*c*). As such, even distributions generally seen as being bounded at zero can be specified with a flexible bound and thus have the potential to signal the presence of wrongly-signed coefficient values[4]. Given the problems caused by the long tails of some the distributions bounded on just one side (e.g. Lognormal), it can be argued that distributions bounded to either side, with estimated bounds[5], such as the Johnson S_B, have an advantage. In the case of flexible underlying distributions, the risk of values with the *wrong* sign being caused by the shape of the distribution, as with the Normal, largely disappears, although problems may still occur in the case of a significant mass at the endpoints, such as in the presence of individuals with zero VTTS (cf. Cirillo & Axhausen 2004, Hess, Bierlaire & Polak 2005*c*).

[4]Here, it is worth noting that a sign change on the attribute can be used in the case of asymmetrical distributions with an a priori constraint on the sign of the skewness (e.g. Lognormal), to act as a mirror function for the distribution of the coefficient.

[5]For distributions with strict domains of definition, an additional offset and range parameter can be used to obtain a distribution bounded on either side, with no a priori assumption on the location of the bounds.

It should be noted that another possible approach comes with the use of empirical or non-parametric distributions. However, such approaches are only beginning to be exploited in the estimation of MMNL models, and difficult issues of implementation and estimation need to be faced. The use of such distributions is not discussed in this work (with the exception of discrete mixtures in Chapter 5), but remains an important avenue for future research.

4.3.2 Data and model specification

The study presented here makes use of data collected as part of a recent value of time study undertaken in Denmark (Burge & Rohr 2004). The same dataset was also used in the non-parametric VTTS study of Fosgerau (2004), but the results of the two studies are not directly comparable, mainly because of the use of a different subsample.

In this study, we make use of data describing a binomial choice process for car-travellers on shopping trips, with alternatives described only in terms of travel cost and travel time. Each respondent was presented with 9 choice-situations, including one with a dominating alternative. After eliminating the observations with a dominating alternative, as well as additional data cleaning[6], a sample of 1,767 observations was obtained, for 230 respondents. With the sole aim of exploring the effects of different distributional assumptions, a very basic utility function was used, such that, in addition to an ASC associated with the first alternative, two coefficients were specified, associated with travel cost (DKK[7]) and travel time (min) respectively, with both attributes entering the utility in linear form. In the MMNL models, the repeated choice nature of the dataset was accommodated under the assumption of tastes varying across respondents, but not across observations for the same respondent. Finally, no treatment of the correlation between the travel cost and travel time coefficients was used in the present analysis.

Aside from a MNL model, estimated as the base model, a high number of different MMNL models were estimated in the current analysis, making use of different continuous distributions for the representation of the variation in the cost and time sensitivity across respondents. Given the high number of possible combinations of distributions, the models were in each case specified with the same choice of distribution for the two coefficients. In total, eleven MMNL models, each with different distributional assumptions, were estimated, making the analysis more comprehensive than most other existing studies looking at the choice of distribution. Several additional distributions, most notably the Beta, could not be used in the present analysis, as it was not possible to estimate an appropriate model[8]. For each type of distribution, the model was coded in Ox version 3.40 (Doornik 2001), and estimated using 1,000 Halton draws per dimension and per individual[9].

[6] Here, any individual who did not choose the dominating alternative in the dominated choice-situation was removed from the sample, as were non-traders.

[7] DKK1≈€0.13

[8] With the Beta, problems with exploding parameters as well as standard errors were encountered, with different implementations, and across a wide range of starting values, and various estimation approaches. Although these problems could be specific to the present application, more work is required to exploit the applicability of the Beta in the context of mixture models.

[9] I.e. leading to the use of 460,000 draws in each model.

For ease of presentation, a common notation was used across distributions, with α and γ giving the main parameters of the distribution, where a and b were additionally used to define the domain of the distribution, where appropriate. We will now look at the various distributions used in the analysis, where details on the actual functional form of the distributions are only given for the Johnson S_B and S_U distributions, with details for the remaining distributions available in the general literature (e.g. Evans et al. 2000).

Normal: specified with mean α and standard deviation γ

Lognormal: specified with mean α and standard deviation γ for the underlying Normal distribution. An additional offset parameter a was estimated, and the attribute entered the utility function under a sign change.

Johnson S_B: specified with offset parameter a, range parameter b, and shape parameters α and γ, with probability density function (*pdf*) given by:

$$f(x) = \frac{\gamma\, b}{(x-a)(a+b-x)}\phi\left(\alpha + \gamma \ln\left(\frac{x-a}{a+b-x}\right)\right), \tag{4.2}$$

where $a < x < a+b$, $\phi()$ is the standard Normal density function, and where $\alpha \in (-\infty, +\infty)$ and $\gamma > 0$. The shape parameter α has an effect on the skewness of the distribution, where negative values give a right-skewed distribution, zero gives a symmetrical distribution, and positive values give a left-skewed distribution. The second shape parameter γ defines the actual shape of the distribution in terms of peak, with values greater than 1 leading to a single, progressively steeper peak, while values lower than 1 will eventually lead to two peaks/modes at the extremes of the domain. With the notation in equation (4.2), a draw from the S_B distribution is obtained as:

$$x = a + \frac{b}{1 + \exp\left(\frac{-(z-\alpha)}{\gamma}\right)}, \tag{4.3}$$

where z is a draw from a standard Normal distribution.

Symmetrical Johnson S_B: specified as in equation (4.2), but with $\alpha = 0$.

Johnson S_U: specified with probability density function

$$f(x) = \frac{\gamma}{\sqrt{(x-a)^2 + b^2}}\phi\left(\alpha + \gamma \ln\left(\frac{x-a}{b} + \sqrt{\left(\frac{x-a}{b}\right)^2 + 1}\right)\right), \tag{4.4}$$

where a draw can be produced as:

$$x = a + b\,\frac{\exp\left(\frac{z-\alpha}{\gamma}\right)^2 - 1}{2\,\exp\left(\frac{z-\alpha}{\gamma}\right)}. \tag{4.5}$$

The S_U distribution is unbounded, but, unlike the Normal distribution, it can be asymmetrical, with the skewness depending on α. With this distribution, the meaning of a and b is different from the *offset* and *range* parameters, given that the distribution is unbounded. Here, they take on the meaning of a location and scale parameter.

Triangular: specified to be symmetrical, with lower bound a, upper bound b, and mode at $\frac{a+b}{2}$.

Gamma distribution: specified with shape parameter α, and scale parameter γ. An additional offset parameter a was used, and the attribute entered the utility under a sign change.

Exponential: specified with scale parameter α. An additional offset parameter a was used, and the attribute entered the utility under a sign change.

Logistic: specified with location parameter α and scale parameter γ.

Weibull: specified with shape parameter γ, offset parameter a, and with the attribute entering the utility under a sign change. In the implementation in Ox, the *pdf* is defined such that the estimate for α is in fact not the standard scale parameter, say η, but represents $\eta^{-\gamma}$.

Uniform: specified with lower endpoint a, and range parameter b.

4.3.3 Estimation results

The results of the estimation are summarised in Table 4.1 for the MNL model and the first group of MMNL models (Normal, Lognormal, Johnson S_B and Johnson S_U), and Table 4.2 for the second group of MMNL models (Gamma, Triangular, Exponential, Logistic, Weibull and Uniform). In each case, the estimation results are shown for the ASC, along with the four parameters (a, b, α, and γ) defining the distribution of the travel cost and travel time coefficient. Additionally, on the basis of the estimated parameters for the distribution of the two coefficients, the probabilities of counter-intuitively signed (i.e. positive) coefficients were calculated in each of the models. Finally, the results also present the mean implied VTTS for each model, along with the standard deviation. In each case, these measures were produced by a simple simulation process, making use of $1,000,000$ random draws for each of the two coefficients, based on the final model estimates for the parameters of the distribution. Here, special care was required in the models showing a non-zero probability of positive values for the travel cost coefficient. The fact that the domain for the denominator of the VTTS ratio straddles zero in such cases leads to extreme values in the simulation, and an overestimation of the variance of the VTTS. For this reason, the upper and lower[10] few percentile points were removed from the distribution of the cost coefficient in the models using the Normal distribution (2%), Triangular distribution (2%) and Logistic distribution (3%). A similar treatment was not used in the case of the travel time coefficient; here, the removal of a sufficient number of percentile points (e.g. 15% for the Normal) would

[10] To minimise the distortion to the mean.

	MNL	Normal	Lognormal	Johnson S_B	Johnson S_B (sym.)	Johnson S_U
Final LL	-1096.64	-1038.00	-1031.85	-1030.60	-1031.97	-1032.77
Parameters	3	5	7	9	7	9
adj. $\rho^2(0)$	0.1022	0.1484	0.1518	0.1512	0.1517	0.1494
Parameter	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)
ASC	0.302 (5.92)	0.367 (6.34)	0.369 (6.35)	0.363 (6.29)	0.361 (6.28)	0.366 (6.46)
Cost (DKK)						
a	-	-	-0.37 (-0.04)	-75.727 (-7.63)	-68.348 (-7.99)	-6.951 (-1)
b	-	-	-	56.686 (6.14)	51.372 (6.02)	0.095 (3.52)
α	-17.295 (-10.44)	-36.146 (-9.88)	3.543 (10.82)	-0.147 (-4.82)	0	5.906 (4.34)
γ	-	17.373 (6.53)	0.7 (3.39)	0.006 (0.29)	0.065 (0.64)	0.953 (4.74)
% positive	0.00%	1.87%	0.00%	0.00%	0.00%	0.00%
Time (min)						
a	-	-	-0.22 (-0.03)	-35.958 (-3.84)	-30.747 (-6)	7.439 (0.38)
b	-	-	-	29.94 (2.57)	25.036 (3.27)	0.284 (1.68)
α	-8.293 (-6.37)	-16.21 (-7.11)	2.687 (4.24)	-0.258 (-0.78)	0	10.385 (1.39)
γ	-	15.341 (6.46)	0.746 (2.48)	0.276 (0.84)	0.152 (0.58)	2.022 (1.5)
% positive	0.00%	14.53%	0.00%	0.00%	0.00%	0.87%
	DKK/hour	DKK/hour	DKK/hour	DKK/hour	DKK/hour	DKK/hour
VTTS (μ)	28.77	41.22	43.37	38.56	38.62	46.27
VTTS (σ)	-	108.11	59.56	34.76	35.52	50.62

Table 4.1: Estimation results for MNL model and MMNL models based on Normal, Lognormal, Johnson S_B and Johnson S_U distributions

have led to a severely underestimated standard deviation. Additionally, by being included in the numerator, the presence of values close to zero causes fewer problems than is the case for the cost coefficient.

We will now look at the results in more detail. The first observation relates to the performance of the various approaches in terms of model fit. All eleven MMNL models lead to significant improvements in LL over the MNL model, ranging from 57.27 units in the model based on the Exponential distribution, to 66.04 units in the model based on the S_B distribution with both shape parameters estimated. The differences in performance between the various MMNL models are very small, and, when taking into account the cost in terms of parameters, some of the other models (e.g. Gamma) in fact score slightly higher in the adjusted ρ^2 measure than the model based on the S_B[11]. Even though the differences in model fit are too small to lend significant weight to any comparisons across models, some interesting observations can be made. As such, all distributions, except for the Exponential, Logistic and Weibull, lead to better model fit than the Normal, from the point of view of the final LL, as well as the adjusted ρ^2 measure. While this should come as no surprise in the case of flexible distributions, such as the Johnson S_B and S_U, and the Gamma, it is striking that the Uniform distribution obtains slightly better fit than the Normal.

A crucial part of the results looks at the implications in terms of the presence of individuals with counter-intuitively signed values for the travel time and travel cost coefficients. A positive value for a travel time coefficient represents a situation where, all else (i.e. travel cost) being equal, a respondent prefers the *slower* alternative.

[11]Given that the different MMNL models cannot be compared with nested likelihood-ratio tests, preference is given to the adjusted ρ^2 measure, where other possibilities include for example the Akaike Information Criterion (AIC).

	Gamma	Symmetrical Triangular	Exponential	Logistic	Weibull	Uniform
Final LL	-1031.61	-1037.74	-1039.37	-1038.81	-1039.18	-1035.75
Parameters	7	5	5	5	7	5
adj. $\rho^2(0)$	0.1520	0.1486	0.1473	0.1478	0.1458	0.1503

Parameter	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)	est. (t-stat.)
ASC	0.368 (6.33)	0.367 (6.34)	0.352 (6.24)	0.366 (6.33)	0.352 (6.25)	0.366 (6.33)
Cost (DKK)						
a	9.636 (1.63)	-80.341 (-7.94)	6.436 (3.4)	-	6.631 (2.34)	-77.714 (-7.69)
b		6.499 (1.38)	-	-	-	73.602 (6.29)
α	1.012 (1.66)	-	0.026 (7.91)	-35.014 (-9.92)	0.028 (1.36)	-
γ	0.03 (2.01)	-	-	9.194 (5.95)	0.986 (5.67)	-
% positive	0.00%	1.12%	0.00%	2.10%	0.00%	0.00%
Time (min)						
a	3.656 (0.6)	-52.673 (-7.87)	16.508 (7.71)	-	16.518 (8.65)	-38.99 (-7.58)
b		19.449 (3.1)	-	-	-	40.811 (4.61)
α	1.011 (0.99)	-	19.541 (0.05)	-15.612 (-7.02)	6.559 (0.83)	-
γ	0.066 (1.47)	-	-	8.875 (6.31)	0.1 (0.33)	-
% positive	0.00%	14.60%	0.00%	14.74%	0.00%	4.68%

	DKK/hour	DKK/hour	DKK/hour	DKK/hour	DKK/hour	DKK/hour
VTTS (μ)	42.46	39.18	42.01	34.49	49.88	44.45
VTTS (σ)	49.32	64.24	33.95	53.04	38.02	59.06

Table 4.2: Estimation results for MMNL models based on Gamma, symmetrical Triangular, Exponential, Logistic, Weibull and Uniform distributions

Similarly, a positive value for the travel cost coefficient represents a situation where, with constant travel time, a respondent chooses the more expensive alternative. No such observations were included in the estimation; as such, results showing a non-zero probability of a positive travel time or travel cost coefficient should be seen as an artefact of the distributional assumptions.

For the travel cost coefficient, a non-zero probability of a positive coefficient was indicated by the models based on the Normal (1.87%), Triangular (1.12%) and Logistic (2.10%) distributions. For the travel time coefficient, the situation is more severe, with high probabilities in the case of the Normal (14.53%), Triangular (14.60%) and Logistic (14.74%) distributions. These results could lead to misleading conclusions in terms of the presence of individuals with negative VTTS. Lower probabilities of a wrongly signed travel time coefficient are observed with the Johnson S_U (0.87%) and the Uniform (4.68%).

Important differences arise between the MNL model and the various MMNL structures in terms of the implied VTTS. Here, the failure to account for the variation in the sensitivity to travel time and travel cost leads to a significant underestimation of the mean VTTS in the MNL model. These findings in terms of differences between VTTS estimates produced by MNL and MMNL models are consistent with a similar observation by Algers et al. (1998); however, in that research, the MNL model produced significantly higher VTTS than the MMNL models, while in the present work, the opposite is the case. This is an indication that the error can act in either direction[12].

Some differences also exist between the eleven MMNL models in the estimated

[12]It should also be noted that Algers et al. (1998) make use of the median instead of the mean in these comparisons.

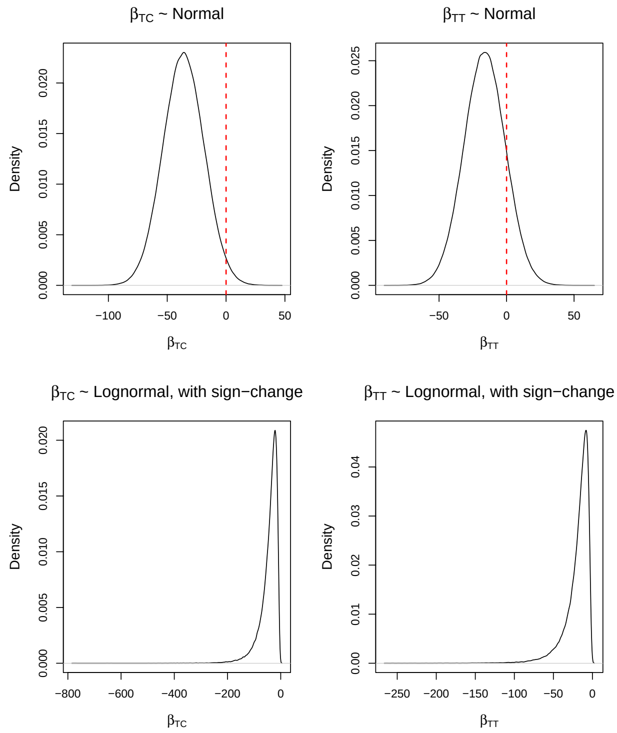


Figure 4.1: Implied distributions for β_{TC} and β_{TT}: MMNL models based on Normal and Lognormal distributions

VTTS. In general, these differences are however relatively small especially when looking at the mean VTTS, which is relatively stable aside from a few outliers, notably with the Logistic (underestimation) and Weibull (overestimation), and to a lesser extent the Johnson S_U. The differences are more significant when looking at the implied variation in the VTTS across respondents. Here, the use of the Normal leads to overestimation, which is partly caused by the presence of negative as well as positive travel time coefficients in the simulation, a factor that also plays a role in the case of the Triangular, Logistic and Uniform distributions. Finally, the long tails in the case of the Lognormal, Johnson S_U and Gamma also lead to higher variation in the trade-off.

We will now look at the implied distribution for β_{TC} and β_{TT} in each of the different MMNL models. For this, plots of the distributions are shown in Figure 4.1 for the Normal and Lognormal, Figure 4.2 for the Johnson S_B (asymmetrical and symmetrical), Figure 4.3 for the Johnson S_U and Gamma, Figure 4.4 for the Triangular and Exponential, and Figure 4.5 for the Logistic and Weibull. The Uniform distribution was excluded from this graphical analysis.

The results for the Normal and Lognormal (Figure 4.1) indicate the presence of a mode relatively close to zero, with a long tail to the left, which is given excessive weight by the Lognormal. In the case of the Normal, the symmetrical nature of the distribution means that, in order to accommodate the strong variation to the left of the mode, the tail to the right extends into the positive part of the domain for both β_{TC} and β_{TT}. Although, with the Lognormal, the additional offset parameter

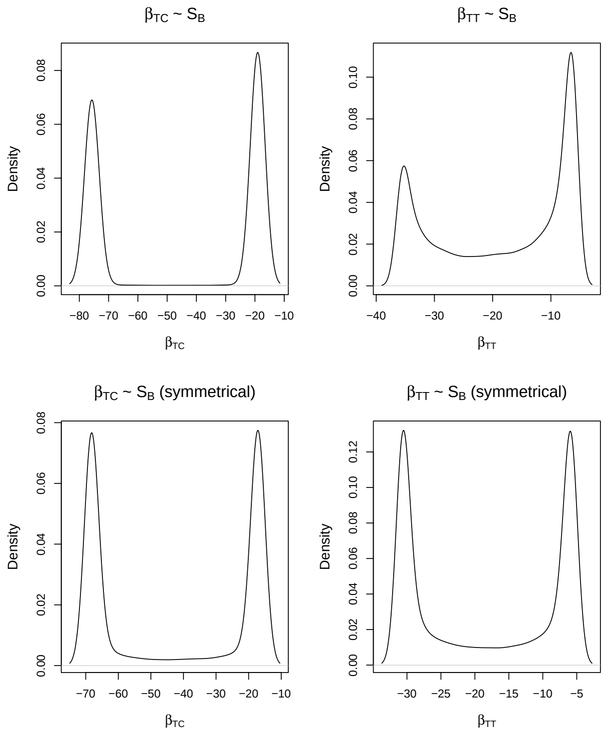


Figure 4.2: Implied distributions for β_{TC} and β_{TT}: MMNL models based on Johnson S_B (asymmetrical and symmetrical) distribution

is negative for both β_{TC} and β_{TT} (and hence positive after a sign-change), the difference to zero is in each case insignificant.

The results for the models based on the asymmetrical and symmetrical Johnson S_B distribution (Figure 4.2) indicate the presence of two modes for β_{TC} and β_{TT}, a phenomenon that none of the other distributions is able to pick up[13]. The difference in model fit between the two approaches is very small, suggesting that the additional constraint on α is acceptable, especially in the case of β_{TT}, where α was not significantly different from zero in the unconstrained model. Here, it should be noted that, in both models, the additional shape parameter γ obtains very low levels of significance, where similar problems with the S_B were already observed by Hess, Bierlaire & Polak (2005*c*), causing some concern[14]. Finally, in both models, the implied range for β_{TC} and β_{TT} is exclusively negative, on the basis of the estimated offset and range parameters.

The results for the Johnson S_U and the Gamma distribution (Figure 4.3) are comparable to those obtained with the Lognormal distribution, a conclusion that also applies in terms of the implied VTTS (mean and standard deviation). However,

[13] A posterior analysis would be of interest, allowing modellers to relate the distribution to socio-demographic information.

[14] The t-statistic is calculated with respect to 0, and not 1, where the difference would be statistically significant. A value of γ tending to 0 leads to a bi-modal distribution. As such, while suggesting problems in terms of the robustness in the estimation of γ, the results can also be seen as a strong indication of the presence of multiple peaks.

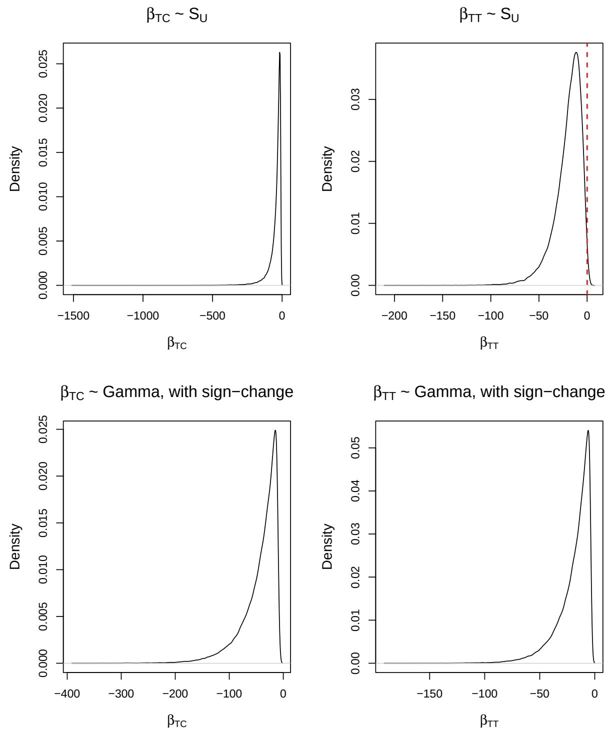


Figure 4.3: Implied distributions for β_{TC} and β_{TT}: MMNL models based on Johnson S_U and Gamma distributions

the S_U seems to lead to an even longer tail than the Lognormal in the case of β_{TC}, while, with the Gamma, the tails are more moderate. The results for the Gamma imply strictly negative values for β_{TC} and β_{TT} (after a sign-change), although the offset parameters are different from zero only at the 90% and 45% level respectively. Additionally, some issues with significance arise for α, significant at the 90% and 68% level for β_{TC} and β_{TT} respectively, while γ is significant only at the 86% level for β_{TT}. For the Johnson S_U, the location parameter is not significantly different from zero at reasonable levels of confidence for either β_{TC} or β_{TT}, where issues with significance also arise for the remaining parameters in the case of β_{TT}, with confidence levels of 91%, 84% and 87% for b, α and γ respectively.

The first observation that can be made from Figure 4.4 is that the model based on the Exponential distribution fails to pick up the variation in β_{TT}. The results in Table 4.2 support the impression that this distribution is inappropriate for β_{TT}, with a very high standard error associated with the α parameter. Aside from that, the offset parameters for both coefficients are significantly different from zero, and, after a sign change, indicate an exclusively negative domain for β_{TC} and β_{TT}. The results for the Triangular distribution are very similar to those obtained with the Normal, in terms of model fit, VTTS estimates[15], and crucially, the conclusions in terms of the presence of individuals with positive values for β_{TC} and β_{TT}, where this is again an effect of the symmetrical nature of the distribution, in conjunction with

[15]Here, the standard deviation is lower in the model based on the Triangular, but is still the second highest across all models.

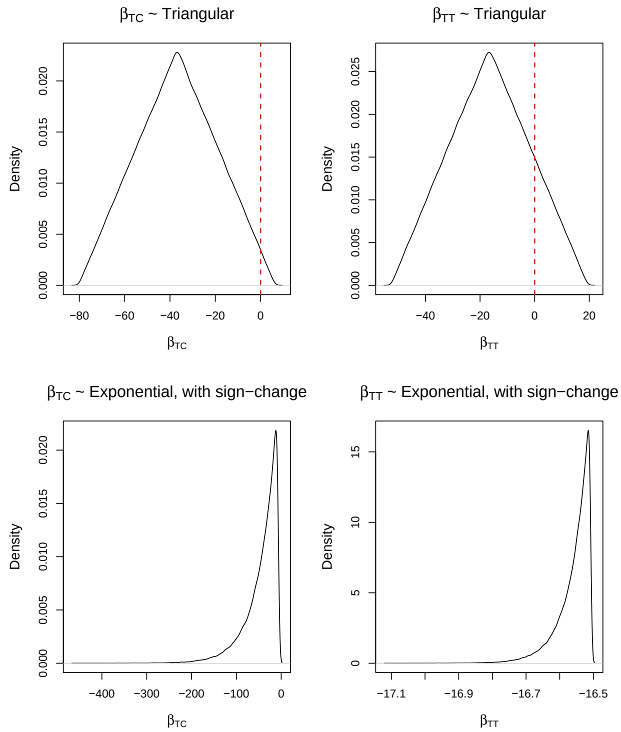


Figure 4.4: Implied distributions for β_{TC} and β_{TT}: MMNL models based on Triangular and Exponential distributions

a mean close to zero, along with high variation. Here, it should also be noted that the upper bound b is not different from zero in the case of β_{TC}, while, for β_{TT}, it is.

The plots for the Logistic and Weibull distribution (Figure 4.5) again show that, while model fits are comparable, there are important differences between symmetrical and asymmetrical distributions in the present application. As such, both distributions pick up significant variation in the sensitivity to β_{TC} and β_{TT}, along with a mean close to zero. While the symmetrical shape of the Logistic distribution leads to probabilities of 2.10% and 14.74% for positive values of β_{TC} and β_{TT} respectively, the offset parameters (a) in the case of the Weibull are significantly different from zero, and, after a sign-change, show an exclusively negative domain for the two coefficients. As was the case with a number of other distributions, issues with parameter significance also arise with the Weibull, namely with α for both coefficients, and γ in the case of β_{TT}.

4.3.4 Summary of findings

The analysis presented in this section has revealed the presence of significant levels of variation in the sensitivity to travel cost and travel time[16] in the population of travellers used in model estimation, allowing the various MMNL models to offer significant improvements in model fit over the MNL model.

The analysis has also shown that the model fit obtained by the eleven MMNL

[16]With the exception of the model based on the Exponential distribution.

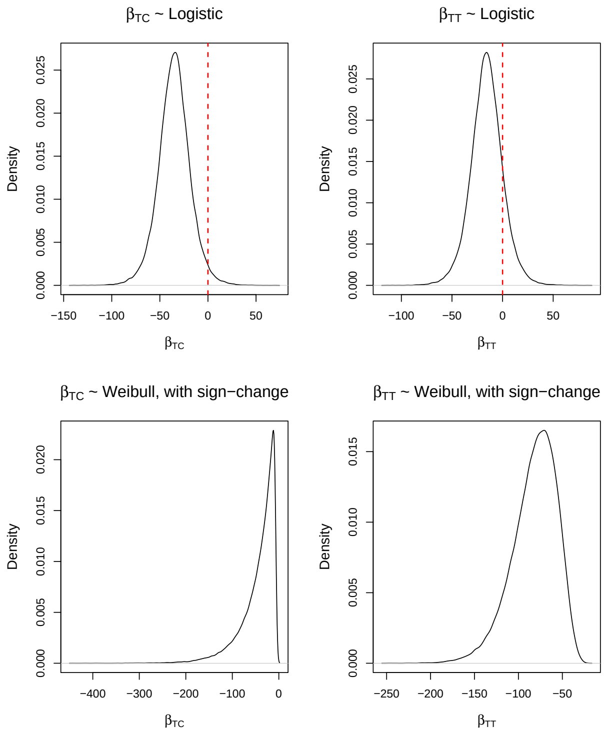


Figure 4.5: Implied distributions for β_{TC} and β_{TT}: MMNL models based on Logistic and Weibull distributions

models is remarkably similar. However, in what is a strong indication that model fit on its own is not a reliable measure when comparing mixture models based on different distributional assumptions[17], the substantial differences across models are quite important. This manifests itself partly in the estimates for the VTTS, where, although the mean values are comparable, there are some differences in the implied variation in the VTTS across respondents. More significant differences arise in terms of the implied shape for the distribution of β_{TC} and β_{TT}. These differences manifest themselves mainly in the form of differences in the weight in the tails, between symmetrical as well as asymmetrical distributions. The most startling result however arises in the case of the S_B distribution, which indicates the presence of two modes[18] in the distribution of β_{TC} and β_{TT}.

In the present context, the most interesting differences arise in terms of the bounds on the distribution of the VTTS. In the absence of a direct estimation of the VTTS, we are constrained to looking at the probability of positive estimates for β_{TC} and β_{TT}. On the basis of the data used in the estimation, the marginal valuations of changes in travel time and travel cost should be exclusively negative, leading to exclusively positive VTTS. However, in the results, several of the models

[17] It can be argued that too little weight is given to the behaviour in the tails of the population. However, it is the tails that are often important in the context of policy analysis, such as in the case of the introduction of tolled facilities, where the upper tail of the VTTS distribution is of interest, and the case of welfare impacts, where the lower tail is of interest.

[18] There is clearly a possibility of more than two modes; the S_B distribution is however limited to two modes.

indicate significant probabilities for negative values of β_{TT}. From the results, it becomes clear that problems arise in the case of symmetrical unbounded distributions (Normal & Logistic), as well as in the case of bounded distributions with a strong shape assumption (symmetrical Triangular[19]). In the present analysis, no problems are observed in the case of flexible distributions bounded either on both sides (S_B), or on the side where a strict truncation occurs in the data (Lognormal, Gamma, Exponential and Weibull). However, the use of distributions bounded on one side can lead to problems with a long tail to the other side, as in the case of the Lognormal. Finally, the Johnson S_U distribution, which is unbounded, is flexible enough to accommodate the *correct* mean and standard deviation, without leading to overestimated weight in the tails[20]. These findings are consistent with the theoretical claims made in Section 4.3.1.

The findings from this case-study are supported by those obtained by Hess, Bierlaire & Polak (2005*c*), who, on the basis of simulated data, show that the use of the Normal distribution can wrongly indicate the presence of individuals with positive values for β_{TT} in the case where the *true* distribution is entirely negative, with a long tail to the *left*. Additionally, results by Cherchi & Polak (2005) support the findings that model fit is not the best indicator when interested in substantive findings such as VTTS, and that results with counter-intuitive signs are often just an artefact of distributional assumptions.

In closing this discussion, it is worth briefly returning to the comparison between the Normal and the Uniform distribution. The results in terms of tail behaviour show far fewer problems for the Uniform distribution in terms of wrongly-signed travel time coefficients than for the Normal (4.68%, compared to 14.53%). This, in conjunction with the results in terms of model fit, could suggest that the Uniform distribution might be a more appropriate choice of *default* distribution in the initial search for random taste heterogeneity, especially when also taking into account that models based on the Uniform distribution are generally easier to estimate than those based on the Normal.

4.4 Interpretation of counter-intuitively signed coefficients

As alluded to in Section 4.3.1, there are several potential reasons why an estimation process can yield a non-zero probability of a positive travel time coefficient, aside from the effects of the shape of the assumed distribution.

Before moving on to potential explanations for positive travel time coefficients, it is worth briefly revisiting the micro-economic framework that governs the concept of VTTS. The currently accepted position is that individuals are assumed to potentially derive utility both from the consumption of goods and from the time they spend in different activities (though of course this may vary across individuals). This is represented by a *direct* utility function that includes both goods consumed and activity time as arguments. Individuals are assumed to organise their consump-

[19] Lower bias could be expected in the case of the asymmetrical Triangular, thanks to heightened flexibility.

[20] The probability of 0.87% for a positive β_{TT} is negligible.

tion of goods and their allocation of time between activities (e.g. work, travel and leisure) such that this direct utility is maximised, subject to constraints on the total amount of time and wealth available, and technical constraints on the minimum amount of time that it is necessary to allocate to a particular activity and/or to the consumption of a good.

The framework in this form was first crystalised in the work of Oort (1969) and, especially, De Serpa (1971), which serves as a useful point of reference for the discussion. A simple version of this framework would consider the allocation of time between say work, leisure and travel. Within this framework, De Serpa defined three concepts of the value of time. The first is the *resource* value of time, which arises because the total amount of time available for allocation to all activities is fixed by the total time constraint. The second is the value of time *allocated* to a particular activity, which arises because time (including travel time) itself is seen as a potential source of (positive or negative) utility, and not simply as a factor contributing to the production of other goods. The third concept is that of the value of *saving time in a particular activity*, which arises because of the technical constraints on the minimum amount of time that must be allocated to particular activities (for example in our case, the minimum time for a trip). This is equal to k/λ, where k is the Lagrange multiplier associated with the minimum travel time constraint, and λ is the Lagrange multiplier associated with the income constraint. It can be shown (see Jara-Diaz 2000) that:

$$\frac{k}{\lambda} = \frac{\partial U/\partial L}{\partial U/\partial G} - \frac{\partial U/\partial t}{\partial U/\partial G}, \tag{4.6}$$

where L is the time allocated to leisure, G is the consumption of goods, and t is the time allocated to travel.

A number of authors (see Jara-Diaz 2000) have shown that the marginal rate of substitution between the time and cost parameters in the (conditional indirect) utility of a discrete choice model is precisely equal to the ratio k/λ. Hence it follows from equation (4.6) that the VTTS which we are considering in this discussion is, from a microeconomic perspective, composed of two distinct components; the value associated with the ability to use time released by reductions in travel time in other activities (such as work or leisure) and the value associated with the change in utility derived directly from the travel experience itself. It seems that there has only been one recent attempt to disentangle these two components of the VTTS, by Jara-Diaz & Guevara (2003), where the empirical results reported suggest that for the sample of Chilean commuters studied, the VTTS was dominated by the strongly negative utility associated with the travel time experience itself.

Working in the above described framework, we should note that the Karush-Kuhn-Tucker optimality conditions guarantee that $k \geq 0$, with the equality condition (i.e., zero VTTS) applying if and only if the individual allocates more than the minimum required amount of time to the trip. For these circumstances to come about, the individual would have to derive a *positive* utility from time spent travelling at a rate exactly equal to μ/λ, where μ is the Lagrange multiplier associated with the total time constraint. That is to say, the traveller would be indifferent as between time spent in leisure and time spent travelling. Note further that in this model, there is no circumstance under which $k < 0$ could be observed. Assuming that $\lambda \geq 0$, this implies no circumstances in which a negative VTTS could be observed.

The above discussion has demonstrated that *if* one accepts the conventional microeconomic time allocation framework as providing an adequate basis for evaluating travel time savings, then positive and zero values of travel time savings are theoretically possible, but not negative ones. However, several recent papers discuss positive elasticity with respect to travel time. There are interesting statements like: "*I'd rather have an hour-plus commute than a five-minute commute. In the morning, it gives me a chance to work through what I'm going to do for the day. And it's my decompression time.*" (Sipress 1999, cited by Redmond & Mokhtarian 2001). Also, the conventional interpretation of travel as a derived demand, implying a disutility for time spent travelling, may be questioned. Mokhtarian & Salomon (2001) discuss the phenomenon of undirected travel, that is cases in which travel is not a byproduct of the activity but itself constitutes the activity, and argue that this may explain the evidences of excess travel observed even in the context of mandatory journeys.

Of course, it could be objected that the empirical results reported in the literature regarding negative VTTS provide prima facia evidence that the existing time allocation theory is incorrect or inadequate. However, while there are certainly many respects in which the existing theory could and should be improved (see for example the discussion in Mackie et al. 2001 and the recent work of Jara-Diaz 2003), it seems rather more likely that some of the recent findings of negative VTTS in the literature can be explained on the basis of inappropriate distributional assumptions.

Aside from the potential effects of inappropriate distributional assumptions, or data problems, two main possible explanations arise in the case of results showing positive travel time coefficients, as discussed by Salomon & Mokhtarian (1998). The first reason is the presence of unobserved objective factors. This is the case when the negative marginal utility of travel time increases is compensated by the gains in utility resulting from simultaneously conducted activities. The problem here is that our existing conceptual frameworks tend to lead us to think of travel and activity participation as distinct, whereas this is clearly not always the case. This topic is set to become increasingly important in the analysis of travel patterns due to the development of mobile data communication tools that massively expand the capacity for conjoining activities and travel in novel ways. The development of models that are able to analyse such conjoint activity patterns is thus an important avenue for future research.

A similar reasoning to that of conjoint activities applies in the case of desirable travel-experience factors (cf. Young & Morris 1981). As an example, commuters walking to work may prefer a slightly longer path through a scenic park to a shorter walk through congested and polluted streets. Similarly, people may prefer to use their car for going shopping for comfort reasons, even though the presence of bus priority lanes would make for a quicker bus journey. The impact of such unobserved attributes is related to the second reason for excess travel cited by Salomon & Mokhtarian (1998); namely the presence of unobserved subjective factors. As an example, the pleasure of driving an automobile, combined with the social positive perception of having and using a car, relayed by the marketing of automobiles, may explain the presence of excess travel.

The impacts of such travel-experience factors can be illustrated relatively easily with the help of suitably generated synthetic data. As such, Hess, Bierlaire & Polak (2004) show that failing to account properly for the impact of travel-experience factors can significantly affect the split between positive and negative coefficients in MMNL models, or even falsely indicate the presence of significant random taste heterogeneity in the case where only fixed coefficients were used in data generation.

Clearly, it is often not possible to unambiguously quantify the impact of conjoint activities or travel-experience factors[21], and there is thus a significant risk of a biased estimate of the travel time coefficient. The possibility of such bias can never be discounted, even in the case of intuitively *correct* results. However, the issues described above should be considered especially in the explanation of positive travel time coefficients (or a positive probability of such coefficient values), and researchers should strive to include as many descriptive attributes as possible, to reduce the impact of the correlation between travel time and unmeasured variables on the estimation of travel time coefficients[22]. If, despite efforts to reduce the impact of unobserved factors, and with the use of flexible bounded distributions, the results still indicate a significant share of positive travel time coefficients, modellers should acknowledge the potential impact of unobservables on their estimates, and an appropriate re-labelling of the coefficients is desirable to avoid any confusion. Here, it should be noted again that a model allowing for a non-zero share of positive β_{TT} may well obtain better model fit, by being able to capture the effects of unobserved attributes. However, if one accepts the validity of the time allocation theory discussed above, then it would be wrong to use this better model fit as a proof of the existence of such valuations, and it should rather be seen as an indication of the extent of the impact of such unmodelled factors.

4.5 Summary and Conclusions

This chapter has discussed issues of specification and interpretation that need to be faced when using mixture models such as MMNL to represent random variations in tastes across respondents.

The main aim of this chapter was to explore the potential of hitherto little used distributions in the estimation of random coefficients models. In this context, a VTTS case-study was conducted, making use of distributions such as the Johnson S_U, Gamma, Weibull and Logistic, in addition to more common choices, such as the Normal, Lognormal and Johnson S_B. The results from this analysis have shown that, while the various distributions lead to similar performance in terms of model fit, they lead to major differences in the substantive results. These relate partly

[21]Factors such as comfort can clearly influence choice behaviour, but are notoriously hard to measure, and hence include in models.

[22]It can be seen that by explicitly accounting for all travel-experience attributes, only the actual cost in time as a resource would remain; this would be constant across alternatives (e.g. modes or activities) for a given person at a specific moment in time. As such, obtaining different VTTS for different alternatives in a mode choice analysis is in fact a sign that some travel-experience attributes have not been included in the utility specification; exploring and exploiting such different VTTS measures is however often one of the main objectives of such studies.

to the mean and standard deviation of the implied distribution of the VTTS, but manifest themselves especially in terms of the findings with regards to the presence of individuals with negative VTTS. Here, the flexible distributions, such as the Johnson S_B, indicate a zero probability for such negative measures, which is consistent with the data. On the other hand, the commonly used Normal distribution indicates a probability of 14.53% for positive valuations of increases in travel time. Similar problems are observed with other symmetrical distributions.

The results presented in this chapter clearly support the notion that the distributional assumptions made during model specification can have a significant impact on model results, a fact that needs to be borne in mind in the interpretation of the results. Furthermore, at least in the present study, the results validate the theoretical claims from Section 4.3.1, showing a lower risk of misspecification with more flexible distributions. This suggests that modellers should increasingly look into the use of alternatives to the Normal distribution for the representation of random taste heterogeneity. While, in some cases, the use of the Normal may be appropriate in studies interested solely in the mean and standard deviation, the presence of strong asymmetries in the *true* distribution can lead to bias even in these more overall measures, as shown in the case of the standard deviation in the application presented in this chapter. In no case however should a distribution like the Normal be used to infer any conclusions in relation to the behaviour in the tails of the population. Here, it seems preferable to use distributions bounded on either side, such as the Johnson S_B, with bounds estimated from the data, to still allow for the effects of data problems or incomplete specifications to manifest themselves.

More work remains to be done, in terms of further tests with the distributions used in this chapter, as well as the use of other distributions, or mixtures of distributions. Additionally, the high standard errors observed for some of the parameters of the more flexible distributions (e.g. Johnson S_B, S_U and Gamma) are a cause for concern, and it remains to be seen whether these findings are specific to the application at hand. Finally, to allow more widespread use of such distributions in mixture models, some effort needs to go into devising efficient and robust approaches for estimating models based on distributions more complex than the Normal.

The second part of this chapter has discussed an important and timely issue, namely the interpretation of results showing a significant share of travellers with negative VTTS, in the case where appropriate measures have been taken to minimise the potential bias caused by inappropriate distributional assumptions. Here, the theoretical discussions in Section 4.4 have shown that, under the microeconomic theory of time allocation, positive as well as zero[23] VTTS measures are possible, but negative measures are not. The discussion has also shown how the estimation can be biased, to the point of indicating high shares of positive travel time coefficients, by the presence of unmodelled factors, such as travel-experience attributes, or conjoint activities. If problems with counter-intuitively signed coefficients occur in practice, then it is critical to acknowledge the potential impacts of such factors, and to interpret the model appropriately. Specifically, the name of the estimated parameter should be changed in order to emphasise that it potentially captures more than one specific effect, and its use to compute VTTS measures, and/or to perform cost-benefit analysis, should be avoided.

[23] See also Richardson (2003).

Here, it should be noted that important further insights into the distribution of taste coefficients can be obtained by conducting a posterior analysis to determine the individual-specific taste coefficients conditional on the observed choices. Indeed, even in the case where the original estimation results indicate a significant probability of positive travel time coefficients, it is conceivable that, in such an analysis, a positive coefficient would only be associated with a very low number of respondents (cf. Sillano & Ortúzar 2004). This further underlines the risk of misinterpretation with MMNL models, and suggests that a model indicating a non-zero probability of positive travel time coefficients should not be used for VTTS calculation or forecasting without first conducting an appropriate posterior analysis.

Finally, in the context of the VTTS discussions in this chapter, it should be noted that a whole range of other effects, besides the distributional assumptions and the presence of unmodelled travel-experience attributes and conjoint activities, can have an influence on the estimation of the VTTS. Such factors, which can play a role in fixed as well as random coefficients models, include for example the assumptions made with regards to model structure in terms of correlation across alternatives and/or observations, the assumptions made with regards to the utility function (i.e. linear vs non-linear), and the design of survey questionnaires in the case of SP data. These issues are discussed for example by Gaudry et al. (1989) and Hensher (2001*b*, 2004).

Chapter 5

Discrete mixture models

5.1 Introduction and context

The discussion in Chapter 4 has highlighted the fact that the important gains in flexibility and accuracy that can be obtained with the use of mixture models come at the cost of having to face major issues in the specification and interpretation of such models.

One of the main complications is the need to specify distributions for random taste coefficients in the relative absence of information about the true underlying pattern of taste variation. Even with the use of the most flexible distributions available, it seems almost inevitable that there will be some discrepancy between the true and postulated distribution; cases will arise in which real-world behaviour cannot be characterised adequately by one of a set of standard statistical distributions.

One case in point arises in the modelling of tastes which may theoretically have a significant mass at zero but be exclusively positively or negatively signed elsewhere (cf. Cirillo & Axhausen 2004). The situation becomes even more complicated in the case of an attribute which some individuals value positively and some individuals value negatively, with a remaining part of the population being indifferent to the attribute. This applies for example in the case of attributes describing discrete qualitative features of an alternative, such as a distinction between forward and backward facing seats for rail-travel. Representing this situation is not possible with the use of standard continuous distributions[1], and the results obtained with such distributions may lead to unwarranted conclusions.

Given these problems, it is of interest to explore alternative ways of representing random variations in tastes across respondents, avoiding some of the issues discussed above.

One possible solution is to use Kernel densities of individual-specific coefficient values in the search for an appropriate distribution. The most basic approach consists of estimating individual-specific MNL models, which is only possible in the presence of multiple observations per individual, and to infer information about the true distribution from plotting the Kernel density of the hence obtained coefficient values. This causes significant problems in practice, given the potential lack of information in the resulting small datasets. In this context, Hensher & Greene (2003)

[1] The notion of a mass at a specific point (especially if not at the extremes of the domain) does not apply.

advocate the use of a jackknife-style procedure that starts with the full sample, and proceeds by eliminating individuals one-by-one, each time estimating a new model. The resulting set of estimates can then be used to produce a Kernel density function. In practice, the applicability of such methods is often limited by high computational cost and data requirements.

A second approach is to use empirical distributions, based on estimating a set of support points with corresponding masses, with linear segments between support points. The success of this approach however not only depends crucially on the number of support-points used, but important issues of implementation need to be faced for the estimation of the support points, where problems with singularities arise.

A final approach comes in the use of non-parametric approaches, which are free of a priori assumptions about the shape of the true distribution. The application of such approaches to the estimation of VTTS is described by Fosgerau (2004). The results show that the non-parametric approaches outperform a set of parametric approaches, but the fact that such methods are very data-hungry leads to problems in recuperating the distribution in the tails of the population, a situation that Fosgerau addresses through the use of a semi-parametric approach, where part of the distribution is accounted for through a set of covariates. While very promising, non-parametric and (to a lesser extent) semi-parametric regression approaches can be difficult to use in practice, and more work is required to allow widespread application.

The three approaches described above are able to deal with the main issue described in Chapter 4, namely the behaviour in the tails of the distribution. Similarly, they do, unlike most standard continuous distributions, have the ability to allow for a multi-modal distribution of a specific taste coefficient. However, it seems that neither of the three approaches can deal adequately with the presence of a heightened mass at a given point, such as a zero VTTS. While the use of Kernel densities (and potentially also the empirical approach) can signal the presence of such mass points, the issue of how to incorporate them in the final model remains.

In this chapter, we explore an alternative approach, based on the idea of replacing the continuous distribution functions by discrete distributions, spreading the mass among several discrete values. Theoretically, such discrete mixtures allow modellers to deal with each of three issues described above (tail-behaviour, multiple modes, inflated mass), although certain issues, notably in estimation, need to be addressed, as described in Section 5.2.

Mathematically, the model structure of a discrete mixture model is similar to that of a latent class model (cf. Kamakura & Russell 1989, Chintagunta et al. 1991), assigning different coefficient values to different parts of the population of respondents, a concept discussed in the field of transport studies for example by Greene & Hensher (2003) and Lee et al. (2003). The work of Gopinath (1995) especially is of interest in the context of the case-study described in this chapter, as it makes use of a latent class model in the analysis of variations in the VTTS across respondents, showing the presence of multiple subgroups in the population. Latent class approaches make use of two sub-models, one for class-allocation, and one for within-class choice. The former models the probability of an individual being assigned to a specific class as a function of attributes of the respondent and possibly of the alternatives in the choice set. The within-class model is then used to compute the class-specific choice probabilities for the different alternatives, conditional on the tastes within that class. The actual choice probability for individual n and alternative i is given by a sum of the class-specific choice probabilities, weighted by the class-allocation choice probabilities for that specific individual.

The latent-class approach is appealing from the point of view that it allows for differences in sensitivities across population groups[2], where the group-allocation can be related to socio-demographic characteristics. However, in practice, it may not always be possible to explain group-allocation with the help of a probabilistic model relating the outcome to observed variables[3]. As such, in this chapter, we explore the use of models in which the class-allocation probabilities are independent of explanatory variables, and are simply given by constants that are to be estimated during model calibration. The resulting model thus exploits the class-membership concept in the context of random coefficients models, with a limited set of possible values for the coefficients. In theory, existing discrete distributions (e.g. Poisson) could be used; however, this comes at the cost of flexibility in terms of an a priori shape assumption. This problem does not exist in the case where a fixed set of coefficient values are used that each have an associated probability, but where the values and associated probabilities are free from any a priori constraints.

Although the properties of discrete mixture models have been discussed by several other authors (e.g. Wedel et al. 1999), the model structure has not received widespread exposure or application, despite its many appealing characteristics. Indeed, thus far, there have seemingly been only two applications of this approach in the area of transport research, by Gopinath (1995), in the context of mode choice for freight shippers, and by Dong & Koppelman (2003), who make use of discrete mixtures of MNL models in the analysis of mode choice for work trips in New-York, referring to the resulting model as the "Mass Point Mixed Logit model".

Given the above discussion, part of the aim of this chapter is to re-explore the potential advantages of discrete mixture models, with the hope of encouraging their more widespread use. However, the main aim, and contribution of this chapter, is to demonstrate how the model structure can be exploited to allow for a part of the population in which people are indifferent to changes in a specific attribute, a treatment that is not generally possible with the use of continuous mixture structures. Although the discussion in this chapter looks specifically at the case of zero valuations of changes in travel time (leading to zero VTTS), the same principle applies in the case of other attributes. Finally, the analysis also aims to investigate the potential bias in coefficient estimates that can result from not allowing for the presence of individuals with such zero valuations.

The remainder of this chapter is organised as follows. The next section sets out the theory behind discrete mixture models. Section 5.3 describes a set of tests of the validity of the model structure conducted with the help of synthetic data, while Section 5.4 presents the main case-study testing for the presence of respondents with zero VTTS. Finally, Section 5.5 summarises the contents of the chapter and

[2]It should be noted that latent class approaches can also be exploited to allow for differences in utility specification or even choice set formation across population groups.

[3]This situation is similar to the case where taste heterogeneity cannot be explained deterministically, leading to a requirement for using random coefficients models.

presents the conclusions of the study. To a large extent, the material covered in this chapter corresponds to that discussed by Hess, Bierlaire & Polak (2005*b*).

5.2 Methodology

We will begin by introducing some general notation, which will be used throughout the remainder of this chapter. Specifically, let $x_{i,n}$ again be a vector defining the attributes of alternative i as faced by respondent n (potentially including interactions with socio-demographic variables), and let β be a vector defining the tastes of the decision-maker, where, in purely deterministic models, β is constant across respondents. Let x_n be a vector grouping together the individual vectors $x_{j,n}$ across the alternatives contained in the choice set of respondent n, and let γ represent an additional set of parameters, which can for example contain the structural parameters (and possibly allocation parameters) used to represent inter-alternative correlation in a GEV context. In a very general form, we can then define $P_n(i \mid x_n, C_n, \gamma, \beta)$ to give the choice probability of alternative i for individual n, with a choice set C_n, conditional on the observed vector x_n, and for given values for the vectors of parameters β and γ (to be estimated). Due to the potential inclusion of socio-demographic attributes in x_n, this notation allows for deterministic variations in tastes across respondents.

This notation can now be used as the building block for models allowing for a distribution of tastes across respondents. In a continuous mixture model, the choice probabilities are then given by:

$$P_n(i \mid x_n, C_n, \gamma, \Omega) = \int_\beta \left[P_n(i \mid x_n, C_n, \gamma, \beta) f(\beta \mid \Omega)\right] \mathrm{d}\beta, \tag{5.1}$$

where the vector β is distributed according to $f(\beta \mid \Omega)$, with vector of parameters Ω. With $P_n(i \mid x_n, C_n, \gamma, \beta)$ giving MNL choice probabilities, equation (5.1) represents the choice probabilities in a MMNL model; however, any other GEV-type choice probability can be used for $P_n(i \mid x_n, C_n, \gamma, \beta)$, with an explicit role for the vector γ, leading to a more general GEV mixture model.

From the point of view of statistics, speaking in the context of mixture densities, the MMNL model is a continuous mixture of MNL models over the distribution of β[4]. In this chapter, we replace these continuous mixtures by discrete mixtures[5], limiting the number of possible values for β. As such, we now divide the set of parameters β into two sets; $\bar{\beta}$ represents a part of β containing deterministic parameters, while $\widehat{\beta}$ is a set of K random parameters that have a discrete distribution. Within this set, the parameter $\widehat{\beta}_k$ has m_k mass points $\widehat{\beta}_k^j$, $j = 1, \ldots, m_k$, each of them associated

[4] A mixture density is a *pdf* which is a convex linear combination of other *pdf*'s. If $f(\varepsilon, \theta)$ is a *pdf*, and if $w(\theta)$ is a nonnegative function such that $\int_a w(a)\mathrm{d}a = 1$, then $g(\varepsilon) = \int_a w(a) f(\varepsilon, \theta)\mathrm{d}a$ is also a *pdf*. We say that g is a mixture of f. With this in mind, the MMNL model is indeed a mixture of MNL models over the continuous distribution of the vector of tastes β.

[5] Returning to the domain of statistics, it is clear that discrete mixtures of *pdf*s are also possible. If $f(\varepsilon, \theta)$ is a *pdf*, and if w_i, $i = 1, \ldots, n$ are nonnegative weights such that $\sum_{i=1}^n w_i = 1$ then $g(\varepsilon) = \sum_{i=1}^n w_i f(\varepsilon, \theta_i)$ is also a *pdf*. We say that g is a discrete mixture of f.

with a probability π_k^j, where we impose the conditions that

$$0 \leq \pi_k^j \leq 1, \quad k = 1, \ldots, K; \; j = 1, \ldots, m_k, \tag{5.2}$$

and

$$\sum_{j=1}^{m_k} \pi_k^j = 1, \quad k = 1, \ldots, K. \tag{5.3}$$

For each realisation $\widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K}$ of $\widehat{\beta}$, the choice probability is given by

$$P_n\left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right), \tag{5.4}$$

where the deterministic part of $\bar{\beta}$ stays constant across realisations of the vector $\widehat{\beta}$.

The unconditional[6] choice probability for alternative i and decision-maker n can now be written straightforwardly as a mixture over the discrete distributions of the various elements contained in $\widehat{\beta}$ as:

$$\begin{aligned} & P_n\left(i \mid x_n, C_n, \gamma, \bar{\beta}, \widehat{\beta}, \pi\right) \\ & \qquad = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P_n\left(i \mid x_n, C_n, \gamma, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right) \pi_1^{j_1} \ldots \pi_K^{j_K}, \end{aligned} \tag{5.5}$$

where $\bar{\beta}$, $\widehat{\beta}$ and π ($\pi = \langle \pi_1^1, \ldots, \pi_1^{m_1}, \ldots, \pi_K^1, \ldots, \pi_K^{m_K} \rangle$) are vectors of parameters to be estimated in a regular maximum likelihood estimation procedure. An obvious advantage of this approach is that, if the probability model (equation (5.4)) used inside the mixture has a closed form, then so does the discrete mixture itself.

In this chapter, we mainly focus on the simple case where the underlying choice model is of MNL form; however, the form given in equation (5.5) is appropriate for any underlying GEV model. The approach can easily be extended to the case of combined discrete and continuous random taste variation, by partitioning β into three parts; the above defined parts $\bar{\beta}$ and $\widehat{\beta}$, and an additional part $\widetilde{\beta}$, whose elements follow continuous distributions. This however leads to a requirement to use simulation, as with all continuous mixture models. Allowing for continuous random terms in addition to discrete random terms not only increases flexibility from the point of view of random taste heterogeneity, but also allows for the use of error-components to represent heteroscedasticity and inter-alternative correlation, where the latter is however also possible with the use of an underlying GEV structure.

Finally, independently of the additional incorporation of continuous random variations in tastes, a treatment of repeated choice observations analogous to the standard continuous mixture treatment[7] is made possible by replacing the conditional choice probabilities for individual observations in equation (5.5) by probabilities for sequences of choices, and by using the resulting discrete mixture term inside the LL function.

[6] On a specific realisation of β, not on the distribution of $\widehat{\beta}$.

[7] Tastes varying across individuals, but not across observations for the same individual.

The approach we use in this chapter clearly offers greater modelling flexibility than an approach based on fixed-point estimates, by allowing for random as well as deterministic variations in tastes. It may also seem tempting to see the approach as an alternative to models using continuous distributions. However, while the approach does have the advantage of being free from any assumptions resulting from the choice of a specific statistical distribution, it is in most cases impractical to use discrete mixtures as an approximation to continuous mixtures, notably because of the resulting over-specification in terms of the number of parameters, which can lead to problems in estimation. In the remainder of the chapter, we therefore rely mainly on the notion that the approach is an extension of a fixed point model, as opposed to an approximation to a continuous mixture model. A detailed comparison between continuous and discrete mixture models, across a number of different datasets, is an important topic for future research[8].

Several issues arise in the estimation of discrete mixture models. Firstly, the non-concavity of the log-likelihood function does not allow the identification of a global maximum, even for discrete mixtures of MNL. Given the potential presence of a high number of local maxima, performing several estimations from various starting points is thus advisable. Also, it is good practice to use starting values other than 0 or 1 for the π_k^j parameters. Secondly, constrained maximum likelihood must be used to account for constraints (5.2) and (5.3). Here, it should be noted that eliminating (5.3) by replacing π_k^1 with

$$\pi_k^1 = 1 - \sum_{j=2}^{m_k} \pi_k^j \tag{5.6}$$

does not help, as the constraint $0 \leq \pi_k^1 \leq 1$ now leads to the new condition $0 \leq \sum_{j=2}^{m_k} \pi_k^j \leq 1$. Thirdly, clustering of mass points (for example around the mode of the true distribution) is a frequent phenomenon with discrete mixture models. Although this can be a sign that the number of mass points is too large[9], it may in some cases also be a feature of the optimisation algorithm, such that the use of additional bounds on the mass points is useful, based on the definition of (potentially mutually exclusive) a priori intervals for the individual mass points.

For the purpose of this analysis, the model was coded into BIOGEME[10] (Bierlaire 2003), where various constraints on the parameters can be imposed to address the issues described above. This also allows modellers to test the validity of specific assumptions, such as a mass at zero for the VTTS. At this point, it should be noted that, in testing for a mass at a specific support point, only zero should be used as a candidate value. Indeed, this equates to explicitly allowing for the presence of individuals for whom the utility functions are unaffected by changes in a specific attribute. For other values, the notion of a heightened mass does not apply, under

[8]Here, some initial results by Hess, Bierlaire & Polak (2005*b*) show that, even though the MMNL model obtains better model fit than a corresponding discrete mixture model with just two support points, the differences are relatively small, suggesting that even a small number of mass points can be sufficient to account for major parts of the variation in tastes.

[9]In this context, a heuristic is needed to determine the optimal number of support points in actual applications.

[10]See also http://roso.epfl.ch/biogeme

Parameter	Value
ASC for car	4
Interchanges	-1.15
Travel cost (CHF)	-0.3
Frequency (per hour)	0.9
Rail travel time (min.)	-0.07

Table 5.1: Generic parameter values used in generation of synthetic data

the standard assumption of the true distribution being continuous.

5.3 Testing the validity of the discrete mixture structure

Before proceeding to the use of discrete mixture models in practice, it is important to investigate the validity of the approach as well as its implementation in BIOGEME, by testing its performance on synthetic data where the *true* values of the parameters are known. For this, a quasi-simulated dataset was produced on the basis of a sample of $1,242$ observations taken from a binomial mode choice survey (car *vs* rail) conducted in the context of the analysis of the VTTS in Switzerland (Axhausen et al. 2004, Koenig et al. 2004). For the present analysis, the sample size was augmented from $1,242$ to $5,000$ through minor random variations on the observed attributes.

The utility specification in this model uses travel cost, travel time, frequency, and the number of interchanges as explanatory variables, where linear specifications are used for all attributes, and where the ASC for rail is normalised to zero. In order to generate the synthetic choices, we assume that, except for the travel time coefficient for the car alternative, the true parameters are fixed as shown in Table 5.1, giving a true VTTS for rail-travel of 14CHF/hour[11].

In the first experiment, we assume that the population is divided into two segments. The VTTS for car-travel in the first segment, composed of 50% of the sample, is assumed to be 16CHF/hour (car travel time coefficient at -0.08), while it is 6CHF/hour for the second segment (car travel time coefficient at -0.03).

The resulting dataset was then used in the estimation of a discrete mixture model with an underlying MNL structure and two support points for the car travel time coefficient, where the results are shown in Table 5.2. The results show a near-perfect recovery of the $50\% - 50\%$ split in VTTS, where the upper VTTS is slightly underestimated, at 14.72CHF/hour, while the lower one is overestimated, at 7.13CHF/hour. The VTTS for rail is also slightly overestimated, at 14.84CHF/hour. These slight biases are however well within acceptable bounds.

In the second experiment, we assume that the segment with the lower VTTS represents only 30% of the population. The estimation results for this dataset are summarised in Table 5.3, showing that the $70\% - 30\%$ split is reproduced almost perfectly. Both VTTS measures are slightly underestimated, at 4.70CHF/hour and

[11] CHF1 ≈€0.65

Sample size:	5,000
Final log-likelihood:	-868.446
Adjusted $\rho^2(0)$:	0.7468

Parameter	est.	t-stat.
ASC for car	4.0265	15.76
Interchanges	-1.2306	-12.70
Travel cost (CHF)	-0.3138	-19.62
Frequency (per hour)	0.9282	14.15
$\beta_{TT,rail}$ (min.)	-0.0776	-13.58
$\beta_{TT,car}(A)$ (min.)	-0.0770	-5.73
$\beta_{TT,car}(B)$ (min.)	-0.0373	-3.54
Mass for $\beta_{TT,car}(A)$	0.5149	2.55
Mass for $\beta_{TT,car}(B)$	0.4851	2.40

Table 5.2: Results for first synthetic data experiment

13.75CHF/hour, instead of 6CHF/hour and 16CHF/hour respectively. The rail VTTS is estimated at 13.73CHF/hour, instead of 14CHF/hour. Again, these biases are acceptable.

Although more testing is required, the two experiments described here have shown that the discrete mixture models are indeed able to recover the values and *market shares* of discretely distributed coefficients. The extension to cases with more than two mass points is possible, although the estimation becomes significantly more complicated, with the presence of several local maxima, and possible degeneracy, that is convergence of two points toward a common value.

5.4 VTTS case-study

We now turn our attention to the analysis exploiting the discrete mixture structure to allow for the presence of individuals with zero VTTS. For these experiments, SP data from the Swiss VTTS study were used, in the form of a binomial route choice survey for rail travellers. The sample used in the present analysis includes 315 observations from business travellers, 1,881 observations from leisure travellers, and 288 observations from travellers on shopping trips[12]. The relatively small sample sizes for the business and shopping groups could decrease reliability of the results in these two groups, although problems with significance were only observed in one case, as detailed later on.

Again, the final utility specification uses travel cost, travel time, frequency, and the number of interchanges as explanatory variables, where linear specifications are used for all attributes. No significant ASCs could be identified in the present model. The analysis first looks at a simple MNL model, estimated separately for each of the three subgroups, with results summarised in Table 5.4. The results show that

[12]This analysis differs from that conducted by Hess, Bierlaire & Polak (2005*b*) in the use of three separate population segments, and a slightly different specification of the utility function.

Sample size:	5,000
Final log-likelihood:	-906.999
Adjusted $\rho^2(0)$:	0.7357

Parameter	est.	t-stat.
ASC for car	4.1307	16.38
Interchanges	-1.2055	-12.90
Travel cost (CHF)	-0.3203	-20.17
Frequency (per hour)	0.9600	14.91
$\beta_{TT,rail}$ (min.)	-0.0733	-14.04
$\beta_{TT,car}(A)$ (min.)	-0.0251	-2.35
$\beta_{TT,car}(B)$ (min.)	-0.0734	-7.25
Mass for $\beta_{TT,car}(A)$	0.2704	2.19
Mass for $\beta_{TT,car}(B)$	0.7296	5.91

Table 5.3: Results for second synthetic data experiment

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood	-124.69		-925.36		-139.30	
Adjusted $\rho^2(0)$	0.4106		0.2872		0.2822	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.1285	-6.60	-1.1737	-18.95	-0.9394	-6.30
Travel cost (CHF)	-0.3051	-5.08	-0.1335	-5.36	-0.5658	-3.75
Frequency (per hour)	0.5970	6.11	0.4188	12.19	0.6603	7.22
Travel time (min.)	-0.1258	-7.71	-0.0300	-7.48	-0.0465	-1.13
VTTS (CHF/hour)	24.73		13.50		4.93	

Table 5.4: VTTS case-study: MNL estimation results

all estimates are of the correct sign, and significant, with the exception of the travel time coefficient for respondents on shopping trips, which is significant only at the 74% level. In terms of substantive results, the estimation does, as expected, show higher VTTS for business travellers, with very low VTTS for shopping trips, where the value does however need to be put into context by noting the high standard error for the travel time coefficient.

We next estimate discrete mixtures of the three MNL models, with results summarised in Table 5.5. With the aim of investigating the presence of individuals with zero valuations of travel time changes, the models are specified with two travel time coefficients, of which one is fixed at zero, while the other is initialised to zero, but estimated freely. Here, it should be noted that the implementation of the models used in the present analysis does not allow for a treatment of the repeated choice nature of the dataset, such that intra-agent and inter-agent variations in tastes are treated in the same way. As in the continuous mixture case, this can be expected

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood	-123.39		-908.13		-137.98	
Adjusted $\rho^2(0)$	0.4074		0.2988		0.2788	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.3298	-5.33	-1.5022	-15.21	-1.1109	-5.58
Travel cost (CHF)	-0.3416	-4.39	-0.2268	-5.86	-0.7190	-3.32
Frequency (per hour)	0.6862	5.00	0.5237	11.14	0.7603	6.00
$\beta_{TT}(A)$ (min.)	-0.1635	-4.07	-0.1570	-4.52	-0.5236	-1.99
$\beta_{TT}(B)$ (min.)	0	-	0	-	0	-
Mass for $\beta_{TT}(A)$	0.9037	10.81	0.3437	7.70	0.1541	1.62
Mass for $\beta_{TT}(B)$	0.0963	1.15	0.6563	14.71	0.8459	8.87
VTTS (A) (CHF/hour)	28.71		41.54		43.69	
VTTS (B) (CHF/hour)	0		0		0	

Table 5.5: VTTS case-study: Discrete mixture MNL estimation results, with one support point fixed at zero

to yield consistent estimates, while the use of the *panel* approach produces efficient estimates.

The results show that, at the cost of two additional parameters, the discrete mixture models offer improvements in LL by 1.30, 17.23 and 1.32 units for respondents on business, leisure and shopping trips respectively. As such, in the present case, the discrete mixture approach leads to significant improvements only in the case of leisure travellers. However, important insights are also gained in the remaining two population segments.

The results show significant differences across the three population groups in terms of the presence of respondents with a zero VTTS. Indeed, in the model for business travellers, the share is very low, at 9.63%, while for leisure travellers, and respondents on shopping trips, the shares are a very high 65.63% and 84.59% respectively. In the case of business-travellers, the share is different from 0% only at the 75% level, while, for shopping trips, it is different from 100% only at the 89% level.

Tests showed that the data did not include any *non-traders*[13], such that the results should not be seen as a simple effect of estimation bias due to captivity. The fact that a much lower share of travellers with zero VTTS is observed in the business models is consistent with intuition. Although it is realistic to assume that, in the absence of a binding time constraint, a non-trivial part of respondents travelling for leisure or shopping purposes are indeed indifferent to travel time changes (either positive or negative)[14], the high shares observed in these two population groups are still striking, and call for a closer investigation, in the form of a comparison with an

[13] Here, the notion of *non-trader* refers to respondents always choosing the same alternative (e.g. alternative A), respondents always choosing the cheapest alternative, and respondents always choosing the fastest alternative.

[14] It must be stressed that this does not imply that travellers seek increases in travel time, but that they are simply equally indifferent to increases as to decreases.

unconstrained model, which is carried out below.

Before proceeding to these additional tests, it is worth looking at the findings in terms of VTTS in the share of the population associated with $\beta_{TT}(A)$. In the model for business travellers, the results are roughly similar to those observed in the model using a fixed travel time coefficient (increase by 16.09%), which was to be expected, given the low probability associated with $\beta_{TT}(B)$. On the other hand, in the models for leisure and shopping trips, the VTTS in the share of the population associated with $\beta_{TT}(A)$ increases dramatically in comparison with the fixed coefficients model, and in fact yield VTTS higher than those observed in the model for business travellers. This however needs to be put into context by noting that the present model specification in effect groups the population into two very crude groups, one for respondents with a zero VTTS, and one for all remaining respondents. Further insights could be expected with the use of a higher number of support points, but this requires additional work to deal with identification issues.

Two interesting further observations can be made from these models. The first observation relates to the model for respondents on shopping trips. Here, the fixed travel time coefficient in the simple MNL model was significant only at the 74% level (cf. Table 5.4). However, when allowing for the presence of respondents with a zero valuation of travel time changes, the coefficient in the remainder of the population is significant at the 95% level, although it should be noted that the associated mass is significantly different from zero only at the 89% level. Again, the results need to be put into context by the small sample size, but the results do suggest that the estimation of a significant common coefficient for the entire population is hampered by the presence of respondents with a zero VTTS. The second observation deals with a related point. In the presence of significant variations in a given coefficient across respondents, the use of a common fixed coefficient can be seen to yield an approximate average value of this coefficient across respondents. In the present case, the simple MNL model is clearly unable to explicitly represent the presence of a part of the population with a zero VTTS, and as such, can be expected to produce a biased fixed-point estimate. This notion is supported by a calculation of the weighted average on the basis of the results from the discrete mixture model. Indeed, using $\pi_{TT}^{(A)}\,\beta_{TT}(A) + \pi_{TT}^{(B)}\,\beta_{TT}(B)$, we obtain values of 25.95, 14.27 and 6.73 CHF/hour in the models for business, leisure and shopping trips respectively, where these values are indeed very close to the fixed-point VTTS obtained with the simple MNL model[15].

We now turn our attention to the comparison between the constrained and unconstrained model. The aim of this process was to test the hypothesis that there is a significant mass at zero, by comparing the model estimated with $\beta_{TT}(B)$ fixed at zero to its unconstrained counter-part. For this, the three models shown in Table 5.5 were re-estimated as shown in Table 5.6, where both $\beta_{TT}(A)$ and $\beta_{TT}(B)$ were estimated freely from the data.

The results are highly interesting. They show that, in the model for business travellers, the unconstrained model leads to a statistically significant improvement in LL by 3.12 units, at the cost of one additional parameter, hence rejecting the constrained model. Furthermore, both estimated support-points are significantly

[15] It is important to note that, because of the non-linearity of the model, this comparison is meaningful at a qualitative level only.

	Business		Leisure		Shopping	
Sample size	315		1,881		288	
Final log-likelihood:	-120.27		-907.41		-137.34	
Adjusted $\rho^2(0)$:	0.4171		0.2987		0.2769	
Parameter	est.	t-stat.	est.	t-stat.	est.	t-stat.
Interchanges	-1.5835	-5.21	-1.5256	-14.66	-1.2352	-4.89
Travel cost	-0.5208	-3.54	-0.2130	-5.71	-0.7048	-3.00
Frequency	0.9033	4.18	0.5292	11.12	0.8573	4.98
Mass for $\beta_{TT}(A)$	0.4718	3.49	0.4121	5.02	0.2852	1.39
Mass for $\beta_{TT}(B)$	0.5282	3.91	0.5879	7.16	0.7148	3.48
$\beta_{TT}(A)$	-0.3408	-3.38	-0.1377	-3.89	-0.4905	-1.92
$\beta_{TT}(B)$	-0.1176	-4.09	0.0119	1.15	0.1157	0.87
VTTS (A) (CHF/hour)	39.26		38.79		41.76	
VTTS (B) (CHF/hour)	13.55		N.S.		N.S.	

Table 5.6: VTTS case-study: Discrete mixture MNL estimation results, with both support points estimated from the data

different from zero, at high levels of confidence. The distribution of the mass between the two support-points is very even, and not significantly different from a $50\%-50\%$ split. Furthermore, the VTTS in group (A) is higher than that produced by the constrained model (cf. Table 5.5), while the weighted average, at 25.68 CHF/hour, is almost identical to that from the constrained model, and again close to the MNL value. Overall, these results reject the hypothesis of a significant mass at zero for the travel time coefficient in this population segment, such that the mass of 9.63% can be explained on the grounds that it captures mass from values close to zero. However, the results also provide proof of heterogeneity, with two different support points for β_{TT}, and better model fit than the MNL model.

While the above process thus rejects the hypothesis of a significant share of travellers with a zero VTTS in the business segment, the situation is very much different in the leisure and shopping segments. Here, the unconstrained model achieves gains in LL by 0.72 and 0.64 units in LL respectively, neither of which is significant, coming at the cost of one additional estimated parameter. Additionally, the estimated values for $\beta_{TT}(B)$ are not significantly different from zero, with confidence levels of 75% and 62% respectively. As such, the positive estimate for the two coefficients is of little importance, and should in no case be seen as a proof of the presence of respondents with a negative VTTS (see also Chapter 4). This is supported by a calculation of the standard error of the actual ratio between $\beta_{TT}(B)$ and β_{TC}, on the basis of a simulation experiment taking into account the correlation between the point estimates as well as their asymptotically Normal distribution (c.f Armstrong et al. 2001), showing significance levels for VTTS(B) of 72.72% and 59.61% in the leisure and shopping groups respectively. The VTTS for respondents in group (A) is quite close to that observed in the constrained models. Overall, the results show that, in these two groups, the unconstrained model does not reject the constrained model, such that the test does not offer convincing proof to suggest that the findings with regards to the high shares for a zero VTTS in the constrained models are incorrect.

5.5 Summary and Conclusions

In this chapter, we have discussed an alternative approach for representing inter-agent variations in tastes, and by extension, choice behaviour. The approach is based on the use of discrete mixtures of choice models, replacing the fixed-parameter choice probabilities by a weighted sum of choice probabilities calculated on the basis of different values for the specific coefficients for which taste heterogeneity is to be introduced. The weights associated with the different *support-points* reflect the market shares of the respective coefficient-values in the sample population. This approach has certain conceptual advantages over continuous mixtures, by being free from any a priori assumption with regards to the shape of the true distribution. Additionally, discrete mixtures can clearly serve as a starting point in the search for an appropriate continuous specification.

The main aim and contribution of this chapter was to demonstrate how discrete mixture models can be used to test for the presence of respondents with zero valuations of changes in a specific travel-attribute, where, in the present case, we looked specifically at the case of a zero VTTS in a route choice experiment. The results, and subsequent *validation* thereof, show that, while no evidence of a significant share of such individuals exists in the case of business travellers, a share of 66% was found for leisure travellers, with a corresponding share of 85% for respondents on shopping trips.

These results are striking, and are possibly in part specific to the data at hand, such that more testing is required. Additionally, it should be noted that, in the case of SP data, another potential reason for results showing zero valuations for changes in a given attribute for some individuals is the design of the surveys, for example in the case of a lack of variation for the concerned attribute for these individuals (i.e. insufficient stimuli). A similar issue arises in the presence of non-traders. As such, further tests should also be conducted on RP data. However, it should be noted that, while, with SP data, multiple possible explanations for zero valuations arise, discrete mixture models maintain their advantage, in terms of being able to highlight the impact of such problems.

Even though the results of this research cannot be generalised without further investigation, certain observations can be made. Indeed, the comparisons between the MNL and discrete mixture models have shown that a failure to account for the presence of individuals with a zero valuation of changes in a travel-attribute can lead to significant bias in the estimated coefficients, and by extension the willingness-to-pay indicators, possibly resulting in misguided policy-measures. Although the discussion in this chapter was limited to the case of changes in travel time, zero valuations potentially play a role for a whole range of attributes, such as for example frequency, and qualitative attributes. Additionally, problems with survey design can potentially also lead to apparent "zero-valuations" in the case of attributes such as cost, where a consistent negative effect would be expected. This problem has seemingly not been addressed in the existing literature, at least not in the context of discrete mixture models. Clearly, the ramifications of this issue are very serious indeed, and the results presented in this chapter call for a thorough investigation into the prevalence of zero valuations, across a host of variables, datasets and data-sources (i.e. RP *vs* SP).

In closing, it should be noted that the same issues in terms of biased results can be seen to apply in the case of continuous mixture models when relying on the use of distributions that are not able to represent a heightened share at zero. Here, the presence of individuals with zero valuations for changes in a specific attribute can potentially also lead to biased results in terms of the existence of a share of respondents with counter-intuitively signed coefficients[16]. In this context, important work remains to be done in terms of exploring the use of model structures allowing for a variation in tastes in the non-zero domain, in addition to the presence of a significant mass at zero, in the spirit of the theoretical distribution discussed by Cirillo & Axhausen (2004), who propose the use of a Normal distribution with a heightened mass at zero. Such an approach can in fact be used in combination with any type of continuous distribution, where a *discrete mixture* is used across two values, one of them equal to zero, while the second value in addition follows a continuous distribution. While straightforward from a conceptual point of view, the approach causes considerable problems in estimation, such that the search for efficient ways of implementing such combined distributions in estimation packages is an important topic for future research.

[16]Returning to the issue of an asymmetrical *true* distribution with a mean close to zero, as discussed in Section 4.3.1.

Chapter 6

Confounding between substitution patterns and random taste heterogeneity

6.1 Introduction and context

As discussed in Section 2.9, modellers have recently begun exploring more advanced structures allowing for the joint representation of inter-alternative correlation and random taste heterogeneity, in the form of combined ECL-RCL models or non-MMNL GEV mixture models. These approaches allow for important gains in flexibility and accuracy in the case where both phenomena potentially have an impact on choice behaviour. Traditionally, such model structures have been seen simply as a tool for jointly representing the two phenomena listed above. However, although they have usually been discussed separately, it should be noted that the differences between these two phenomena are not necessarily that *clear-cut*, and that there is a significant risk of confounding. As such, advanced mixture structures are not only a tool allowing for the joint representation of the two phenomena, but potentially also a means of avoiding misleading results caused by confounding in models allowing only for either of the two phenomena to have an effect.

The aim of this chapter is to explore the issue of confounding between inter-alternative correlation and random taste heterogeneity and to illustrate how advanced mixture structures can be used to reduce the risk of biased results. The main motivation for this study lies in the work of Hess, Bierlaire & Polak (2005*a*), who, in the estimation of mode choice models, observe that when jointly allowing for the two phenomena, the impact of either is reduced significantly, and the substantive conclusions change, suggesting that confounding does indeed occur in those models using a treatment only for one of the two phenomena.

The technical reasons for confounding between inter-alternative correlation and random inter-agent taste variations can be explained straightforwardly by looking separately at the cases of unmodelled random taste heterogeneity and unmodelled inter-alternative correlation, in the context of a mode choice scenario involving three alternatives $(\mathcal{A}, \mathcal{B}, \mathcal{C})$.

We first look at the case where the prevalence of random taste heterogeneity masks the findings in terms of inter-alternative correlation. Let us assume that the sensitivity to a given attribute in the utility of alternatives $\mathcal{A}$ and $\mathcal{B}$ varies randomly across respondents. Clearly, these random disturbances would lead to correlated error-terms for the utilities of these two alternatives, which would be picked up by a GEV model nesting together alternatives $\mathcal{A}$ and $\mathcal{B}$. In a model allowing for inter-alternative correlation but not inter-agent taste heterogeneity, results showing heightened correlation between alternatives $\mathcal{A}$ and $\mathcal{B}$ are then at least partly biased by the unexplained random variations in tastes relating to attributes included in the utility functions of these two alternatives.

The opposite scenario, in which the presence of simple inter-alternative correlation can mask the findings in terms of random taste heterogeneity, is most easily explained from the point of view of an ECL model. Let us assume that correlation between the unobserved utilities for alternatives $\mathcal{A}$ and $\mathcal{B}$ exists due to the presence of shared unobserved attributes for these two alternatives. The standard way of accounting for such correlation is through the use of a GEV structure, nesting together the two alternatives. The correlation can however also be accounted for with the help of an ECL approach, where an additional randomly distributed error-component is added to the utilities of alternatives $\mathcal{A}$ and $\mathcal{B}$. Additionally, as noted by Walker (2001), the correlation in a two-nest model with three alternatives can be accommodated by including an error-component only in the utility of that alternative which is nested on its own, i.e. alternative $\mathcal{C}$ in the present context. As such, it can be seen that, when allowing for random variations in a coefficient included either jointly in the utilities of alternatives $\mathcal{A}$ and $\mathcal{B}$, or solely in the utility of alternative $\mathcal{C}$, the resulting model in fact approximates an ECL formulation, where the dummy variables associated with the error-components however no longer take on simple $0 - 1$ values. This means that the random taste heterogeneity can in fact simply be an artifact of the inter-alternative correlation caused by the unobserved shared attributes. The scope for confounding is clearly increased significantly in the case where the attribute associated with the concerned coefficient exhibits little variation across observations, or where, in the case of a coefficient included in multiple utility functions, the associated attributes are highly correlated across alternatives.

The above discussion has shown that allowing only for either inter-alternative correlation or random taste heterogeneity can mask the findings with regards to the other phenomenon. The effects of this can be quite significant. Although, in some cases, it is possible for the wrongly specified model to attain similar model fit and even reproduce similar behaviour, the risk of misinterpretation of results persists. This relates to the implied cross-elasticities as well as to implications in terms of behaviour in the tails of the population. While testing separately for the two phenomena, say with a GEV and a RCL model, can alert the modeller to the relative performance of the two approaches, it does not remove the risk of biased findings and a joint model should be used, in order to minimise the potential risks of confounding.

The issue of confounding in discrete choice models has been discussed in various contexts in the existing literature. As such, Kitamura & Bunch (1990) look at the dangers of confounding unobserved heterogeneity and state dependence, and discuss the difficulty of a specification search in such a context. Another discussion of the issues with confounding in the context of state dependence is given by Heckman (1981). Swait & Bernardino (2000) look at the confounding between correlation structure and (deterministic) taste heterogeneity, and illustrate how accounting jointly for the two effects in NL models can allow the effects of the two phenomena to be separated. While the discussion by Swait & Bernardino looks exclusively at closed form models, issues with confounding also arise in the case of mixture models. Indeed, the results of Cherchi & Ortúzar (2004) on MMNL models suggest some confounding between correlation, random taste heterogeneity and heteroscedasticity, while results by Hess, Bierlaire & Polak (2004) show that not accounting for correlation in the unobserved utility terms can lead to erroneous conclusions with regards to the presence of random variations in tastes across respondents, a point strongly related to the issue of unmodelled travel-experience attributes discussed in Section 4.4.

While the above results are an indication that such problems with confounding can arise, they cannot easily be generalised, given the use of real-world data, where the *true* error-structure is not known, in the case of Swait & Bernardino (2000), Cherchi & Ortúzar (2004) and Hess, Bierlaire & Polak (2005*a*), and the use of a very basic and small-scale synthetic data experiment by Hess, Bierlaire & Polak (2004). What is needed in this case is a large-scale systematic comparison, using synthetic data, across a range of different scenarios. This is the approach taken in this chapter. We concentrate on the case of random instead of deterministic taste heterogeneity, and the potential for confounding thereof with *simple* inter-alternative correlation. The discussion in this chapter focusses on the use of GEV mixture models; the use of combined RCL-ECL models on the same data is discussed in Appendix A. Finally, an additional source for confounding, namely the presence of heteroscedasticity, is not explored here.

The remainder of this chapter is organised as follows. Section 6.2 presents six case-studies illustrating the risk of confounding when not jointly allowing for the two phenomena, while Section 6.3 presents three forecasting exercises, showing the impact of wrongly specified models on predicted changes in market shares under hypothetical policy-changes. Finally, Section 6.4 summarises the findings and presents the conclusions of this chapter.

6.2 Case-studies

This section presents a number of case-studies illustrating the issue of confounding and showing how the use of a joint modelling approach can help reduce the problems.

The most reliable way of conducting such an analysis is based on the use of simulated data, such that the *true* error structure is known. For the present study, separate quasi-simulated datasets were generated for the different case-studies, making use of data from an SP survey conducted to estimate the hypothetical demand for a new high-speed transit system in Switzerland; the Swiss Metro (cf. Abay 1999, Bierlaire et al. 2001). This dataset provides us with *good* attribute-level data, avoiding the issues caused by the randomness in purely-simulated data. The choice-vectors used in the various case-studies are entirely independent of the original SP survey responses.

Three alternatives were included in the choice set; car, rail and Swiss Metro (SM). For the present study, a subset of $3,000$ individuals were used, each with all three alternatives available to them. Only three attributes, namely travel time, travel cost, and headway (for rail and SM) were used here. In each case-study, separate travel time coefficients were used for the three modes ($\beta_{TT,car}$, $\beta_{TT,rail}$, and $\beta_{TT,SM}$), in conjunction with a common travel cost coefficient (β_{TC}), a joint headway coefficient for rail and SM (β_{HW}), and two ASCs, for car and SM (δ_{car} and δ_{SM}). The simulated datasets were all generated on the basis of a purely cross-sectional approach, making use of the original level-of-service data presented to respondents in the SP survey.

The experimental design used for the case-studies presented in this section reflects the three separate scenarios in which issues with confounding of the type discussed in this chapter can arise. The first group contains two examples in which the *true* model is of closed GEV form, reflecting the case where the true error-structure leads to heightened inter-alternative correlation, while the *target* model allows only for random taste heterogeneity. The opposite scenario acts as the basis for the second group of case-studies, which contains three examples in which the *true* model is a RCL model with random variation in at least one of the coefficients. Finally, the third group contains a single example, where the *true* model is a mixture of a two-nest NL model, representing the situation in which both phenomena play a role, with the *target* model allowing only for one of the two effects to be modelled. In each case, models allowing for randomly distributed taste coefficients make use of the Normal distribution, reducing computational cost, but also reflecting the current state-of-practice. All models presented in this chapter were estimated using BIOGEME.

6.2.1 True model: NL with two nests

In the first case-study, we assume that all taste coefficients take on fixed values across all respondents, and that the rail and SM alternatives are nested together in a simple two-level NL model, with an associated structural parameter of 0.5, leading to a correlation of 0.75 between the unobserved utility terms for the two alternatives. Four types of model were estimated on this dataset; MNL, NL, RCL, and NL mixture. The results of the analysis are summarised in Table 6.1, which also gives the log-likelihood of the true model as calculated on the data used in the analysis, allowing us to establish the relative performance of the different target models.

All three possible two-nest NL structures were estimated on the data, but heightened correlation was only found in the model using the same structure as the *true* model, i.e. nesting together rail and SM. No further gains could be made by estimating a CNL model on the data. The results show that the NL model manages to reproduce the behaviour of the *true* model, especially with regards to the structural parameter, although all three VTTS measures are overestimated, with the maximum bias arising in the VTTS for car, which is overestimated by 9.8%. This bias can probably be explained on the grounds of sampling error, leading to an understated cost-sensitivity in the data. Further investigation is required to establish why no such understating occurs for the travel time coefficients. The results also show that, by not allowing for the correlation between rail and SM, the MNL model obtains lower model fit than the NL model, and leads to bigger bias in the three VTTS measures, with overestimation by rates of between 13% and 18.7%.

	True model	MNL		NL		RCL		NL mixture	
Final LL	-903.35	-931.47		-900.20		-916.77		-899.15	
adj. $\rho^2(0)$	0.7235	0.7153		0.7244		0.7182		0.7232	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-5.624	-16.70	-4.079	-13.22	-6.644	-12.54	-4.222	-10.99
δ_{SM}	-3.000	-4.918	-19.63	-3.208	-12.82	-5.722	-15.37	-3.319	-10.33
β_{TC} (μ)	-0.100	-0.126	-20.50	-0.105	-20.47	-0.170	-11.13	-0.110	-13.93
β_{TC} (σ)	-	-	-	-	-	0.015	1.13	0.008	0.90
β_{HW} (μ)	-0.020	-0.033	-17.13	-0.022	-13.21	-0.039	-13.88	-0.023	-10.70
β_{HW} (σ)	-	-	-	-	-	0.001	0.48	0.001	0.73
$\beta_{TT,car}$ (μ)	-0.030	-0.045	-14.88	-0.035	-13.79	-0.061	-10.47	-0.036	-10.33
$\beta_{TT,car}$ (σ)	-	-	-	-	-	0.017	5.14	0.004	1.06
$\beta_{TT,rail}$ (μ)	-0.040	-0.059	-22.15	-0.043	-16.82	-0.073	-14.00	-0.045	-11.88
$\beta_{TT,rail}$ (σ)	-	-	-	-	-	0.001	0.95	0.002	1.58
$\beta_{TT,SM}$ (μ)	-0.035	-0.050	-14.15	-0.039	-14.28	-0.064	-10.94	-0.040	-11.06
$\beta_{TT,SM}$ (σ)	-	-	-	-	-	0.001	0.87	0.001	0.67
$\lambda_{rail,SM}$	0.50	1.00	-	0.51	5.96	1.00	-	0.47	4.15
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	18.00	21.37		19.77		21.62		19.82	
VTTS (car) (σ)	-	-		-		$6.47^{(*)}$		$2.45^{(**)}$	
VTTS (rail) (μ)	24.00	27.97		24.59		26.07		24.64	
VTTS (rail) (σ)	-	-		-		$2.31^{(**)}$		$2.09^{(**)}$	
VTTS (SM) (μ)	21.00	23.74		22.13		22.83		22.13	
VTTS (SM) (σ)	-	-		-		$2.03^{(**)}$		$1.57^{(**)}$	

$^{(*/**)}$ One/both involved dispersion parameters not significant at 95% level

T-statistics for structural parameters calculated wrt 1

Table 6.1: Estimation results on synthetic data generated with two-nest NL model

The simple RCL model, allowing only for random variations in tastes and using no explicit treatment of inter-alternative correlation, offers an improvement in log-likelihood (LL) by 14.7 units over the MNL model. A closer inspection shows this to be due to a single additional statistically significant parameter, namely the standard deviation for the travel time coefficient for car. As such, the RCL model falsely indicates significant variations in the sensitivity to car travel time across the population, where no such variations exist in the actual data. Here it can be seen that the RCL model is giving an approximation to an ECL approach, using a single error-component associated with the car alternative, where this is less successful than the correct NL approach (lower LL), given the high level of variation in the associated attribute, which hinders the approximation. Without prior knowledge of the true error-structure, which can yield the above explanation, these results would suggest the presence of significant levels of random taste heterogeneity, where this is entirely due to the issue of confounding described in Section 6.1. The final model estimated on the data, a NL mixture structure, obtains a model fit very similar to that obtained with the *true* structure, a two-nest closed form NL model. In this model, the structural parameter is again closely reproduced, as was the case in the NL model. Furthermore, none of the five standard deviation parameters is statistically significant at the usual 95% level. It should be noted that, for two of the parameters, namely the standard deviations for β_{HW} and $\beta_{TT,rail}$, the level of significance is higher than in the RCL model. In both cases however, the relative level of the standard deviation when compared to the mean value remains very small, such that the model does not show any significant levels of variation. This was not the case for the standard deviation of $\beta_{TT,car}$ in the RCL model, where a 95% confidence interval would lead to a 50% spread to either side of the mean.

The VTTS measures reproduced for the two mixture structures were generated by simple simulation, using the final estimates shown in Table 6.1. As such, the simulation exercise also made use of standard deviations that were not statistically significant. However, in these cases, the relative size of the dispersion parameter when compared to the mean value was small, and a separate analysis showed that using fixed values for the concerned coefficients did not change the mean estimates in any significant fashion. Additionally, it was not necessary to use special treatment for β_{TC} (such as removing upper percentiles) to avoid extreme values caused by a division by a value close to zero, given the statistically insignificant standard deviation for this coefficient. The results show that the NL mixture model reproduces essentially the same mean VTTS measures as the two-nest NL model, which further reinforces the findings that this structure avoids issues of confounding. The use of the RCL model leads to higher overestimation of the three mean VTTS measures than is the case for the NL mixture model, where, except for car, the bias is however lower than in the MNL model. Finally, the model does suggest a high level of variation in the VTTS for car across respondents, with lower and upper 95% confidence limits of 8.94CHF/hour and 34.30CHF/hour respectively[1]. The extent of this variation, which is purely an artifact of confounding, could lead to misguided policy implications in real-world scenarios where policy-makers are often interested in the behaviour in the tails of the population, such as travellers with very high VTTS in the case of road-pricing or tolls.

6.2.2 True model: CNL with two nests

In the second case-study, we again assume that all taste coefficients take on fixed values across all respondents. However, in addition to the correlation between rail and SM, the rail alternative is now nested with car, with an allocation by equal shares of the rail alternative to the two nests, where the nesting parameters in the rail-car and rail-SM nests are set at 0.5 and 0.33 respectively.

The results of this analysis are summarised in Table 6.2, showing the estimates for two closed form models, MNL and CNL, and their mixture counter-parts. Simple NL and NL mixture models (without cross-nesting) were also estimated on the data, showing that not allowing for cross-nesting produces biased results in terms of trade-offs, as well as in terms of random taste heterogeneity in the mixture model. The results reproduced in Table 6.2 show that the CNL model slightly underestimates the three VTTS measures; the bias is biggest for rail-travel, where, incidentally, the bias produced by the MNL model is smaller, as is the case for car-travel. Nevertheless,

[1]Here, a simple confidence interval based on a Normal distribution was used, which is not fully appropriate, given that the actual distribution of the VTTS is given by a ratio of two Normals. However, with the standard deviation of β_{TC} being close to zero, the distribution of the ratio $\frac{\beta_{TT,car}}{\beta_{TC}}$ does indeed approximate a Normal distribution.

	True model	MNL		CNL		RCL		CNL mixture	
Final LL	-860.82	-882.30		-855.90		-874.26		-855.87	
adj. $\rho^2(0)$	0.7355	0.730		0.737		0.731		0.735	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-6.125	-16.82	-3.932	-13.71	-7.880	-7.37	-3.937	-13.52
δ_{SM}	-3.000	-4.879	-18.46	-3.100	-13.51	-6.218	-8.13	-3.100	-13.42
β_{TC} (μ)	-0.100	-0.148	-20.60	-0.100	-15.62	-0.206	-6.72	-0.101	-14.37
β_{TC} (σ)	-	-	-	-	-	0.045	3.28	0.004	0.21
β_{HW} (μ)	-0.020	-0.031	-15.74	-0.018	-11.23	-0.040	-7.16	-0.018	-11.11
β_{HW} (σ)	-	-	-	-	-	0.006	0.60	0.000	0.01
$\beta_{TT,car}$ (μ)	-0.030	-0.043	-16.26	-0.029	-12.50	-0.060	-6.57	-0.029	-12.32
$\beta_{TT,car}$ (σ)	-	-	-	-	-	0.012	4.16	0.000	0.10
$\beta_{TT,rail}$ (μ)	-0.040	-0.061	-21.95	-0.040	-14.54	-0.082	-7.30	-0.040	-14.29
$\beta_{TT,rail}$ (σ)	-	-	-	-	-	0.008	2.78	0.000	0.06
$\beta_{TT,SM}$ (μ)	-0.035	-0.050	-14.30	-0.033	-11.74	-0.066	-6.46	-0.033	-11.53
$\beta_{TT,SM}$ (σ)	-	-	-	-	-	0.008	1.54	0.000	0.13
$\lambda_{rail,car}$	0.50	1.00	-	0.42	3.63	1.00	-	0.42	3.53
$\lambda_{rail,SM}$	0.33	1.00	-	0.36	4.44	1.00	-	0.36	3.94
$\alpha_{rail,rail-car}$	0.5	-	-	0.50	0.07	-	-	0.50	0.06
$\alpha_{rail,rail-SM}$	0.5	-	-	0.50	-0.07	-	-	0.50	-0.06
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	18.00	17.65		17.17		18.55		17.17	
VTTS (car) (σ)	-	-		-		6.93		$0.73^{(**)}$	
VTTS (rail) (μ)	24.00	24.94		23.75		25.25		23.75	
VTTS (rail) (σ)	-	-		-		7.93		$1.01^{(**)}$	
VTTS (SM) (μ)	21.00	20.16		19.64		20.53		19.65	
VTTS (SM) (σ)	-	-		-		$6.83^{(*)}$		$0.85^{(**)}$	

$^{(*/**)}$ One/both involved dispersion parameters not significant at 95% level

T-statistics calculated wrt **1** for structural parameters and wrt **0.5** for allocation parameters

Table 6.2: Estimation results on synthetic data generated with two-nest CNL model

by not allowing for the correlation between rail and car, and rail and SM, which is reproduced closely in the CNL model, the MNL model obtains a poorer model fit.

The RCL model estimated on this dataset suggests the presence of significant variations for β_{TC}, $\beta_{TT,car}$ and $\beta_{TT,rail}$, with additional variation, significant at the 88% level, for $\beta_{TT,SM}$. In the model used to generate the data, all coefficients were kept fixed, suggesting that this is due entirely to confounding[2]. This is supported by the findings for the CNL mixture model, where all standard deviations obtain very low levels of statistical significance. While the RCL model obtains better fit than the MNL model, its fit is lower than for the CNL and CNL mixture models, which obtain almost exactly the same final LL, along with indistinguishable results, suggesting that confounding is not an issue in the CNL mixture model, which is thus able to correctly interpret the error-structure. The risk of reaching misleading conclusions on the basis of the RCL model are further highlighted by the fact that, while the

[2] Here, it should be noted that, while significant, the estimated dispersion parameters are in general small compared to the associated mean parameter.

	True model	MNL		NL		RCL		NL mixture	
Final LL	-1371.12	-1377.09		-1371.15		-1365.76		-1365.23	
adj. $\rho^2(0)$	0.5816	0.5801		0.5815		0.5832		0.5830	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-3.450	-14.46	-2.932	-11.26	-3.662	-13.63	-3.387	-9.46
δ_{SM}	-3.000	-2.762	-13.93	-2.055	-8.10	-2.891	-13.81	-2.565	-6.99
β_{TC}	-0.100	-0.069	-20.74	-0.065	-19.98	-0.085	-14.37	-0.080	-11.57
β_{HW}	-0.020	-0.016	-10.38	-0.012	-7.66	-0.017	-10.32	-0.015	-6.91
$\beta_{TT,car}$ (μ)	-0.030	-0.021	-10.20	-0.018	-8.72	-0.027	-9.85	-0.025	-7.00
$\beta_{TT,car}$ (σ)	0.015	-	-	-	-	0.011	5.35	0.010	4.08
$\beta_{TT,rail}$	-0.040	-0.033	-19.09	-0.027	-12.35	-0.038	-17.10	-0.034	-9.04
$\beta_{TT,SM}$	-0.035	-0.027	-10.64	-0.024	-10.08	-0.033	-10.46	-0.030	-8.02
$\lambda_{rail,SM}$	1.00	1.00	-	0.71	2.79	1.00	-	0.88	0.95
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	18.00	18.03		16.55		19.19		18.48	
VTTS (car) (σ)	9.00	-		-		7.91		7.18	
VTTS (rail)	24.00	28.57		25.34		26.53		25.77	
VTTS (SM)	21.00	23.84		22.52		23.27		22.85	

T-statistics for structural parameters calculated wrt 1

Table 6.3: Estimation results on synthetic data generated with RCL model with randomly distributed car travel time coefficient

mean VTTS measures[3] are acceptable, the model suggests very wide confidence intervals, which can again cause problems in policy-analysis.

6.2.3 True model: RCL with single random coefficient

In the first of the RCL case-studies, we allow for random variations in a single coefficient, namely that associated with travel time for the car-alternative. The fact that this coefficient is used only in the utility of a single alternative increases the scope for an approximation to an ECL model (cf. Section 6.1). However, it should be noted that there is significant variation in the associated attribute across respondents, such that the study still allows us to highlight the issues of confounding in a fairly general case.

The results of this analysis are summarised in Table 6.3, showing the results for a MNL model, a NL model nesting rail with SM, and the mixture counterparts of these two models. The remaining two NL models collapsed back to a MNL structure, while no further gains could be made with the use of a CNL model. The only coefficient for which it was possible to identify any random variations in either of the two mixture models is $\beta_{TT,car}$, such that all other coefficients are kept fixed in the presentation of the results. As such, with the only randomly distributed coefficient entering the VTTS calculation in the numerator, no simulation was required in the computation of the VTTS measures.

The actual results show that, while the MNL model retrieves the mean VTTS

[3]Simulation was used in the computation of the VTTS measures in the mixture models. Again, no special treatment was required in the case of β_{TC}, given that the upper limit of reasonable confidence intervals for β_{TC} was well below zero.

for car-travel, it overestimates the VTTS measures for rail and SM, despite the fact that the associated travel time coefficients were kept fixed in the generation of the data. On the other hand, the NL model nesting together rail and SM underestimates the mean VTTS for car-travel, but offers a better approximation to the true VTTS for rail and SM. However, the model incorrectly retrieves heightened correlation between rail and SM. This can be seen to be an effect of the random variation in the sensitivity to car travel time in the data, such that the true model has similarities to an ECL structure in which a single error-component is used in the utility of the car alternative[4]. Given the fact that the associated attribute varies significantly across observations, it should come as no surprise that the NL model offers a lower model fit than either of the two mixture structures, which correctly retrieve the variation in the sensitivity to car travel time, although some bias remains in the VTTS measures, which, except for the standard deviation of the VTTS for car-travel, is lower in the NL mixture model. Finally, it can be seen that the NL mixture model retrieves a structural parameter which, although being smaller than unity, has a high associated standard error, making the difference statistically insignificant, unlike in the NL model, where the difference was also more than twice as large.

This application thus highlights the fact that the findings in terms of inter-alternative correlation can be biased by the presence of unexplained random variations in taste coefficients, and suggests that the use of mixture models, jointly allowing for the two phenomena, can help to reduce these risks of confounding.

6.2.4 True model: RCL with single random coefficient shared by all three alternatives

In the second of the RCL case-studies, the travel cost coefficient, which is common to all three alternatives, is specified to follow a random distribution.

The results of this analysis are summarised in Table 6.4, reporting the estimates of a MNL model, the NL model nesting together rail and car, and their two mixture counterparts. The remaining two NL structures and their mixture counterparts collapsed back to MNL and RCL structures respectively. The only coefficient for which it was possible to identify any random variations in either of the two mixture models is β_{TC}, such that all other coefficients are kept fixed in the presentation of the results. The first observation that can be made is the much better model fit obtained by the mixture models, which illustrates the impact of not allowing for random variations in the travel cost coefficient in the two closed form models. The difference is much bigger than was the case in the example described in Section 6.2.3, which is partly due to the fact that the travel cost coefficient is used in the utilities of all three alternatives. The next observation relates to the results of the NL model, which shows heightened correlation between the error-terms for the utilities of the rail and car alternatives; this is an effect of confounding between taste heterogeneity and inter-alternative correlation, given that the structural parameters in the data generation process were all fixed to a value of 1[5]. While the NL mixture model also retrieves some correlation between the rail and car alternatives, the associated

[4]It can be seen that this case-study is the counterpart of the one presented in Section 6.2.1.

[5]It is worth noting here that the travel cost attributes for car and rail are not strongly correlated, reducing the scope for an approximation to an ECL structure.

	True model	MNL		NL		RCL		NL mixture	
Final LL	-1326.49	-1775.25		-1631.61		-1322.96		-1320.93	
adj. $\rho^2(0)$	0.5951	0.4592		0.5025		0.5962		0.5965	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-2.638	-8.38	-0.948	-8.99	-3.812	-14.61	-3.516	-12.51
δ_{SM}	-3.000	-2.718	-14.57	-1.682	-12.46	-3.009	-14.14	-2.950	-14.43
β_{TC} (μ)	-0.100	-0.026	-2.37	-0.015	-12.13	-0.096	-18.14	-0.086	-12.54
β_{TC} (σ)	0.035	-	-	-	-	0.035	17.89	0.033	12.38
β_{HW}	-0.020	-0.015	-10.61	-0.005	-7.54	-0.018	-11.02	-0.017	-10.59
$\beta_{TT,car}$	-0.030	-0.017	-7.30	-0.007	-8.78	-0.032	-14.39	-0.029	-12.67
$\beta_{TT,rail}$	-0.040	-0.025	-8.35	-0.010	-9.65	-0.041	-19.66	-0.038	-15.69
$\beta_{TT,SM}$	-0.035	-0.018	-4.74	-0.004	-3.35	-0.036	-12.43	-0.033	-10.29
$\lambda_{rail,car}$	1.00	1.00	-	0.20	7.45	1.00	-	0.82	1.89
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	20.33	39.77		30.04		22.83		23.93	
VTTS (car) (σ)	8.36	-		-		10.29		11.87	
VTTS (rail) (μ)	27.11	56.90		40.52		29.41		30.91	
VTTS (rail) (σ)	11.14	-		-		13.26		15.34	
VTTS (SM) (μ)	23.72	40.36		17.61		26.08		26.84	
VTTS (SM) (σ)	9.75	-		-		11.76		13.31	

T-statistics for structural parameters calculated wrt 1

Table 6.4: Estimation results on synthetic data generated with RCL model with randomly distributed travel cost coefficient

structural parameter is much closer to a value of 1, although it is significantly different from it at the 94% level. This suggests that the NL mixture model is able to avoid most, but not all of the confounding[6].

Finally, in terms of the recovery of the true VTTS distribution, the results show very strong overestimation of the mean VTTS measures in the two closed form models, except for the mean VTTS for travel on SM, which is underestimated in the NL model. The two mixture models[7] offer acceptable performance in the recovery of the mean and standard deviation for the three VTTS measures, where some bias remains, which is bigger in the NL mixture model, and which can again be seen as a sign of sampling error.

6.2.5 True model: RCL with two random coefficients

The final of the three RCL case-studies combines the approaches from Section 6.2.3 and Section 6.2.4 in that both $\beta_{TT,car}$ and β_{TC} are assumed to vary randomly across individuals.

The findings of the analysis are summarised in Table 6.5, showing results for a MNL model, a NL model nesting rail with car, and the mixture counterparts of these two models. No correlation could be retrieved with either of the two remaining

[6]The results could also suggest some errors in data generation, but the experiments were repeated with different sets of draws in the generation process, and similar results were obtained.

[7]Here, it should also be noted that, to avoid extreme values due to a division by a value close to zero, the lower and upper two percentile points were removed from the distribution of β_{TC} in the simulation of the VTTS measures.

	True model	MNL		NL		RCL		NL mixture	
Final LL	-1550.59	-1882.74		-1804.19		-1543.84		-1543.13	
adj. $\rho^2(0)$	0.5268	0.4266		0.4502		0.5288		0.5288	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-2.485	-8.06	-1.063	-8.59	-3.732	-11.91	-3.532	-10.75
δ_{SM}	-3.000	-2.652	-14.75	-1.848	-12.92	-3.006	-13.66	-2.983	-13.97
β_{TC} (μ)	-0.100	-0.021	-2.70	-0.012	-11.51	-0.085	-10.92	-0.080	-9.37
β_{TC} (σ)	0.035	-	-	-	-	0.031	10.84	0.029	9.31
β_{HW}	-0.020	-0.013	-9.80	-0.005	-6.77	-0.016	-9.47	-0.016	-9.08
$\beta_{TT,car}$ (μ)	-0.030	-0.012	-8.88	-0.006	-7.40	-0.025	-8.41	-0.025	-8.28
$\beta_{TT,car}$ (σ)	0.020	-	-	-	-	0.016	5.91	0.016	5.74
$\beta_{TT,rail}$	-0.040	-0.022	-9.88	-0.010	-9.12	-0.038	-14.66	-0.036	-13.40
$\beta_{TT,SM}$	-0.035	-0.013	-4.90	-0.004	-2.97	-0.031	-9.35	-0.030	-8.62
$\lambda_{rail,car}$	1.00	1.00	-	0.30	6.43	1.00	-	0.86	0.99
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	20.34	34.27		28.15		20.59		21.34	
VTTS (car) (σ)	16.94	-		-		17.36		17.88	
VTTS (rail) (μ)	27.15	63.30		49.24		30.49		31.06	
VTTS (rail) (σ)	11.23	-		-		13.83		14.03	
VTTS (SM) (μ)	23.76	38.22		18.99		25.25		25.47	
VTTS (SM) (σ)	9.83	-		-		11.45		11.50	

T-statistics for structural parameters calculated wrt 1

Table 6.5: Estimation results on synthetic data generated with RCL model with randomly distributed travel cost and car travel time coefficients

NL structures or their mixture counterparts. Furthermore, the only coefficients for which it was possible to identify any random variations in either of the two mixture models are $\beta_{TT,car}$ and β_{TC}, such that all other coefficients are again kept fixed in the presentation of the results.

The findings from this analysis are consistent with those reported in Section 6.2.3 and Section 6.2.4. The model fit obtained by the two mixture structures is superior to that obtained with the closed form counterparts, which is a direct result of allowing for random variations in $\beta_{TT,car}$ and β_{TC}. The NL model outperforms the MNL model, by retrieving some of the effects of the random variations in $\beta_{TT,car}$ and β_{TC} as correlation in the error-terms of the utilities for the car and rail alternatives, again highlighting the issue of confounding. Interestingly, like in the model reported in Section 6.2.4, the NL model again underestimates the mean VTTS for SM travel, while overestimating those for car and rail, and the MNL again overestimates all three of the measures. Again, the two mixture models offer good approximation of the true mean and standard deviation of the VTTS measures[8], where the remaining bias is again slightly larger in the NL mixture model. Finally, in this example, the nesting parameter in the NL mixture model is different from unity only at the 68% level, showing reduced risk of confounding.

[8]In the simulation, the lower and upper two percentile points of the distribution of β_{TC} were once again removed.

	True model	MNL		NL (rail-SM)		NL (rail-car)		RCL		NL mixture	
Final LL	-1151.42	-1562.29		-1560.55		-1525.55		-1163.73		-1147.78	
adj. $\rho^2(0)$	0.6467	0.5239		0.5241		0.5347		0.6433		0.6478	
		est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4.000	-3.548	-15.96	-3.548	-12.92	-2.197	-11.25	-5.629	-11.92	-3.558	-11.35
δ_{SM}	-3.000	-3.592	-17.11	-3.585	-10.91	-2.766	-14.54	-4.824	-13.50	-2.628	-10.14
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	-
	CHF/hour	CHF/hour		CHF/hour		CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	20.37	35.39		35.52		33.17		25.23		21.69	
VTTS (car) (σ)	8.48	-		-		-		16.46		11.95$^{(*)}$	
VTTS (rail) (μ)	27.16	54.77		54.86		47.84		30.79		28.63	
VTTS (rail) (σ)	11.31	-		-		-		17.26		15.56$^{(*)}$	
VTTS (SM) (μ)	23.76	37.71		37.92		29.21		27.72		26.12	
VTTS (SM) (σ)	9.9	-		-		-		15.52$^{(*)}$		14.20$^{(*)}$	

$^{(*)}$ Dispersion parameter for travel time coefficient not significant at 95% level

T-statistics for structural parameters calculated wrt 1

Table 6.6: Estimation results on synthetic data generated with two-nest NL mixture model

6.2.6 True model: NL mixture

The final case-study, in which the *true* model allows for random taste heterogeneity as well as inter-alternative correlation, makes use of a NL mixture structure, nesting together rail and SM with a nesting parameter of 0.5, and letting β_{TC} vary randomly across respondents, hence combining the approaches from Section 6.2.1 and Section 6.2.4.

The results of this analysis are summarised in Table 6.6. Five models were retained; one MNL model, the NL models nesting rail with SM, and rail with car, a RCL model, and a NL mixture nesting rail with SM. Both NL models outperform the MNL model, although the improvement offered by the model nesting rail with SM is only marginal. Here, a striking observation can be made. In the true model, rail is nested with SM. However, in the simple NL models, which do not allow for the additional variation in β_{TC}, better performance is obtained by nesting rail with car, while, in the model nesting rail with SM, the structural parameter is significantly different from unity only at the 78% level. The notion that this is most probably an effect of confounding between correlation and taste heterogeneity is supported by the observation that, in the corresponding mixture models, only the model nesting rail with SM manages to retrieve significant amounts of correlation, where the estimated structural parameter is virtually indistinguishable from that used in the generation of the data. The use of CNL structures, in the closed form as well as mixture models, did not yield any improvements over the simple NL or NL mixture models. The same applies for the two other possible NL mixture structures not included in Table 6.6.

Both mixture models correctly retrieve the variation in β_{TC}. However, the simple RCL model additionally retrieves significant variations in β_{HW}, $\beta_{TT,car}$ and $\beta_{TT,Rail}$, while, for $\beta_{TT,SM}$, the dispersion parameter is significant only at the 76% level. Also, it should be noted that, for $\beta_{TT,rail}$ and $\beta_{TT,SM}$, the estimated dispersion parameters are very small when compared to the associated mean parameter. In terms of VTTS measures, the three closed form structures overestimate all three measures, where the bias is smallest in the model nesting rail with car. In the two mixture models, the lower and upper two percentile points of the distribution of β_{TC} were again removed in the simulation of the VTTS measures. The results show much lower bias than in the closed form models when looking at the retrieval of the mean VTTS measures, which is an effect of allowing for the variation in β_{TC}. Some bias remains, which is lower in the NL mixture model than in the RCL model. The higher standard deviations for the VTTS measures in the RCL model are an effect of the additional levels of variation retrieved for the travel time coefficients for car, as well as the overestimated variation in β_{TC}, which also plays a role in the NL mixture model.

6.3 Impact of confounding on model forecasts

The final part of the analysis consisted of illustrating the impact of confounding between random taste heterogeneity and inter-alternative correlation on model forecasts. This gives an account of the potential risk of misleading policy decisions, complementing the earlier comments relating to the problems caused by incorrect results in terms of the variation in behaviour across respondents (cf. Section 6.2.1).

The forecasting exercises make use of one example from each of the three groups of case-studies described at the beginning of Section 6.2, namely the simple two-level NL data (Section 6.2.1), the RCL data with randomly distributed β_{TC} (Section 6.2.4), and the NL mixture data (Section 6.2.6). In each case, the forecasting scenario looks at a hypothetical increase in the cost of rail-travel by 20%, permitting us to gauge the impact of allowing for heightened substitution between rail and SM, and random variations in the sensitivity to travel cost. The use of the data generated in Section 6.2.4, where the NL mixture model also reveals some *incorrect* correlation between two of the alternatives, allows us to look at forecasts produced in the case where flexible structures are also affected by some confounding. In each of the forecasting exercises, we look at the choice probabilities for a representative individual in addition to the overall market shares. In both cases, the results show the original and forecasted choice probabilities respectively market shares, along with the relative change in these measures for the three alternatives, and the bias in the predicted change, calculated as $\frac{\Delta-\Delta_T}{\Delta_T}$, where Δ gives the proportional change in choice probability or market share in the target model, and Δ_T gives the proportional change in choice probability or market share in the *true* model. We will now look at the results of the three forecasting examples in turn.

The results of the forecasting analysis making use of the data generated by a two-level NL model (cf. Section 6.2.1) are summarised in Table 6.7. The results with the *true* model for the representative individual show a decrease in probability for the rail alternative by 58.45%, as a result of an increase in rail-fares by 20%. They also

Choice probabilities for representative individual

Original choice probabilities

	Car	Rail	SM
True model	46.74%	21.63%	31.63%
MNL	50.07%	24.46%	25.47%
NL	49.37%	20.02%	30.61%
RCL	50.41%	21.79%	27.80%
NL mixture	50.01%	19.54%	30.45%

Forecasted choice probabilities

	Car	Rail	SM
True model	50.73%	8.99%	40.28%
MNL	57.14%	13.79%	29.07%
NL	53.39%	7.90%	38.71%
RCL	57.41%	10.00%	32.59%
NL mixture	53.94%	7.19%	38.87%

Relative change in choice probabilities

	Car	Rail	SM
True model	+8.54%	-58.45%	+27.35%
MNL	+14.13%	-43.63%	+14.13%
NL	+8.15%	-60.53%	+26.44%
RCL	+13.89%	-54.10%	+17.22%
NL mixture	+7.85%	-63.21%	+27.67%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	+65.43%	-25.35%	-48.35%
NL	-4.61%	+3.56%	-3.31%
RCL	+62.70%	-7.44%	-37.03%
NL mixture	-8.04%	+8.14%	+1.18%

Overall market shares

Original market shares

	Car	Rail	SM
True model	55.69%	23.69%	20.62%
MNL	55.73%	23.43%	20.83%
NL	55.73%	23.49%	20.79%
RCL	55.74%	23.47%	20.87%
NL mixture	55.78%	23.50%	20.73%

Forecasted market shares

	Car	Rail	SM
True model	58.46%	13.22%	28.32%
MNL	58.54%	14.90%	26.65%
NL	58.47%	13.12%	28.41%
RCL	58.55%	13.92%	27.80%
NL mixture	58.53%	13.01%	28.46%

Relative change in market shares

	Car	Rail	SM
True model	+4.97%	-44.19%	+37.36%
MNL	+5.04%	-36.40%	+27.91%
NL	+4.92%	-44.16%	+36.67%
RCL	+5.04%	-40.70%	+33.24%
NL mixture	+4.92%	-44.62%	+37.31%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	+1.56%	-17.62%	-25.30%
NL	-1.01%	-0.06%	-1.86%
RCL	1.40%	-7.90%	-11.03%
NL mixture	-0.90%	+0.98%	-0.12%

Table 6.7: Forecasting exercise using data generated by two-level NL model

Choice probabilities for representative individual

Original choice probabilities

	Car	Rail	SM
True model	38.64%	27.65%	33.71%
MNL	36.84%	37.00%	26.16%
NL	39.77%	29.05%	31.17%
RCL	39.11%	28.46%	32.43%
NL mixture	40.23%	27.12%	32.65%

Forecasted choice probabilities

	Car	Rail	SM
True model	43.79%	18.32%	37.89%
MNL	38.79%	33.66%	27.55%
NL	46.03%	22.13%	31.84%
RCL	44.29%	19.32%	36.39%
NL mixture	46.10%	17.94%	35.96%

Relative change in choice probabilities

	Car	Rail	SM
True model	+13.31%	-33.72%	+12.39%
MNL	+5.31%	-9.03%	+5.31%
NL	+15.73%	-23.84%	+2.15%
RCL	+13.23%	-32.12%	+12.24%
NL mixture	+14.60%	-33.86%	+10.14%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	-60.15%	-73.22%	-57.19%
NL	+18.11%	-29.31%	-82.63%
RCL	-0.63%	-4.74%	-1.27%
NL mixture	+9.66%	+0.41%	-18.20%

Overall market shares

Original market shares

	Car	Rail	SM
True model	53.94%	23.78%	22.28%
MNL	53.67%	24.27%	22.07%
NL	52.73%	25.21%	22.07%
RCL	53.64%	24.27%	22.09%
NL mixture	53.61%	24.25%	22.13%

Forecasted market shares

	Car	Rail	SM
True model	56.68%	15.02%	28.30%
MNL	55.54%	20.42%	24.04%
NL	56.11%	20.53%	23.36%
RCL	56.41%	15.63%	27.97%
NL mixture	56.47%	15.91%	27.62%

Relative change in market shares

	Car	Rail	SM
True model	+5.08%	-36.82%	+26.99%
MNL	+3.49%	-15.86%	+8.96%
NL	+6.41%	-18.55%	+5.87%
RCL	+5.15%	-35.61%	+26.64%
NL mixture	+5.34%	-34.40%	+24.77%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	-31.32%	-56.93%	-66.81%
NL	+26.28%	-49.62%	-78.27%
RCL	+1.36%	-3.27%	-1.33%
NL mixture	+5.15%	-6.55%	-8.24%

Table 6.8: Forecasting exercise using data generated by RCL model with randomly distributed travel cost coefficient

show a disproportional shift towards the SM alternative as a result of the heightened correlation between rail and SM, while, in the MNL model, the *IIA* assumption leads to a proportional shift in probability towards car and SM, and hence significant bias in the forecasted choice probabilities. The RCL model also underestimates the shift towards the SM alternative, with a corresponding overestimation of the shift towards the car alternative, where the bias is slightly smaller than in the MNL model, thanks to the fact that the model approximates an ECL structure in representing the correlation between rail and SM with the help of a randomly distributed car travel time coefficient, which takes a role similar to an error-component. Here, the confounding thus actually means that the model gives a more accurate performance than the MNL model, but the risk of misguided interpretation remains, given that the model implies the presence of significant variations in the sensitivity to changes in car travel time. The most accurate prediction performance is offered by the NL model and its mixture counterpart.

At the population level, the results from Table 6.7 show a decrease in market share for rail in the true model by 44.19%. The heightened correlation between rail and SM leads to a greater shift in market share from rail to SM than from rail to car. The estimated NL and NL mixture models produce forecasts that are very similar to those obtained with the *true* model, reflected in the low bias reported in the final part of Table 6.7, while the use of the MNL model leads to underestimated changes in the market shares of rail and SM. Here, the averaging across observations means that the *IIA* assumption, which holds at the individual level, does not hold at the population level. Finally, the results do show that the use of the RCL model leads to some bias in prediction, by underestimating the changes in the market shares for rail and SM. The bias is in this case rather small, which is again a reflection of the fact that the RCL model offers an approximation of an ECL approach with a single error-component in the utility function of the car alternative.

The results of the forecasting experiment using the data generated with a RCL model with randomly distributed β_{TC} are summarised in Table 6.8.

The results for the representative individual show a decrease in choice probability for rail by 33.72%, with a slightly bigger shift towards car. In the MNL model, the *IIA* assumption guarantees proportional shifts, but they are massively underestimated, with a lower than warranted decrease in the choice probability for the rail alternative. The NL model nesting rail with car overestimates the shift towards car, as well as underestimating the actual decrease in probability for the rail alternative. The RCL model, and to a lesser degree the NL mixture model, manage to replicate the behaviour of the true model, where the remaining bias in the NL mixture model is down to the confounding which is also exhibited by this model.

In the results for the overall market shares, there is a much bigger relative shift in market share towards the SM alternative, which can be explained on the grounds of the low initial market share for SM, when compared to that of car. The actual results in terms of bias are consistent with those observed in the case of the representative individual, showing that the MNL and especially the NL model significantly underestimate the shift in market share towards the SM alternative. In the NL model, this is a direct effect of the nesting of car with rail, which is caused by confounding. Again, some bias remains in the NL mixture model, which is caused by the fact that this model picks up some of the taste heterogeneity in the form of

Choice probabilities for representative individual

Original choice probabilities

	Car	Rail	SM
True model	46.79%	21.37%	31.84%
MNL	41.64%	34.43%	23.93%
NL (rail-car)	44.66%	28.18%	27.16%
NL (rail-SM)	42.39%	35.29%	22.33%
RCL	53.29%	20.36%	26.35%
NL mixture	48.69%	19.10%	32.21%

Forecasted choice probabilities

	Car	Rail	SM
True model	50.65%	9.49%	39.86%
MNL	44.35%	30.17%	25.48%
NL (rail-car)	49.19%	22.56%	28.25%
NL (rail-SM)	45.18%	30.68%	24.15%
RCL	57.74%	12.00%	30.26%
NL mixture	51.84%	9.10%	39.06%

Relative change in choice probabilities

	Car	Rail	SM
True model	+8.25%	-55.58%	+25.18%
MNL	+6.50%	-12.38%	+6.50%
NL (rail-car)	+10.14%	-19.93%	+4.00%
NL (rail-SM)	+6.59%	-13.07%	+8.15%
RCL	+8.36%	-41.06%	+14.83%
NL mixture	+6.47%	-52.36%	+21.27%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	-21.21%	-77.73%	-74.18%
NL (rail-car)	+22.91%	-64.15%	-84.12%
NL (rail-SM)	-20.15%	-76.48%	-67.63%
RCL	+1.29%	-26.13%	-41.10%
NL mixture	-21.54%	-5.78%	-15.53%

Overall market shares

Original market shares

	Car	Rail	SM
True model	55.12%	23.65%	21.23%
MNL	55.83%	22.73%	21.43%
NL (rail-car)	55.83%	22.73%	21.43%
NL (rail-SM)	56.00%	23.14%	20.86%
RCL	55.71%	22.60%	21.68%
NL mixture	55.73%	22.70%	21.57%

Forecasted market shares

	Car	Rail	SM
True model	57.82%	13.45%	28.73%
MNL	57.80%	18.67%	23.53%
NL (rail-car)	58.45%	18.22%	23.33%
NL (rail-SM)	57.99%	18.87%	23.13%
RCL	58.36%	13.72%	27.92%
NL mixture	58.32%	12.88%	28.80%

Relative change in market shares

	Car	Rail	SM
True model	+4.91%	-43.13%	+35.31%
MNL	+3.53%	-17.88%	+9.77%
NL (rail-car)	+4.69%	-19.84%	+8.84%
NL (rail-SM)	+3.56%	-18.44%	+10.90%
RCL	+4.76%	-39.30%	+28.74%
NL mixture	+4.65%	-43.25%	+33.48%

Bias in predicted change

	Car	Rail	SM
True model	-	-	-
MNL	-28.11%	-58.55%	-72.32%
NL (rail-car)	-4.49%	-53.99%	-74.95%
NL (rail-SM)	-27.42%	-57.24%	-69.13%
RCL	-2.99%	-8.88%	-18.62%
NL mixture	-5.16%	+0.28%	-5.18%

Table 6.9: Forecasting exercise using data generated by two-level NL mixture model

correlation between car and rail. However, the bias is much lower than in the NL model. The best performance is obtained by the RCL model, which retrieves the behaviour from *true* model almost perfectly, although it underestimates the decrease in the market share of rail.

The results of the final forecasting example, making use of data generated with a NL mixture model (cf. Section 6.2.6), are summarised in Table 6.9.

The results for the representative individual show a decrease in the choice probability for rail by 55.58%, with a bigger shift towards SM than car, which is an effect of the nesting structure in the true model, which ensures correlation between rail and SM. In the MNL model, the *IIA* assumption leads to proportional changes, as well as a major underestimation of the actual changes. The NL model nesting car with rail, whose parameters are heavily influenced by confounding, underestimates the decrease in probability for rail, and falsely predicts a larger shift towards the car alternative. The NL model nesting rail with SM also underestimates the changes in probability, and the confounding in this model leads to an underestimation of the correlation between rail and SM, although it does show a bigger than proportional shift towards the SM alternative. The RCL model offers better performance than any of the closed form models, yet underestimates the changes in probability for rail and SM. However, it correctly predicts a bigger shift towards SM, which illustrates the *positive* effect of confounding in this case, leading to an approximation of an ECL structure. The bias in the change of the probability for car in the NL mixture model needs to be put in context by the small change in this probability.

The population-level results show a decrease in market share for rail by 43.13%, with a much bigger shift towards SM than to rail. Thanks to the averaging across individuals, the MNL model and the NL model nesting rail with car now correctly indicate a bigger shift in market share towards SM, but underestimate the extent of the disproportionality as well as of the actual changes in the market shares, especially for rail and SM. A similar issue with underestimation affects the model nesting rail with SM, which is again an effect of confounding leading to an attenuation of the nesting parameter in this model. As was the case with the representative individual, the RCL model again underestimates the changes in market share for rail and SM, while the best overall performance is obtained by the NL mixture model.

6.4 Summary and Conclusions

The aim of this chapter was to highlight a major issue that can affect estimation results in models allowing for an analysis of the error-structure, namely that of confounding between simple inter-alternative correlation and random inter-agent taste heterogeneity.

The theoretical discussions have shown that it is possible for the estimates relating to random taste heterogeneity to be biased by the presence of unexplained inter-alternative correlation, and conversely, for the estimates relating to inter-alternative correlation to be biased by the presence of unexplained random inter-agent taste heterogeneity.

Two possible scenarios arise in which this issue of confounding can play a major role. The first case is one where, in the true model, only one of the two phenomena plays a role, but where the estimated model allows only for the other phenomenon to have an effect. Here, the effects of the unexplained phenomenon can lead to erroneous results showing an effect of the other phenomenon. In the second scenario, both phenomena play a role, but the model employed in estimation allows only for the presence of either of the two phenomena. Here, the presence of the second, unexplained phenomenon, can lead to biased estimates in relation to the other phenomenon.

The six case-studies presented in this chapter have presented examples of each of the cases discussed above. The first two case-studies have illustrated how the presence of unexplained inter-alternative correlation can lead to erroneous results with regards to the prevalence of random taste heterogeneity (Section 6.2.1 and Section 6.2.2). The following three case-studies have illustrated the converse, showing how the presence of unexplained random taste heterogeneity can lead to erroneous results with regards to the presence of inter-alternative correlation (Section 6.2.3, Section 6.2.4, and Section 6.2.5). Finally, the case-study presented in Section 6.2.6 has shown that, in the case where both phenomena play a role, not accounting for the effect of one of the two phenomena can lead to biased results in relation to the second phenomenon. Each of the six case-studies has also shown how, by using models allowing jointly for the effects of the two phenomena, the risk of confounding is much reduced, although, in some cases, minor issues with confounding can still exist even with the use of such models (cf. Section 6.2.4).

The forecasting examples presented in Section 6.3 have shown that the use of models affected by confounding can lead to biased forecasts of market shares, which can in turn lead to misguided policy decisions. While the issue with misleading forecasts arises especially in the case of incorrect results in relation to inter-alternative correlation, problems are also caused in the case of incorrect results in terms of random taste heterogeneity, for example by giving an inadequate account of variations in willing-to-pay measures across individuals, which can lead to major problems in cost-benefit analysis.

The results discussed in this chapter offer strong evidence that modellers should acknowledge the potential risk of confounding, especially given the lack of a priori knowledge as to the *true* nature of the error-structure. While testing separately for the two phenomena, say with a GEV and a RCL model, can alert the modeller to the relative performance of the two approaches, it does not remove the risk of biased findings. As such, the findings from this chapter suggest that modellers should always allow for the effects of both phenomena in a joint fashion, either in a GEV mixture structure, as described in this chapter, or with the help of a combined ECL-RCL formulation, as discussed in Appendix A.

Chapter 7

Mixed Covariance models

7.1 Introduction and context

While the developments discussed in Chapter 2 in relation to closed form GEV as well as GEV mixture models have led to gradual gains in modelling flexibility, by allowing modellers to accommodate correlation across alternatives as well as deterministic and random taste heterogeneity across respondents, little effort has gone into the development of model forms allowing for a representation of heterogeneity across respondents in the correlation structure in place between the different alternatives. Such correlation heterogeneity is however potentially a crucial factor in the variation of choice-making behaviour across decision-makers. As an example, in an airline choice scenario, travellers' behaviour can be strongly affected by their membership in a given airline's frequent flier programme, to the point that, in the case where seats on their desired flight are not available, they are more likely to switch to a different flight on the same airline than to choose a flight by an alternative airline. In many cases, it may not be possible to accommodate the effects of airline allegiance directly, mainly for data reasons (cf. Chapter 8). In these circumstances, the greater substitution between flights on the same airline can be accommodated through a nesting structure that allows for correlation between flights on the same airline. It is clearly possible that the effects of airline allegiance, and hence the level of correlation, vary across travellers, meaning that the use of an approach which imposes covariance homogeneity potentially leads to biased model results.

While some of the covariance heterogeneity can conceivably be accommodated through an appropriate segmentation of the population (using separate models), it is likely that some within-segment heterogeneity remains. The existing literature seems to contain only two examples of a model allowing for such heterogeneity. The first of these comes in the form of the Covariance Nested Logit (COVNL) model discussed by Bhat (1997). In the COVNL model, the structural parameters themselves (and hence the pattern of substitution between alternatives) are a function of socio-demographic attributes of the decision-makers, such that the correlation heterogeneity is explained with the help of these attributes. Koppelman & Sethi (2005) later expand this approach by incorporating covariance heterogeneity in a GNL model[1], where they additionally allow for heteroscedasticity across respondents through a parameterisation of the scale factor, describing the resulting model

[1] Thus also allowing for cross-nesting.

as the Heterogeneous Generalized Nested Logit (HGNL) model.

While it is highly desirable to explain any covariance heterogeneity in a deterministic way, this is clearly not always possible. The aim of this chapter is therefore to develop a model structure that can accommodate random covariance heterogeneity in addition to deterministic covariance heterogeneity. The discussion presented in this chapter is based on an underlying GEV model for representing the correlation between alternatives; it is similarly possible to do this with the help of an ECL structure, and the development of such a framework is described in Appendix B. The model structure developed in this chapter was first discussed by Hess, Bolduc & Polak (2005).

The remainder of this chapter is organised as follows. The methodology for the Mixed Covariance GEV model is introduced in Section 7.2. Section 7.3 presents an application showing how one specific example of a Mixed Covariance GEV model works in practice. Finally, Section 7.4 presents the conclusions of the research.

7.2 Methodology

We will now develop the structure for our Mixed Covariance GEV model, where the derivation described here looks mainly at the case of a simple two-level NL model; the extension to multi-level as well as cross-nesting structures is possible, and several notes to that extent are made in the text. The exposition of the theory is divided into three parts. We first look at the general model form, in Section 7.2.1, before moving on to the cases of purely random variation (Section 7.2.2) and combined deterministic and random variation (Section 7.2.3).

7.2.1 General model form

The choice probabilities in a nested model are represented through a product of successive choice probabilities that represent a chain from the root of the tree (uppermost node) to the elementary alternative for which the probability is calculated. In a two-level NL model, the choice probability of alternative i (belonging to nest m) for individual n is then given by:

$$\begin{aligned} P_n(i) &= P_n(S_m)\, P_n(i \mid S_m) \\ &= \frac{e^{\lambda_m I_{m,n}}}{\sum_{l=1}^{M} e^{\lambda_l I_{l,n}}} \frac{e^{\frac{V_{i,n}}{\lambda_m}}}{\sum_{j\in S_m} e^{\frac{V_{j,n}}{\lambda_m}}}, \end{aligned} \tag{7.1}$$

with logsum term

$$I_{m,n} = \ln \sum_{j\in S_m} e^{\frac{V_{j,n}}{\lambda_m}}, \tag{7.2}$$

where $V_{j,n}$ gives the observed utility for alternative j and individual n, λ_m is the structural parameter associated with nest m, S_m defines the set of alternatives contained in nest m, and M gives the total number of nests. The extension of the choice probability from equation (7.1) to the multi-level case is straightforward, with details given for example by Koppelman & Sethi (2000).

The COVNL model of Bhat (1997) expands on the standard NL model, by parameterising the structural parameters λ as:

$$\lambda_{m,n} = F\left(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n}\right), \tag{7.3}$$

where α is a constant, $\boldsymbol{z_n}$ is a vector of attributes of decision-maker n, and $\boldsymbol{\gamma}$ is a vector of coefficients. In this notation, $\lambda_{m,n}$ is the structural parameter for nest m and decision-maker n. Both α and $\boldsymbol{\gamma}$ are to be estimated.

To ensure consistency with utility maximisation, $F\left(\right)$ needs to be specified so as to produce values in the $0-1$ interval. Furthermore, Bhat (1997) states that increases in $\boldsymbol{z_n}$ should have a monotonic effect on λ_n (where this ensures consistency in the case of multi-level structures, cf. equation (7.7)). This double requirement can be satisfied by using a function $F\left(\right)$ with:

$$\begin{aligned} F\left(-\infty\right) &= 0 \\ F\left(+\infty\right) &= 1 \\ f\left(x\right) &= \frac{\partial F\left(\right)}{\partial x} > 0 \end{aligned} \tag{7.4}$$

These conditions are met by the use of a continuous cumulative probability distribution function, where Bhat (1997) suggests the use of the logistic distribution.

We now extend this approach to the case where λ_m follows a random distribution across individuals. Conditional on a given set of values for the vector (of length M) of structural parameters $\boldsymbol{\lambda}$, the NL choice probabilities are given by equation (7.1). We now assume that the vector $\boldsymbol{\lambda}$ is distributed according to $f\left(\boldsymbol{\lambda} \mid \Omega\right)$, where Ω is a vector of parameters of the distribution of the different elements of $\boldsymbol{\lambda}$. This specification is general, and can be adapted for the special cases presented in Sections 7.2.2 and 7.2.3.

The conditional choice probability in equation (7.1) is now replaced by the unconditional choice probability:

$$P_n\left(i\right) = \int_{\boldsymbol{\lambda}} P_n\left(i \mid \boldsymbol{\lambda}\right) f\left(\boldsymbol{\lambda} \mid \Omega\right) \mathrm{d}\boldsymbol{\lambda} \tag{7.5}$$

$$= \int_{\boldsymbol{\lambda}} \frac{e^{\lambda_m I_{m,n}}}{\sum_{l=1}^{M} e^{\lambda_l I_{l,n}}} \frac{e^{\frac{V_{i,n}}{\lambda_m}}}{\sum_{j \in S_m} e^{\frac{V_{j,n}}{\lambda_m}}} f\left(\boldsymbol{\lambda} \mid \Omega\right) \mathrm{d}\boldsymbol{\lambda}, \tag{7.6}$$

where $\boldsymbol{\lambda} = \{\lambda_1, ..., \lambda_M\}$. Here, equation (7.6) is specific to the two-level NL model given in equation (7.1), while equation (7.5) shows the general form, where $P_n\left(i \mid \boldsymbol{\lambda}\right)$ can represent the conditional choice probability for any GEV model[2]. The logsum term I_m is defined as in equation (7.2), and it should be noted that this logsum term is conditional on a given value of λ_m, and hence $\boldsymbol{\lambda}$, by being inside the integral. The behaviour of the model depends crucially on the specification used for $f\left(\boldsymbol{\lambda} \mid \Omega\right)$,

[2]In the case of cross-nesting structures, there is an additional dependency on a vector of allocation parameters, which is not explicitly stated in equation (7.5). There is in that case also a possibility of allowing for deterministic as well as random variations across agents in the allocation parameters.

where the requirements on the range of the structural parameters need to be borne in mind. This issue is discussed in more detail in the description of the two special cases in Sections 7.2.2 and 7.2.3.

The approach becomes more complicated in the case of multi-level structures, due to the condition that the structural parameters need to decrease as we move down the tree. In the COVNL, this is made possible by specifying the structural parameter of a lower-level nest, λ_l, as in equation (7.3), and by adapting the specification of the upper-level nesting parameter as:

$$\lambda_{u,n} = F\left[\left(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n}\right) + G\left(\delta + \boldsymbol{\eta}'\boldsymbol{w_n}\right)\right], \tag{7.7}$$

where $\boldsymbol{w_n}$ is an additional vector of individual characteristics, which can be the same as $\boldsymbol{z_n}$, and where δ and $\boldsymbol{\eta}$ are a constant and a vector respectively that need to be estimated. Finally, $G\left(\right)$ is a monotonically increasing function mapping real numbers onto the space of positive real numbers, such as for example with the exponential distribution.

In the case of the Mixed Covariance NL model, the issue becomes more complicated, as the different structural parameters are now random variables. To ensure consistency with utility maximisation, the distribution of the structural parameters must be specified such that structural parameters belonging to the same link in a tree are no longer distributed independently. As it is desirable not to have to impose a constraint of equality of the structural parameters on a given level[3], it is preferable to use a top-down approach in the notation for the Mixed Covariance NL model, given that a specific node may have multiple *descendants*, while, in a model without cross-nesting, each node has only one direct *ancestor*.

One possible way of ensuring decreasing structural parameters is to specify the values as follows. With an upper-level structural parameter being given by:

$$\lambda_u \sim f\left(\lambda_u \mid \Omega_u\right), \tag{7.8}$$

the structural parameter of one of its *descendants*, λ_{li}, is given by:

$$\lambda_{li} = \lambda_u \, \widehat{\lambda_{li}}, \tag{7.9}$$

with

$$\widehat{\lambda_{li}} \sim f\left(\widehat{\lambda_{li}} \mid \Omega_{\widehat{\lambda_{li}}}\right), \tag{7.10}$$

where, in either case, the subscript imposed on Ω refers to the subelements linked to the structural parameter in question. This approach avoids the need to specify a complete joint density for the structural parameters.

The structural parameter at a lower level is thus given by multiplying the structural parameter at the level above it with a draw from the distribution used for the structural parameter at the lower level. As this draw is contained between 0 and 1, the resulting product is necessarily constrained between 0 and λ_u, giving $0 \leq \lambda_{li} \leq \lambda_u \leq 1$. If, at a given level, the draw from the distribution approaches 1, such that the resulting structural parameter takes the same value as its *ancestor*,

[3]This approach is taken by Bhat (1997).

this level of the tree becomes obsolete in that link, and the nests below it can be attached directly to the *ancestor* node. Extension of this theory to models with more than three levels is straightforward.

Extensions to models allowing for cross-nesting is also possible, although slightly more tedious. In this case, a given node can have multiple ancestors, and the condition of decreasing structural parameters needs to apply for each of the links to an ancestor. This means that the structural parameter at a given node needs to be less than or equal to that of the direct *ancestor* with the lowest structural parameter. Hence, in equations (7.9) and (7.10), λ_u is accordingly replaced by the structural parameter of this specific *ancestor* node. As it is thus possible to adapt this approach for models allowing for cross-nesting as well as for models allowing for multi-level nesting, it can be seen that the approach should be applicable for all existing GEV structures.

The final step in the theoretical development of our proposed model form is the representation of taste heterogeneity across individuals, where this heterogeneity relates to the coefficients multiplying the attributes of the alternatives, as opposed to the structural parameters. The above framework clearly already allows for deterministic variations in tastes; additional random variation can be accommodated very easily in the present model form, through integration of the choice probabilities that are conditional on β over the assumed distribution of the taste coefficients. This comes in addition to the integration over the distribution of the structural parameters.

Let $P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda})$ give the choice probability of alternative i for individual n, conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. Following the theory described in this section, we then have:

$$P_n(i \mid \boldsymbol{\beta}) = \int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}) f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda}. \tag{7.11}$$

By assuming that the tastes are distributed randomly across decision-makers according to $g(\boldsymbol{\beta} \mid \Theta)$, with parameter vector Θ, we obtain the unconditional choice probability[4]:

$$\begin{aligned} P_n(i) &= \int_{\boldsymbol{\beta}} P_n(i \mid \boldsymbol{\beta}) g(\boldsymbol{\beta} \mid \Theta) \mathrm{d}\boldsymbol{\beta} \\ &= \int_{\boldsymbol{\beta}} \left(\int_{\boldsymbol{\lambda}} P_n(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda}) f(\boldsymbol{\lambda} \mid \Omega) \mathrm{d}\boldsymbol{\lambda} \right) g(\boldsymbol{\beta} \mid \Theta) \, \mathrm{d}\boldsymbol{\beta}. \end{aligned} \tag{7.12}$$

7.2.2 Model with purely random covariance heterogeneity

We now look at the case where any variation in the structural parameters (and hence the correlation) across individuals is purely random. Two possible approaches arise in this case.

[4] Although beyond the scope of the present discussion, it is possible to expand this approach to the case where $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ follow some form of joint distribution.

In the first approach, we rewrite the choice probabilities in equation (7.5) as:

$$P_n(i) = \int_{\boldsymbol{y}} P_n(i \mid \boldsymbol{\lambda} = T(\boldsymbol{y}))\, f(\boldsymbol{y} \mid \Omega) \mathrm{d}\boldsymbol{y}, \tag{7.13}$$

where $T(\boldsymbol{y})$ is a transform that maps the elements in $\boldsymbol{y}$ from the real space of numbers into the $0-1$ interval. With this approach, any choice of statistical distribution can be used for $f(\boldsymbol{y} \mid \Omega)$, and a transform such as the logistic distribution can be used for $T(\boldsymbol{y})$.

The second approach avoids the use of the additional transform $T(\boldsymbol{y})$, and draws for the structural parameters are produced directly from the function $f(\boldsymbol{\lambda} \mid \Omega)$, as shown in equation (7.5). In this case, the condition on the range of the structural parameters applies directly at the level of $f(\boldsymbol{\lambda} \mid \Omega)$, leading to a requirement to use distributions bounded on either side, with the left bound being greater than 0, and the right bound being smaller than 1. The vector Ω now contains the parameters of the actual distribution of the structural parameters, as opposed to the distribution of the random vector $\boldsymbol{y}$ used as the base of the transform described in the first approach. A number of different statistical distributions can be used with this approach, including basic examples such as the *Uniform* or *Triangular*, or more advanced ones, like the Johnson S_B distribution.

It is not clear a priori which of the two approaches is preferable. The former approach allows for greater freedom in the choice of distribution for $f(\boldsymbol{y} \mid \Omega)$, while the latter approach provides more control over the actual shape of the distribution of the structural parameters. The merits of the two approaches potentially need to be evaluated on a case-by-case basis.

7.2.3 Model with deterministic and random covariance heterogeneity

While the description in Section 7.2.2 has shown that the framework developed in Section 7.2.1 can be adapted straightforwardly to allow for a purely random distribution of structural parameters across individuals, the use of this approach leads to similar issues of interpretation as in the case of randomly distributed taste coefficients in a GEV mixture model. Indeed, this approach provides little information about the values of the structural parameters for a given individual or a given population group, although posterior methods can be used to infer some such information. It is thus clearly preferable to as much as possible explain this covariance heterogeneity in a deterministic manner, as in the COVNL model of Bhat (1997). As mentioned in the introduction, this is not always possible, such that the Mixed Covariance models presented in this chapter present a useful alternative. However, it is conceivable that there are cases in which it is possible to explain some of the variation in a deterministic way, while some remaining part of covariance heterogeneity can only be explained in a random way, along the lines of $\lambda = F(\alpha + \boldsymbol{\gamma}'\boldsymbol{z_n} + \epsilon)$, where ϵ is a random component. Two approaches are possible in this case; one is to use a mixed version of a formulation analogous to the COVNL formulation (but within a top-down approach), while the other is to parameterise the parameters of the distribution used to represent covariance heterogeneity in the Mixed Covariance

GEV model. We will now look at these two approaches in turn.

Extension of COVNL approach

We begin the description of this approach by rewriting the choice probabilities in equation (7.5) as:

$$P(i) = \int_{\boldsymbol{\theta}} P\left(i \mid \boldsymbol{\lambda} = T\left(H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)\right)\right) f\left(\boldsymbol{\theta} \mid \Omega\right) \mathrm{d}\boldsymbol{\theta}, \tag{7.14}$$

In this notation, $T\left(\right)$ is defined as previously as a transform mapping independent elements from the real space of numbers into the $0-1$ interval. The function $H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)$ is used to generate a vector of length m of real numbers, as a function of the parameters contained in the vector $\boldsymbol{\theta}$ and the vector of individual-specific attributes $\boldsymbol{z_n}$, with $\boldsymbol{\theta}$ being distributed according to $f\left(\boldsymbol{\theta} \mid \Omega\right)$. This model can be seen to be an extension of the COVNL model described in Section 7.2.1 as follows. Let us assume that we have a model with a single structural parameter λ_m. It can be seen that, by specifying $T\left(\right)$ to be the logistic transform, $H\left(\boldsymbol{z_n}, \boldsymbol{\theta}\right)$ to yield $\alpha + \boldsymbol{\gamma}' \boldsymbol{z_n}$, and setting $f\left(\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}) \mid \Omega\right) = 1$, the model reduces to the COVNL model. In this case, the parameters contained in the vector $\boldsymbol{\theta}$ are fixed across individuals. However, the model uses a top-down approach, which makes for easier adaptation in the case of multi-level structures or cross-nesting structures (see Section 7.2.1).

By removing the assumption that $f\left(\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}) \mid \Omega\right) = 1$, we obtain a model with random variation in the structural parameters across individuals. Depending on the specification of $f\left(\boldsymbol{\theta} \mid \Omega\right)$, only some of the elements in θ will be random, allowing for example for a random offset α across individuals, with purely deterministic variation on top of it, or a fixed offset point with random and deterministic variation on top of it, or both. Different choices for $H\left(\right)$ and $T\left(\right)$ (with appropriate domain conditions) lead to differences in model behaviour. Finally, it can be seen that by setting all elements in $\boldsymbol{z_n}$ to be zero, we obtain a model with purely random variation as in the first approach described in Section 7.2.2 if α is distributed randomly, while the model reduces further to NL if α is kept fixed across respondents. This completes the extension of the COVNL framework to the case with random parameters.

Parameterisation of distributional parameters

We will base our derivation of the parameterisation method on the second approach described in Section 7.2.2, such that draws for λ_m are obtained directly from an appropriate distribution with an acceptable domain, as opposed to requiring the use of a transform (which is also possible). Let us assume that we have $\omega_m \in \Omega$, such that ω_m represents for example the mean used in the distribution function of structural parameter λ_m, with a corresponding variable $\sigma_m \in \Omega$ giving the dispersion parameter of the distribution of structural parameter λ_m. For now, let us assume that σ_m stays constant across individuals; extension to the case where it varies (deterministically across individuals) in addition to ω_m is straightforward. We now look at the case where some of the variation in λ_m is explained by random variation (through using the distribution $f(\lambda_m \mid \omega_m, \sigma_m)$ and some variation is explained by the attributes of the decision-maker, by parameterisation of ω_m. Specifying $\omega_{m,n}$ to be the mean value of the distribution of λ_m for decision-maker n, we can then simply use:

$$\omega_{m,n} = \alpha_{\omega_m} + \boldsymbol{\gamma_{\omega_m}}'\boldsymbol{z_n}, \tag{7.15}$$

where $\boldsymbol{z_n}$ represents a vector of attributes of decision-maker n, and α_{ω_m} and $\boldsymbol{\gamma_{\omega_m}}$ represent a constant and a vector of coefficients respectively, both of which are specific to the parameter ω_m.

In the case where no parameterisation of the parameters of the distribution is (or can be) used, only the constant α_{ω_m} will be estimated. In this case, $\omega_{m,n}$ stays the same across respondents, and the only differences in the value of λ_m across respondents are due to random variation. On the other hand, a model version that is very similar to the COVNL model can be obtained by only using one distributional parameter for each structural parameter, i.e. by setting

$$P(\lambda_{m,n} = \omega_{m,n} \mid \Omega) = 1 \tag{7.16}$$

This is equivalent to setting the dispersion term σ_m to be equal to zero. In this case, different structural parameters are still used for different individuals, but they no longer vary randomly across individuals; the variation is entirely deterministic. By further setting $\boldsymbol{\gamma_{\omega_m}} = 0$ for all m, the model reduces to the NL model.

Discussion

It is of interest to briefly discuss the differences between the two approaches. Both approaches attain the goal of jointly introducing deterministic and random covariance heterogeneity. The former approach has the advantage of easier interpretation, and possibly simplifies more easily to models with purely deterministic covariance heterogeneity, as well as models with fixed covariances. The only apparent advantage of the second approach is that it can avoid the need for additional transforms in the case where strictly bounded statistical distributions are used. Although, like the first approach, this variant also allows for an effect by an unlimited number of socio-demographic attributes, their impact needs to be gauged simultaneously for a minimum of two separate values, giving the mean and dispersion of the associated statistical distribution. As such, the former approach is probably preferable, although a detailed empirical comparison would be needed to reach a definitive answer.

7.3 Application

In this section, we present an application of one specific type of Mixed Covariance GEV model, namely a discrete mixture of a two-level NL model, with two possible levels of correlation in the population, leading to a Discrete Mixture Covariance NL (DM-COVNL) model. As such, the work presented here relates to the discussion on discrete mixture models in Chapter 5, where, in the present context, the mixture allows for covariance heterogeneity, as opposed to taste heterogeneity[5].

The justification for using the DM-COVNL model instead of a continuous mixture in this application is primarily a pragmatic one. Indeed, while it can simply be seen as a special case of a continuous mixture, it has the clear advantage of not requiring simulation in estimation. However, the discrete approach also has some advantages in terms of illustration of the differences with a homogeneous covariance model, as well as having conceptual advantages in terms of the notion of an unobserved attribute leading to inter-alternative correlation for only part of the population of decision-makers.

Using a notation similar to that used in Chapter 5, the choice probability for alternative i and individual n in a model with K nests is given by:

$$P_n\left(i \mid \boldsymbol{\beta}\right) = \sum_{m_1=1}^{M_1} \cdots \sum_{m_K=1}^{M_K} P_n\left(i \mid \boldsymbol{\beta}, \boldsymbol{\lambda} =< \lambda_1^{m_1}, \ldots, \lambda_K^{m_K} >\right) \pi_1^{m_1} \ldots \pi_K^{m_K}, \quad (7.17)$$

where the structural parameter λ_k, associated with the k^{th} nest, takes on M_k separate values, defined as λ_k^1 to $\lambda_k^{M_k}$, where each has an associated probability (or mass), with $0 \leq \pi_k^{m_k} \leq 1 \; \forall k, m_k$, and where $\sum_{m_k=1}^{M_k} \pi_k^{m_k} = 1 \; \forall k$. Here, in addition to the taste coefficients, estimates need to be produced for the different levels for all the structural parameters, as well as for the associated probabilities.

With the aim of illustrating the ability of the model to recover covariance heterogeneity, and to show the bias resulting from an inappropriate assumption of covariance homogeneity, the application presented here makes use of simulated data. The data used are again based on the Swissmetro dataset (cf. Section 6.2). This time, a sample of $9,000$ observations was used, based on an original sample of $3,000$ observations, where the data augmentation was based on small random variations of the original attribute levels. The base specification used for the utility function is the same as that in Section 6.2, with separate travel time coefficients for rail, SM and car, a common cost-coefficient, a headway coefficient common to rail and SM, and two ASCs, associated with SM and car.

Unlike in the applications conducted in Chapter 6, the generation of the data is now based on the principle of nine observations per individuals, as opposed to a purely cross-sectional approach. In the generation of the data, the $1,000$ individuals were split into two groups. In the first group, representing 30% of the population, there is high correlation between the error-terms for the rail and SM alternatives, with a structural parameter equal to 0.3. In the remaining 70% of the population, the structure equates to a MNL model. The allocation to the two groups is performed on a purely random basis (taking into account the $30\% - 70\%$ split), such that a deterministic segmentation of the population cannot be used to account for the differences in correlation structure. This construction represents a situation in which, for example, for some individuals, an unobserved attribute leads to heightened substitution between rail and SM, while, for the remainder of the population it does not[6].

[5]Here, it should be noted that the issue of confounding between the two, as discussed in Chapter 6, can also apply in the case of mixed covariance models.

[6]This could for example simply reflect an inherent dislike of car-travel for some respondents.

On the basis of the resulting individual-specific structural parameters, and the coefficient values reported for the *true* model in Table 7.1, the choice probabilities for the three alternatives were calculated for each individual, on the basis of a two-level NL structure nesting rail with SM, where, for those individuals with $\lambda = 1$, the probabilities correspond to a MNL structure. A Monte-Carlo exercise was then used to determine the chosen alternative. As such, for each individual, the actual structural parameter applying for that respondent was used. This is more correct, and consistent with the underlying *true* model, than an approach which uses simulation over the two values, assigning to each individual the weighted choice probability across the two values for λ. As such, the resulting dataset reflects a real-world situation (in which a single value applies for each individual), rather than a DM-COVNL approximation to such a real-world situation. This in turn means that the estimation can show how well the DM-COVNL model, which does use a weighted average across the two values for λ, can replicate the *true* model.

Three separate models were estimated on the resulting dataset; a MNL model, a simple NL model nesting together rail and SM, and a DM-COVNL model. All three models were coded in Ox. In the DM-COVNL model, we estimate two distinct structural parameters for the rail-SM nest, specified as λ_a and λ_b. As such, with λ_a, the choice probability of rail in the t^{th} choice situation for individual n is given by:

$$P_{n,t}\left(rail \mid \lambda_a\right) = \frac{e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \frac{e^{\frac{V_{rail,n,t}}{\lambda_a}}}{e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}}, \tag{7.18}$$

where $V_{rail,n,t}$, $V_{SM,n,t}$ and $V_{car,n,t}$ give the observed utility for rail, SM and car respectively, for individual n, in choice situation t. The corresponding choice probabilities for SM and car are given by:

$$P_{n,t}\left(SM \mid \lambda_a\right) = \frac{e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \frac{e^{\frac{V_{SM,n,t}}{\lambda_a}}}{e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}}, \tag{7.19}$$

and

$$P_{n,t}\left(car \mid \lambda_a\right) = \frac{e^{V_{car,n,t}}}{e^{V_{car,n,t}} + e^{\lambda_a \ln\left(e^{\frac{V_{rail,n,t}}{\lambda_a}} + e^{\frac{V_{SM,n,t}}{\lambda_a}}\right)}} \tag{7.20}$$

On the basis of equations (7.18), (7.19) and (7.20), the probability of the observed sequence of choices for individual n, conditional on λ_a, is given by:

$$L\left(n \mid \lambda_a\right) = \prod_{t=1}^{T_n} \left[\delta_{n,t,rail} P_{n,t}\left(rail \mid \lambda_a\right) + \delta_{n,t,SM} P_{n,t}\left(SM \mid \lambda_a\right) + \delta_{n,t,car} P_{n,t}\left(car \mid \lambda_a\right)\right], \tag{7.21}$$

where T_n gives the number of choice-situations for respondent n (equal to 9 in this application), and where the dummy variable $\delta_{n,t,rail}$ is equal to 1 if respondent n chooses rail in the t^{th} choice-situation, and zero otherwise, with a corresponding definition for $\delta_{n,t,SM}$ and $\delta_{n,t,car}$.

With an equivalent notation in the case of the second structural parameter, λ_b, the contribution by individual n to the likelihood function is given by:

$$L(n) = \pi_{\lambda_a} L(n \mid \lambda_a) + \pi_{\lambda_b} L(n \mid \lambda_b), \tag{7.22}$$

where π_{λ_a} and π_{λ_b} give the mass for λ_a and λ_b respectively, with $0 \leq \pi_{\lambda_a} \leq 1$, $0 \leq \pi_{\lambda_b} \leq 1$, and $\pi_{\lambda_a} + \pi_{\lambda_b} = 1$. The fact that the weighting over the two support points occurs at the level of $L(n \mid \lambda_a)$ and $L(n \mid \lambda_b)$, rather than at the level of individual choice probabilities, reflects the notion that the level of correlation stays constant across replications for the same individual.

Finally, on the basis of equation (7.22), the log-likelihood function for the DM-COVNL model used in this example is given by:

$$\begin{aligned} LL &= \ln\left(\prod_{n=1}^{N} L(n)\right) \\ &= \sum_{n=1}^{N} \ln\left[\pi_{\lambda_a} L(n \mid \lambda_a) + \pi_{\lambda_b} L(n \mid \lambda_b)\right], \end{aligned} \tag{7.23}$$

where N gives the total number of individuals, with, in the present application, $N = 1,000$.

The estimation results for the three models are summarised in Table 7.1, together with the coefficient values used in the generation of the data. The results show that the use of the NL model leads to statistically significant improvements in model fit over the MNL model, by 20.89 units, at the cost of one additional parameter. The DM-COVNL model leads to the best model fit overall, offering an improvement by 29.62 units over the NL model, with two additional estimated parameters (λ_b and π_{λ_a})[7]. It should be said that, although statistically significant, these improvements are not dramatic, suggesting that the likelihood function is relatively unaffected by the treatment of correlation. Additionally, it can be seen that the results in terms of willingness to pay indicators are very similar across the three models. Indeed, the recovery of the *true* values is very good, and the differences in bias are very small, where the lowest bias is obtained with the DM-COVNL model[8].

More significant differences however arise when looking at the implications in terms of correlation between the unobserved utility components for the rail and SM alternatives. The MNL model, by definition, offers no treatment of the correlation, and as such fails to allow for the heightened substitution between rail and SM. The simple two-nest NL model is based on the assumption of a homogeneous correlation structure. Here, the estimate produced for the unique nesting parameter in this model, at 0.78, is virtually indistinguishable from the weighted average of the two

[7]In the code written in Ox, only π_{λ_a} was estimated, with π_{λ_b} given by $1 - \pi_{\lambda_a}$.

[8]The fact that the bias decreases as we move from the MNL model to the NL model and on to the DM-COVNL model does suggest some interaction between observed and unobserved utility components, where a proper treatment of the unobserved utility components in the DM-COVNL model results in lower impact on the observed utility components.

	True model	MNL		NL		DM-COVNL	
Final LL	-	-7136.16		-7115.27		-7085.65	
Parameters	-	7		8		10	
adj. $\rho^2(0)$	-	0.2776		0.2796		0.2824	
		est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-4	-4.3977	-34.31	-4.0194	-30.96	-3.9547	-30.77
δ_{SM}	-3	-3.5073	-33.80	-3.0582	-28.01	-3.0376	-28.51
β_{TC}	-0.1	-0.1082	-51.03	-0.0999	-42.17	-0.0994	-41.82
β_{HW}	-0.02	-0.0233	-32.06	-0.0205	-27.62	-0.0202	-28.27
$\beta_{TT,car}$	-0.03	-0.0331	-39.36	-0.0302	-33.17	-0.0300	-32.97
$\beta_{TT,rail}$	-0.04	-0.0446	-47.61	-0.0402	-37.32	-0.0399	-37.31
$\beta_{TT,SM}$	-0.035	-0.0382	-37.14	-0.0350	-33.32	-0.0347	-33.42
λ_a	1	-		0.78	7.52	1.00	0.00
λ_b	0.3	-		-		0.32	12.50
π_{λ_a}	0.7	-		-		0.71	9.96
Monetary value	CHF/hour	CHF/hour		CHF/hour		CHF/hour	
TT_{car}	18.00	18.36		18.13		18.11	
TT_{rail}	24.00	24.74		24.17		24.11	
TT_{SM}	21.00	21.21		21.05		20.96	
HW	12.00	12.95		12.33		12.22	

T-statistics for structural parameters calculated wrt 1

Table 7.1: Estimation results on mixed covariance data

structural parameters present in the *true* population ($0.3 \cdot 0.3 + 0.7 \cdot 1.0 = 0.79$). This result is consistent with a similar observation made in the case of discrete mixture models for taste heterogeneity (cf. Section 5.4), reflecting the fact that single parameter models yield estimates that are weighted averages of the actual values present in the population. It should be noted that, in the current example, with only two parameters, this approximation is made relatively easy, and more bias could be expected in the presence of more than two values for a parameter. Finally, the DM-COVNL is able to essentially perfectly recover the nature of the correlation structure in place in the *true* data; λ_a obtains a value equal to 1.0, while, for λ_b, the estimated value is very close to the true value of 0.3, with the difference being significant only at the 28% level. Similarly, the estimated shares for the two structural parameters, at $\pi_{\lambda_a} = 0.71$ and $\pi_{\lambda_b} = 0.29$ are indistinguishable from the true $70\% - 30\%$ split. In an actual application, it would, after model estimation, be of interest to proceed with a posterior analysis, to produce the most likely structural parameter for each individual. The same approach would be used in the case of a continuous mixture model. On the basis of the results from such an analysis, attempts could then be made to relate the correlation to socio-demographic attributes, and to use an appropriate segmentation in later forecasting applications.

In practice, posterior analyses of this nature are used very sparsely; in the absence of the resulting insight into the *actual* structural parameters, the mixture model, in this case the DM-COVNL model, would thus potentially be used directly in

	Original probabilities			Forecasted probabilities		
	Rail	SM	Car	Rail	SM	Car
True model	17.49%	33.01%	49.49%	4.07%	38.38%	57.55%
MNL	15.65%	30.76%	53.59%	3.15%	35.32%	61.53%
NL ($\lambda = 0.78$)	14.41%	32.41%	53.18%	2.17%	38.67%	59.16%
DM-COVNL ($\lambda = 1$)	17.38%	32.23%	50.39%	4.07%	37.42%	58.50%
DM-COVNL ($\lambda = 0.32$)	4.94%	35.06%	60.00%	0.03%	38.98%	60.98%
DM-COVNL (total)	13.73%	33.06%	53.21%	2.89%	37.88%	59.23%
	Relative change			**Bias in predicted change**		
	Rail	SM	Car	Rail	SM	Car
True model	-76.76%	+16.28%	+16.28%	-	-	-
MNL	-79.87%	+14.82%	+14.82%	+4.06%	-8.96%	-8.96%
NL ($\lambda = 0.78$)	-84.93%	+19.31%	+11.26%	+10.65%	+18.62%	-30.84%
DM-COVNL ($\lambda = 1$)	-76.55%	+16.10%	+16.10%	-0.27%	-1.07%	-1.07%
DM-COVNL ($\lambda = 0.32$)	-99.30%	+11.18%	+1.65%	+29.37%	-31.29%	-89.89%
DM-COVNL (total)	-78.96%	+14.57%	+11.32%	+2.87%	-10.48%	-30.46%

Table 7.2: Forecasting on mixed covariance data: representative individual with $\lambda_a = 1.0$ (observation 2,044)

forecasting[9]. As such, it is of interest to compare the forecasting performance of the three models. To illustrate the differences in performance depending on the correlation structure in place in the *true* data, two representative individuals were selected, one belonging to the group with λ_a (respondent 228), and one belonging to the group with λ_b (respondent 812). In each case, a single observation was selected, where, for respondent 228, the first observation was used (observation $2,044$), while, for respondent 812, the second observation was used (observation $7,301$). The forecasting analysis looks at the changes in the choice probabilities for the three alternatives following an increase in the cost of rail-travel by 10%. The results of the forecasting exercise are summarised in Table 7.2 for observation $2,044$ and Table 7.3 for observation $7,301$. The bias measure used as an indicator of the correct recovery of the behaviour implied by the *true* model is defined as in Section 6.3. In each case, the results for the DM-COVNL are split into three parts, showing the results for the part of the model that uses λ_a, the part of the model that uses λ_b, and the results for the combined model. This gives an idea of the gains in performance that could be expected if the forecasting exercise was preceded by a posterior analysis that was able to yield an appropriate segmentation, while also giving an account of the bias introduced by using the actual DM-COVNL, instead of its sub-parts.

The results for observation $2,044$ (Table 7.2) show a decrease in the choice probability of rail from 17.49% to 4.07%, following an increase in rail-fares by 10%[10]. The fact that λ_a is used for this individual implies an equal relative shift of probability towards SM and car. The same applies in the MNL model and the DM-COVNL sub-model with λ_a, resulting in the lowest bias for these two models, where the fact

[9] As opposed to using a posterior segmentation.

[10] Lower decreases were observed at the population level (-35.89%), but the individual-observation results are used here, as they provide more insight into substitution patterns.

	Original probabilities			Forecasted probabilities		
	Rail	SM	Car	Rail	SM	Car
True model	37.96%	25.03%	37.01%	20.86%	38.64%	40.50%
MNL	36.60%	33.01%	30.38%	29.22%	36.86%	33.92%
NL ($\lambda = 0.78$)	36.92%	32.33%	30.75%	28.63%	37.34%	34.03%
DM-COVNL ($\lambda = 1$)	38.13%	33.83%	28.05%	31.17%	37.63%	31.20%
DM-COVNL ($\lambda = 0.32$)	36.56%	25.01%	38.43%	20.61%	37.46%	41.93%
DM-COVNL (total)	37.67%	31.24%	31.09%	28.07%	37.58%	34.35%

	Relative change			Bias in predicted change		
	Rail	SM	Car	Rail	SM	Car
True model	-45.07%	+54.39%	+9.45%	-	-	-
MNL	-20.16%	+11.64%	+11.64%	-55.26%	-78.60%	+23.17%
NL ($\lambda = 0.78$)	-22.45%	+15.51%	+10.65%	-50.18%	-71.49%	+12.70%
DM-COVNL ($\lambda = 1$)	-18.25%	+11.24%	+11.24%	-59.51%	-79.33%	+18.98%
DM-COVNL ($\lambda = 0.32$)	-43.62%	+49.75%	+9.11%	-3.21%	-8.53%	-3.56%
DM-COVNL (total)	-25.47%	+20.29%	+10.47%	-43.47%	-62.69%	+10.80%

Table 7.3: Forecasting on mixed covariance data: representative individual with $\lambda_b = 0.3$ (observation 7,301)

that the bias in the DM-COVNL sub-model is lower than in the MNL model (with the same treatment of correlation) can potentially be explained on the basis of more accurate estimates for the marginal utility coefficients. This is a result of the fact that the overall DM-COVNL model accounts for the correlation in the second sub-group, which the MNL model does not, where interaction between the observed and unobserved utility components leads to the bias in the estimates. The effects of the correlation structure become most visible when looking at the forecasts produced by the NL model, with $\lambda = 0.78$, and the DM-COVNL sub-model with $\lambda_b = 0.32$. Here, either approach leads to biased forecasts, by falsely indicating heightened substitution from rail to SM, where, due to the higher implied correlation, the bias is bigger in the DM-COVNL sub-model with λ_b than in the NL model. Here, it should also be noted that the DM-COVNL sub-model with λ_b significantly underestimates the original choice probability for rail. Finally, the combined DM-COVNL model leads to lower bias than the NL model, where it should also be said that the DM-COVNL model performs quite well overall for the changes in the probability for rail and SM, with the only major bias, when compared to the MNL model[11], arising for the change in the probability of the car alternative.

The results for observation $7,301$ (Table 7.3) show a decrease in the choice probability of rail from 37.96% to 20.86%, following an increase in rail-fares by 10%. With individual 812 belonging to the 30% of the population with heightened correlation between rail and SM, the true model shows a much bigger relative shift from rail to SM than to car, a situation that is recovered almost perfectly in the DM-COVNL sub-model using $\lambda_b = 0.32$. The MNL model wrongly predicts equal relative shifts in probability from rail to SM and from rail to car, where the same applies for the DM-COVNL sub-model using $\lambda_a = 1.0$. While the NL model correctly recovers the fact that there is a bigger than proportional shift towards SM than towards car,

[11]Which has the clear advantage in this case in terms of the correct correlation structure.

it underestimates the extent of the differences, through underestimating the correlation between the unobserved utility terms for rail and SM. The same occurs in the overall DM-COVNL model, where the underestimation is however less severe than in the NL model[12]. It should also be said that all models, except the DM-COVNL sub-model with $\lambda_b = 0.32$, significantly underestimate the decrease in the probability of the rail alternative, where this bias is however smallest in the overall DM-COVNL model, which also obtains the lowest overall bias out of the three full models.

In summary, this application has shown that the DM-COVNL model is able to recover the distribution of the covariance in the simulated dataset arbitrarily closely, while the simple NL model produces a weighted mean of the *true* values, on the basis of an assumption of covariance homogeneity. The forecasting application has also shown that the DM-COVNL model leads to lower bias than the NL model. Here, it should be noted that, in the special case described here, the MNL model performs well for the part of the population with no correlation between rail and SM, whereas it leads to significant bias in the remaining part of the population[13]. The fact that, in each case, the lowest bias is obtained by the appropriate DM-COVNL sub-model again illustrates the potential gains that could be obtained by conducting a posterior analysis to attempt to relate the difference in correlation structure to socio-demographic attributes with the aim of obtaining an appropriate segmentation for use in the actual forecasting exercise.

7.4 Summary and Conclusions

The aim of this chapter was to extend the standard discrete choice modelling framework so as to allow for random variations in the covariance structure across respondents. The discussion in this chapter has centred on the case of an underlying GEV model, and specifically, a two level NL model. The extension to other underlying GEV structures poses no major difficulties, as described in the text, while the use of an alternative approach, based on an underlying ECL structure, is described in more detail in Appendix B.

The development of the Mixed Covariance GEV structure in this chapter has shown how it is possible to allow jointly for random as well deterministic variations in the covariance structure across respondents. Additionally, it is possible, by adding an extra layer of integration, to allow for random taste heterogeneity, in addition to covariance heterogeneity. Here, it should also be noted that additional random terms can be added to allow for heteroscedasticity across alternatives, leading to additional dimensions of integration.

The application presented in Section 7.3 has described one special case of a Mixed Covariance GEV model, in which the mixture is discrete rather than continuous. The results have shown that the DM-COVNL structure is able to recover the covariance structure in place in the data very closely, and leads to lower bias

[12]The shift from rail to SM is close to twice as big as the shift from rail to car, while, in the NL model, the ratio is below 1.5. In the true model, the ratio is close to 6.

[13]Much poorer overall performance would be obtained in the case where, in the *true* model, both structural parameters are inferior to 1, or if the share for $\lambda_a = 1$ was smaller.

in forecasting than the simple NL model, which is based on the assumption of a homogeneous covariance structure.

Much work remains to be done, including the development of more sophisticated mixed covariance structures, the testing of continuous mixture structures on simulated data, and the use of discrete and continuous mixture structures with real data. Here, it should be noted that the discussion in this chapter has focussed primarily on variations in the extent of correlation across respondents, rather than variations in the actual correlation structure. The latter applies for example in the case where, for individual A, there is correlation between alternatives 1 and 2, while, for individual B, there is correlation between alternatives 2 and 3. Such variations in the actual structure can, in the absence of an appropriate segmentation, be accommodated in a cross-nesting framework, with the variation in structure accounted for primarily through variations in the allocation parameters.

In closing, it should be said again that mixed covariance models should in part be seen as an explanatory tool, which, unlike other models, have the power to highlight the presence of variations in inter-alternative correlation across respondents. On the basis of such results, the modeller can then attempt to refine the model to accommodate some covariance heterogeneity in a deterministic fashion, either through a segmentation of the data, or by parameterising the covariance structure, as described by Bhat (1997), potentially with additional random covariance heterogeneity, as described in Section 7.2.3. If such attempts at a deterministic approach fail, it is still desirable, for interpretation as well as forecasting reasons, to try to link the variations to socio-demographic information through a posterior analysis[14]. However, if this is not possible, then it is clearly preferable to account for the variation in a random way (in interpretation as well as forecasting), as opposed to maintaining the assumption of covariance homogeneity. Either way, the modelling approach described in this chapter is thus a valuable tool for the analysis of choice behaviour.

[14]Here, it should be said that the same reasoning applies in the case of mixture models looking for taste heterogeneity; again, a deterministic treatment is clearly preferable for interpretation as well as forecasting reasons, and the mixture model can thus be seen as an explanatory tool.

Chapter 8

Modelling air-travel choice-behaviour

8.1 Introduction

As illustrated by the discussions in the theoretical part of this thesis, the area of discrete choice modelling has seen a significant increase in activity over recent years, with the development of ever more flexible model structures that allow for an increasingly realistic representation of complex choice behaviour. As also alluded to however, at the same time, a wide gap has opened between the state-of-the-art, i.e. the theoretical developments, and the state-of-practice, i.e. the actual applications of the model structures to the analysis of real-world problems.

The applied part of this thesis aims to at least partly bridge this gap in one specific area, namely the field of air-transport[1], which, from a topical as well as a methodological angle, is one of the most interesting domains for analysing travel behaviour. Indeed, from a policy point of view, the continuing precarious financial situation of the air-travel industry, together with the long-term nature of any policy-changes, means that reliable forecasts of passenger behaviour are a crucial component of transport planning in this area, especially given that important decisions need to be faced in many areas over the coming years, with a view to capacity extension, as well as a host of other measures, such as the possible introduction of congestion-charging and other noise or air-pollution related surcharges.

Secondly, the very nature of the behavioural processes, which involve decisions along a multitude of dimensions of choice, influenced by a very high number of factors, makes air-transport one of the most appealing and challenging areas for analysing choice behaviour, and an ideal area for deploying some of the new, highly flexible model structures. Additionally, the fact that the nature of air-travel is still evolving leads to a constant need for new research. Here, one example of recent changes comes in the increase in activity by low-cost carriers in Europe, which have opened air-travel to a much wider part of the population, and have led to increased use of regional airports (cf. Barret 2000). To further complicate matters, the product offered by *standard* network carriers is also changing, with ongoing

[1]It should be noted that this work is solely concerned with the field of passenger transport, and ignores the separate dimension of air-cargo. Additionally, the work looks only at the choices made by passengers, as opposed to industry-agents, such as airports and airlines.

consolidation and the shaping of new alliances (cf. Dennis 2005). Finally, the fact that the role of travel agents is increasingly being taken over by internet bookings may well lead to changes in behaviour, as the impact of travel agents on the actual choices slowly disappears.

Although some progress has recently been made to address this gap between the state-of-the-art and the state-of-practice in air-transport research (as discussed in Section 8.3), a lot remains to be done. Indeed, the number of advanced modelling applications is still very limited when compared to other areas of transport analysis, and they almost universally rely on assumptions that often significantly simplify the complexity of the choice process.

In the case-studies presented in this part of the thesis, we look at the modelling of one specific choice, namely that of departure airport in multi-airport regions. The analysis of the relationship between changes in level-of-service attributes and shifts in demand between airports is an important component of long-term transport strategies in such areas[2]. Indeed, changes in demand at the individual airports not only have an effect on the commercial viability of the single airports, but can have significant effects on the support structure of the airports (auxiliary businesses), the local transport network, as well as on seemingly less related businesses (e.g. local hotels). One scenario in which forecasts of passenger behaviour can become necessary is the expansion of airport-capacity in multi-airport regions[3]; as any such work is a costly and long-term project, it is important to *a priori* forecast the effects of the different schemes under consideration, both on traffic at the airports as well as on traffic in the associated ground-level airport.

The work presented in this thesis recognises the fact that passengers additionally make a choice of airline, and access-mode, and the results highlight the importance of modelling these choices jointly. Three separate case-studies of airport choice behaviour are presented in this thesis[4]; the present chapter acts as an introduction to the applied part of the thesis, and sets the stage for the detailed discussion of these three studies.

The remainder of this chapter is organised as follows. In Section 8.2, we discuss the choice processes undertaken by air-travellers. This is followed by a review of previous research using discrete choice models in the field of air transport in Section 8.3. Section 8.4 presents an overview of the three case-studies conducted in this thesis, looking separately at the scope and aims of the studies, the model approach that was used, and some of the issues that had to be faced. The chapter closes with a brief summary in Section 8.5.

[2]See de Neufville (1995) for a discussion of the issues involved in the management of multi-airport systems.

[3]Despite the difficulties faced by the industry, air-travel is predicted to continue its growth at average annual rates of around 5% (cf. Boeing 2004). This rapid increase in the number of passengers, flights and routes has not only led to environmental concerns, but has also resulted in significant problems with congestion, leading to urgent needs for capacity expansion, partly with a view to ensuring the continuing important contribution of aviation to the economy (cf. OEF 1999).

[4]Chapters 9, 10, and 11.

8.2 Air-travel choice-behaviour

8.2.1 General framework

Before discussing the scope of the air-travel research described in this thesis, it is worth reconsidering the actual choice behaviour of air-travellers. Broadly speaking, outside a mode-specific context, and without aiming to define the order of choices, travellers can, for a given trip[5], be seen to take decisions along three main *upper-level* dimensions of choice:

- Destination
- Timing (time & date)
- Main mode of travel

Here, it can be seen that the decisions along the *destination* dimension and the *main mode* dimension potentially involve a number of sub-choices, especially in the case of long-distance travel. As such, the destination dimension can for example also encompass choices of sub-destinations for journeys involving travel to more than one destination. The division along the *main mode* dimension is far more extensive, containing for example choices of itinerary, and fare-class in the case of public transport. Additionally, in the case of a combination of modes, there is a choice of the auxiliary modes, for example for the journeys to and from the departure and arrival location of the main mode.

It should also be noted that the above three dimensions are strongly interrelated. As such, while trip timing is clearly influenced by outside factors such as work commitments, the choice set in terms of possible departure times (and in some cases even departure dates) depends on mode-specific attributes for all but self-operated modes. The choice of destination can clearly be seen to have a significant impact on the other two dimensions of choice, for example by limiting the number of possible modes of travel. However, even within this general framework, it can occasionally be argued that the choice of destination is not in fact made a priori, but is itself a function of other choices. As an example, a traveller who takes a decision to rely on public transport will be limited in the number of potential destinations. An even stronger example is given by the case of people refusing to travel by air, or by sea. Clearly, such factors come into play mainly in the case of leisure travel, where, depending on the circumstances, they can play a major role.

8.2.2 Dimensions of choice in air-travel

In the case of air-travel, the situation becomes significantly more complicated. Indeed, not only are the three main dimensions of choice listed above again strongly inter-related, but the choice of air as the main mode of travel leads to a high number of sub-choices, potentially more so than with any other mode. Essentially, on top of the choice of air as the main-mode, the choices made by an air-traveller can be divided into three main subcategories, which we will now look at in turn, before touching on the potential inter-dependencies between the various choices.

[5]Hence not looking at the frequency of travel, or indeed the decision to travel.

The choices are described for the outbound-leg of a return journey. In general, for passengers on their return-leg, the majority of journey-factors are pre-determined by the choices made on the outbound leg, although some factors, such as timing and possibly also routing are determined separately[6]. Finally, for passengers on one-way journeys, the choice process is very similar to the one described below for the outbound journey of return passengers, though outside factors and personal priorities may change significantly.

Origin-side

The choices on the origin-side of an air-journey can be divided into two main parts, namely the choice of departure-airport, and the choices made for the ground-level journey to this departure-airport.

In many cases, the choice set for the departure airport is very limited, and dominated heavily by the airport closest to the passenger's ground-level origin. However, for passengers living near major urban centres, there will often be a choice between a number of airports located at similar distances from a given passenger's ground-level origin. In some rare cases, passengers may even be faced with a choice between airports located in separate nearby multi-airport regions, such that there is a choice of *departure-city*.

Passengers take multiple decisions along the access-journey dimension, which are dominated by the choice of access-mode, or combinations thereof. Depending on the mode(s) chosen, there is the additional choice of a route, while, for journeys involving car, there is often a choice to be made between self-drive and drop-off, and, in the former case, a choice between different parking options. Additionally, passengers do make a choice of departure time, which, although dependent on personal preferences, is highly influenced by the departure time of the actual air-journey.

Destination-side

In many ways, the destination-side choices are the mirror-image of those made at the origin-side. Aside from the actual choice of ground-level destination[7], these include the choice of destination airport[8], and ground-level transport between this airport and the final destination.

However, there are some subtle differences. Indeed, from the point of view of a passenger on the outbound leg, there is in general an issue of a lower level of knowledge than at the origin-side, relating partly to the geographical location of the different possible destination airports, but also to choice set formation along the egress-journey dimension, in terms of ground-level transport modes, as well as routes. Here, another point needs to be taken into account in that, for the majority of travellers, private car is not an option at the destination end, but is, for at least some of these travellers, replaced by rental car.

[6]Although not central to this part of the discussion, it should be noted that these potential differences between residents and visitors call for a separate treatment of the two groups of travellers, in addition to any segmentation along the purpose dimension.

[7]Which is a sub-choice of the more general choice of destination in Section 8.2.1.

[8]Again, the choice set here depends heavily on the actual destination, where there is, in some cases, only one realistic option in terms of destination airport.

Aside from the above discussion about different choice set formation in the ground-level dimension, the point about a lower level of information would suggest a less rational behaviour from an outside perspective[9], except for the more regular traveller. Additionally however, it should be noted that the set of priorities at the destination end may be different from those at the origin end, for example in terms of a higher reluctance to accept long ground-level journeys than might be the case for the departure end.

Air-side

Except for the questions of origin and destination, and ground-level transport, the air-side category contains all remaining choices describing the journey. Aside from *spontaneous* choices made at different stages of the journey (such as what to do while at the airport), these choices all relate to the actual travel from the origin airport to the destination airport. Apart from the choice of a specific class or travel[10], these can in turn be subdivided into three very much interrelated dimensions of choice.

The first choice is that of an airline operating a route to the chosen destination. In most cases, passengers will travel on a single airline for the duration of their journey. However, on some routings, there is the possibility of a combination of airlines[11]. The choice of an airline is one of the factors that makes air-travel different from other areas of transport analysis, given the importance of the carrier choice dimension, which is less prominent with other modes, such as rail.

The choice of an airline or combination of airlines is strongly related to the choice of a routing. The first level along this dimension of choice divides flights into direct and connecting flights, with the possibility of a third category, for flights involving a stopover without a change of aircraft. The second level applies only to connecting flights, and involves the choice between a number of different possible routes, which includes a decision on the number of connections, and the choice of connecting airports.

The final dimension of choice for the actual air-journey is that of timing, i.e. the choice of a departure time and a departure date, which is again strongly interrelated with the *upper-level* choice of timing in Section 8.2.1. For some passengers, the most important factor will be the departure time, while for others, it will be the arrival time. In practice, this equates to the choice of a specific flight.

There are ways to consider further subdivisions of choice-dimensions in the air-side category. However, in general, such factors, which include for example the type of aircraft, can in fact be seen as an attribute of a specific flight, which is thus accounted for by the other dimensions of choice listed above, and as such, can be included in models as a simple explanatory variable.

[9] This does not per se suggest irrational behaviour. It simply means that, had the traveller been in possession of all information, he might have been expected to behave differently.

[10] While passengers make a choice between *cabin-type*, the additional choice of *fare-type*, in terms of restrictions, is in general less important, and depends heavily on availability and hence time of booking.

[11] The situation has in recent years increased in complexity, given that a large share of routes are now operated under *code-share* agreements.

8.2.3 Choice processes in air-travel

The above discussion has illustrated that the process of putting together a trip from a ground-level origin to a ground-level destination, with an intermediary air-journey, is a complicated one, involving decisions along a multitude of dimensions. What makes the analysis of such processes even more complicated is the fact that there potentially exists a highly complex structure of interdependencies between the various dimensions of choice, which is likely to vary across individuals as well as across situations. The aim of this section is to briefly look at the main interactions, as well as touch on some less-obvious ones[12].

It should be clear that the *upper-level* decisions, in terms of destination and trip timing, have a strong influence on the air-travel specific decisions, such as the choice of departure airport, airline, and routing, on the basis that not all destinations are served from all airports and by all airlines. Additionally, the link between the choice of destination and the choice of destination airport needs no further explanation. A similar, though less strong reasoning applies in the case of timing, where the choice of a specific departure date or time will have an influence on the choice set in terms of departure airports, airlines, flight-routings, and even destination airport.

At the same time, it can be seen that the various dimensions of choice at the actual *air-travel* level are also strongly inter-related. As such, the choices of departure and destination airport, airline(s) and routing[13] all depend on each other, such that the choice of a specific airline can for example limit the choice set in terms of possible departure airports, and vice-versa. Here, it is not immediately obvious which of the decisions takes priority, and the order may indeed vary across travellers[14]. In the context of the description of the dimensions of choice in Section 8.2.2, it can clearly also be seen that the choice of a specific departure or destination airport can have an influence on the choices taken along the ground-level choice-dimensions, where it is also possible for these dependencies to act in the other direction[15]. On the basis of this discussion, it can be seen that the *upper-level* choices, which have an impact on the choice of departure and destination airport, also have an indirect impact on the choices along the ground-level dimensions.

The above two paragraphs have described the effect of *upper-level* choices on *lower-level* choices, and the inter-dependencies between *lower-level* choices. At this point, it is worth noting that there may also be cases where the *lower-level* choices take precedence over the *upper-level* choices. As such, it is possible that some leisure travellers make their choice of destination dependent on their choice of departure airport or airline, a principle that can be seen to apply especially in the context of

[12] Given the complexity of the choice processes, this description is based on a number of non-trivial hypotheses, excludes certain possibilities, and should in no way be seen as a *definitive* description of the inter-dependencies between choice-dimensions for air-travellers.

[13] I.e. direct or connecting, with the additional choice of connecting airport(s).

[14] Some travellers may be captive to a certain airline, while others may be captive to a certain airport.

[15] Here an example comes in the case of travellers who rely on public transport for their ground-level journey, which can, in conjunction with their ground-level origin, eliminate some airports from consideration. In fact, it can be suggested that, in some regions, the high allegiance by passengers to a given mode of transport, principally car, leads to a higher probability of accepting a change of airport or airline than a change of access-mode.

regional airports, and low-cost airlines[16]. While such *upwards* effects can be seen to apply primarily in the case of leisure-travellers, it can also be argued that they play a role for business travellers, for example in the case where the choice of destination for a meeting is taken conditional on other factors such as flight availability, especially in the context of meetings taking place at airports[17]. The detailed exploration of such *upwards* interactions is beyond the scope of this thesis, but is an important area for future research.

8.2.4 Discussion

The above description of the choice processes undertaken by air-travellers has shown that such journeys not only involve decisions along a multitude of choice-dimensions, but that there exist complex inter-dependencies among choice-dimensions. Given that a number of these dependencies potentially act in both dimensions, it is clearly inappropriate to attempt to model the decision-making as a sequential choice process, but rather, that simultaneous analysis is required in the absence of information on the relative level of priorities across travellers[18].

One final point needs to be addressed. Indeed, the discussion so far could suggest that the question of main mode of travel is taken at a separate level. As such, travellers would gauge the overall product offered by the various modes, and then make a choice of main mode before moving on to mode-specific choices (e.g. airline, route, ...). Clearly, for some journeys, a mode choice decision is taken a priori, such that the above discussion holds[19]. However, there are situations where such a straightforward approach does not apply. In the present context, aviation, the problems arise in the case of short to medium distances, where the competition from car and especially rail needs to be taken into account[20]. In such cases, it is not necessarily clear whether passenger make an a priori mode choice between air and rail, before, if applicable, making within-mode decisions in the case of air-travel. Rather, it can be imagined that, for at least part of the travelling population, the alternative modes, such as high-speed rail, appear on the same level as the various air-travel alternatives. As such, the alternatives are evaluated in parallel, with all intra-modal considerations taken into account at the same time as the cross-modal comparison. Although such a parallel analysis does not pose any major problems from a methodological point of view (in terms of model structure), it

[16]These operators can be seen to induce new demand, such that some of their passengers would not travel at all (or at least not by air), if it wasn't for the presence of the specific airline.

[17]It is interesting to note that airports are diversifying their products, and no longer merely transport hubs, they are turning into shopping and business centres.

[18]The use of such information is highly hypothetical in any case, as it is not clear whether there are situations in which it is possible to define a clear sequential choice process involving decisions along all of the above listed dimensions.

[19]As in the case of long-haul travel, where the advantages of air as a mode outweigh all other factors.

[20]A timely case in point is the competition between air and high-speed rail on medium-distance travel in Europe, where improved rail services have led to the closure of certain air-routes, such as Paris to Brussels (cf. Kerlouegan & Gelie 2001) and Paris to Grenoble (cf. Cabret 2004), while on other routes, such as Paris to Marseilles, London to Paris and London to Brussels, a veritable *fare war* has erupted between air and rail, and air has lost a significant part of its market share to rail (cf. Baret 2001).

does come at the cost of increased data requirements, where it is now necessary to obtain detailed data for ground-level modes, in addition to the air-travel data, the procurement of which already causes problems on its own, as described in Section 8.4.4. Additionally, it should be noted that for some routes, the number of possible ground-level options (in terms of combinations and routes) is so high that the data requirements can become insurmountable. As such, it comes as no surprise that the majority of studies make an assumption of an a priori decision to travel by air. While this does not per se invalidate the results of the analyses in question, it is important to acknowledge the possible shortcomings, at least in the presence of routes where there is potentially high inter-mode competition which is not characterised by an a priori choice of mode before proceeding to intra-mode decisions. This issue is touched upon again in Section 8.4.1.

8.3 Literature review

8.3.1 Introduction

This section presents a comprehensive review of existing research on the modelling of air-travel choice behaviour. The review centres on work in the *academic* domain. There is also a large body of work of a less independent nature, conducted or commissioned by governmental organisations. Aside from not always being widely accessible, such work generally makes use of more basic model structures, albeit inside very complex forecasting systems, and is thus of lower interest in the present context. The review is divided into several parts, according to the scope (and in some cases geographical context) of the different research projects.

8.3.2 Air as a modal alternative

A number of discrete choice studies have used air as one of the alternatives in a wider mode choice analysis, without looking in detail at elementary air-travel choices (i.e. airport, airline, ...). Such work relates to the discussion in Section 8.2.4 with regards to the choice of main mode.

One example of such a study is given by Bhat (1995), who looks at the choice of mode for business travellers in the Toronto-Montreal corridor, using the Heteroscedastic Extreme Value (HEV) model, which allows for different scale parameters across alternatives. The same dataset was also used by Bhat (1998*a*) in a study looking at variations in tastes across respondents, and by Wen & Koppelman (2001), in an application comparing various GEV structures of different nesting complexity. Another analysis of mode choice behaviour, where air is one of the available alternatives, is conducted by Mandel et al. (1997), who show the advantages of using non-linear formulations for travel-attributes with the help of Box-Cox transforms in MNL models, and highlight the significant impacts of the utility specification on forecasting in terms of greatly different market shares for the high speed rail alternative. Finally, González-Savignat (2004) estimates a MNL model on SP data for the choice between air and a hypothetical high-speed rail alternative in Spain, producing forecasts which show that, on journeys with train times up to three hours, the introduction of high speed rail services can generate a significant shift of travellers away from air-travel.

8.3.3 Airport choice

Given the high number of existing studies of airport choice, the discussion of previous work is arranged by geographical context, grouping research into three sets; studies conducted in the United States (relating to the case-study in Chapter 9), studies conducted in the United Kingdom (relating to the case-study in Chapter 10), and studies conducted elsewhere.

Studies of airport choice in the United States

The number of studies of airport choice in the United States is much larger than in other areas; this is at least partly due to the greater availability of appropriate passenger-survey data, which is less governed by commercial considerations than in other areas.

One of the first studies of airport choice was conducted by Skinner (1976), who uses a MNL model for airport choice in the Baltimore-Washington DC area (3 airports). This study reveals significant effects of flight frequency and ground accessibility, with travellers being more sensitive to the latter. Windle & Dresner (1995) also use a MNL model in this area, and find significant effects for flight frequency and airport access time. The results also show a high level of significance for a repetitive choice dummy variable; the more often a traveller uses a certain airport in a year, the more likely the traveller is to choose the same airport again. A later MNL study in this area is conducted by Pathomsiri et al. (2004), with broadly similar results.

Lin (1977) uses a very basic binary choice model in an analysis looking at airport choice in a low demand region in the North of New York State, near the border with Canada, where passengers have a choice between several small regional airports, but are heavily influenced by the presence of Montreal's major airport across the Canadian border. The study finds that the international boundary plays an important role in choice behaviour, but that, otherwise, a majority of trip makers are willing to travel considerable distances by ground in order to depart from an airport with better service, such as for example higher frequency.

A large number of studies of airport choice (and related aspects) have been undertaken in the San Francisco Bay (SF-bay) area. Harvey (1987) uses a MNL model for airport choice, and finds that airport access time and flight frequency are significant for both leisure and business travellers, with lower values of time for leisure travellers, and with all passengers preferring direct flights over connecting flights. More recently, Pels et al. (2001) have used a NL model for airport and airline choice in the SF-bay area, showing that for both business and leisure travellers, the choice of airline should be nested within the choice of airport. Pels et al. (2003) model the joint choice of airport and access mode in a NL model, with airport choice at the top level and access mode choice at the lower level, showing high values of time, especially for business travellers.

The analysis by Basar & Bhat (2004), who look at airport choice in the SF-bay area, differs from other studies in that it uses a two-stage model as proposed by

Manski (1977); in the first stage, the choice set is generated[21], while in the second stage, a choice of airport is made from this choice set. This study thus acknowledges the fact that not all airports are considered by all travellers; as the inclusion of unconsidered alternatives in a choice set can lead to biased results (see for example Williams & Ortúzar 1982), the decision to incorporate choice set generation is thus certainly warranted. The results obtained with the parameterised choice set consideration model (PCMNL) show that flight frequency is the most important aspect in choice set composition, surprisingly dominating the also significant access time factor. In terms of the actual choice of airport, after elimination of non-considered airports, access time is the most important factor.

Studies of airport choice in the United Kingdom

There have also been a number of studies of airport choice in the United Kingdom; the difficulty involved in securing appropriate data has however somewhat limited this number in recent times.

A frequently cited example is that of Ashford & Bencheman (1987), who use a MNL model for airport choice at five airports in England (Heathrow, Manchester, Birmingham, East Midlands and Luton), and find that access time and flight frequency are significant factors. In addition, air-fares play a role for all domestic passengers and for international leisure travellers.

Ndoh et al. (1990) compare MNL and NL models for passenger route choice in central England and find that the NL model is superior. The modelling results suggest that it is preferable to nest the choice of route type above the choice of hub airport, and the choice of departure airport (see also Caves et al. 1991).

Thompson & Caves (1993) use a MNL model to forecast the market share for a new airport in North England; access time, flight frequency and the number of seats on the aircraft[22] are found to be significant, with access time being most important for travellers living close to the airport and frequency being more important for travellers living further afield.

In another study which includes one of the five London airports (Heathrow), Brooke et al. (1994) use MNL models in the analysis of passenger distribution between airports in an area centring on the Midlands, finding flight frequency to be most important attribute, followed by access time.

Studies of airport choice in other areas

Ozoka & Ashford (1989) use a MNL model to predict the effect of building a third airport in a multi-airport region in Nigeria and find access time to be significant, such that the choice of location plays an important part in the success of a new airport, along with the provision of good ground-access facilities.

Innes & Doucet (1990) use a binary Logit model to predict the choice between airports in Canada, and find that the type of aircraft plays a role; travellers have a higher desire for jet services than turboprop services, suggesting that the quality of service (journey time and comfort) provided is very important. The results also show a preference for direct flights over connecting flights.

[21] With seven possible choice sets, on the basis of the three airports used in the study.

[22] Reflecting comfort, but possibly also capturing visibility and availability effects.

Furuichi & Koppelman (1994) use a NL model for departure airport and destination choice for passengers on international routes from Japan, and find significant effects of access time, access journey cost and flight-frequency. Here, the choice of departure airport is nested within the choice of destination, hence also acknowledging the effect of choosing a specific destination on the choice set of departure airports. The authors suggest that the use of the NL model is only made possible by the strong relationship between departure airports and specific international destinations in this dataset, and hint at the fact that if this was not the case, CNL structures would have to be used.

Mandel (1998) uses Box-Cox transforms in a MNL framework looking at modelling the competition between airports in Germany, with the specific aim of forecasting demand changes at airports following changes in service frequency and fare levels, where the modelling framework also takes into account the interaction of multi-modal, inter-modal and intra-modal effects. The model is applied to several scenarios, including the introduction of new routes, and the development of a secondary hub.

Veldhuis et al. (1999) produce the comprehensive Integrated Airport Competition Model (IACM), which uses a sequence of NL choice processes that model the choice of main mode (e.g. air, train,...), followed by the choice of air route (i.e. direct vs indirect), the choice of airport, and finally the choice of access-mode at the chosen airport. They apply this model to Amsterdam's Schiphol airport, where competition with other airports is allowed for by acknowledging the effects of airports in the wider surrounding area. The model is further used by feeding logsum terms from the top of the tree (main mode) into an elasticity based propensity-to-travel-long-distance model that is destination-specific. The aim of this work is to develop a model system that is readily transferable to other airports in Europe, as also discussed by Kroes et al. (1994). In other work looking at Schiphol, Ashley et al. (1995) develop a tool for forecasting traffic at the airport, which can predict the effects of policy changes on demand at the airport.

Suzuki et al. (2003) look at the issue of airport leakage, that is, the phenomenon of travellers avoiding their local airport, and giving preference to larger airports which are further away from their ground-level origin. The analysis calibrates a MNL model on survey data collected in Iowa in 2001, showing that leisure travellers are more likely to leak to the larger airports, while the utility of an airport seems to be related to the quality of service experienced previously by the traveller at this airport. However, even in the case of negative experiences, a traveller is still more likely to choose the airport than one where he has no experience at all, ceteris paribus.

8.3.4 Airline and fare-class choice

Kanafani & Sadoulet (1977) use a MNL model to model the choice among fare types for passengers on long-haul journeys. To counter the independence assumptions of the MNL model, the observed utility for alternative i contains attributes of similar alternatives, as in a *Mother Logit* type model. The model is applied to aggregate North Atlantic vacation traffic data, with results showing significant impacts of relative fares, but also strong seasonal variations.

Proussaloglou & Koppelman (1995) use a MNL model for the choice of airline on travellers' most recent air-trips, where the data were collected through a mail-in survey. The strongest impact on the utility of an airline was found to be the membership in its frequent flier programme, where the effect is even more significant for *very active* members. Other factors that increase the attractiveness of an airline include the convenience of the schedule, low fares, on-time reliability, and market presence by a carrier.

Chin (2002) uses binary Logit and Probit choice models in an analysis looking at the effects of frequent flier programmes, and finds that they have a positive effect on choice probabilities for the associated airline, where this is however not as significant as the effect of scheduling convenience.

8.3.5 Access-mode choice

Harvey (1986) looks at the choice of access-mode for journeys to the SF-bay area airports, and finds that, as expected, journey time and cost are the strongest determinants in this choice process.

One of the crucial forecasting scenarios in access-mode choice research is the introduction of a new mode. This requires special model structures, taking into account the likely correlations with existing modes to acknowledge the differential substitution effects. Bates et al. (1987) look at this issue in the case of the introduction of a dedicated railway service for London's Heathrow airport, which is now operated as the *Heathrow Express*. They propose the use of an iterative version of the NL model and discuss how this can be applied in the case of airport-access modelling.

Bondzio (1996) uses NL models for the joint choice of access mode and departure airport in Germany. Interestingly, the optimal structure for business travellers was found to be one nesting access mode choice above departure airport choice, showing the importance of the access-mode dimension. This is the opposite of the generally used approach. On the other hand, for leisure travellers, the NL models did not lead to significant improvements over a MNL structure.

Monteiro & Hansen (1997) use MNL and NL models to forecast the impact of the now completed expansion of BART[23] to San Francisco International (SFO) airport, predicting a slight strengthening in the dominant position of SFO in the SF-bay area.

Psaraki & Abacoumkin (2002) use a MNL model in conjunction with clustering analysis to predict ground access modal split at the new Athens airport. The results show for example that the attributes of parking options at the airport play a significant role in the choice of access mode.

8.3.6 SP data

The majority of studies of air-travel choice behaviour rely solely on the use of RP data. While this avoids problems with regards to issues of understanding of hypothetical choices, it also means that the results are in many cases affected by rather strong assumptions which are required because of the quality of the (level-of-service)

[23]Bay Area Rapid Transit.

data, especially in the case where the data comes in the form of survey data collected directly from passengers. Fewer issues arise in the case where actual bookings data are available, but this is rarely the case.

Nason (1980) demonstrates how the MNL model can be used to analyse the choice between different fare classes, using SP data looking at a binomial choice between a full-fare ticket, with a guaranteed seat, and a hypothetical new standby ticket.

Another example of an application using SP data is given by Bradley (1998), who uses binary Logit models in the analysis of the choice of departure airport and route, with data collected from passengers at Schiphol, Brussels, and Eindhoven airports. The most significant impact on choice behaviour is found to be air-fare, where a log-transform was used, and where differences exist across different groups of travellers. Other factors with significant effects include access time and transfer time, in addition to a dummy variable associated with connecting flights.

Proussaloglou & Koppelman (1999) use a telephone survey resembling a booking process, for passengers from whom information about actual trips had previously been collected. Respondents then made a choice of carrier, flight and fare class for their specific route. The results show negative impacts of fare, especially for leisure travellers, as well as for schedule delay, with positive impacts for frequent flier programmes. Similarly, increased market presence of the carrier, and quality of service had positive effects. The results suggest that business travellers are willing to pay a premium of \$21 to travel on an airline for which they hold a frequent flier account. These values increase in the case of the airline in whose programme they participate most actively, with valuations of \$52 for *low-frequency travellers* and \$72 for *high-frequency travellers*. A similar pattern is observed for leisure travellers, although the willingness to pay is much lower, at \$7, \$18 and \$26 respectively.

Algers & Beser (2001) discuss the modelling of the choice of flight and booking class. They acknowledge the limitations of RP data in this context[24], but also stress that issues with SP bias need to be borne in mind. As such, they propose to use both RP and SP data in the analysis, with the RP data being used to correct the scale of the utility function obtained with the SP data. Given that complications regularly arise with the use of RP data (cf. case-studies in Chapters 9 and 10), the use of a combination of RP and SP data is indeed an important avenue for future research[25], as the estimates obtained on the basis of SP data alone are not reliable.

Hensher et al. (2001) use SP data for airline choice between New Zealand and Australia, but focus primarily on methodological issues (survey design). Nevertheless, they find interesting effects, such as a for example a significant positive influence of frequent flier programmes.

The study of Adler et al. (2005) is of special interest, given that they use the same SP data used in the case-study discussed in Chapter 11. They show positive effects of airline and airport allegiance, and better on-time performance, with negative

[24]In this context, the issue is not so much one of availability, as the data come in the form of bookings data, rather than survey data. However, the study by Algers & Beser (2001) is interested in the behaviour of passengers in the case where the *desired* ticket class is not available; here, SP data have an advantage in terms of allowing for insights into changes in behaviour under different hypothetical scenarios.

[25]Independently of whether the RP data come in the form of bookings data or survey data.

effects for fares, flight time, access time, and connections (see also Section 11.1).

8.3.7 Other air-travel applications

Outside the choice of airport, airline and access-mode, choice models have also been used in other areas of air-travel behaviour research.

A very important aspect of air-travel choice behaviour, especially from the airlines' point of view, is the rescheduling, standby, and no-show behaviour of passengers, an issue that is discussed in detail by Garrow (2004). The approach is described by Garrow & Koppelman (2004*b*,*a*), where the first publication is limited to MNL models, while the follow-up additionally uses NL models, and shows that the exogenous sampling maximum likelihood (ESML) estimator, which is necessary due to the weighted nature of the data, can also be used with NL models for choice-based samples in conjunction with an adequate transformation of the estimated constants.

The modelling of route choice in air-travel is a very complex undertaking, mainly for data reasons, meaning that it is often necessary to rely on aggregate level data. As such, Coldren et al. (2003) use aggregate Logit models for modelling the market share of different itineraries, finding that, as expected, itineraries with stop-overs and connections are less popular than non-stop flights. Other effects include the quality of the connections, the type of aircraft, and the time of day. This work is later extended by Coldren & Koppelman (2005) to the case of more general model structures, including multi-level NL models, and NL approximations to cross-nesting structures (referred to as Weighted Nested Logit). The results show that the nesting structures reject the MNL structure, indicating that correlation (and hence heightened competition) exists between itineraries sharing a common carrier or with departure times that are close to each other.

Wei & Hansen (2005) use aggregate NL models for the market share and demand in non-stop duopoly markets. They use 13 separate markets (i.e. routes), and estimate the models on quarterly data collected over a period of 10 years. The upper level contains the decision to travel, with the lower level containing the choice between the two airlines. The authors find that frequency is a better tool for increasing market share than increases in aircraft size, an observation that can possibly be explained on the grounds of smaller gaps between adjacent departure times. Additionally, they find that airline market share is super-proportional to frequency share, which can be seen as a reflection of the notion that frequency increases visibility.

As mentioned repeatedly in this thesis, the absence of information on the availability of specific fare classes at the time of booking leads to major problems in model estimation. In the rare cases where such information is available, it is important to take it into account. However, this will often only be possible at an aggregate level, given further data limitations, as discussed by Battersby (2004), who uses concepts of expected utility and finds that expected seat-availability has a positive impact on the choice probability of a given flight class.

8.3.8 Summary

The review presented in this section has shown that there exists a large body of work on the modelling of airport choice in multiple-airport regions, while there is also a substantive number of studies looking at airline choice and access-mode choice.

The discussion has also highlighted that, while some of the studies have made use of the more advanced models that are available, the majority of research has relied on fairly basic modelling techniques, with a heavy bias towards the MNL model. Additionally, existing research has generally used major simplifications of the choice process along at least one of the dimensions of choice, making the explicit joint analysis of the three dimensions of airport, airline and access-mode choice an important avenue for research.

In terms of actual substantive results, a rather consistent pattern emerges. Almost universally, the studies show that access time and flight frequency play a determining role in air-travel choice behaviour. Here, it should be noted that the marginal utility of frequency is a very arbitrary concept, as it is not taken into account per se by travellers[26], but can be seen to capture a variety of factors, including a visibility effect, and an approximation to schedule delay (under the considerable assumption of a relatively even spread of departure times.

Only a small subset of the RP studies were able to recover a significant impact of air-fares, while no such problems were observed in the SP studies, which generally show fare to be one of the strongest determinants of choice (e.g. Bradley 1998). Additionally, in such studies, it is often possible to retrieve significant effects of factors such as airline allegiance, a treatment of which is again generally not feasible in the case of RP data. This observation highlights an important issue with the use of RP data[27] in the analysis of air-travel choice-behaviour, which is discussed in more detail in Section 8.4.4.

8.4 Overview of research

Three separate case-studies are described in this thesis, using data from two RP surveys, collected in the SF-bay area (Chapter 9) and Greater London (Chapter 10), and data from an internet-based SP survey collected in the US (Chapter 11). In the discussion that follows, we set the stage for the description of these three case-studies. We first look at the scope of the applications (Section 8.4.1), before setting out the aims of the research (Section 8.4.2), and discussing the choice of model structure (Section 8.4.3). Finally, we highlight a number of issues that had to be faced in the modelling analyses (Section 8.4.4).

8.4.1 Scope of applications

While the overall approach used with the three datasets is very similar, there are some differences between the three datasets that change the scope between case-studies. Detailed description of the three datasets are presented in the respective chapters; the aim of the discussion that follows is simply to describe the approach taken along the various dimensions of choice with the separate datasets.

Although the discussions in Sections 8.2.2 and 8.2.3 have highlighted the fact that air-travellers take decisions along a multitude of dimensions, with potentially

[26] With the possible exception of travellers on very flexible tickets, who can just turn up at the airport and pick any flight, and as such are interested in headway.

[27] Again, this applies principally in the case of survey data, as opposed to bookings data.

important interactions between these various dimensions, in practice, it is generally only possible to look at a subset of these choice-dimensions. This is mainly a reflection of the formidable requirements in terms of data that would arise in the case of a study looking jointly at all the choice-dimensions described in Section 8.2.2. However, it should also be noted that, for an adequate analysis of the interactions between choice-dimensions, dynamic model structures would almost certainly be required, in conjunction with repeated choice data.

The three case-studies described in this thesis look at the joint choice of airport and airline, where the two RP studies additionally look at the choice of access-mode. While the detailed study and modelling of the interactions between choice-dimensions is an important avenue for future research, it should be clear that this is a learning process, and that, before attempting such an analysis, there is a requirement to first look at the joint modelling of even a subset of these choices. As such, the work described in this thesis is an initial stepping stone in the development of a more accurate framework for analysing air-travel choice behaviour.

We will now briefly revisit the various dimensions of air-travel choice behaviour, and discuss their treatment in the three case-studies.

Upper-level choice dimensions

Upper-level choices, such as the decision to travel, the choice of destination, trip timing, and the decision to travel by air, are not modelled in any of the three case-studies. In the SP case-study, these issues do not apply, and as such, any upper-level choice-dimensions can be safely ignored. In the RP studies, the assumption was made that the decision to travel and the choice of destination are taken a priori, as the analysis of these two dimensions would have required the use of a destination choice model and a trip generation model respectively, which, aside from causing significant problems in terms of data needs, is beyond the scope of the present research. Similarly, as respondents have been observed to travel on a given day, the choice of travel date cannot be modelled, with a similar reasoning applying for timing, where no information is available on desired departure time. The main problem arises in the treatment of the choice of main-mode. Here, the inclusion of a number of short-haul and medium-haul destinations[28] means that a non-trivial number of respondents potentially had the possibility of using ground-level modes of transport as an alternative to air-travel, where, for data reasons, this upper-level choice cannot be modelled in the present context. This in turn leads to the requirement to work on the notion that any passenger included in the study has already taken the decision to travel by air, either a priori, or by comparing the air-travel options to those on other modes. In either case, the traveller clearly still faces a choice between different air-travel options, and the modelling of this choice process is the topic of the present work. For the purpose of the two RP analyses, the non-air alternatives can in fact be seen as never having been chosen by any of the travellers included in the data, and as such, are excluded from the analysis. Either way, the additional analysis of the choice of main-mode, in parallel to the air-travel specific choices (or even as an a priori choice), remains an important avenue for future research.

[28] The inclusion of such destinations was inevitable, given their high weight in the choice data.

Departure & destination airports

This research looks at the choice of departure airport on a single leg of an air-journey, for passengers departing from multi-airport regions. With the exception of several destinations in the SF-bay area study (Section 9.2.1), passengers faced a choice between more than one airport only at one end of their journey, and accordingly, the choice between airports is modelled only in the main study city, leading to a choice between departure airports for the outbound leg of resident passengers, and for the return leg of visiting passengers[29]. Except for those passengers travelling on an open-jaw ticket, the choice of departure airport on the return leg can be seen to equate to the choice of destination airport on the outbound leg.

While it would be desirable to look jointly at the choice of departure and destination airport, this is hampered by the relatively low number of routes where passengers have an actual choice of airport in the origin and destination area, and where the two choices are independent. Additionally, such a modelling approach would lead to significant increases in data requirements. Finally, a major issue would arise in terms of deciding whether either of the two choices (departure or destination airport) takes precedence, where the order of preferences potentially varies across respondents[30].

Access & egress journeys

Ground-level journey related choices are only modelled in the RP case-studies, given that this dimension of choice was not represented explicitly in the SP survey. For data reasons, only the choice of access-mode to the departure airport on the current leg was modelled in the two RP studies. An additional simplification arises in the analysis of the access-journey choices in that the studies look only at the choice of main-mode, and ignore the possibility of trip-chaining, as well as the choice of different routes. The effects of this restriction are very limited in the case of the SF-bay area study, given the high market share for car, while in the London study, the possibility remains open to re-estimate the models with more detailed level-of-service data, where issues of route choice however need to be addressed.

Routing

In the case of flight routing, an important difference arises again between the two RP case-studies and the SP case-study. In the former two, the topic is left untreated, and the studies look only at the choice between direct flights, a decision based on the low share of connecting passengers in the data, and the lack of detailed data on connecting flights. The effects of the elimination of alternatives with connections from the choice set are negligible, given that, for the specific set of destinations used, the real-world share for connecting flights was comparatively low (and in some cases zero). In the SP dataset, connecting flights are included in the choice set, allowing for an analysis of the relative valuations of direct and indirect flights. However, the choice of air-routing is not modelled, where, given the complexity of the task,

[29] Only in the RP studies.

[30] This complication is also the basis for the attempts to exclude destinations located in multi-airport regions from the analysis, as discussed in more detail in Section 8.4.4.

and the high data requirements, this process is commonly modelled at an aggregate rather than disaggregate level, as for example in the work of Coldren et al. (2003).

Airline

Aside from the very first SF-bay area models, the three case-studies all acknowledge the fact that passengers make a choice of airline in addition to the choice of airport. However, no combinations of airlines are allowed for, given the use of direct flights only in the RP studies, while the connecting flights used in the SP survey only involved a change of plane, and not airline.

8.4.2 Aims of research

As alluded to in Section 8.1, the overall aim of the three air-travel case-studies conducted in this thesis is to attempt to at last partly bridge the gap between the state-of-the-art in discrete choice modelling and the state-of-practice in air-travel behaviour research. This is achieved through the use of advanced model structures, but also through attempts to more adequately represent the true nature of the choice processes undertaken by air-travellers. Several major sub-aims can be identified, and these are described hereafter.

Recognise the multi-dimensional nature of the choice process

It should be clear from the discussion in Section 8.2.2 that air-travellers take decisions along a multitude of choice-dimensions. As discussed in Section 8.4.1, the three case-studies all look at the combined choice of departure airport and airline, where the two RP studies additionally look at the choice of access-mode. A main aim of this research is to define how these various dimensions of choice can be treated in parallel, as opposed to using a sequential approach. This applies especially in the case of the two RP studies, where the choice set for each individual is defined manually, unlike in the SP study, where it is explicitly defined by the survey design. As such, the alternatives chosen by respondents in the RP case-studies are defined as *combined* alternatives, where each such alternative is made up of *three* elementary alternatives (airport, airline and access-mode), and where the utility function for a *combined* alternative is made up of joint terms as well as terms specific to the three *elementary* alternatives. This approach enables us to model the three choices simultaneously as opposed to sequentially, where such a simultaneous treatment is preferable in the absence of information on the priorities of the three dimensions of choice. In a way, the approach used in combining alternatives thus turns a three-dimensional choice process into a single-dimensional one.

Analysis of variations in choice behaviour

One of the main aims of this study is the analysis of variations in choice behaviour across respondents, in the form of variations in tastes between separate population segments, as well as within separate population segments. The main advancement of the state-of-practice in this case comes in the exploitation of model structures allowing for random variations in tastes across respondents, which are only slowly beginning to be applied in the area of air-travel[31]. Another innovative method used in the analysis of taste variations in the present work[32] is that of continuous interactions between taste coefficients and socio-demographic or trip-related attributes.

Use advanced structures for representing complex substitution patterns

By understanding the multi-dimensional nature of the choice process, it becomes evident that some of the *combined* alternatives share the attributes of other alternatives along one or more of the choice dimensions[33]. While some of these commonalities can be accounted for through the attributes included in the observed part of utility, it is almost inevitable that there is also some correlation between unobserved utility terms along these dimensions. Classically, research in the area of air-transport has accounted for this correlation with the help of multi-level NL structures. As described in Section 8.4.3, it can be seen that with such structures, it is not possible to account jointly for the correlation along all 3 dimensions of choice. As such, one of the aims and contributions of this work[34] is to make use of cross-nesting structures, allowing for the joint analysis of correlation along all three dimensions[35].

Avoid over-aggregation in level-of-service data

Another major problem with existing studies has been the use of an insufficient level of disaggregation in the level-of-service data. This is highly correlated with the decision in a lot of previous work to use simplifications along a number of choice-dimensions; as an example, the use of airport-specific attributes, as opposed to airline-specific attributes, leads to high aggregation error in the face of product differentiation across airlines. In the two RP case-studies presented in this thesis[36], the aim was always to use the highest possible level of disaggregation. As such, the only main aggregation used was to group together flights by the same airline on the same route (departure airport $\mathcal{A}$ to destination airport $\mathcal{B}$). This grouping was performed on a daily basis, where the days of week and the time of year were taken into account, leading to more detail than in approaches disregarding the day of week

[31] Here, the work described in Chapter 9, now published in Hess & Polak (2005*b*) and Hess & Polak (2005*a*), provides some of the first applications of the MMNL structure in this area of research.

[32] SP case-study, Chapter 11.

[33] In the context of the combined choice of airport, airline, and access-mode, let K, L and M define the number of airports, airlines and access-modes respectively, and let us assume that all combinations of airports, airlines and access-modes are possible, leading to a total of KLM *combined* alternatives. It can then be seen that a given alternative shares the same airport (and hence the related attributes) with $LM-1$ alternatives, the same airline with $KM-1$ alternatives, and the same access-mode with $KL-1$ alternatives. Furthermore, an alternative shares the same airport and airline with $M-1$ alternatives, the same airport and access-mode with $L-1$ alternatives, and the same airline and access-mode with $K-1$ alternatives.

[34] London case-study, Chapter 10.

[35] The correlation across dimensions, such as the correlation between different airlines, is not explored in this work, but remains an important avenue for future research.

[36] This discussion does not apply in the case of the SP study, where exact information on the alternatives was available.

or time of year. The aggregation was performed across all flight-specific attributes, such as fare, flight time and aircraft type[37]. This approach prevents an analysis of schedule delay sensitivities, but this was in any case not possible, given the lack of data on preferred arrival times. On the other hand, the aggregation yields daily frequencies, allowing for the use of this variable as a proxy for inter-departure gaps, as well as a *visibility* effect.

Conduct a study of airport choice in London

Aside from the more methodological aims, the research also has a topical aim, namely that of conducting an analysis of airport choice in the Greater London area. The Greater London area is arguably the most competitive multi-airport region in the world (cf. Section 10.1), making it a more appropriate candidate for airport choice studies than other regions used previously. Additionally, the continuing discussions with regards to how to expand capacity in this area make the study very timely. However, although some of the London airports had previously been used as alternatives in applications looking at airport choice in the wider geographical area (as discussed in Section 8.3), there has thus far not been a public domain study of airport choice among the five main London airports[38].

8.4.3 Model structure

One of the most important questions arising in the analysis of air-travel choice behaviour is the structure used for the models, and more specifically, for representing the commonalities between alternatives. This applies even more so in the case of multi-dimensional choice processes, such as those described in the two RP case-studies (Chapters 9 and 10). In this section, we look specifically at the issue of the nesting structure used in the analysis of these three-dimensional choice processes[39]. Another model-structure question is that of the representation of random taste heterogeneity; here, simple MMNL models were used, which need no further exposition at this point (cf. Section 2.9.1).

It is clearly a major and probably unwarranted assumption to rule out the presence of heightened correlation in the unobserved utility terms along any of the choice dimensions. In order for such an assumption to be valid, any commonalities between two alternatives sharing the same airport, airline or access-mode would need to be explained in the observed part of utility[40]. This is clearly not possible in general, especially in the case of RP data, which, in aviation, are often characterised by a

[37]Here, the minimum and maximum aircraft size used on a given route by a given airline was retained, where there was little variation across flights on the same airline-route pairing.

[38]There have been a number of *official* studies of airport choice behaviour in the UK (with sub-models looking at the South-East), all based on MNL models, aimed at producing systems for forecasting passenger levels. The most recent version of the Department for Transport's model, SPASM, is described by Scott Wilson Kirkpatrick (2004). It is an extension of the earlier *Second Passenger Allocation Model* (SPAM), developed by NATS (1998), which itself has several predecessors, all using MNL structures.

[39]No nesting approaches are used in the SP case-study (Chapter 11), which makes use of binomial data.

[40]The same applies in the case of alternatives differing only along a single dimension.

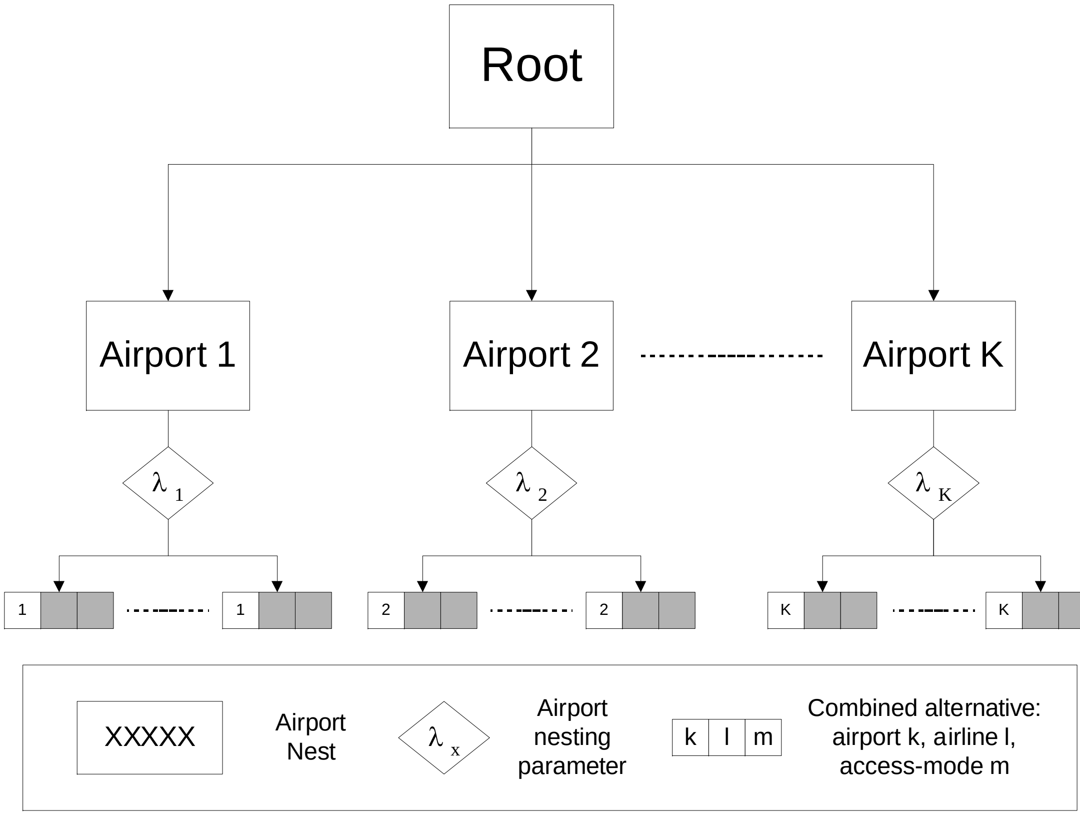


Figure 8.1: Structure of two-level NL model, using nesting along airport dimension

lack of information on crucial factors such as fares and frequent flier programmes. The likely resulting correlation in the unobserved part of utility makes the use of the MNL model almost surely inappropriate, especially in forecasting.

In this thesis, GEV structures were used for the representation of the correlation in the unobserved part of utility; while the use of ECL models might have advantages in terms of flexibility[41], the high number of error-components required to represent the complex substitution patterns makes the approach inapplicable from a purely computational perspective.

The most basic GEV nesting approach that can be used in the analysis of air-travel choice behaviour is a simple two-level NL model, where, in the context of the present research, three main possibilities arise, using nesting along a single dimension of choice (airport, airline or access-mode), with one nest per elementary alternative represented in that dimension of choice. As an example, the appropriate structure for the NL model using nesting by airport is shown in Figure 8.1, with K mutually-exclusive nests, one for each airport, and where each nest has its own nesting parameter, λ_k, allowing for different substitution patterns in the different nests. Only a subset of the composite nests and of the *combined* alternatives is shown in the graph. The same logic applies in the case of a two-level NL treatment

[41]For example by allowing for heteroscedasticity, or additional random taste heterogeneity.

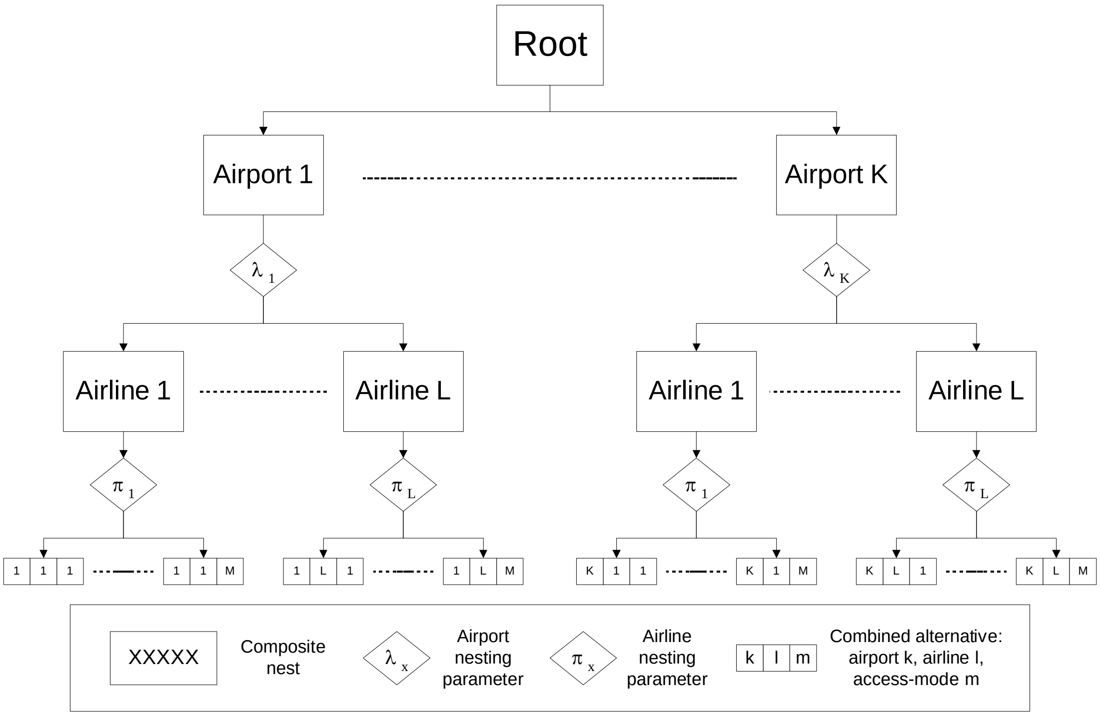


Figure 8.2: Structure of three-level NL model, using nesting along airport dimension and airline dimension

of the correlation along the airline or access-mode dimensions.

The NL model can be adapted to allow for correlation along more than one dimension, by using a multi-level structure. A common example in the case of air-travel is to nest the choice of airline within the choice of airport. It is important to stress that this should not be seen as representing a sequential choice process. Rather, it means that there is correlation between two alternatives that share the same airport, but that the correlation is larger if they additionally share the same airline. The structure of such a model is illustrated in Figure 8.2, where λ_k is the nesting parameter associated with airport nest k, and π_l is the nesting parameter associated with airline nest l. Again, only a subset of the composite nests and of the *combined* alternatives is shown.

By noting that a model nesting airport choice above airline choice is not the same as a model nesting airline choice above airport choice, it can be seen that six possible two-level structures arise in the present context. While NL structures can, in this form, thus be used for analysing correlations along two dimensions of choice, it should be noted that multi-level NL models have two important shortcomings which limit their potential for the analysis of choice processes of the type described in this work.

The main shortcoming in the present context is that the structures can be used for the analysis of correlation along at most two dimensions of choice. Indeed, using the example shown in Figure 8.2, it can be seen that, by adding in an additional level of nesting by access-mode below the airline-level, each access-mode nest would

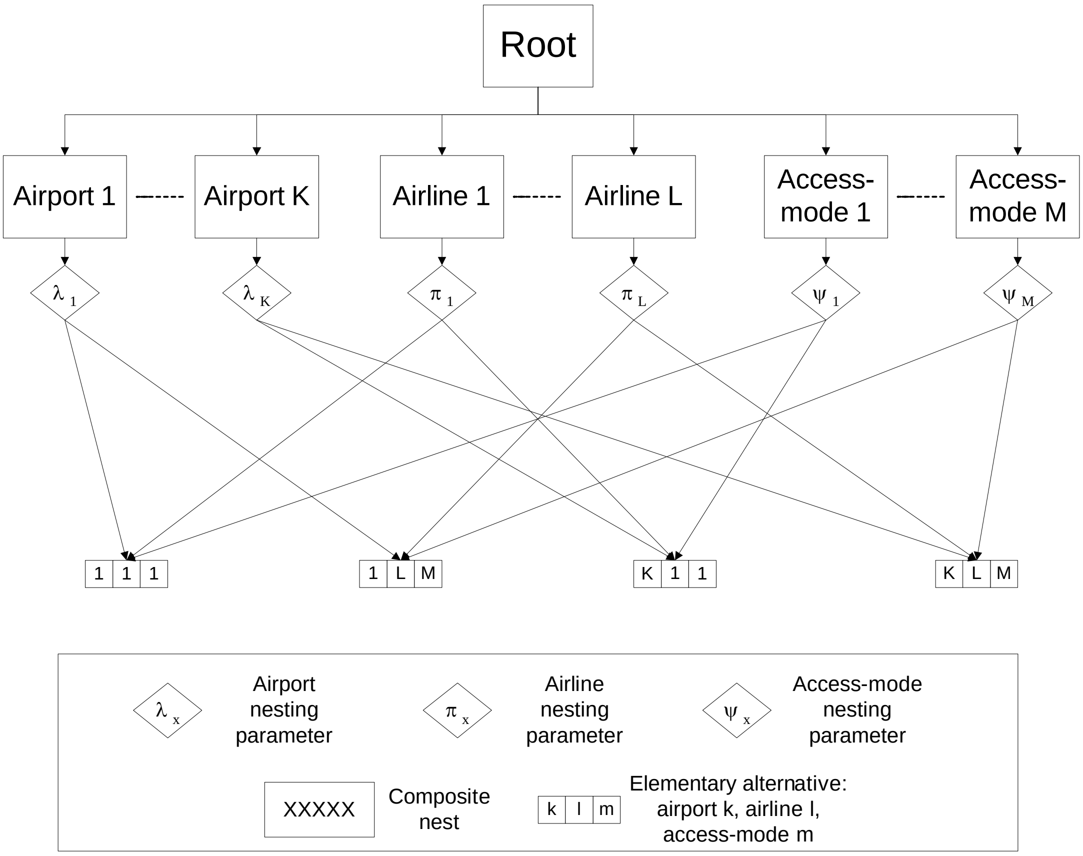


Figure 8.3: Structure of CNL model for the joint treatment of correlation along the airport, airline and access-mode dimensions

contain a single alternative, as the airline nest just above it would contain exactly one alternative for each access-mode, such that the structural parameter for the access-mode nest would cancel out. The same principle applies in the case of the other possible four-level structures, where, in each case, the lower level of nesting becomes obsolete.

While a three-level NL model can be used to analyse the correlation along two out of the three dimensions of choice, the second shortcoming of the structure means that problems arise even with this task. In fact, it can be seen that the full extent of correlation can only be taken into account along one dimension, with a limited amount along the second dimension. Indeed, by nesting the alternatives first by airport, and then by airline, the nest for airline l inside the nest for airport k will only group together the options on airline l for that airport k (cf. Figure 8.2). The same reasoning applies for other nests. As such, the model is not able to capture correlation between alternatives using airline l at airport k_1 and alternatives using airline l at airport k_2, which is clearly a restriction. This problem also applies in the other multi-level nesting approaches. Aside from being a major shortcoming, this is also another reflection of the above comment that the order of nesting matters.

These deficiencies of multi-level nesting structures are the motivation for the efforts made in the London case-study (Chapter 10) to use cross-nesting structures. In the present context, a CNL model is specified by defining three groups of nests, namely K airport nests, L airline nests and M access-mode nests, and by allowing each alternative to belong to exactly one nest in each of these groups. As such, the structure addresses both of the shortcomings described above for the three-level NL model, by being able to accommodate correlation along all three dimensions, and by doing so in a simultaneous fashion. This means for example that the model is able to capture the correlation between all alternatives sharing airline l, independently of which airport they are associated with. At the same time, the correlation will be higher between alternatives that additionally share the same airport.

An example of such a model is shown in Figure 8.3, where, in addition to the previously defined λ_k and π_l, Ψ_m is used as the structural parameter for access-mode nest m. Again, only a subset of the composite nests and of the *combined* alternatives is shown. Additionally, the allocation parameters, governing the proportion by which an alternative belongs to each of the three nests, are not shown in Figure 8.3. Here, an interesting observation can be made. Indeed, it can be seen from the above discussion that the CNL model has an advantage in this case, as it avoids the issue of the ordering of nesting levels. This makes the model similar to the PD model discussed in Section 2.5.7. In fact, the conceptual similarities between the PD model and the cross-nesting approach adopted here are further highlighted by noting that, from a structural point of view[42], each alternative is described by three attributes, or principles of differentiation, an airport, an airline, and an access-mode, where each attribute can take on a set of different values (e.g. the K airports in the case of the airport-*attribute*).

Two points merit some further attention. The above discussion has looked exclusively at nesting alternatives along one or more of the three choice-dimensions. It should be noted that an infinite number of other nesting approaches are possible. Here, one promising approach is to nest low-cost carriers *against* network carriers. Such an approach was not explored in this thesis, for two main reasons; in the SF-bay area study, only a single major low-cost carrier was included, while, in the London study, the age of the data (1996) meant that the impact of low-cost carriers was rather modest. Here, the use of a more recent version of the London data is an important avenue for future research, given the high level of activity by low-cost carriers in this region over recent years.

The other point relates to the use of GEV mixture models. While the SF-bay case-study in Chapter 9 makes use of NL as well as MMNL models, the joint analysis of the two phenomena was not explored in this thesis, given the prohibitive cost of estimation, where, even with the MMNL models in the SF-bay study, individual estimations took several days. Here, it should also be noted that the cost of estimation prevented the use of MMNL models in the London case-study in Chapter 10. These issues are an illustration of the high cost of estimating mixture models on large *real-world* problems[43], a fact that is clearly one of the reasons for the prevailing gap between the state-of-the-art and the state-of-practice. In this context, and given the discussion in Chapter 6 about the issue of confounding between simple inter-alternative correlation and random inter-agent variations in tastes, it is thus

[42] Other differences are accommodated in the observed part of utility.
[43] Despite improvements in estimation efficiency, as discussed in Chapter 3.

important to note that the individual models potentially capture a mixture of both phenomena, and as such, may overstate the extent of simple inter-alternative correlation (in the GEV models) or random taste heterogeneity (in the MMNL models). Although the present work does not attempt to correct this *bias*, previous work has, seemingly without exception, failed to mention this issue altogether.

8.4.4 Issues

Two main issues that had to be addressed in the context of the two RP studies merit some further discussion. These relate to the selection of destinations for inclusion in these studies, and the approach taken to deal with certain limitations of the data.

Selection of destinations in RP studies

The selection of destinations to be included in RP studies is heavily influenced by the choice data used, and the destinations represented therein[44]. As such, only destinations with a sufficient number of observations in the choice data can be included in the analysis. The number of destinations is further decreased by the requirement that only destinations that can be reached from at least two airports in the study area should be included, for obvious reasons[45]. If a decision is taken to include only destinations reachable by direct flights, as in the two RP studies conducted in this research, then this clearly reduces the number of eligible destinations further. Finally, in the absence of an explicit treatment of the choice of destination airport, it is desirable to include only destinations that are served by a single airport, to avoid biased results in the case where the unmodelled choice of destination airport takes precedence over the choice of departure airport. In the London case-study, the large sample size allowed for the selection of a set of destinations located exclusively in regions served by a single main airport. In the SF-bay study, this was not possible, and several destinations located in multi-airport regions had to included. It is in this case important to maximise the probability of there being a conscious choice of airport in the study area, for residents as well as visitors. This issue is discussed in more detail in the context of the selection of destinations for the SF-bay area case-study (Section 9.2.1).

Data issues

Almost certainly the single biggest issue that needs to be faced in the analysis of air-travel behaviour is that of the quality of the available data. This has already been alluded to in various places in this chapter, but several points remain to be addressed. Issues with data quality in air-travel behaviour research arise especially in the case of RP survey data, where several factors, including the design of the

[44] We ignore the case where the modeller himself is responsible for the collection of the choice data, and hence in a position to define quotas so as to obtain an adequately sized sample for an a priori defined set of destinations.

[45] Although it should be noted that the inclusion of destinations served from a single airport still provides information along the airline and access-mode dimensions.

survey[46] and the compatibility with auxiliary datasets play a role. Here, we look specifically at two main issues that affect the RP case-studies presented in this thesis. The issues described here apply to a lesser extent in the case of RP bookings data, or RP survey data collected by the modeller himself. However, issues with regards to attributes of the unchosen alternatives generally remain.

The first major data problem that needs to be faced in RP studies based on survey data is the relative lack of information on the unchosen alternatives, in terms of attributes as well as availabilities. The main issue is that, in such studies, disaggregate choice data is in general used in conjunction with aggregate level-of-service data, for at least some of the attributes. While this may be acceptable for some characteristics, such as frequency and flight-time, it does create significant problems in the treatment of air-fares, and flight availability by extension. Indeed, although information on the booked class may be available, it is not generally possible to obtain information on the availability of the different fare-classes on unchosen flights, or indeed unchosen fare-classes on the chosen flight. This significantly complicates the characterisation of unchosen alternatives. As such, in the absence of availability data, or an appropriate probabilistic treatment of availability, this leads to a requirement for an assumption that tickets on all other possible flights were available at the time of booking, hence almost surely including some alternatives in the choice set that were actually unavailable. It should be noted that existing RP studies have often failed to discuss this issue. The fact that air-fare information is generally only available in aggregate form for given airline-route pairings (i.e. making no distinction between fare-classes) essentially leads to the additional assumption of equal ticket selling speeds across all flights (routes as well as departure times)[47], which is clearly not necessarily the case. Unfortunately, these assumptions cannot be avoided, and it is not clear what effect the inadequate treatment of air-fares has on model results. However, it should come as little surprise that, in the majority RP studies, it has not been possible to recover a meaningful marginal utility of fare changes (cf. Section 8.3). This can for example be explained by the example of a respondent having to choose a more expensive option because the cheap flights have all sold out. In the absence of information on flight or fare-class availability, this will, from the modeller's perspective, imply cost-prone behaviour. Here, it should also be noted that, although less of an issue in the present research, given the age of the data, the entrance of low-cost carriers on certain routes has also led to a complicated relationship between the fares of low-cost carriers and network carriers (cf. Pels & Rietveld 2004), and the increasingly dynamic nature of air-fares makes the use of aggregate fares even less reliable.

Another complication in RP studies arises with respect to the treatment of airline allegiance. It is well known that passengers are influenced in their choice of airline by their membership in frequent flier programmes (cf. Chin 2002, Adler et al. 2005), either on a personal basis, or as part of a company-wide scheme[48]. However,

---

[46]Often the datasets used in modelling air-travel choice behaviour were collected with the aim of conducting basic analyses, and as such, are lacking several vital ingredients that would allow more assumption-free research to be conducted.

[47]I.e. if a given passenger purchased an APEX ticket for his flight, then we need to work on the assumption that APEX tickets were also available for alternative flights, at the time of booking.

[48]For a discussion of frequent flier programmes, and their relative benefits to different types of travellers, see Suzuki (2003).

information on frequent flier memberships is not generally collected in passenger surveys, while information on actual benefits is heavily governed by data protection issues. As such, this potentially crucial influence on choice behaviour cannot usually be taken into account in RP studies. In the case of a dataset including a large number of international flights, operated by a variety of airlines, there is however an alternative way of modelling travellers' loyalty behaviour, by analysing their allegiance to their national carrier. Results by Bruning & Prentice (2002) suggest that such allegiance to the national carrier does indeed play a significant role[49], but that there are great variations across nationalities, as well as across the evaluation of airlines of different foreign nationalities.

The two issues described above apply almost exclusively in the case of RP data. Indeed, with SP data, detailed information on all alternatives faced by the respondent is available to the modeller, such that availability need not be considered. Additionally, data issues relating to air-fares and frequent flier membership[50] no longer apply, which is reflected in the much greater success in terms of recovering effects of such attributes in the case of SP data, as discussed in Section 8.3.6. Nevertheless, it should also be stressed that the use of SP data does pose some additional methodological problems in making inferences about behaviour from responses to hypothetical choice situations[51]. Additionally, the use of too many alternatives or attributes is likely to lead to an overloading of information, while the use of a restricted number of alternatives or attributes will not do justice to the complexity of the real-world choice processes. The use of a high number of attributes also leads to a high number of combinations of attribute levels, and consequently, a high number of choice experiments are required for each individual. This thus again leads to potential problems of complexity[52]. Encouragingly however, results by Hensher et al. (2001) in the context of airline choice suggest that, even with as many as 32 choice sets, fatigue effects were not significant.

Despite the limitations of SP data, it can be argued that, to some extent, in the case of air-travel research, these issues are outweighed by the advantages in terms of adequate information relating to the attributes of the unchosen alternatives. As mentioned previously, an interesting approach in this context is to combine RP and SP data, as discussed by Algers & Beser (2001), hence correcting for the bias inherent to models estimated on SP data[53]. The problem in this case however is one of obtaining compatible RP and SP datasets.

8.5 Summary

This chapter has acted as an introduction to the applied part of the thesis, which deals with the modelling of air-travel behaviour.

The discussion has highlighted the complexity of the choice processes undertaken by air-travellers, and has shown how they differ from behavioural processes in other areas of transport research. The discussion in Section 8.4.4 has also highlighted

[49]See also Yoo & Ashford (1997)

[50]If included in the survey.

[51]See for example Louviere et al. (2000).

[52]See Caussade et al. (2005) for a recent discussion of the effects of survey design on SP estimates.

[53]See also Morikawa (1989).

some of the issues that need to be faced in such research, notably with regards to the quality of the data in RP studies based on survey data.

Section 8.4.1 has discussed the scope of the three case-studies conducted as part of this research, which look at the choice of airport and airline, and, in the case of the two RP studies, also the choice of access-mode. This specific modelling context has seen a lot of research in the past, yet, as highlighted in the review of the existing literature in Section 8.3, much work remains to be done. As such, a number of aims for the present research have been identified in Section 8.4.2, which can briefly be summarised as follows.

- Explicitly model the multi-dimensional nature of the choice process
- Allow for deterministic, random, and continuous variations in choice behaviour
- Use advanced structures for correlation along all dimensions of choice
- Avoid over-aggregation in level-of-service data
- Conduct a study of airport choice in London

Two omissions in the present research are worth noting, one relating to choice set formation, as discussed by Basar & Bhat (2004), and the other relating to an appropriate treatment of availability of flights, and flight classes by extension, as discussed by Battersby (2004). These were found to be beyond the scope of the present research, but their inclusion within an advanced modelling framework, taking into account the various developments described here, is an important area for future research. The problem is that, especially for the latter of the two issues, appropriate auxiliary datasets are required.

Chapter 9

San Francisco Bay area case-study

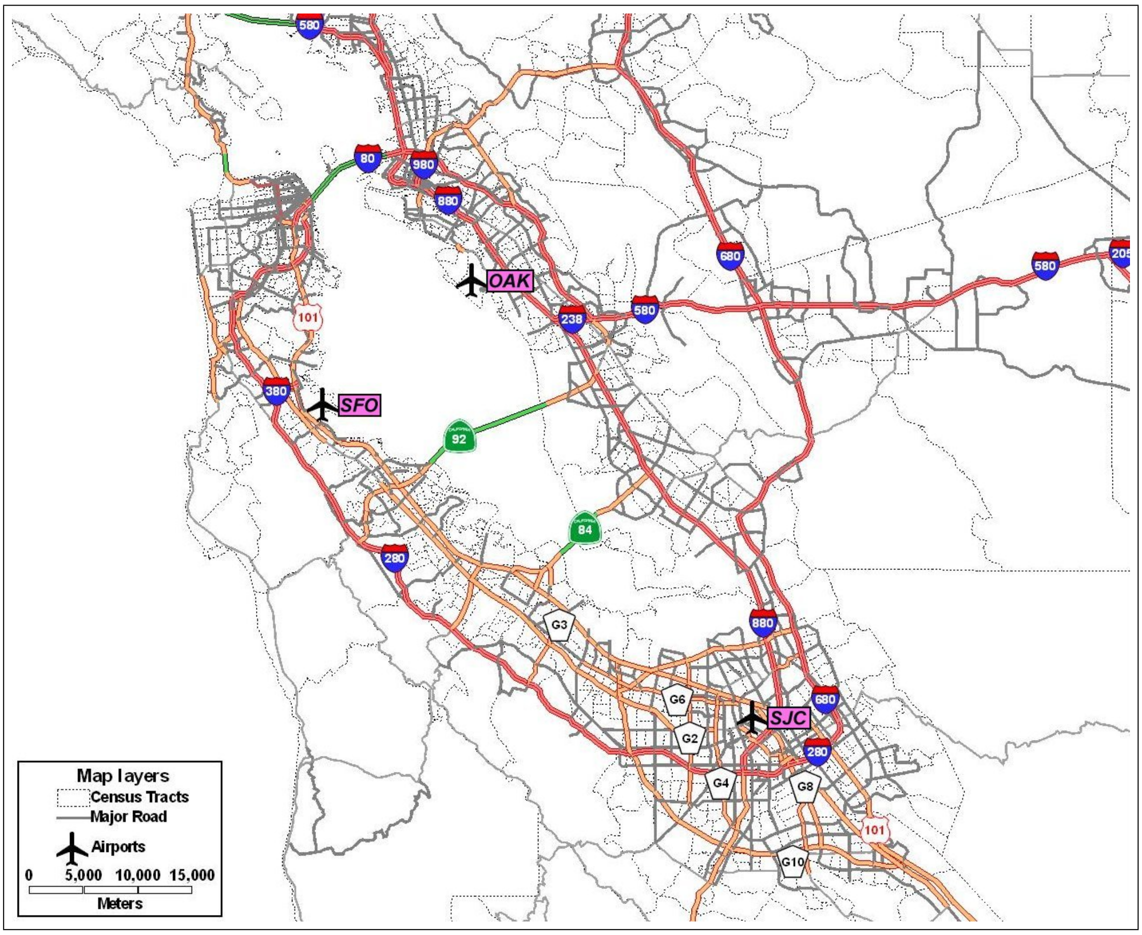
9.1 Introduction and context

This chapter describes the case-study conducted in the San Francisco Bay area, which is served by three major airports; San Francisco International (SFO), Metropolitan Oakland International (OAK), and Mineta San José International (SJC). The geographical location of the three airports is illustrated in Figure 9.1, which additionally shows the main road links serving the Bay area, and, by extension, the different airports. The map gives an indication of the strong geographical captivity, with each of the three airports being in relatively close proximity to one of the main urban centres in the region, something that applies especially in the case of SJC.

Air-traffic in the area has grown significantly over the past two and a half decades; this is illustrated in Table 9.1, which shows the number of passengers per year between 1990 and 2004, and includes connecting passengers[1]. The values show that, although SFO is still by far the largest of the three airports, with more than half the total number of passengers, its market share has decreased over time. Additionally, while both SFO and SJC suffered reductions in traffic after 2001, this has not been the case for OAK.

Forecasts by MTC (2000) predict significant rises in traffic in the SF-bay area, which will increase passenger levels at all three airports, though the relative share of SFO can be expected to decrease further. The forecasts for the years 2010 and 2020 are shown in Table 9.2, together with the year 1998, which was used as a reference value, where the differences with the values shown in Table 9.1 are down to a different counting approach (the actual predicted growth rates should be unaffected). The forecast show strong increases in traffic levels, with average annual growth rates of 3.17% between 1998 and 2010, and of 3.05% between 2010 and 2020. While there are thus only minor differences between the two periods in terms of overall growth rates, the differences for specific airports are more significant. Indeed, for OAK and SJC, the growth rate decreases from an annual 5.53% to 3.54%, and from 4.90% to 3.29%, while, for SFO, it increases from 1.91% to 2.76%. This also means that the decrease in the market share for SFO loses in intensity.

[1] There are some discrepancies between available passenger counts, given that passengers on stop-over flights not involving a change of aircraft are not always included in the counts. Nevertheless, the overall scale of the values remains relatively unaffected, given the low share of such passengers. Wherever possible, passenger counts were obtained directly from the airports.



©2005, Caliper Corporation. Caliper and TransCAD are registered trademarks of Caliper Corporation

Figure 9.1: Map of the San Francisco Bay area, with main airports and ground-level transport network

Like in many other major airport systems, demand in the SF-bay area is close to capacity. This is especially the case at SFO, where poor weather conditions, in conjunction with a close alignment of parallel runways, have meant that full runway capacity was only available 62% of the time between 1996 and 1999, with only a single runway being in use during 26% of the time (cf. RAPC 2000). A similar problem, though less intense, applies at OAK, while SJC is the only of the three airports that has recently embarked on major expansion work, with the construction of a new runway. At SFO, capacity will be exceeded by demand during good weather after 2010, where today, it is already exceeded during poor weather. At OAK, capacity will be exceeded during good and poor weather sometime between 2010 and 2020, while at SJC, the expansion work has guaranteed sufficient capacity for the time being (cf. RAPC 2000).

These problems with capacity have led to an investigation into possible ways of increasing capacity in the SF-bay area, as described by RAPC (2000). While the implementation of new air traffic control measures will ease the situation during poor

	OAK		**SFO**		**SJC**		**Total**
Year	**mppa**	**share**	**mppa**	**share**	**mppa**	**share**	**mppa**
1990	5.512	12.62%	31.060	71.14%	7.090	16.24%	43.662
1991	6.181	13.74%	31.775	70.65%	7.020	15.61%	44.976
1992	6.610	14.27%	32.610	70.39%	7.109	15.34%	46.329
1993	7.494	15.85%	32.770	69.32%	7.012	14.83%	47.276
1994	8.382	16.34%	34.648	67.52%	8.282	16.14%	51.312
1995	9.835	17.87%	36.263	65.87%	8.953	16.26%	55.051
1996	9.735	16.50%	39.252	66.53%	10.010	16.97%	58.997
1997	9.145	15.28%	40.494	67.66%	10.214	17.07%	59.853
1998	9.231	15.43%	40.101	67.01%	10.512	17.57%	59.844
1999	9.880	15.99%	40.331	65.29%	11.561	18.72%	61.772
2000	10.621	16.40%	41.049	63.38%	13.097	20.22%	64.767
2001	11.417	19.30%	34.643	58.57%	13.091	22.13%	59.151
2002	12.724	23.01%	31.450	56.88%	11.116	20.10%	55.290
2003	13.548	25.30%	29.313	54.75%	10.678	19.94%	53.539
2004	14.098	24.32%	32.835	56.63%	11.047	19.05%	57.980

Table 9.1: Annual passenger counts for three main SF-bay area airports, 1990 − 2004 (mppa = million passengers per annum)

Year	OAK	SFO	SJC	Total
1998	9.159 (16.19%)	37.107 (65.59%)	10.308 (18.22%)	56.574
2010	17.472 (21.23%)	46.545 (56.55%)	18.294 (22.23%)	82.311
2020	24.740 (22.26%)	61.116 (54.99%)	25.278 (22.75%)	111.134

Growth	OAK	SFO	SJC	Total
1998-2010	+90.76%	+25.43%	+77.47%	45.49%
2010-2020	+41.60%	+31.31%	+38.18%	35.02%

Table 9.2: Forecasts of passenger levels in the SF-bay area, reference year 1998, in mppa

weather at SFO for the time being, new runway construction work at SFO and OAK seems inevitable. Indeed, other measures, such as the construction of a high-speed rail system, are not expected to draw sufficient demand away from air-travel, while the development of a rapid ground-level or water-level connection system between SFO and OAK, which will be more expensive than the construction of new runways, cannot be expected to help manage the distribution of passengers more effectively in the absence of airline and airfare regulation. In addition, new airport construction is not an option, given the lack of potential sites, and the associated air-traffic control complications. Expansion at SFO and OAK is an expensive and complex undertaking, almost certainly leading to a requirement to reclaim land from the sea.

On the basis of the above discussion, the SF-bay area is an ideal candidate for a study of airport choice, and air-travel behaviour by extension. Three main aims apply in the SF-bay case-study; an analysis of the advantages of accounting for the multi-dimensional structure of the choice process, a study of the correlation along the three dimensions of choice, and an investigation into the prevalence of random taste heterogeneity.

The remainder of this chapter is organised as follows. The following section presents the data used in the analysis. This is followed by a discussion of the study looking at the choice of airport in Section 9.3, and a discussion of the study looking at the joint analysis of airport, airline and access-mode choice in Section 9.4. Finally, Section 9.5 summarises the findings of the case-study.

9.2 Description of data

9.2.1 Air-passenger survey data

Data on passengers' choice behaviour were obtained from the Airline Passenger Survey conducted by the Metropolitan Transport Commission (MTC) in August and October 1995[2]. This contained information on over 21,000 departing air-travellers. Passenger interviews were conducted at the three main SF-Bay area airports, as well as at the minor Sonoma County airport (STS), which was not included in the present study. For a detailed description of the survey, see Franz (1996).

The number of passengers interviewed at the three main airports is not entirely representative of the real-world traffic at the airports; indeed, SJC is over-sampled, while OAK is under-sampled. This needs to be taken into account in the modelling analysis. Aircraft occupancy data was used to calculate the total traffic on the different routes used in the analysis, for each of the carriers. From this, relative weights were assigned to each airport-airline pair[3]. A similar process was used to calculate corresponding weights for the sample data used in the present analysis. The individual pairs of weights were then used to calculate multiplicative weights for use in the estimation (weighted maximum likelihood[4]), where separate sets of weights were calculated for each of the different subsets of the data, as well as for the validation sample.

On the basis of the survey data, and the discussion in Chapter 8, a total of 14 destinations were included in the study, all of which were served by direct flights from all three airports on every day of the week during the study period. All 14 destinations are located in the continental US, with 5 in California alone, highlighting the high density of traffic in the California corridor. After the selection of the destinations, an initial sample of 9,924 respondents was obtained. This contained some 3,246 travellers who indicated that they could not have flown out of a different airport. Possible reasons for this include unavailability (at the time of booking) on flights from other airports on the chosen flight date and time (especially likely for travellers with inflexible timing), misinformation of the traveller, or an a priori

[2]In the analyses presented in this thesis, a division of the data by collection period was avoided on the grounds of the resulting small sample sizes.

[3]The access-mode choice dimension, along which the sampling was random, does not need to be taken into account in this reweighting process.

[4]In the application looking only at the choice of airport (Section 9.3), the weights used were route-specific, using summation over airlines. Unlike the three-dimensional analysis, this application uses a full set of constants, such that a correction of the constants could have been used for the MNL models. However, the weighting approach was used for reasons of consistency, given the need for such an approach in the MMNL models.

Destination airport

Dept. Apt.	Burbank, CA	Chicago, OHare, IL	Dallas, Ft. Worth, TX	Denver, CO	Las Vegas, NV	Los Angeles, CA	Ontario, CA	Orange County, CA	Phoenix, AZ	Portland, OR	Reno, NV	Salt Lake City, UT	San Diego, CA	Seattle, WA	Total
SFO	55	89	36	65	57	199	35	37	128	140	1	42	258	213	1,355
SJC	167	58	91	71	163	367	111	247	133	106	156	61	248	169	2,148
OAK	211	1	25	9	68	381	135	177	51	101	39	43	139	208	1,588
Total	433	148	152	145	288	947	281	461	312	347	196	146	645	590	5,091

Table 9.3: Summary of choice data for SF-bay area case-study

decision not to consider any of the other airports. A separate analysis showed that the inclusion of these travellers produces biased results, leading to the decision to exclude these observations from the analysis. In a way, this acts as an approximation to a model that incorporates choice set generation.

From the resulting sample of $6,678$ travellers, a further $1,587$ passengers were excluded during data-cleaning, mainly because of missing socio-demographic information, but also because of issues along the access-mode dimension[5], and the difficulties of assigning individuals from minor purpose segments to the main purpose groups[6]. This led to a final sample of $5,091$ observations, with flights to 14 destinations.

The data used are summarised in Table 9.3, which clearly shows the over-sampling of SJC. The specific choice of destinations had little effect on the distribution of observations across other dimensions, such as journey purposes and household income. Clearly, the sampling has an effect on the market shares for the different airlines; this was taken into account in the calculation of weights, as described above. The resulting dataset was split into two parts, a dataset used in the actual analysis (4,582 observations), and a 10% sample retained for later validation of the models (509 observations).

As discussed in Chapter 8, special care is required in the case of destinations that are themselves located in multi-airport regions. Given the composition of the choice data, it was in the present study not possible to rely solely on the use of destinations with a single airport, and several airports from multi-airport regions had to be included in the study. These can be divided into two groups; airports in the wider Los Angeles (LA) area[7], and cities that have secondary airports, such as Chicago, Dallas, Las Vegas, and Phoenix.

[5]Here, it was necessary to eliminate 111 passengers whose chosen access-mode was a hotel-courtesy-shuttle, where it was not possible to unambiguously define the availability of this mode for all passengers.

[6]Here, 299 respondents travelling for *extraordinary events* were excluded, in addition to 360 respondents travelling for some *other purpose*.

[7]Burbank, Los Angeles, Ontario, Orange County and San Diego.

In the second group of airports, the negligible number of observations in the choice data for the secondary airports meant that only the main airport could be included. This is clearly a simplification of the actual choice process, and assumes that passengers have made an a priori decision to travel to the main airport in the destination area. Unfortunately, this assumption could not be avoided, for the above reasons. Similarly, the destinations could not be excluded from the analysis, given their high representation in the choice data. The bias caused by including these destinations should however be acceptably small. Indeed, the share of traffic to these secondary airports is so small that the assumption of an a priori choice of destination airport can be seen to apply for a large share of the travelling population.

The situation with the five airports in the wider LA area is slightly more complicated. Again, the decision to include the destinations in the study was imposed by the high weight they carry in the choice data. However, unlike in the case of secondary airports discussed above, these five airports are all served via direct flights of relatively high frequency from each of the three SF-bay area airports. In this case, it is important to establish whether passengers are likely to make a choice of airport in the San Francisco area, besides the choice made in the LA area, especially so for passengers whose return journey started in the LA area. Here, the fact that frequent daily direct flights were available between each of the three SF-Bay area airports and each of the five airports in the wider LA region can be seen to increase the probability that passengers make a specific choice of airport in the SF-Bay area, independently of the choice of airport in the LA area (where this choice may however take precedence, especially for visiting passengers). Nevertheless, it is still possible that the inclusion of these destinations, without a direct analysis of destination airport choice (which was not possible for data reasons), produces biased results. This is in this case not avoidable, but it is important to at least acknowledge the potential bias this introduces, something that previous studies have often failed to do.

9.2.2 Air-travel level-of-service data

For the present analysis, air-travel level-of-service data were obtained from BACK Aviation Solutions[8], containing daily information for each operator serving the selected routes in August and October 1995. Eight airlines were used in the analysis, and these are hereafter referred to as airline $A1$ to airline $A8$. Besides the frequencies for the different operators, the dataset contains information on flight times and the type of aircraft used. Information on fares is available only in aggregate form (for specific airline-route pairings), leading to the problems of unreliable fare data discussed in Chapter 8. Finally, the dataset was complemented by information on the on-time performance of the different airlines used in the analysis, and the overall on-time performance of airlines at the three airports[9].

9.2.3 Ground-access level-of-service data

Ground-access level-of-service information was obtained from the MTC in the form of origin-destination (O-D) travel time and cost matrices for the $1,099$ travel area

[8] www.backaviation.com

[9] Available from the Bureau of Transport Statistics, via www.bts.gov/programs/oai

zones (TAZ) used for the SF-Bay area[10]. The dataset contains information on travel distance, travel time and tolls for car travel, under peak and off-peak conditions, and for varying car-occupancy (which has an effect on tolls). Similarly, the dataset contains information on access time, wait time, travel time, egress time and fares for public transport journeys. Corresponding data for other modes, such as taxi, limousine and special airport bus services were calculated separately, based on current prices[11] and the changes in the Consumer Price Index for California from August and October 1995 to September 2003.

Complication arose with regards to the costs of rental cars, in addition to generic car costs. Indeed, attempts to include parking cost, marginal running cost, and rental cost in the models led to inconsistent results; it was thus decided to exclude these costs from the model, allowing us to merge private car and rental car into a generic car mode, where the only cost is that of any toll incurred. This led to six remaining access-modes; car, public transport (transit), scheduled airport bus services, door-to-door services, taxi and limousine. It was assumed that taxi and limousine services are available for each origin, while the availability of door-to-door and scheduled services depends on the distance to the airports. The availability of public transport was obtained from the MTC O-D matrices, and, in the absence of any information on the availability of the car mode, it had to be assumed that car is always available.

9.3 Choice of airport

We first present the findings of the study looking only at the choice of airport. A detailed account of this analysis is published in Hess & Polak (2005*b*); here, only a brief description is presented, with the aim of allowing a comparison to be made between the performance of this *single-dimensional* approach and the *multi-dimensional* approach used in the remainder of this chapter. Before proceeding to the presentation of the results, it should be noted that, in this application, a slightly different final sample size was obtained when compared to the multi-dimensional approaches. This is the result of using different criteria in the data cleaning process, where factors specific to access-modes and airlines carry less importance than in the fully disaggregate case. This way, a final sample of $5,097$ individuals was obtained, divided into $1,268$ resident business travellers, $1,500$ resident leisure travellers, $1,269$ visiting business travellers, and $1,060$ visiting leisure travellers. In each group, a random sub-sample of roughly 10% was retained for later model validation.

9.3.1 Model specification and estimation

Two separate groups of models were estimated in this analysis; MNL models looking for the *optimal* specification of the observed utility function, and MMNL models exploring the prevalence of random taste heterogeneity. In both cases, Kenneth Train's Gauss code was used.

[10] www.mtc.ca.gov/maps_and_data/

[11] September 2003.

On the basis of the above division of the population into residents and visitors, as well as business and leisure travellers, four separate models were estimated, where this was found to be preferable to the use of separate coefficients within the same model. During model specification, the influence of a number of attributes was explored. These attributes included fare, frequency, access time, access cost, flight time, the number of operators serving a route, the type of aircraft used, and the on-time performance at the different airports. In the context of an analysis looking only at the choice of airport, any information is airport-specific, such that combination of the level-of-service data across airlines was used, assigning to each passenger the industry-level information on frequencies, fares and other airline-specific attributes for flights from each of the three airports to the desired destination on the actual date of travel.

The analysis showed that, of all the attributes included in the initial specification search, only fare, frequency and access time were found to have a significant effect. Even here, some qualification is needed. As such, it was not possible to estimate a significant fare coefficient for visiting business travellers, and, while, for both groups of leisure travellers, a consistent effect of fare could be identified, this was only possible in the lowest of three defined income groups in the case of resident business travellers. The problems with retrieving universally significant fare-effects could reflect the comparatively low sensitivity to fare for business travellers, but is almost surely also partly due to the use of highly aggregate fare information. No interactions with income could be identified for the access time coefficient, while, for frequency, a difference between income groups could only be retrieved for visiting leisure travellers, where a separate coefficient applies in the high income group. The analysis also showed clear gains in performance for a specification that allows frequency to enter the utility function in a non-linear fashion, where a log-transform was used in the present context. Similar non-linear treatments for the remaining two attributes did not lead to any gains in model performance.

On the basis of this utility specification, MMNL models were then estimated four the four groups. Here, significant random variation was retrieved in each of the four sub-models for the access time coefficient, and, except in the model for visiting business travellers, also for the ASC associated with SFO[12]. For the marginal utility of access time changes, the best results were obtained with the use of a Lognormal distribution (in conjunction with a sign change), while, in the absence of an a priori sign assumption, a Normal distribution was used for the distribution of the ASC of the SFO alternative. The levels of variation observed for the marginal utility of fare and frequency, as well as for the ASC of SJC, were not significantly different from zero, such that fixed coefficients were used. In fact, it can be seen that this model acts as an ECL approximation to a NL model nesting together OAK and SJC.

The results of the analysis are summarised in Table 9.4, where the full estimates are only shown for the MMNL models, along with the model fit for the MNL models[13], and where, for the access time coefficient, the actual mean and standard deviation are given in addition to the parameters of the underlying Normal distribution, using the transformation from equations (3.20) and (3.21), and taking

[12]The ASC for OAK was normalised to zero.

[13]The estimates of the MNL models are of little interest in the present context; they are discussed in detail by Hess & Polak (2005*b*).

	Resident business	Resident leisure	Visitor business	Visitor leisure
Parameter	**est. (t-stat.)**	**est. (t-stat.)**	**est. (t-stat.)**	**est. (t-stat.)**
Fare (common)	-	-0.0475 (-3.8)	-	-0.0477 (-3.7)
Fare (inc. $\leq \$21,000$)	-0.043 (-2.55)	-	-	-
Freq. (common)	1.9469 (5.6)	1.8333 (5.7)	1.8881 (7.7)	-
Freq. (inc. $\leq \$44,000$)	-	-	-	1.9701 (5.2)
Freq. (inc. $> \$44,000$)	-	-	-	3.0328 (5.2)
Access time c	-1.8571 (-15.5)	-1.8916 (-17.1)	-1.9706 (-20.6)	-1.9669 (-13)
Access time s	0.6742 (4.3)	0.5102 (3.6)	0.9373 (5.4)	0.6934 (5.5)
Access time μ	-0.196	-0.1718	-0.2163	-0.1779
Access time σ	0.1487	0.0937	0.2566	0.1398
ASC SFO μ	1.1563 (4.2)	0.9289 (3.9)	0.3632 (2.5)	0.5028 (1.9)
ASC SFO σ	2.026 (3.6)	1.365 (2.7)	-	1.6019 (2.2)
ASC SJC	-0.1045 (-0.5)	-0.1515 (-0.8)	-0.7767 (-3.7)	0.7784 (2.8)
Observations	1,140	1,347	1,142	952
LL (MMNL)	-604.03	-659.67	-573.67	-514.62
adj. $\rho^2(0)$ (MMNL)	0.5121	0.5495	0.5388	0.5003
LL (MNL)	-615.53	-666.22	-592.05	-519.92
adj. $\rho^2(0)$ (MNL)	0.5045	0.5464	0.5249	0.4972

Table 9.4: Estimation results for airport choice models in the SF-bay area

into account the sign change of the associated attribute. In each case, the use of the MMNL specification led to statistically significant gains in model fit over the corresponding MNL structure, with the most significant gain being obtained by the model for visiting business travellers, despite the fact that this model has only one randomly distributed coefficient. It should be noted that, although the gains in model fit obtained by the MMNL models are statistically significant, they are quite modest. As such, the differences in prediction performance between the two model structures are very low in the present context.

The main reason for presenting the results for the airport choice study in this context is to allow for an illustration of the differences in performance with the multi-dimensional models used in the remainder of this analysis. As such, the substantive results of these single-dimensional models are of little interest; trade-offs calculated on the basis of the MNL and MMNL estimates are discussed at great length by Hess & Polak (2005*b*), and are not reproduced here.

9.3.2 Model prediction performance

To give an account of the prediction performance of the models estimated above, the four MMNL models were applied to the estimation and validation samples[14]. From this, probabilities were obtained for each respondent for each of the three alternatives, on the basis of which the average probability of correct prediction was calculated. It is important to note that this is different from the unreliable *percentage right* measure, which determines implied choices on the basis of the highest choice probabilities, and calculates the percentage of correct predictions. This latter measure completely misrepresents the notion of a random utility model.

[14]Very similar performance was obtained with the MNL models, which was to be expected on the basis of the small differences in model fit. As such, these results are not reproduced here.

	Resident business	Resident leisure	Visitor business	Visitor leisure
Estimation sample	64.3%	68.0%	66.5%	65.9%
Validation sample	67.6%	66.1%	67.0%	68.3%

Table 9.5: Prediction performance of MMNL models for airport choice

The results of this exercise are summarised in Table 9.5, showing that, except for the model for resident leisure travellers, the correct prediction performance on the validation sample is actually slightly higher than that obtained with the estimation sample, suggesting that the models have not been overfitted to the estimation data, and are capable of offering good performance on unknown data. This notion is also supported by additional results by Hess & Polak (2005*b*) which show good performance in the recovery of the market shares in the validation sample.

9.3.3 Summary of findings for airport choice study

In line with previous research, the analysis described in this section has shown that there exist significant influences on airport choice due to access time and frequency of service. Additionally, for some parts of the population, the analysis has retrieved a significant effect associated with air-fares. The results also indicate that there are differences across travellers in their sensitivity to these factors, and that while differences in sensitivity to fare and frequency can be adequately accommodated by deterministic market segmentation, the sensitivity to access time additionally varies randomly within these market segments.

9.4 Multi-dimensional choice process

In this section, we describe the models fitted during the various analyses using a three-dimensional formulation of the choice set. We first discuss the structure of the choice set in Section 9.4.1, and the base specification of the utility function in Section 9.4.2. We then turn our attention to the different models used, namely MNL in Section 9.4.3, the different nesting structures in Section 9.4.4, and MMNL in Section 9.4.5. The substantive results for the different model structures are compared in Section 9.4.6, with model validation carried out in Section 9.4.7. All models presented in this section were estimated with the help of ALogit[15], where, in the case of the NL models, appropriate "dummy" correction levels were introduced to ensure consistency with utility maximisation (cf. Koppelman & Wen 1998)[16].

[15]See www.hcg.nl

[16]As an additional check, the models were later re-estimated with BIOGEME, which uses an implementation of the UMNL model. The results obtained with the UMNL specification were identical to those obtained with the NNNL specification with additional *dummy*-levels in ALogit.

9.4.1 Structure of choice set

The final sample contains data on 3 departure airports, 8 airlines, and 6 access-modes, leading to 144 distinct triplets of alternatives. Given the three-dimensional choice set, any given alternative shares the attributes of 73 other alternatives along a single dimension of choice, and shares the attributes of 14 alternatives along two such dimensions. For each observation, data on the attributes and availability of the elementary alternatives (i.e. airport, airline and access-mode) were appended to the survey data. The attributes and availability of the access-modes depend on the ground-level origin of a traveller, while the attributes and availability of the different airline options depend on the choice of destination (where not every airline operates from each airport to all 14 destinations used). The days of week were taken into account in the definition of the attributes and availability of the different flight options, as was the season (August or October), while peak and off-peak aspects were taken into account for the access-journey attributes. After adding in airport-specific attributes, the combination into triplets of alternatives was performed via the specification of utilities, where the availability of a triplet of alternatives is given by simple multiplication of individual availabilities.

9.4.2 Base specification of utility function

The base specification of the utility function was identical across the different model structures used, but there are some differences across population segments in terms of what attributes had a significant impact on choice behaviour. These differences are highlighted in the discussion of the estimation results for the MNL models in Section 9.4.3.

At this point, it should be noted that the multi-dimensional study uses six such population segments, instead of the four used in the simple study of airport choice. This is based on preliminary results showing differences between respondents on holiday-related travel, and those visiting friends or relatives (VFR). As such, three purpose-specific segmentations were used (business, holiday, VFR), which, together with the division into residents and visitors, led to six separate groups, and by extension, models.

For each model, attempts were made to include coefficients showing travellers' sensitivity to various attributes of the airports, airlines and access-modes. Thanks to the disaggregation into three-dimensional data, the number of attributes for which significant differences existed across alternatives (and which could thus be expected to influence choice behaviour) was much higher. The set of potential explanators used in the specification search included factors such as flight frequency, flight time (block time, which indirectly takes into account airport congestion), fare and aircraft type (jet vs turboprop), as well as access time (in-vehicle), walk time to access-mode (e.g. to public transport station), wait time for access-mode and access cost. In each case, attributes are specific to individual alternatives (e.g. no aggregation across airlines). However, two simplification were used in model specification, coming in the use of generic coefficients along the access-mode dimension, and along the airline dimension, such that any characteristics specific to a given access-mode or aircraft-type are captured solely by the appropriate constants. The use of separate coefficients for different modes along the access-mode dimension and different aircraft-types along the air-travel dimension remains an avenue for future research.

Both linear and various non-linear specifications of the various explanatory variables were tested. The best results were obtained with the use of a logarithmic transform, this however only led to an improvement in model fit when applied to flight frequency, whereas non-linear specifications of flight time, in-vehicle time, access walk time, wait time and fare led to unsatisfactory results. Also, some potentially important influences, such as carrier loyalty, could not be explored, due to lack of data (e.g. no information on frequent flyer programmes), while, in the presence of national flights only, the notion of allegiance to the national carrier does not apply. Similarly, it was not possible to identify a significant direct effect of the on-time performance of airlines or airports on the respective choice probabilities. Attempts were made to segment the population by income, for example in order to show different values of time in different income-classes. Three income groups were defined, segmenting the population into low income ($< \$21,000$ per annum), medium income (between $\$21,000$ and $\$44,000$ per annum) and high income (above $\$44,000$ per annum).

A further specification issue that was explored was the influence of past choices on choice behaviour (cf. Windle & Dresner 1995). In the present analysis, we had information on the number of flights a given traveller took from each of the three SF-bay airports in the past twelve months. For each one of the three airports, a coefficient in the utility function was thus associated with the inertia variable for that airport, where, to account for cross-effects, coefficients in a given airport's utility function were also associated with the inertia variables of the remaining two airports. Clearly, some normalisation is required in this case, so that, aside from three airport-specific inertia coefficients, inertia coefficients associated with SJC and OAK were included in the utility of SFO, while a coefficient associated with SFO was also included in the utility of SJC, and no cross-coefficients were included in the utility of OAK-alternatives. The inclusion of these variables did in each case, as expected, lead to dramatic improvements in model fit, where the gains were even more significant when using a log-transform, such that this approach was adopted. It should of course be noted that the inclusion of these coefficients could lead to problems with endogeneity, as the values of the past choice indicators may be closely correlated with the other explanatory variables and with unobservables. The dependence on past choices would also make this approach inapplicable in the case where the model was used for forecasting. However, this is not the main purpose of the present analysis; furthermore, in each one of the models used, the values of the remaining coefficients remained largely unaffected, suggesting that the inclusion of these inertia terms did not introduce major bias.

Finally, it should be noted that a non-standard specification of constants was used in these models. With a full set of ASCs, 144 constants would be needed, of which 143 could be estimated. This approach however led to severe identification problems, which are at least partly caused by the low representation of certain triplets of alternatives, in terms of inclusion in the choice sets, and even more so in terms of actual choices. It was found that better performance, with notably fewer identification problems, was obtained by an alternative approach, which takes into account the multi-dimensional nature of the choice set. As such, separate sets of ASCs are associated with each of the three dimensions, leading to 3 airport

		Business		Holiday		VFR$^{(\dagger)}$	
		est.	**t-stat.**	**est.**	**t-stat.**	**est.**	**t-stat.**
	Access cost	-	-	-0.0208	-2.21	-0.0223	-2.29
	Access cost, $> \$44,000$ **p.a.**	-0.0244	-2.86	-	-	-	-
	Access cost, $< \$44,000$ **p.a.**	-0.0358	-4.17	-	-	-	-
	In-vehicle time	-0.0522	-12.13	-0.0594	-12.94	-0.0490	-9.43
	Walk time, $> \$44,000$ **p.a.**	-0.1531	-2.97	-	-	-	-
	Walk time, $< \$44,000$ **p.a.**	-0.1139	-2.47	-	-	-	-
	Fare	-	-	-0.0131	-1.9	-0.0267	-3.03
	Flight time	-0.0471	-2.37	-	-	-	-
	Flight frequency	1.3183	10.77	1.3235	9.35	1.4447	7.87
	Turboprop	-2.5296	-3.2	-4.2294	-2.7	-	-
Inertia var.	**OAK on OAK**	1.9993	9.44	2.1024	5.09	2.2919	5.24
	SFO on SFO	1.1830	9.62	1.1887	7.89	2.0488	8.83
	SJC on SJC	1.9641	8.49	2.5909	5.04	3.1690	5.87
	OAK on SFO	0.6620	3.37	0.8328	1.98	0.4413	1.02
	SJC on SFO	0.7845	3.68	1.4302	2.71	0.5574	1.1
	SFO on SJC	0.1731	1.07	0.1618	0.79	0.0292	0.09
	Observations	1,098		831		641	
	Log-likelihood	-1551.62		-1384.81		-1050.84	
	Adj. $\rho^2(0)$	0.5861		0.5112		0.5046	

$^{(\dagger)}$ Visiting friends or relatives

Table 9.6: MNL estimation results for travellers resident in the SF-bay area

constants, 8 airline constants, and 6 access-mode constants, where, in each group, one constant was normalised to a value of zero, leading to 14 estimated ASCs. This approach enables identification of the models, but this comes at the cost of a violation of the zero-mean assumption for the unobserved part of utility, which is however not avoidable.

9.4.3 MNL models

In the following paragraphs, we describe the findings of the analysis fitting MNL models to the six separate estimation datasets. The results of the various models are summarised in Table 9.6 for residents and Table 9.7 for visitors, where in each case, all estimated parameters aside from the ASCs are reproduced. The models presented here are the same as those estimated by Hess & Polak (2004*b*). Where not mentioned otherwise, the absence of a segmentation by income means that differences in the estimated coefficients were not significant across groups, while the absence of a specific coefficient means that a significant effect for the associated attribute could not be retrieved.

MNL model for business trips by residents

The estimation dataset contains information on $1,098$ business trips by residents. The estimation process revealed significant effects of walk access time, access cost, in-vehicle access time, flight time and frequency. Additionally, a significant negative effect could be associated with turboprop flights. No meaningful and significant effect of fare could be identified, even after taking into account income. This can

		Business		Holiday		VFR[†]	
		est.	**t-stat.**	**est.**	**t-stat.**	**est.**	**t-stat.**
	Access cost	-	-	-0.0145	-1.66	-	-
	Access cost, $>$ \$44,000 **p.a.**	-0.0219	-2.55	-	-	-	-
	Access cost, $<$ \$44,000 **p.a.**	-0.0286	-3.94	-	-	-	-
	In-veh. time	-	-	-0.0769	-13.22	-0.0698	-11.06
	In-veh. time, $>$ \$21,000 **p.a.**	-0.0820	-14.43	-	-	-	-
	In-veh. time, $<$ \$21,000 **p.a.**	-0.0496	-7.18	-	-	-	-
	Wait time	-0.2507	-3.28	-	-	-	-
	Fare, $<$ \$21,000 **p.a.**	-	-	-	-	-0.0501	-3.55
	Fare, [\$21,000, \$44,000] **p.a.**	-	-	-	-	-0.0267	-1.95
	Flight time	-0.0293	-1.39	-0.0908	-3.42	-0.1522	-5.12
	Flight frequency	1.3066	11.34	1.0783	7.51	0.7244	4.41
Inertia var.	**OAK on OAK**	1.1881	6.57	1.2529	2.9	1.3899	2.96
	SFO on SFO	1.9324	9.39	0.7514	3.97	1.0991	3.35
	SJC on SJC	1.3973	6.1	2.0564	4.42	2.2569	4.17
	OAK on SFO	-0.7172	-3.36	-0.4741	-0.99	0.1887	0.35
	SJC on SFO	0.0075	0.03	0.8318	1.86	-0.1219	-0.17
	SFO on SJC	0.5032	2.38	-0.1084	-0.34	0.1809	0.42
	Observations	1,057		534		421	
	Log-likelihood	-1517.68		-1018.25		-621.81	
	Adj. $\rho^2(0)$	0.4379		0.3725		0.5044	

[†] Visiting friends or relatives

Table 9.7: MNL estimation results for travellers not resident in the SF-bay area

mainly be explained by the poor quality of the fare data, but could also signal indifference to fare increases on the part of business travellers. It was possible to segment the sensitivity to walk time and access cost by income, although, given very low differences between the estimates in the low and medium income group, only two coefficients were retained, one for people earning less than \$44,000 per annum, and one for the remaining travellers. The results show lower sensitivity to cost for people with higher income, along with higher sensitivity to increases in walk time. In terms of the airport-inertia variables, the estimates show positive direct effects for all three airports, with positive cross-effects of past usage of SJC and OAK on the utility of SFO, and a positive (but not significant) cross-effect of past usage of SFO on the utility of SJC.

MNL model for business trips by visitors

The estimation dataset contains information on 1,057 business trips by visitors. Just as in the case of resident business travellers, no significant impact of fares could be identified. In-vehicle access time and access cost are again significant, and negative, with increasing sensitivity to in-vehicle access time with higher income (only two groups could be used) and lower sensitivity to cost with higher income (two groups only). Whereas it was not possible to estimate a significant effect of wait time for resident business travellers, a significant negative effect could be identified for their non-resident counterparts. However, the estimate for flight time is no longer significant at the conventional 95% level, and it was not possible to include an effect of equipment type, as flights using turboprop planes were never chosen. Also, with this model, no effect could be associated with access walk time, while frequency again has a strong positive effect. Finally, unlike in the model for resident business travellers, the inertia cross-effect of past flights at OAK has a negative effect on the utility of SFO, while the cross-effect of past flights at SJC on the utility of SFO is now no longer significantly different from zero, while there is a positive cross-effect of SFO acting on the utility of SJC.

MNL model for holiday trips by residents

The model estimated on the 831 observations for residents' holiday trips suggests a lower utility for flights using turboprop aircraft, negative impacts by access cost and in-vehicle time, and a positive effect of flight frequency. All inertia coefficients are positive, though the cross-effect of past flights at SFO on the utility of SJC is not significant. Finally, for this group of travellers, a negative effect could be identified for fare (although of lower statistical significance) while no effect could be associated with flight time and access walk time. No significant gains could be made through segmenting the population by income for any of the coefficients.

MNL model for holiday trips by visitors

For the 534 visitors on holiday trips, no significant effect of fare could be identified, and the effect of access cost, although of the correct sign, is not significant at the 95% level. In-vehicle time has a significant negative effect, as has flight time, while increases in frequency lead to increases in utility. The inertia variables show a positive cross-effect of past flights at SJC on the utility of SFO, while the other two cross-effect estimates are not significant. Finally, the aircraft-type coefficient had to be excluded from the model (turboprop flights never chosen), while no effect could be identified for wait time, and segmentations by income did not lead to any gains in model fit.

MNL model for VFR trips by residents

The estimates for the model fitted to the sample of 641 residents on VFR trips show significant negative effects of access cost, in-vehicle time and air-fare, along with positive effects of flight frequency. The cross-effect inertia variables are not significant, equipment-size could not be included and no effects could be identified for walk time, wait time and flight time, while segmentations by income led to a loss of information in the model.

MNL model for VFR trips by visitors

The final subsample used in the estimation of the MNL models contains information on 421 VFR trips by non-residents. The results show negative impacts of fare in the medium and low income classes (with higher sensitivity in the low income class), while the effect for high-earners was not significant and was dropped from the model. In-vehicle time and flight time have a negative effect, with a positive effect for frequency increases. Again, the cross-effect inertia variables are not statistically significant, while no effect could be associated with access walk time, wait time, and access cost, and the turboprop coefficient had to be excluded.

Summary of MNL results

The discussion of the MNL results has revealed that there are important differences across the six segments in the optimal specification of utility. The common point across all segments is that a logarithmic specification is always preferable to a linear specification in the case of the frequency and inertia coefficients, and that in-vehicle access time and flight frequency are the main factors influencing choice. Significant effects of air-fare could only be identified for resident holiday and VFR travellers, as well as for visiting VFR travellers, where there are also differences across income groups in fare-sensitivity. In terms of model fit (on the basis of the adjusted ρ^2 statistic), the models for residents perform better than those for visitors for business and holiday trips, while for VFR trips, there is no significant difference in performance. This relates to the point made in Section 8.2.2 about the greater difficulties involved in analysing choice-behaviour by non-residents.

9.4.4 Nesting structures

We next turn our attention to the estimation of NL models, where the findings described in this section are those reported by Hess & Polak (2004*b*). The discussion reports the results of three separate sets of NL models, nesting the alternatives by airport, airline and access-mode respectively. The use of the six possible multi-level structures (cf. Section 8.4.3) was also explored, but none of them led to satisfactory results, where in general, the models collapsed back to a two-level structure, suggesting that a multi-level structure is not applicable with the current data and specification of the choice set[17]. As such, the analysis is limited to the use of two-level structures, nesting the alternatives either by airport, airline, or access-mode. In each case, the specification of the NL model corresponds to that of the associated MNL model, although there is an occasional drop in the significance level of the estimated parameters, which does suggest some interaction between observed and unobserved utility. Given the high number of models and estimated parameters, and the fact that the specification corresponds to that of the MNL model, the results presented in this section are limited to the findings in terms of nesting structure and model performance. The substantive differences between the NL and MNL models are illustrated in the comparison of trade-offs in Section 9.4.6, which also allows for a comparison with the MMNL results.

Nesting by airport

The first set of models nest the alternatives by airport, leading to 48 alternatives per nest (8 airlines and 6 access-modes). The results are summarised in Table 9.8 (using λ_k to define the structural parameter of airport nest k), with t-statistics for the structural parameters (calculated wrt 1) given in brackets. For comparison, the table again gives the final log-likelihood of the corresponding MNL models, and

[17]It should be noted that Pels et al. (2003) succeeded in estimating three-level models on the same data. This is an indication that the differences that arise between studies in terms of auxiliary data and utility specification can have a significant impact in terms of the processes that can be included in the observed part of utility, hence also changing the structure in the unobserved part of utility.

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
LL(MNL)	-1551.62	-1384.81	-1050.84	-1517.68	-1018.25	-621.81
LL(NL)	-1545.14	-1372.19	-1039.67	-1487.71	-999.51	-621.62
adj. $\rho^2(0)$ (MNL)	0.5861	0.5112	0.5046	0.4379	0.3725	0.5044
adj. $\rho^2(0)$ (NL)	0.5873	0.5148	0.5088	0.4481	0.3826	0.5038
λ_{SFO}	1	1	1	1	1	1
λ_{SJC}	0.78 (4.02)	0.76 (4.08)	0.67 (5.5)	0.53 (10.64)	0.44 (8.79)	0.93 (0.63)
λ_{OAK}	0.89 (1.64)	0.73 (4.61)	0.78 (3)	0.72 (3.7)	0.74 (2.24)	1

T-statistics wrt 1

$^{(\dagger)}$ Visiting friends or relatives

Table 9.8: Estimation results for NL model using nesting by airport on SF-bay area data

the adjusted ρ^2 measure for the MNL and NL models, which takes into account the extra cost in terms of estimated parameters. The results show that, for every single model, the structural parameter of the nest containing the SFO alternatives had to be constrained to a value of 1, as it would otherwise have exceeded this value, becoming inconsistent with utility maximisation. This suggests that there is no heightened correlation between the different alternatives available from SFO. Except for the case of visitors on VFR trips, where the structural parameter for OAK had to be constrained to 1, the estimates for the structural parameters of the other two airports are always below 1, although there are cases where the difference is not statistically significant.

There are differences across models in the values of the structural parameters, and also in the relative values of the structural parameters for the SJC and OAK nests (although SJC is generally lower than OAK), suggesting differences between the different groups of travellers. In terms of model fit, the use of the NL models leads to a significant increase in log-likelihood, except in the case of visitors on VFR trips, where the log-likelihood is virtually identical to that of the MNL model, as is the NL model itself, given that the structural parameters for SFO and OAK are equal to 1, while the structural parameter for SJC is very close to 1. Except for VFR trips, the improvements in model fit are more important for visitors than for residents, and the lower structural parameters for visitors on business and holiday[18] trips suggest a lower substitution effect between airports (i.e. higher correlation for alternatives sharing an airport) than is the case for residents.

Nesting by airline

The lack of information on frequent-flier programme membership and other airline-specific attributes means that there should be some correlation in the unobserved part of utility between different alternatives that refer to the same airline. As such, it is of interest to attempt to use a nesting structure that uses a single nest for each airline, leading to 8 nests, with 18 alternatives each. The results of this analysis, which are summarised in Table 9.9 (using π_l to define the structural parameter of

[18] Only for SJC.

	Resident			Visitor		
	Business	Holiday	VFR[†]	Business	Holiday	VFR[†]
LL(MNL)	-1551.62	-1384.81	-1050.84	-1517.68	-1018.25	-621.81
LL(NL)	-1536.66	-1371.21	-1034.07	-1507.62	-1003.93	-620.24
adj. $\rho^2(0)$ (MNL)	0.5861	0.5112	0.5046	0.4379	0.3725	0.5044
adj. $\rho^2(0)$ (NL)	0.5890	0.5141	0.5105	0.4394	0.3770	0.5018
π_{A1}	0.95 (0.25)	0.92 (0.32)	1	0.96 (0.14)	0.7 (1.34)	1
π_{A2}	0.61 (4.59)	0.78 (1.05)	0.87 (1.47)	0.98 (0.16)	0.62 (4.62)	0.86 (1.17)
π_{A3}	1	1	0.86 (0.43)	0.89 (0.36)	0.77 (1.17)	0.85 (0.61)
π_{A4}	1	1	1	0.65 (2.22)	0.72 (1.07)	0.68 (1.25)
π_{A5}	0.74 (3.35)	0.74 (2.66)	0.63 (3.92)	0.63 (2.22)	0.39 (4.97)	1
π_{A6}	1	0.9967 (0.03)	1	1	0.68 (2.44)	0.79 (2.13)
π_{A7}	1	1	1	1	1	1
π_{A8}	0.84 (0.9)	0.72 (3.28)	0.67 (1.35)	0.79 (1.13)	0.53 (7.01)	0.84 (0.71)

T-statistics wrt 1

[†] Visiting friends or relatives

Table 9.9: Estimation results for NL model using nesting by airline on SF-bay area data

airline nest l), show that a comparatively high number of structural parameters had to be constrained to a value of 1, while many others are not statistically different from 1. Nevertheless, again except for the model for visitor VFR trips, the use of the NL model resulted in a significant improvement in model fit over the MNL model. Also, the great variability in the values of the structural parameters for given airlines across the different models suggests significant differences in the cross-elasticities in the different models. It can be observed that airlines $A5$ and $A8$ on average have lower structural parameters than the other airlines. This could at least be partly related to the fact that these two carriers run a low-cost airline scheme; the product offered by these airlines is different from that offered by other airlines, which increases the scope for correlation. As suggested in Section 8.4.3, an interesting avenue for future research is to explore the correlation between alternatives on different low-cost airlines; here, the scope for such an extension was very limited, given the low number of observations for airline $A5$ relative to $A8$, and the limited route-overlap of the two airlines.

Nesting by access-mode

The results of the analysis using nesting by access-mode are summarised in Table 9.10 (using Ψ_m to define the structural parameter of access-mode nest m). In many regards, nesting by access-mode proved to be the most promising approach, as, unlike the approaches using nesting by airport and airline, the present nesting approach leads to significant increases in model fit across groups, including for VFR trips by visitors. Also, in total, only three of the structural parameters had to be constrained to a value of 1, although some of the estimated structural parameters are not statistically different from 1, while it should also be noted that some of the structural parameters are surprisingly close to zero. Except for the model for business trips by visitors (for whom the car and rental car market shares are lower

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
LL(MNL)	-1551.62	-1384.81	-1050.84	-1517.68	-1018.25	-621.81
LL(NL)	-1520.42	-1351.18	-1007.20	-1508.79	-1004.26	-603.07
adj. $\rho^2(0)$ (MNL)	0.5861	0.5112	0.5046	0.4379	0.3725	0.5044
adj. $\rho^2(0)$ (NL)	0.5930	0.5207	0.5224	0.4390	0.3774	0.5149
Ψ_{car}	0.18 (15.6)	0.13 (20.9)	0.13 (21.6)	0.45 (7.4)	0.16 (11.8)	0.09 (22.0)
$\Psi_{scheduled}$	0.19 (10.5)	0.18 (8.9)	0.05 (39.9)	0.64 (1.2)	0.15 (7.6)	0.8 (0.3)
$\Psi_{transit}$	0.31 (5.3)	0.3 (5.1)	1	0.25 (4.6)	0.33 (2.6)	0.02 (49.1)
$\Psi_{door-2-door}$	0.29 (6.3)	0.18 (12.3)	0.18 (9.2)	0.5 (1.6)	0.16 (11.8)	0.12 (12.6)
Ψ_{taxi}	0.13 (19.7)	0.09 (29.3)	0.17 (10.5)	0.38 (7.2)	0.16 (11.8)	0.05 (27.9)
$\Psi_{limousine}$	1	0.22 (5.6)	0.31 (5.1)	0.36 (4.6)	0.25 (3.9)	1

T-statistics wrt 1

$^{(\dagger)}$ Visiting friends or relatives

Table 9.10: Estimation results for NL model using nesting by access-mode on SF-bay area data

than for other groups), the structural parameter for car is always very low, partly reflecting travellers' strong allegiance to car as an access-mode. A comparably constant low structural parameter is observed for the taxi nest, while the structural parameter for the scheduled nest especially varies widely across models. Finally, it should be noted that, for holiday trips by visitors, the structural parameters of the car, door-to-door and taxi nests were constrained to have the same value, given that the initial estimates were almost indistinguishable. This led to a drop in the log-likelihood by a mere 0.028 points.

Summary of NL results

The analysis has shown that some gains in model fit can be obtained by using a nesting structure, although these gains are often not as significant as expected. This could be due to two very distinct reasons. NL models differ from the MNL model in that they accommodate correlation between the unobserved components of utility. The first explanation interprets the similarity in the performance of the two models as an endorsement of the MNL models, suggesting that the (observed) utility specification used in the models captures almost all of the correlation in utility across alternatives, reducing the scope for the NL model to capture any correlation patterns in the remaining unobserved part of utility. An alternative explanation is based on the reasoning that the specific nesting structures used are little better that the MNL model in capturing the true structure of the underlying correlations in the unobserved component of utility. The same conclusion would extend to the multi-level NL structures initially explored. It is not clear from the empirical results alone which of these potential explanations is most appropriate.

Although the gains in model fit were not as important as expected, several conclusions can be drawn from the analysis discussed above. First, there seem to be important differences across population groups in the values of the structural parameters. This suggests that gains in performance could be made by using a modelling structure in which the structural themselves vary across respondents, either deterministically, as in the COVNL model of Bhat (1997), or in a random fashion, as described in Chapter 7. Secondly, the results indicate differences in performance between the three nesting structures across the six datasets used. As such, the models using nesting by access-mode lead to the best performance for the three datasets with resident travellers, while for visitors, this is only the case for VFR trips, with nesting by airport leading to the best performance for business and holiday trips. This again suggests differences in behaviour between residents and visitors. The fact that nesting by airport produces the best results in two out of the three models for visitors could for example reflect the presence of additional effects linked with access-distance, which could suggest that these travellers often simply choose the airport that is closest to their intended ground-level destination (keeping in mind that the chosen airport is actually their arrival airport from the outbound leg). Finally, nesting by airline never leads to the best performance.

Cross-nesting structures

A final issue that needs to be discussed in the context of nesting models is the use of cross-nesting structures for the joint analysis of correlation along the three dimensions of choice, as discussed in Section 8.4.3. Given the above results in terms of correlation along either of the three dimensions, and the inability to estimate multi-level models, the use of cross-nesting structures has a lot of appeal in this context.

However, efforts by Hess (2004) to estimate a CNL model on the present data (resident business travellers only) led to inconclusive results, and as such are not reproduced here, rather giving preference to the more in-depth discussion of CNL structures in the London case-study (Chapter 10). The experiments conducted on this data showed that it is almost inevitable to constrain the allocation parameters to take on values of $\frac{1}{3}$, such that an alternative belongs in equal parts to one airport, one airline and one access-mode nest[19]. Furthermore, while the results indicate heightened correlation along each of the three dimensions of choice, hence justifying the use of a cross-nesting approach, the model only offers significant improvements in performance over the NL models using nesting by airport or by airline. Indeed, the improvements offered over the NL model using nesting by access-mode are not statistically significant, when taking into account the higher number of estimated parameters, even after constraining the allocation parameters. This result reinforces the findings from the NL analysis, in that nesting by access-mode is the most fruitful approach. On the other hand, the fact that heightened correlation was found jointly along multiple dimensions shows the structural advantages of the CNL model over NL models, with which, in the present analysis, it was not possible to retrieve correlation along more than a single dimension.

[19]The actual estimation of the allocation parameters not only leads to huge increases in computational cost, but the associated rise in the number of estimated parameters means that any improvements in model fit are not likely to be statistically significant.

9.4.5 Mixture structures

The final part of the estimation analysis looks at the use of mixture models, where, in the present context, the analysis was limited to the use of MMNL models, and where the random structure of the model was used solely for expressing random taste heterogeneity across respondents, and not heteroscedasticity or inter-alternative correlation. This analysis is partly based on work by Hess & Polak (2005*a*), where that study was however limited to the segment of resident business travellers, and also made use of a less detailed specification of the utility function, notably in the form of no interactions with income.

The specification of the MMNL models for the six separate population segments was based on the respective MNL specifications. With this approach, two important points need to be addressed. Firstly, it can be seen that, with the present specification, the use of randomly distributed ASCs would have led to a model approximating a nesting structure. This was not the aim of the analysis, such that the ASCs were kept fixed, and the same normalisation could continue to be used[20]. The second point that needs to be addressed is the potential of attributes having a significant effect in MMNL models where this was not the case in the MNL models. This can be explained on the grounds that the mean effect of an attribute in the population might not be significant, while the attribute does however have a significant effect for part of the population in such a way that a simple segmentation cannot account for it. No such effects were identified in the specification search, such that the utility function in the MMNL model is indeed based on the MNL specification, a fact that facilitates model comparison.

The main aim of this analysis is to test for the prevalence of random taste heterogeneity, and to demonstrate the applicability of the MMNL model to the joint analysis of airport, airline and access-mode choice. As such, the policy implications are of lesser importance, enabling certain simplifying assumptions to be made for the analysis. The first of these comes in the use of the Normal distribution for all parameters that follow a random distribution. This can lead to problems with interpretation (cf. Chapter 4), and these are highlighted at appropriate places in the text. In the present context, the advantages of the Normal distribution from a computational point of view are very significant, where, in the one-dimensional application described in Section 9.3.1, the use of the Lognormal distribution was made possible by the much lower number of alternatives (3 compared to 144). Even with the use of the Normal distribution, the computational costs were non-trivial[21], as a sufficiently high level of precision for the simulation process had to be guaranteed. The second major simplification used in the analysis is an absence of an explicit investigation into the correlation between randomly distributed coefficients; again, this would be an important component of a more policy-oriented analysis.

In the analysis, attempts were made to identify variations across respondents in all estimated parameters except for the ASCs. In practice, only a subset of parameters exhibited significant levels of variation. As an example, no variation was found in any of the inertia coefficients, across all six models. Higher levels of

[20]In models using a random distribution for some of the ASCs, the normalisation is no longer arbitrary.

[21]Of the order of several days.

		Business		Holiday		VFR[†]	
		est.	**t-stat.**	**est.**	**t-stat.**	**est.**	**t-stat.**
Access cost		-	-	-0.0220	-2.36	-0.0255	-2.62
Access cost, $>\$44,000$ p.a.	μ	-0.0590	-3.33	-	-	-	-
	σ	0.0402	3.38	-	-	-	-
Access cost, $<\$44,000$ p.a.		-0.0441	-4.60	-	-	-	-
Access in-vehicle time	μ	-0.0600	-9.33	-0.0691	-11.19	-0.0637	-7.98
	σ	0.0328	2.82	0.0355	5.29	0.0510	5.56
Walk time, $>\$44,000$ p.a.		-0.1532	-2.93	-	-	-	-
Walk time, $<\$44,000$ p.a.		-0.1203	-2.55	-	-	-	-
Fare		-	-	-0.0139	-1.92	-0.0324	-3.19
Flight time		-0.0546	-2.57	-	-	-	-
Flight frequency	μ	1.4892	8.58	1.3454	9.25	1.5080	7.77
	σ	0.5808	2.13	-	-	-	-
Turboprop	μ	-5.7843	-1.65	-4.4751	-2.79	-	-
	σ	3.6785	1.84	-	-	-	-
Inertia var. OAK on OAK		2.1207	8.83	2.1780	4.99	2.5157	5.11
Inertia var. SFO on SFO		1.2454	9.30	1.2125	7.80	2.1368	8.41
Inertia var. SJC on SJC		2.1619	8.09	2.6980	5.11	3.6434	6.05
Inertia var. OAK on SFO		0.7037	3.24	0.8982	2.02	0.5855	1.21
Inertia var. SJC on SFO		0.7825	3.47	1.4323	2.65	0.6153	1.14
Inertia var. SFO on SJC		0.1063	0.59	0.0870	0.40	-0.1085	-0.29
Observations		1,098		831		641	
Log-likelihood		-1543.93		-1379.36		-1044.11	
Adj. $\rho^2(0)$		0.5871		0.5127		0.5073	

[†] Visiting friends or relatives

Table 9.11: MMNL estimation results for travellers resident in the SF-bay area

variation would be expected with the use of SP data, partly because of the greater quality of the level-of-service data, but also because of the presence of multiple observations for each respondent (cf. Chapter 11), where this latter point can also apply in the case of "travel-diary" RP data.

The results of the MMNL analysis are summarised in Table 9.11 for residents, and Table 9.12 for non-residents. The findings in terms of specification are discussed in the remainder of this section, with the substantive conclusions presented in conjunction with those for the other models in Section 9.4.6.

MMNL model for business trips by residents

In the MMNL model for resident business travellers, 4 coefficients were found to exhibit significant levels of random variation across respondents, namely those associated with access cost in the high income group, in-vehicle access time, flight frequency, and the turboprop dummy variable. With 4 additional parameters, the MMNL model offers an improvement in LL by 7.69 units, which, with an associated χ^2_4 p-value of 0.004 for the LR-test, is statistically significant, yet far from spectacular.

Several points deserve some further attention. The first one of these relates to the mean coefficient associated with the turboprop dummy, where the level of significance has dropped to the 90% level, with the associated standard deviation

			Business		Holiday		VFR[†]	
			est.	**t-stat.**	**est.**	**t-stat.**	**est.**	**t-stat.**
	Access cost		-	-	-0.0145	-1.67	-	-
	Access cost, > $44,000 p.a.		-0.0227	-2.65	-	-	-	-
	Access cost, < $44,000 p.a.		-0.0302	-4.09	-	-	-	-
	Access in-vehicle time	μ	-	-	-0.0878	-12.16	-0.0920	-7.86
		σ	-	-	-	-	0.0493	3.79
	In-vehicle time, > $21,000 p.a.	μ	-0.0997	-9.21	-	-	-	-
		σ	0.0437	3.13	-	-	-	-
	In-vehicle time, < $21,000 p.a.	μ	-0.0760	-5.28	-	-	-	-
		σ	0.0600	3.50	-	-	-	-
	Wait time		-0.2722	-3.53	-	-	-	-
	Fare, < $21,000 p.a.	μ	-	-	-	-	-0.0984	-3.50
		σ	-	-	-	-	0.0827	2.37
	Fare, [$21,000, $44,000] p.a.	μ	-	-	-	-	-0.0518	-2.32
		σ	-	-	-	-	0.0644	1.81
	Flight time	μ	-0.0362	-1.50	-0.0980	-3.09	-0.1844	-4.96
		σ	0.0838	2.08	0.1869	4.99	-	-
	Flight frequency	μ	1.3712	10.90	1.0978	7.05	1.1720	4.46
		σ	-	-	-	-	1.2995	3.51
Inertia var.	**OAK on OAK**		1.2962	6.08	1.4686	2.85	1.7573	3.03
	SFO on SFO		2.1303	8.61	0.8648	4.07	1.3733	3.34
	SJC on SJC		1.6048	5.76	2.3851	4.45	3.0187	4.20
	OAK on SFO		-0.7744	-3.11	-0.4359	-0.77	0.3912	0.60
	SJC on SFO		-0.0852	-0.29	0.8448	1.71	-0.0203	-0.02
	SFO on SJC		0.4245	1.72	-0.0629	-0.18	0.2029	0.39
	Observations		1,057		534		421	
	Log-likelihood		-1512.00		-1010.58		-614.66	
	Adj. $\rho^2(0)$		0.4389		0.3766		0.5068	

[†] Visiting friends or relatives

Table 9.12: MMNL estimation results for travellers not resident in the SF-bay area

being significant at the 93% level. The significance level of many of the other coefficients can in fact be observed to have increased slightly when compared to the MNL model.

The second point relates to the impact of using a Normal distribution for the randomly distributed coefficients. The effects in this case are benign, given the purely research-oriented nature of the analysis, with, in the worst affected case, namely the access cost coefficient in the high income group, a probability of 7% of a wrongly-signed coefficient.

One final point that needs addressing is the effect of using a random distribution for the access cost coefficient for only part of the population. Indeed, by comparing the results to the MNL results, it can be seen that the mean value of the affected coefficient increases substantially (also from a relative point of view). This means that the mean cost-sensitivity in the high income group now exceeds the fixed cost-sensitivity in the low income group, which is not consistent with intuition. Attempts to account for heterogeneity in cost-sensitivity in the low income group led to a comparable increase in the mean sensitivity, hence redressing the balance. However, the associated dispersion coefficient was not statistically significant at any reasonable level of confidence, such that a fixed coefficient was used. This observation thus serves as an explanation for the counter-intuitive results.

MMNL model for business trips by visitors

In the model for business trips by visitors, significant random heterogeneity could be identified for three coefficients, namely the two coefficients associated with in-vehicle access time, and the coefficient associated with flight time. At the cost of these 3 additional parameters, the model offers improvements in LL over the MNL model by 5.68 units, giving a χ^2_3 p-value of 0.0099 for the associated LR-test. In this model, the effects of using the Normal distribution are at their worst, leading to a probability of a wrongly-signed coefficient of 33% for the flight time coefficient, such that extra caution is again required in the interpretation of the model results. Similarly, there is a 10% probability of a wrongly-signed coefficient for the access time coefficient in the low income group.

MMNL model for holiday trips by residents

The MMNL model for resident holiday travellers retrieves significant random taste heterogeneity for a single coefficient, namely the sensitivity to access time changes. The model obtains gains in model fit over the MNL model by 5.45 units, with an associated χ^2_1 p-value of 0.00096 for the LR-test. The effects of using the Normal distribution are again benign, with a probability of 3% of a wrongly-signed coefficient.

MMNL model for holiday trips by visitors

In the model for holiday trips by visitors, significant random heterogeneity could only be retrieved for a single coefficient, namely that associated with flight time. While the gains in model fit are again statistically significant, with an increase in LL by 7.67 units at the cost of just one parameter, the effects of using a Normal distribution are again more severe, with a probability of 30% of a wrongly-signed coefficient.

MMNL model for VFR trips by residents

The observations in the case of VFR trips by residents are very similar to those for resident holiday travellers, with significant amounts of random taste heterogeneity identified only for the access time coefficient, and statistically significant gains in LL by 6.73 units over the corresponding MNL model. In this model, the effects of using the Normal distribution are however more severe, with a probability of 11% of a wrongly-signed coefficient, where it is thus important to recognise that this is almost surely a simple artefact of the Normal distribution, and does not actually reveal the presence of respondents with negative VTTS for the access-journey.

MMNL model for VFR trips by visitors

In the final MMNL model, estimated on the data for VFR trips by visitors, significant levels of random taste heterogeneity were retrieved for 4 coefficients, namely that associated with access time, the two fare coefficients[22], and the frequency coefficient. This leads to an improvement in LL by 7.15, with an associated χ^2_4 p-value of 0.0064 for the LR-test. In terms of the effects of using the Normal distribution, the probability of a wrongly-signed value for the access time coefficient is 3%, with corresponding probabilities of 12% and 21% for the fare-coefficients in the low and middle income groups. Finally, for the frequency coefficient, the probability of a negative value is 18%. Again, these observations should be seen as an effect of using the Normal distribution, and would lead to requirements for extra care in policy-oriented research.

Summary of MMNL results

The presentation of the MMNL results has shown that, in each of the six population segments, there is sufficient variation in tastes across respondents to allow the mixture models to obtain statistically better model fit than their MNL counterparts. However, it should be said that, just as with the improvements offered by the nesting approaches, the gains in model fit are relatively modest. Nevertheless, the models provide some further insight into choice behaviour, and reduce the risk of biased trade-offs[23].

The discussion has also shown that the success of the mixture structures varies across population segments, with the number of random coefficients ranging from 1 to 4, where the common point in all models, except for visiting holiday travellers, is the prevalence of significant variations in the sensitivity to access time.

In closing, it should be stressed again that the aim of the present analysis was mainly one of exploration, and not one of providing adequate trade-offs for use in policy analysis. As such, the use of the Normal distribution was warranted, given its computational advantages. Nevertheless, it is important in this case not to misrepresent the findings in terms of implied bounds on the coefficients, but to acknowledge the potential impacts of the distributional assumptions on these results.

9.4.6 Comparison of substantive results

Even though the differences in model fit between the five structures are relatively modest, it is conceivable that the actual substantive results vary more significantly across models. To illustrate this, a brief analysis was conducted with the aim of comparing a common trade-off across models, as well as across population subgroups. The only two coefficients that are included in every single model are frequency and in-vehicle time, allowing the computation of the willingness to accept increases in access time in return for increases in flight frequency. A trade-off that is of more practical interest is the VTTS; however, in the present context, the calculation of reliable values for this measure is hampered by the use of different income segmentations, the absence of an access cost coefficient for visitors on VFR trips, and the unreliable access cost information[24].

[22] The standard deviation in the middle income group is significant only at the 93% level.

[23] This is not necessarily the case with the present specification, as it relies on restrictive distributional assumptions, by making use of the Normal.

[24] This is a common problem with RP data in the context of airport-access journeys, as illustrated also in the London case-study in Chapter 10.

The results of this analysis are summarised in Table 9.13, where, given that frequency enters the utility under a log-transform, K is used to represent the inverse of the current frequency, and where a sign change has been used to represent the willingness to accept *increases* in access time in return for *increases* in frequency. For the closed form models, the trade-off is simply given by the ratio of the two point-estimates, multiplied by K. The same applies in the MMNL model for holiday trips by visitors, where both frequency and access time were treated as fixed coefficients. However, in the remaining five population segments, the variation in the coefficients needs to be taken into account, especially for the access time coefficient, which forms the denominator of the trade-off. Here, a basic simulation approach was used, where the aim was simply to produce an estimate of the mean value for the trade-off. The process is made considerably easier by the fact that correlation between random coefficients was not taken into account. However, a major issue arises because the use of the Normal distribution, and specifically, the presence of significant shares of counter-intuitively signed coefficient values in some segments. The incorporation of such values in the simulation process would lead to significant bias in the calculated mean trade-offs (due to cancelling out effects). For this reason, the following approach was adopted. For each randomly distributed coefficient, a sample of $1,000,000$ random draws from the appropriate Normal distribution was produced. This was then censored to exclude counter-intuitively signed values, with the same censoring applied in the other tail, to guarantee that the symmetry remains unaffected. To ensure equally sized vectors for both coefficients involved in a trade-off, the maximum censoring across the two coefficients was used for both coefficients. In most cases, only the upper and lower few percentile points had to be removed, with the main exception being the model for VFR trips by visitors, where 18 percentile points had to be removed from either side. Given the use of this censoring approach, the estimated standard deviation for the trade-off is unreliable, such that only the mean values are presented in Table 9.13.

The first observation that can be made from Table 9.13 is that, overall, the results show a higher willingness to accept increases in access time for residents than for visitors. The differences are especially significant in the case of VFR trips, where the relative value of frequency increases is at its highest for residents, while it is at its lowest for visitors. In terms of purpose-related differences, the results suggest higher relative sensitivity to frequency for business travellers than for leisure travellers in the models for visitors, while for residents, frequency is valued less highly for holiday travellers. To some extent, these conclusions are however potentially influenced by the quality of the access-journey level-of-service data.

Given the aims of this study, the more interesting differences arise when comparing the results across model structures. Here, it is important to note that no major issues with parameter significance arose in any of the models for the coefficients used in the trade-offs, increasing the reliability of the comparisons. The results show that, although there is some overall consistency in the trade-offs, there are also some differences, for example when looking at the results for MMNL, which overall give greater weight to the frequency coefficient, highlighting the effects of accounting for random taste heterogeneity. However, there are also some differences between the MNL model and the NL models, and also across NL models, such as for example in the case of the model using nesting by access-mode for resident VFR

MNL

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
common	$25.28K$	$22.3K$	$29.47K$	-	$14.01K$	$10.38K$
Inc. $< \$21,000$	-	-	-	$26.32K$	-	-
Inc. $> \$21,000$	-	-	-	$15.93K$	-	-

Nesting by airport

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
common	$24.27K$	$19.81K$	$26.51K$	-	$12.74K$	$10.25K$
Inc. $< \$21,000$	-	-	-	$21.41K$	-	-
Inc. $> \$21,000$	-	-	-	$13.32K$	-	-

Nesting by airline

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
common	$26.35K$	$23.59K$	$32.98K$	-	$16.81K$	$11.38K$
Inc. $< \$21,000$	-	-	-	$27.58K$	-	-
Inc. $> \$21,000$	-	-	-	$16.74K$	-	-

Nesting by access-mode

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
common	$20.08K$	$19.05K$	$16.52K$	-	$12.45K$	$7.7K$
Inc. $< \$21,000$	-	-	-	$24.69K$	-	-
Inc. $> \$21,000$	-	-	-	$15.16K$	-	-

MMNL

	Resident			Visitor		
	Business	Holiday	VFR$^{(\dagger)}$	Business	Holiday	VFR$^{(\dagger)}$
common	$34.13K$	$25.03K$	$38.78K$	-	$12.5K$	$13.6K$
Inc. $< \$21,000$	-	-	-	$28.63K$	-	-
Inc. $> \$21,000$	-	-	-	$17.63K$	-	-

$^{(\dagger)}$ Visiting friends or relatives

Table 9.13: Trade-offs between flight frequency and access time (min/flight) in models for combined choice of airport, airline and access-mode ($K = 1/f$, with f giving current frequency)

travellers. Overall, these findings show that, although the differences between models in terms of LL may be relatively modest, the actual substantive conclusions are quite different. This highlights a relative flatness of the LL function, but also shows the impact of model structure on substantive results, making it an important issue in policy-oriented research.

Although this study does not aim to produce reliable trade-offs for use in policy analysis, it is worth noting that, at the average observed flight frequency (in the data) of 10 flights, the resulting value of K (0.1) leads to very low willingness to accept access time increases in return for increases in flight frequency. This should however be put into context by noting that the average observed access time was only around 30 minutes. Finally, the high values for K in the case of routes with low frequency (e.g. at $f = 2$, $K = 0.5$) imply a willingness to accept significant increases in access time in return for increases in flight frequency on routes with big gaps between individual flight departures[25].

9.4.7 Model validation

To complete the analysis, the five sets of models were applied to the validation sample of 519 observations, which was divided into sets of 114, 93 and 74 observations for resident business, leisure and VFR travellers respectively, and sets of 119, 54 and 55 observations for visiting business, leisure and VFR travellers respectively. For each of the models, the final coefficient values produced during the estimation process were used in the apply runs. On the basis of this, the validation approach produces, for every observation, a choice probability for each of the 144 triplets of alternatives; this can be used to calculate the average probability of correct prediction for the actual chosen alternative across respondents. Aside from this probability for the choice of the actual triplet of airport, airline and access-mode, it is also of interest to look at the probability of correct choice for just the airport, just the airline, and just the access-mode. These probabilities can be obtained through summing the probabilities of the single elementary alternatives falling into the given group.

A separate analysis reported by Hess & Polak (2004*b*) looked at the recovery of market shares for the different airports, airlines and access-modes, indicating very good performance across all models used. It can be seen that this measure is highly correlated with the performance in terms of correct prediction probability[26], and as such, these results are not reproduced here.

The results of the validation process are summarised in Table 9.14. The first observation that can be made from this table is the surprisingly high probability of correct prediction of the actual chosen alternative. Indeed, even in the population segment with the worst performance (holiday trips by visitors), the probability of correct prediction is close to 30%, which is very high when one takes into account the extent of the choice set, where, on average (across choice-situations), 31 of the 144 alternatives were available.

In terms of the correct prediction of airport choice, the probabilities range from 68.51% to as high as 85.39%. This performance compares well with the results in other studies, and the rates obtained in some of the models in fact exceed those obtained in previous studies. As an example, in one of the more recent studies in the SF-bay area, Basar & Bhat (2004) obtain an average correct prediction rate of 72.9% on their validation sample.

The performance in terms of the choice of access-mode is also very good, although it is significantly lower than the performance along the airport dimension for residents on VFR trips, while it is also slightly lower for visiting holiday travellers. On the other hand, it is marginally better than the performance along the

[25]Where frequency is again used as a proxy for the gap in departure times between flights.

[26]The market shares are calculated on the basis of the estimated choice probabilities and the weights used to correct for sampling effects. With higher probabilities of correct prediction, the *unweighted* market shares approach the sample market shares. As such, it can be seen that, after reweighting, which uses the weights to relate the actual real-world market shares to the sample shares, the market shares in a model with high correct prediction will be less biased than in a model with lower correct prediction rates.

MNL

	Resident			Visitor		
	Business	Holiday	VFR(†)	Business	Holiday	VFR(†)
Choice	47.13%	30.56%	36.58%	34.33%	27.21%	36.83%
Airport	84.04%	69.58%	80.83%	70.69%	69.53%	73.20%
Access mode	84.04%	67.72%	66.47%	70.18%	63.22%	77.08%
Airline	60.68%	54.93%	60.26%	55.39%	53.31%	60.97%

Nesting by airport

	Resident			Visitor		
	Business	Holiday	VFR(†)	Business	Holiday	VFR(†)
Choice	48.02%	31.39%	36.74%	36.19%	28.97%	36.81%
Airport	83.69%	69.16%	80.07%	70.69%	68.51%	73.13%
Access mode	85.22%	68.91%	67.50%	72.39%	66.41%	77.25%
Airline	61.06%	55.03%	60.08%	55.90%	54.34%	60.73%

Nesting by airline

	Resident			Visitor		
	Business	Holiday	VFR(†)	Business	Holiday	VFR(†)
Choice	47.90%	31.82%	36.50%	35.00%	27.78%	36.93%
Airport	84.18%	70.24%	80.36%	71.21%	68.61%	73.26%
Access mode	84.92%	68.64%	67.26%	71.08%	64.24%	76.96%
Airline	60.30%	54.79%	59.41%	55.27%	51.60%	60.52%

Nesting by access-mode

	Resident			Visitor		
	Business	Holiday	VFR(†)	Business	Holiday	VFR(†)
Choice	48.41%	31.38%	39.60%	34.65%	27.78%	37.83%
Airport	85.39%	70.98%	84.97%	71.11%	72.41%	74.46%
Access mode	83.76%	67.29%	66.16%	70.25%	62.11%	76.98%
Airline	61.33%	55.46%	61.36%	55.49%	53.49%	61.04%

MMNL

	Resident			Visitor		
	Business	Holiday	VFR(†)	Business	Holiday	VFR(†)
Choice	47.53%	30.89%	37.45%	34.52%	27.73%	37.59%
Airport	84.21%	69.60%	81.59%	70.71%	68.78%	73.81%
Access mode	84.09%	67.77%	66.80%	70.21%	63.18%	77.15%
Airline	60.85%	55.07%	60.77%	55.52%	53.08%	60.69%

(†) Visiting friends or relatives

Table 9.14: Prediction performance of NL models on SF-bay area validation data

airport dimension for visiting VFR travellers. The variation in performance does suggest that the choice process is less deterministic in some segments than in others, but could also be an indication that the data problems in terms of the availability of the car mode, and the lack of information on parking behaviour, play a bigger role in some segments than in others.

The performance of the models in terms of correctly predicting the choice of airline is poorer than that for the choice of airport and access-mode; however the values still always exceed 50%, despite the complete absence of a treatment of airline allegiance. Again, superior performance could be expected if better data were available, notably with regards to fare structures and frequent flyer programmes.

In terms of differences between population segments, the best average performance across all choice dimensions is obtained for resident business travellers, where an argument can be made that such travellers behave in a more rational manner (from the modeller's point of view), due to better information. The comparatively poor performance of the models for holiday trips can partly be explained by the fact that at least some of the travellers on such trips have purchased a package holiday (or special flight deal); for such deals, the choice of departure airport and airline is potentially influenced by factors that were not directly measurable and could thus not be included in the models.

Given the relatively modest differences in performance between the five model structures in the actual estimation processes, it should not be surprising that there are no systematic differences in prediction performance on the validation sample. Even though there are some outliers, such as the performance of the NL model using nesting by access-mode in the models for residents on VFR trips[27] (prediction of actual choice, and prediction of choice of airport), the average differences in performance are too small to come to any conclusions in terms of advantages for one of the model structures. This is further reinforced by the fact that it is not directly clear what measure of error should be associated with these probabilities.

A final aim of a validation process is to establish whether the models have not been overfitted to the estimation data. Tests by Hess & Polak (2004*b*) which involved applying the models to the actual estimation sample produced very similar performance to that obtained here with the validation sample, suggesting that the models have indeed not been overfitted to the estimation data.

9.5 Summary and Conclusions

This chapter has presented the findings of the case-study conducted in the SF-bay area. The study had three main aims, as set out in Section 9.1; an analysis of the advantages of accounting for the multi-dimensional structure of the choice process, a study of the correlation along the three dimensions of choice, and an investigation into the prevalence of random taste heterogeneity.

In a direct comparison between the results from Section 9.3 and Section 9.4, it is important to recognise the bigger sample sizes in the common leisure models in the single-dimensional analysis, which will likely lead to more stable analysis. Nevertheless, when looking at the holiday segment, the performance along the airport

[27]For this model, there were also significant differences in the trade-off in Table 9.13.

dimension in the multi-dimensional study is comparable to the performance in the overall leisure segment in the one-dimensional models, while that in the VFR segment is better. The biggest indication of the advantage of the more disaggregate approach is given when looking at the business segments; here, the multi-dimensional models produce rates between 83.69% and 85.39% for residents, with rates between 70.69% and 71.21% for visitors. While the performance for visitors is comparable in the one-dimensional models, at 67.00%, the performance for residents is much poorer, at 67.60%. Although the more detailed utility specification used in the disaggregate models (e.g. inclusion of inertia coefficients) can in part account for these improvements, the difference in performance is such that it can indeed be suggested that important gains can be made by using dimension-specific level-of-service information in the modelling of air-travel behaviour (i.e. avoid the use of measures of overall service at an airport). It should also be noted that the disaggregate approach has advantages in terms of the important insights it provides into choice behaviour along the additional dimensions of airline and access-mode choice. Finally, in a forecasting application, the multi-dimensional approach has the potential to produce market shares for specific airlines or access-modes, in addition to combinations of alternatives, such as the market share of a given airline at a specific airport.

The investigation into the use of nesting and mixture structures has shown that both approaches can lead to statistically significant gains in model performance. Although the gains are very modest, which is reflected in the similarity in validation performance across models, the differences in the estimated trade-offs (cf. Section 9.4.6) do suggest that there are some significant differences in the substantive results across models. Additionally, it should be stressed again that the advanced models are more intuitively correct, and as such should be preferred. This is particularly important in the context of forecasting applications, where the differences across models, especially in the case of nesting structures, can be expected to be quite significant. Finally, the more complicated models also provide useful further insights into choice-behaviour, in the form of substitution patterns, but also in terms of an indication of the differences across respondents in their response to changes in level-of-service variables.

In terms of actual substantive results, the analysis has revealed significant effects of access time and frequency, across all population groups. Here, it is important to put the findings in terms of frequency into context by remembering that this coefficient potentially captures a host of effects, including visibility and scheduling convenience. In common with many previous RP studies of air-travel choice behaviour, it was, in the present analysis, not possible to retrieve a significant effect of air-fares across all population subgroups. While, in the case of business travellers, this could be an indication of actual low fare sensitivity, it is more likely that the problems with the data, in terms of availabilities as well as disaggregate air-fare information, are the main reason for this result.

Chapter 10

Greater London case-study

10.1 Introduction and context

This chapter describes the case-study conducted in Greater London, an area which has by far the highest levels of air traffic in Europe, with, in 2002, some 117.13 million passengers using the five main airports. The area is dominated by Heathrow (LHR), the world's busiest international airport (measured in terms of the number of passengers on international routes), and the main hub in Europe. Additionally, a large number of routes are offered from Gatwick (LGW), the world's busiest single-runway airport, while Stansted (STN), Europe's fastest growing major airport, and Luton (LTN) act mainly as bases for holiday and low-cost operators. Finally, the centrally located London City (LCY) airport caters primarily to business travellers, and, due to its short runway, is restricted to short-haul flights operated by turboprop planes and small jet aircraft. The geographical location of the five airports, with respect to central London, is illustrated in Figure 10.1, which additionally shows the main road and rail links serving London, and, by extension, the different airports.

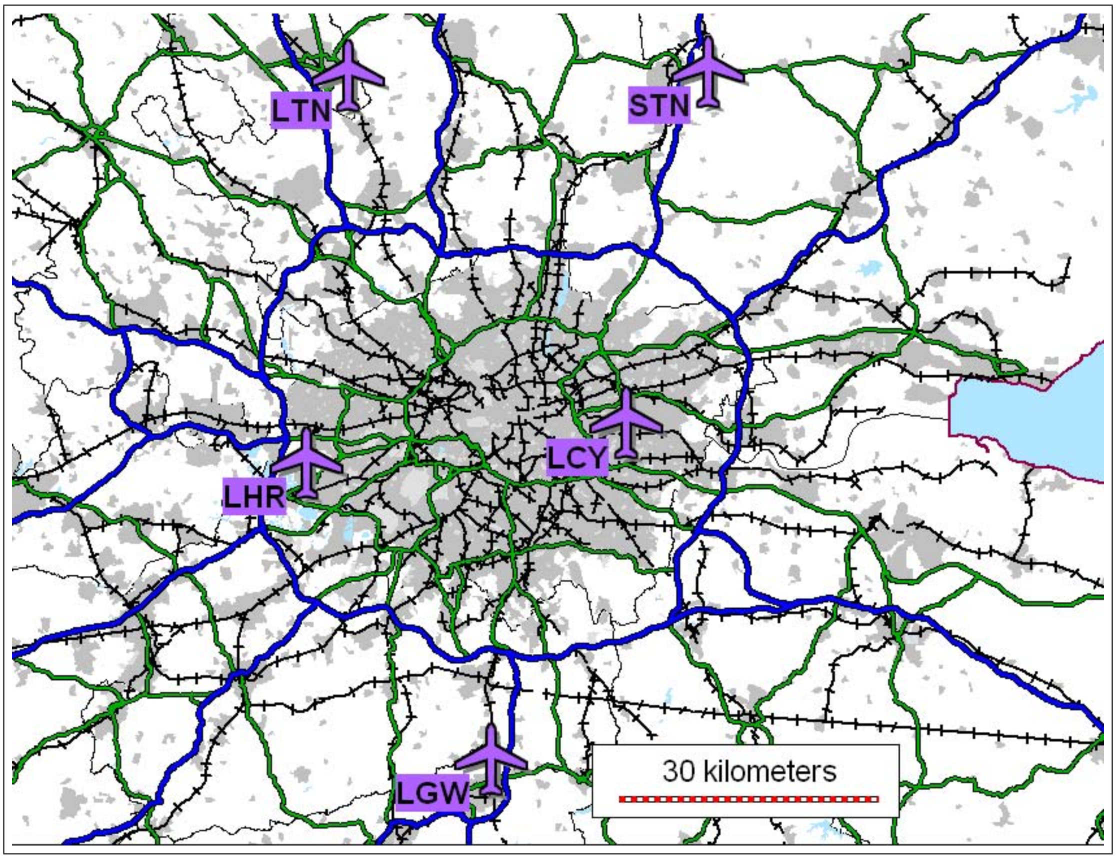
Table 10.1 shows the annual number of passengers handled[1] at the five London airports between 1990 and 2004. The figures show that LHR is easily the busiest of the five airports (ahead of LGW), capturing more than half of the total number of passengers. At the same time, the passenger counts however also reveal that, while still experiencing growth in traffic (with a few exceptions, notably LHR in 2001, and LGW in 2002), the two main airports have seen their share of the market reduced, while that of STN has grown dramatically, averaging an annual growth rate of over 21% between 1997 and 2004[2]. This is due mainly to increases in activities by low-cost airlines at this airport. As such, given the continuing discussion of induced *vs* shifted demand[3], it is unclear whether the gain in market share should be seen as a draw away of passengers from LHR and LGW, or the development of new demand. Given the extent of the increase, it is conceivable that both phenomena apply.

Even though London, and especially LHR, has always had the largest share of

[1] The counts include transit passengers, i.e. passengers who do not change aircraft during their stopover at the airport. These can be argued not be *handled* by the airport, but play an insignificant role in the totals in any case, accounting for under 1% of all passengers.

[2] The corresponding rate of around 2% at LHR and LGW is biased downwards by the effects of September 11^{th}, which were felt much more heavily at these airports.

[3] Dennis (2004) suggests that around 40% of low-cost traffic is generated by the airlines, with the remaining 60% shifted away from network and charter carriers.



©2004, Saturn Technologies, United Kingdom

Figure 10.1: Map of Greater London, with main airports and ground-level transport network

air-passengers in the UK, it is still striking to note that the number of passengers handled at LHR in 2002 (63.4 million) is higher than the total annual number of passengers handled at UK airports up to 1985 (61.6 million). The importance of the London area is further underlined by the fact that the total number of passengers handled at the five London airports in 2004 (128.9 million) accounts for 58.8% of the total number of passengers handled at all UK airports (219.2 million).

Like in many other major aviation centres, the effects of September 11^{th} and the global economic downturn on air-traffic were also felt in London. However, the effects were concentrated in certain subsets of the market, and thanks to the success notably of STN, overall traffic decreased by only 2.2% in 2001, and recovered a strong growth rate afterwards, reaching 7% in 2004.

Forecasts show that air-travel in the United Kingdom can be expected to continue growing at a very high rate. As an example, forecasts produced by the Department for Transport in 1997 predicted almost a doubling in air passenger numbers by the year 2010 (cf. DfT 1997). This study used 1995 as the base, with 129.6 mppa, and predicted figures of 167.2 mppa, 205.3 mppa, 253.1 mppa and 310 mppa for the years 2000, 2005, 2010 and 2015 respectively (summed over all UK airports). This thus assumes decreasing annual growth rates, ranging from 5.2% between 1995 and

Year	LHR		LGW		STN		LTN		LCY		Total
2004	67.343	52.24%	31.453	24.40%	20.911	16.22%	7.536	5.85%	1.675	1.30%	128.918
2003	63.495	52.70%	30.005	24.90%	18.722	15.54%	6.797	5.64%	1.471	1.22%	120.49
2002	63.362	54.09%	29.627	25.29%	16.055	13.71%	6.487	5.54%	1.602	1.37%	117.133
2001	60.765	53.40%	31.182	27.40%	13.665	12.01%	6.555	5.76%	1.619	1.42%	113.786
2000	64.62	55.54%	32.069	27.56%	11.88	10.21%	6.191	5.32%	1.584	1.36%	116.343
1999	62.268	57.15%	30.564	28.05%	9.447	8.67%	5.285	4.85%	1.386	1.27%	108.95
1998	60.684	59.37%	29.173	28.54%	6.863	6.71%	4.133	4.04%	1.36	1.33%	102.213
1997	58.185	61.27%	26.959	28.39%	5.427	5.71%	3.239	3.41%	1.161	1.22%	94.971
1996	55.982	63.46%	24.226	27.46%	4.849	5.50%	2.429	2.75%	0.726	0.82%	88.211
1995	54.407	65.37%	22.510	27.05%	3.913	4.70%	1.843	2.21%	0.554	0.67%	83.227
1994	51.666	65.89%	21.142	26.96%	3.289	4.19%	1.832	2.34%	0.478	0.61%	78.407
1993	47.954	65.71%	20.217	27.70%	2.702	3.70%	1.865	2.56%	0.245	0.34%	72.984
1992	45.305	64.88%	20.008	28.66%	2.362	3.38%	1.963	2.81%	0.186	0.27%	69.824
1991	40.563	64.01%	18.912	29.85%	1.739	2.74%	1.982	3.13%	0.172	0.27%	63.367
1990	42.972	62.89%	21.209	31.04%	1.217	1.78%	2.696	3.95%	0.230	0.34%	68.324

Table 10.1: Annual passenger counts (including connecting passengers) at London's five main airports, 1997-2004 (mppa)

2000 to 4.1% between 2010 and 2015. Although an upper limit of 180.6 million was given for the forecasted passenger numbers for the year 2000, the fact that passenger numbers in 2000 actually topped 181 million shows that such forecasts are likely to underestimate the growth in passenger numbers. A revised version of this forecast, produced in 2000, predicts an increase in passenger numbers to 2.5 times the level observed in 1998, by the year 2020 (c.f DfT 2000). Finally, it has been suggested that, with unconstrained growth, the annual number of passengers could rise to around 500 million by the year 2030, with around 300 mppa for the South East airports (cf. DfT 2003*a*). This forecast assumes a declining growth rate, especially for the South East, as the market becomes saturated. Even so, the South East would still have around 60% of the total number of passengers, with most of this distributed across the London airports.

The forecasts show that, in the case of unconstrained growth, London would be likely to further strengthen its role as one of the world's most important multi-airport regions, keeping its position as the prime hub in Europe. However, given the limits in capacity, unconstrained growth is clearly a purely hypothetical situation and merely reflects the potential in terms of demand. Given that demand at the London airports, especially at LHR, already exceeds capacity, concerns have been raised that London could lose its status as the main European hub, given the extra capacity available in competing regions, such as Paris, Frankfurt and Amsterdam. For that reason, a major consultation has taken place to consider different schemes for expanding airport capacity in the South East, and especially in the London area (cf. DfT 2003*a*).

One major airport expansion scheme is already in progress, with the construction of a fifth terminal at LHR, which should increase the airport's capacity to 89 mppa by 2008. Assuming an annual increase in passenger numbers by 4%, the new capacity limit would however be reached already within 4 years of the opening of the new terminal. It should be noted that, given the additional constraints in terms of the maximum number of possible take-offs and landings (especially in the absence of mixed-mode operations), such growth rates would at some point in the near future only be possible with the use of larger aircraft. In any case, both restrictions (runway and terminal capacity) lead to a need for further increases in capacity.

Several options for airport expansion in the Greater London area were considered in the consultation. The most popular proposal with airlines was the construction of a third (short) runway at LHR, which would increase capacity to 116 mppa (DfT 2003*a*); this project is however facing fierce opposition by local residents, who, on the basis of the Terminal 5 Inquiry Inspector's report, had understood that no further runways would be built. Another possibility, which would not require any new building work, is the introduction of *mixed mode* operations at LHR, allowing both runways to be used simultaneously for take-off and landing (whereas currently one runway is used for take-offs and one runway is used for landings). This would not only improve runway utilisation and hence capacity, but would also alleviate taxi-way congestion, and it is expected that it would allow an increase by 10% in the number of take-offs and landings. Public opposition to this scheme is also very high, given the added air, and especially noise pollution that would result from such a move[4].

A major problem in the search for alternative ways of increasing capacity is the agreement signed in 1979 between the British Airports Authority (BAA) and West Sussex County Council that a second runway at LGW would not be built before 2019. While the use of larger planes could increase capacity from a current 40 mppa to 46.5 mppa (DfT 2003*a*), this agreement is a major hindrance. Estimates by DfT (2003*a*) show that capacity could be increased dramatically, were the construction of new runways considered. Indeed, the construction of a second (close parallel) runway would increase capacity to 62 mppa, while the construction of an additional wide-spaced runway would increase capacity to 83 mppa. Although initially excluded from the consultation, a High Court ruling in November 2002 (in favour of Medway, Kent County, and Essex County councils) judged that these options should not be excluded from consideration, still leaving the door open for expansions of capacity at LGW.

For STN, DfT (2003*a*) estimate that capacity could be increased to $25-35$ mppa (depending on success of planning applications) without new runway development, while capacity could be increased to 82 mppa, 102 mppa and 129 mppa respectively with one, two or three additional runways. The capacity of LTN is currently set at 10 mppa; DfT (2003*a*) estimate that replacing the existing runway by a new runway could increase this to 31 mppa. Given its location and layout, no major new development can be expected for London City airport, and DfT (2003*a*) estimate that capacity will be reached by 2030.

DfT (2003*a*) also consider other airports, at Norwich and Southampton for example; given their size, these are however of little importance for the present research. Finally, the government also considered the development of a new four-runway airport at Cliffe in North Kent, which could have had a capacity of 113 mppa. However, these plans have now been rejected, mainly due to environmental concerns, where

[4]Under the present arrangements, people living at the end of either runway are affected by take-off and landing-related noise only during 50% of the time. Given that the westerly runways are generally in use, landing-related noise is the biggest factor, given the density of housing in the areas east of LHR.

there were also concerns about the heightened risk of bird-strike related accidents given the large bird population in this area (cf. Bell et al. 2003).

The main consultation is now closed, and the recent government *White paper* (DfT 2003*b*) has recommended the construction of a single new runway at STN by 2012. There are also plans to extend the capacity of LHR between 2015 and 2020 with the construction of a new (short) runway and possibly a sixth terminal. However, new EU limits on pollution will come into effect in 2010, and it is not clear whether any further expansion at LHR would be possible without violating these constraints. For this reason, the construction of a second runway at LGW, after 2019, is still kept open. The options of multiple new runways at STN and LGW have been rejected, as has the option of runway replacement at LTN, and, as mentioned above, the development of the Cliffe airport.

The major airlines have expressed their satisfaction at the decision to include a new runway at LHR in the plans. However, they also remain fiercely opposed to the idea of taxes at LHR being used to cross-subsidise the developments at STN, and have indicated that they have no desire to shift a significant share of their traffic from LHR to STN. As such, deliberations are set to continue, and it is thus still of interest to gauge the attractiveness of the different airports, and to analyse how the attractiveness of airports with additional capacity could be improved (e.g. by cutting access time). This makes the London area a prime candidate for a study of air-travel choice behaviour in multi-airport regions. The congestion in the ground-level network also makes the analysis of airport-access an important topic. Furthermore, unlike in many other areas that have been the topic of studies of airport choice, there are very high levels of competition between the different airports, and lower captivity by specific geographical areas to a given airport, due their arrangement at roughly equal distances from the centre of London (aside from LCY)[5]. Finally, unlike in the case of studies in the US, where the market share of car can exceed 75% (as in the SF-bay area), the modal split for the access-journey is more diverse, increasing interest in the analysis of choices along that dimension.

Aside from conducting a study of the joint choice of airport, airline and access-mode in the London area, the main aim of this chapter is to explore the potential of cross-nesting structures for the joint analysis of correlation along the three dimensions of choice. The discussion presented in this chapter is limited to closed-form models, mainly on the grounds of estimation complexity in the presence of the very large sample sizes (and choice sets), but also given the extensive treatment of random taste heterogeneity in Chapters 9 and 11. The use of mixture models on the London data remains an avenue for future research, including in a GEV mixture context, given the conclusions with regards to correlation presented in this chapter.

The remainder of this chapter is organised as follows. Section 10.2 describes the data used in the analysis, Section 10.3 discusses model specification, with the estimation results summarised in Section 10.4. The chapter concludes with a model validation exercise in Section 10.5, and a presentation of the conclusions in Section 10.6.

[5]See also Figure 10.1

10.2 Description of data

10.2.1 Air-passenger survey data

For the present analysis, data from the 1996 passenger survey[6] were obtained from the Civil Aviation Authority (CAA 1996). Although slightly dated, this is the most recent large-size full survey available for this region (containing data collected across an entire year, as opposed to being limited to a few months). The dataset also has the advantage that the effects of September 11^{th} need not be taken into account. On the other hand, the age of the data prevents a detailed analysis of the impact of low-cost carriers on air-travel choice behaviour, given that their operations in 1996 were far more limited than is the case nowadays. It should also be noted that the analysis of the access-mode choice dimension is simplified by the fact that the premium *Heathrow Express* service only started its operations in 1998. The use of a more recent version of the dataset is an important avenue for future research.

The original sample obtained from the CAA contained responses from $47,831$ passengers, for 31 destinations (reachable by direct flights from at least two of the five London airports), and 37 airlines. After data-cleaning (missing data, compatibility between datasets), a usable sample of $33,527$ passengers was obtained. This compares favourably to the sample of $5,091$ available for the SF-bay analysis (cf. Section 9.2.1). For the present analysis, the sample was split into four subsets, dividing the population into residents[7] and visitors, and using a purpose split of business *vs* leisure. This led to samples of $7,059$ resident business travellers, $8,704$ resident leisure travellers, $7,587$ visiting business travellers and $10,177$ visiting leisure travellers. In each sub-group, a 95% subsample was used for model calibration, with the remaining observations retained for model validation. Additional subdivisions of the business or leisure groups did not lead to any gains in explanatory power.

Of the 31 destinations used in the analysis, 5 are in Great Britain (Aberdeen, Edinburgh, Leeds, Manchester and Newcastle), 1 on the Channel Islands (Guernsey), 3 in Ireland (Cork, Dublin and Shannon), 3 in the Benelux (Amsterdam, Brussels and Rotterdam), 3 in Scandinavia and the Nordic countries (Copenhagen, Gothenburg and Helsinki), 3 in Germany (Düsseldorf, Hamburg and Munich), 3 in Austria and Switzerland (Geneva, Vienna, Zurich), 1 in France (Nice), 3 in Spain (Barcelona, Madrid, Malaga), 2 in the South East of Europe (Athens and Larnaca), 1 in the Middle East (Tel Aviv), and 3 in the United States (Boston, Detroit and Miami). The most popular destination for business travellers in the sample is Amsterdam, ahead of Edinburgh, Dublin and Brussels. For leisure travellers, the most popular destination is Dublin, ahead of Amsterdam, Edinburgh and Boston.

All destinations included in the sample are served by a single main airport, avoiding the problem with multi-airport destination areas described in Section 9.2.1. This was one of the main factors used in the selection of appropriate destinations, a luxury that was allowed by the large sample size. Nevertheless, some destinations remain where there is competition between air and ground-level transport, namely the 5

[6]Data collected from departing passengers at the airports.

[7]Respondents were considered as residents if they reside in the Greater London area and immediately adjacent counties for domestic and short-haul European flights, while respondents on medium-distance European flights and intercontinental flights were considered as residents if they reside in Great Britain.

destinations in Great Britain (especially the three in England), and Brussels, where there is competition with Eurostar. As described in Section 8.4.1, this competition is not taken into account in the present study, where we work on the basis of an a priori choice of main mode.

The dataset is summarised in Table 10.2, giving the number of passengers for each destination in the four subsamples (prior to the further division into estimation and validation samples). The passenger counts show great variations across destinations, along with some variations across the four subgroups. From Table 10.2, it can be seen that a sufficient number of observations are included for each destination, with the exception of Detroit. The very low number of observations for this destination is a direct result of incompatibilities between the survey and level-of-service datasets for this destination[8], leading to the exclusion of a high number of observations. The 4 remaining observations were retained in the analysis, to avoid having to restructure the choice set. Their inclusion does not lead to any bias in the results, given that the actual analysis is not destination-driven[9].

10.2.2 Air-travel level-of-service data

Air-side level-of-service data were again obtained from BACK aviation. The dataset contains daily airline-specific information[10] for all routes used in the analysis, including information on flight frequencies, departure times, flight times (block times, thus taking into account airport congestion), aircraft types used and available seat capacity. The main item of information missing from this dataset is that of the fares for the different routes and airlines. Such data were compiled from two sources; the International Passenger Survey[11] (ONS 1996) and the fare supplement of the Official Airways Guide for 1996 (OAG 1996). Information on travel-class as well as ticket type (single or return) was taken into account in assembling the data. As was the case with the fare data used in the SF-bay study, the resulting dataset is of highly aggregate nature[12], leading to similar problems in the estimation of the marginal utility of air-fares. Again, no information is available on frequent flier programmes. However, unlike in the case of the SF-bay area study, the inclusion of international flights allows for an analysis of allegiance to the national carrier. Finally, the dataset was completed by adding in information on-time performance, obtained from the Civil Aviation Authority[13].

[8] No such problems were encountered for any of the other destinations.

[9] The choice of destination is not modelled in this analysis. It only plays a role in the generation of the choice sets along the other dimensions.

[10] Information on the dates of operation of individual flights was used to compile disaggregate information for each single day in the year 1996.

[11] It should be noted that there are potential problems of endogeneity in using data on fares actually paid, given the likelihood that passengers choose the cheapest fares. However, in the face of incomplete listed fare data, this is not avoidable, and the same problem occurs with information from the *10% sample*, as generally used in air-travel modelling studies in the US.

[12] The ONS dataset was made available at the individual passenger level for a high number of respondents, such that statistics on the distribution of fares could be calculated. However, given the lack of such detailed information for national flights and some international destinations, where only aggregate OAG data were available, the decision was taken to use only the mean fares, for reasons of consistency.

[13] www.caa.co.uk

	Resident		Visitor		
	Business	Leisure	Business	Leisure	TOTAL
Aberdeen	294	218	315	283	1,110
Amsterdam	1,074	811	1,266	1,253	4,404
Athens	120	222	84	246	672
Barcelona	117	331	75	150	673
Boston	206	518	176	414	1,314
Brussels	447	136	453	195	1,231
Copenhagen	121	122	213	430	886
Cork	55	326	61	280	722
Detroit	0	1	1	2	4
Dublin	714	1,628	819	1,681	4,842
Düsseldorf	217	160	243	273	893
Edinburgh	752	542	813	513	2,620
Geneva	273	319	248	454	1,294
Gothenburg	57	35	96	128	316
Guernsey	123	417	119	265	924
Hamburg	128	100	200	203	631
Helsinki	68	28	47	71	214
Larnaca	44	310	39	160	553
Leeds/Bradford	65	23	93	22	203
Madrid	237	462	214	412	1,325
Malaga	51	398	20	61	530
Manchester	429	80	394	90	993
Miami	57	242	35	77	411
Munich	237	191	287	472	1,187
Newcastle	307	76	219	78	680
Nice	124	290	80	233	727
Rotterdam	302	55	391	287	1,035
Shannon	38	176	52	199	465
Tel Aviv	56	161	75	231	523
Vienna	103	99	164	520	886
Zurich	243	227	295	494	1,259
TOTAL	7,059	8,704	7,587	10,177	33,527

Table 10.2: Passenger counts in survey data, by data-subset and destination

10.2.3 Ground-access level-of-service data

For the analysis of the ground-level choice dimension, data from the National Airport Access Model (NAAM) were obtained for the base year 1999 (Scott Wilson Kirkpatrick 1999). Corresponding cost information for 1996 was produced with the help of the retail price index, while assuming that relative travel times have on average stayed constant. This dataset contains level-of-service information for travel between 455 different travel area zones and the five airports. Six different modes are considered in the analysis; private car, rental car, public transport (rail, bus, local transport), long-distance coach, taxi and minicab (MC). Respondents observed to have used hotel shuttles were again excluded from the analysis, for the same reasons as given in Section 9.2.1. The use a high level of disaggregation in the non public transport modes alongside aggregated public transport information might be criticised given the continuing focus on competition between premium dedicated airport rail services and other forms of public transport (e.g. Gatwick Express versus local train services, airport coach services versus local bus and Tube). The division used in the present analysis reflects the highest common factor between the survey and level-of-service datasets, and it is hoped that future work can rely on a higher level of disaggregation. The NAAM dataset did not contain information on taxi and minicab services; this was produced independently, on the basis of data for the year 2004, with appropriate transformations to obtain usable data for 1996. In terms of availability, taxi and minicab are assumed to be available for all possible ground-level and airport combinations, while the availability of public transport (PT) and long-distance coach (LDC) was determined on the basis of the NAAM data. Finally, rental car is assumed to be available to all travellers above the age of 18 (in the absence of license-holding information), while car is assumed to be available to all residents, and those visitors who chose it (in the absence of information on the availability of kiss-and-fly options for such travellers).

No combinations of modes were considered in the present analysis, and the final mode indicated in the survey was used as the chosen mode. This is a major simplification of the actual choice process, given the high incidence of access-journeys using a combination of different modes. However, in the absence of detailed route choice information, this simplification was not avoidable. Clearly, there is a risk that this approach can lead to biased results, and this needs to be taken into account in the interpretation of the findings of the study.

For each mode, information was included on travel time, wait time, and the number of interchanges (where appropriate). For the cost information, a fixed one-day charge of £35 was used for rental cars (in the absence of cost-bearing party information) in addition to marginal running costs (fuel only), while fare information was used for PT, LDC, taxi and minicab. For private car, two specifications were retained, one using only the marginal running costs in terms of fuel, with a second also including depreciation. Finally, the dataset was completed by adding parking cost information for the different airports, for short as well as long-term parking, where this was computed on the basis of current parking fees, which were transformed to 1996 levels using the change in the retail price index.

10.3 Model structure and specification

Several important issues relating to model specification deserve some further attention. These relate to the definition of the choice set, the re-weighting of the sample for model estimation, the specification of the constants used in the model, and the way attributes enter the utility function.

10.3.1 Choice set

With the use of 5 departure airports, 37 airlines, and 6 access-modes, a total of 1,110 combinations of airports, airlines and access-modes arise. However, not all airlines operate from all airports, and the total number of airport-airline pairs is actually 54 (instead of 185), which reduces the number of alternatives (airport, airline, access-mode triplets) to 324, compared to 144 in the SF-bay area study[14]. The number of available alternatives for specific individuals in the estimation sample ranges from 6 to 58, with a mean of 31. The approach used in assembling the utilities is the same as that described in the SF-bay study (cf. Section 9.4.1), using combinations of sub-utilities for the 54 airport-airline pairings, and the 30 airport-access-mode pairings.

10.3.2 Re-weighting of survey data

Given that the survey data are choice-based, some form of re-weighting needs to be performed in order for the estimation to represent population-level market shares as opposed to sample-level shares (influenced by survey quotas), thus avoiding biased results. In the present analysis, multiplicative weights were once again used in the specification of the log-likelihood function, where, for a given respondent, the weight is proportional to the ratio between the population weight and the sample weight for the corresponding group, and where group allocation was based on a host of criteria, dominated by route and airline choice. In the case of the GEV models discussed in this section, the correction can in fact be performed at the level of the ASCs when a full set of constants is used (cf. Bierlaire et al. 2003), but given the unconventional specification of constants used in this analysis (cf. Section 10.3.3), and the aim to reuse the models inside later mixture structures, preference was given to a weighting approach, in conjunction with the use of robust estimators.

10.3.3 Specification of constants

An important question arises with regards to the specification of the constants in the model. In one-dimensional choice processes, a single ASC is associated with each alternative, with all but one of the constants being estimated (normalisation ensuring identification). In the case of multi-dimensional choice processes, the situation becomes slightly more complicated. In the SF-bay study, a separate set of ASCs was used in each of the three choice-dimensions, with one normalised ASC in each group. The problem with this approach is that it ignores the potential impact of interactions between the choice-dimensions. To address this deficiency, an alternative specification was attempted in the London study, using a single constant for each airport-airline pair. This increases the total number of airport and airline related constants from 42 (37 airlines at 5 airports) to 54, of which 53 are estimated. Separate experiments showed that this approach leads to very significant gains in model performance, suggesting some interaction between choice dimensions. While it is in theory possible to further improve the specification by using a separate constant for each airport, airline and access-mode triplet, the gains from this approach are no longer significant, coming at the cost of an increase in the number of constants

[14]A similar simplification (which complicates the coding for automatically generating utility functions) of the choice set was not used in the SF-bay study, where the number of alternatives was at an acceptably lower level. Furthermore, of the 24 possible airport-airline combinations in the SF-study, 20 were actually in use, such that the choice set would only have reduced from 144 to 120.

from 60 to 324 (respectively 58 to 323 estimated constants), equating to a full set of constants. Furthermore, this approach again led to severe problems with identification, for the same reasons given in Section 9.4.2[15]. Attempts to use airport-access constants in combination with separate constants in the airline dimension also led to gains in model fit, which were however less significant than those obtained with the airport-airline specification, which was thus retained. It can be seen that the approach used here thus again leads to a violation of the zero mean assumption for the unobserved part of utility, where this was however again not avoidable (cf. Section 9.4.2). Finally, it should be noted that, depending on the population segment, the number of estimated constants can be lower than 58, in the case where some options are never chosen (or available).

10.3.4 Non-linearities

The final point that deserves some discussion is the way in which explanatory variables enter the utility function, in terms of the use of non-linear transforms. In the present analysis, like in the SF-bay case-study, the log-transform was used for this purpose. A preliminary analysis was conducted to determine which attributes benefited from the use of a non-linear specification. These experiments showed that important gains in model performance could be obtained by using a log-transform for flight frequency, flight time, in-vehicle access time (IVT), and access cost, such that this approach was adopted. Any remaining attributes were treated in a linear fashion.

10.4 Model estimation

This section discusses the findings of the modelling analysis. It starts with a preliminary analysis of the stated reasons of airport choice in Section 10.4.1, and a discussion of the specification of utility in Section 10.4.2. This is followed by a presentation of the results for MNL (Section 10.4.3), NL (Section 10.4.4) and CNL (Section 10.4.5). The section concludes with a comparison of the substantive results across models in Section 10.4.6. All models presented in this chapter were estimated with BIOGEME.

10.4.1 Preliminary data analysis

Before the actual discrete choice experiments, a brief analysis was conducted to look into passengers' stated primary reason for choosing their specific departure airport. The results, which are summarised in Table 10.3, show that access-distance (as a function of ground-level origin) outranks all other factors. Unsurprisingly, other factors that play a strong role are flight availability and timing; given the lack of appropriate data, these can however only be included in the form of flight frequency information. Just over a tenth of passengers stated that the decision had been taken by a third party; a small-scale model fitting exercise excluding these observations led

[15] This relates directly to the issues described by Hess, Polak & Bierlaire (2005) in the case of models overloaded with constants.

RESIDENT			
Business travellers		**Leisure travellers**	
Near home	41.80%	Near home	34.87%
Flights available	11.45%	No answer	19.39%
No answer	11.35%	Flights available	13.48%
Third party decision	11.29%	Third party decision	10.58%
Near business	7.96%	Economic/cheaper	9.72%
Timing of flights	6.71%	Prefer airport	4.14%
Prefer airport	3.73%	Other	3.03%
Economic/cheaper	2.69%	Timing of flights	2.55%
No local services	1.55%	Near business	1.19%
Other	1.46%	No local services	1.06%

VISITOR			
Business travellers		**Leisure travellers**	
Near business	36.50%	No answer	26.38%
Flights available	14.40%	Near leisure	18.84%
No answer	11.88%	Flights available	14.66%
Third party decision	11.68%	Third party decision	13.64%
Timing of flights	6.71%	Economic/cheaper	10.43%
Prefer airport	4.73%	Other	4.45%
Economic/cheaper	4.58%	Near home	4.41%
Other	4.31%	Prefer airport	3.64%
Near home	3.71%	Timing of flights	2.17%
Near leisure	1.50%	Connecting flights	1.37%

Table 10.3: Stated main reason for choice of airport in London survey data

to very comparable results, such that these observations were retained to increase the overall sample size. This result would suggest that similar choice processes apply for self-bookings and third-party bookings.

Several other points need to be addressed. The first of these is fare, which is indicated as the primary reason for airport choice by fewer than one in twenty business travellers, and by around one in ten leisure travellers. In fact, for business travellers, fare plays the main role for a smaller share of passengers than does a personal penchant for the specific airport. The restrictions of an approach asking for the main reason for choosing an airport are clear; it is quite likely that fare plays a contributing, though not primary role for a much larger share of passengers. Nevertheless, the results do suggest that, at the time the data were collected (1996), business travellers especially were not too concerned about airfares. Another interpretation is that, at the time of the survey, there was far less variation in the fares of flights to business destinations. The increased activity by low cost carriers has changed this; if fare differences are large enough, even business travellers become sensitive. Another point that deserves some attention is the high number of passengers providing no answer. While some of these non-responses can be seen as a simple refusal to provide a response, they may, in some cases, also indicate a lack of choice, or a lack of information. This applies primarily to leisure travellers, and especially so in the visiting group. The final point that needs addressing is the lack of responses relating to airline allegiance, which could suggest that this plays only a contributing role in the choice process. However, it is also possible that some effect of airline allegiance is reflected in the "*Other*" and "*No answer*" replies.

10.4.2 Utility specification

A comprehensive set of variables were used in the initial modelling analysis. These included attributes relating to the air journey (frequency, fare, flight time, aircraft type, seat capacity, on-time performance of the airport and airline) and the access-journey (access cost, in-vehicle access time, out-of-vehicle access time, wait time, number of interchanges, parking cost). Like in the SF-bay study, no treatment of the distribution of departure times was used in the present analysis. In the absence of frequent flier information, attempts were made to account for airline allegiance by including a UK-airline dummy variable in the models for resident travellers, and a foreign-airline dummy variable in the models for visiting travellers. No further gains could be made by using separate dummy variables for all different foreign nationalities. Unlike in the SF-bay study, no information on past choices was available, such that a treatment of "airport allegiance" was not possible. Finally, attempts to model further interactions with socio-demographic attributes aside from purpose and residency status were not successful. This is characterised notably by the absence of an income effect, which can be blamed partly on the discrete nature of the income information. This increases the scope for later analyses allowing for random variations in tastes within the four groups used in the present study. Also, like in the SF-bay study (cf. Section 9.4.2), taste-coefficients along the access-mode dimension and along the air-travel dimension were generic, rather than being linked to a specific access-mode or type of aircraft.

Before proceeding to the actual estimation results, some generic conclusions can be presented. Indeed, the actual modelling analysis showed that only a small set of the attributes listed above have a statistically significant impact on choice behaviour, at least with the present sample and model specification. As such, no effect could be identified for parking cost (possibly due to the absence of cost-bearing information), seat capacity, out-of-vehicle access time, wait time, on-time performance, and the number of ground-level interchanges[16]. Furthermore, aircraft size, in the form of a dummy variable for turboprop planes, showed no effect; here however, the highly correlated flight time attribute had a significant negative effect across models. Furthermore, a significant effect of air-fare could only be identified for visiting leisure travellers, while allegiance to the national carrier played a role only for visiting business travellers. The inability to estimate a consistent fare effect across population segments can again be explained mainly on the basis of the low quality of the fare data, but should also be put into context by considering the observations made from Table 10.3. Finally, the analysis showed that the use of the combined fuel and depreciation cost for car journeys is preferable to the use of fuel cost on its own.

[16]More detailed ground-access level-of-service information could alleviate these problems, and is, as mentioned in Section 10.2.3, an important avenue for future research.

	Resident				Visitor			
	Business		Leisure		Business		Leisure	
Observations	6,706		8,269		7,207		9,667	
Parameters	55		57		57		58	
Final LL	-14945.3		-17627.1		-15278.1		-20553.8	
Adjusted $\rho^2(0)$	0.3416		0.3529		0.3549		0.3418	
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\beta_{\text{air}-\text{fare}}$	-	-	-	-	-	-	-0.0026	-1.93
$\beta_{\text{LN(access cost)}}$	-1.2831	-8.05	-0.9083	-7.20	-0.9004	-7.74	-0.7097	-5.04
$\beta_{\text{LN(flight time)}}$	-2.2963	-3.17	-2.7678	-3.18	-2.1878	-2.01	-4.3711	-4.48
$\beta_{\text{LN(frequency)}}$	0.5641	2.42	0.9776	4.73	0.5070	1.85	0.7024	3.22
$\beta_{\text{LN(IVT}^\dagger)}$	-1.4440	-6.21	-1.6898	-10.75	-1.6319	-10.67	-1.4025	-7.48
$\delta_{\text{national carrier}}$	-	-	-	-	0.4653	1.69	-	-

† IVT = in-vehicle access time

Table 10.4: MNL estimation results for London data

10.4.3 MNL models

The estimation results for the four MNL models are shown in Table 10.4. The results show consistent negative effects of increases in access cost, flight time and in-vehicle access time, with positive effects of increases in flight frequency (significant at the 94% level for visiting business travellers). In each case, a log-transform was used. Additionally, there is a negative effect of increases in air fare (linear) for visiting leisure travellers, and a positive dummy variable is associated with non-UK carriers for visiting business travellers, though this is significant only at the 90% level.

10.4.4 NL models

In this section, we look at the three sets of NL models fitted to the four subsamples, using nesting by airport, airline and access-mode respectively. The specification of utility used in the NL models was the same as in the MNL models. As such, this discussion will centre primarily on the conclusions in terms of nesting, with the substantive differences between the models discussed in Section 10.4.6.

NL model using nesting by airport

The results for the first group of NL models, which use nesting by airport, are shown in Table 10.5 (using λ_k to define the structural parameter for airport nest k). The results show that the four NL models lead to improvements in LL over their MNL counterparts by 49.2, 120.7, 97.7 and 151.3 units respectively, at the cost of 4 additional parameters in the case of resident business travellers, and 3 additional parameters in the remaining three models. In each case, the improvement in model fit is statistically significant, with χ^2 p-values of 0 for the associated LR tests. Several parameters have experienced a drop in significance when compared to the MNL model, notably the fare-coefficient for visiting leisure travellers, which is now only significant at the 91% level.

	Resident				Visitor			
	Business		Leisure		Business		Leisure	
Observations	6,706		8,269		7,207		9,667	
Parameters	59		60		60		61	
Final LL	-14896.1		-17506.4		-15180.4		-20402.5	
Adjusted $\rho^2(0)$	0.3436		0.3572		0.3589		0.3465	
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\beta_{\text{air-fare}}$	-	-	-	-	-	-	-0.0021	-1.73
$\beta_{\text{LN(access cost)}}$	-1.1807	-7.83	-0.8556	-8.01	-0.8370	-8.48	-0.7292	-5.94
$\beta_{\text{LN(flight time)}}$	-2.1002	-2.91	-2.0361	-2.48	-1.9097	-1.86	-3.6898	-3.90
$\beta_{\text{LN(frequency)}}$	0.5446	2.40	0.9460	4.95	0.4883	1.86	0.7335	3.59
$\beta_{\text{LN(IVT}^\dagger)}$	-1.4610	-6.75	-1.5081	-10.54	-1.5896	-11.15	-1.2530	-7.42
$\delta_{\text{national carrier}}$	-	-	-	-	0.4223	1.57	-	-
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
λ_{LCY}	0.8730	0.43	1.00	-	1.00	-	1.00	-
λ_{LGW}	0.8266	1.90	0.7205	4.16	0.8114	1.68	0.6939	4.28
λ_{LHR}	1.00	-	1.00	-	1.00	-	1.00	-
λ_{LTN}	0.5470	2.14	0.7029	1.17	0.7312	2.10	0.8377	1.36
λ_{STN}	0.7568	1.27	0.6519	2.82	0.4415	3.64	0.6773	1.55

† IVT = in-vehicle access time

T-statistics calculated with respect to 0 for taste coefficients, and with respect to 1 for structural parameters.

Table 10.5: NL estimation results for London data, using nesting by airport

In terms of nesting conclusions, a consistent pattern emerges. The nesting parameter for LHR needs to be constrained to a value of 1 across all models, indicating no heightened correlation in the unobserved utility terms between alternatives associated with LHR along the airport dimension. A similar conclusion applies for LCY, where, although for resident business travellers, the original value does not exceed 1, it is not significantly different from 1. For the remaining three airports, the values are consistently below 1, indicating heightened correlation, although, in some cases, the significance level is below the usual 95% limit. The use of additional constraints however led to significant drops in model performance.

The lowest values for the nesting parameters, and hence the highest levels of correlation, are observed for LTN and STN (with the exception of LTN for visiting leisure travellers). These two airports are different from the remaining three in terms of their route network, their outlying location, and in terms of being used extensively by low-cost airlines (even back in 1996). Some of these characteristics are not captured in the observed part of utility, explaining the high levels of correlation between alternatives in the unobserved part of utility. A potential topic for further investigation is the analysis of the correlation between alternatives at the two airports, given their similarities, especially with more recent data.

NL model using nesting by airline

The results for the second group of NL models, which use nesting by airline, are shown in Table 10.6 (using π_l to define the structural parameter for airline nest l). The results show that the four NL models lead to improvements in LL over their MNL counterparts by 74.6, 183.9, 173.4 and 288.4 units respectively, at the cost of 19 additional parameters in the two models for resident travellers, 16 additional parameters for visiting business travellers, and 24 parameters for visiting leisure travellers. Again, all four improvements are statistically significant, with χ^2 p-values of 0 for the associated LR tests. Again, there are some changes in parameter significance, with the significance of the fare-coefficient for visiting leisure travellers decreasing to the 87% level.

Of the 37 nesting parameters, 6 had to be constrained to 1 in each of the four models, and as such, are not listed in Table 10.6. These relate to airlines $A10$, $A23$, $A26$, $A33$, $A35$, and $A36$. In addition to the six overall constraints, a number of other nesting parameters initially took on unacceptable values in some of the population segments. As such, out of the 37 possible parameters, 18 had to be constrained in the two models for residents, along with 21 in the model for visiting business travellers, and 13 in the model for visiting leisure travellers. A large number of the estimated structural parameters are not significantly different from a value of 1, but additional constraints led to significant drops in model performance. For this group of models, it is difficult to infer conclusions about the nesting structure, given the high number of nests, and low overall significance of the structural parameters, although it can be noted that the structural parameters for nests associated with low-cost airlines tend to indicate consistently high levels of correlation, which can be an indication of product differentiation, and as such, could help to explain allegiance by passengers to such airlines.

NL model using nesting by access-mode

The results for the final group of NL models, which use nesting by access-mode, are shown in Table 10.7 (using Ψ_m to define the structural parameter for access-mode nest m). The results show that the four NL models lead to improvements in LL over their MNL counterparts by 128.6, 45.3, 163.5 and 131.1 units respectively, at the cost of 5 additional parameters in the model for resident business travellers, and 4 additional parameters in the remaining three models. Again, all four improvements are statistically significant, with χ^2 p-values of 0 for the associated LR tests.

The common observation across models is that the structural parameter associated with the public transport nest needs to be constrained to a value of 1. Although this suggests a lack of correlation between public transport alternatives, the low level of disaggregation along the public transport dimension could play a role in this (cf. Section 10.2.3), and more similarities could be expected in subgroups of public transport modes. Consistently low values for the structural parameters are obtained for car and taxi, showing high correlation within these nests, which can reflect mode-allegiance for these alternatives. Again, not all estimated structural parameters are significantly different from a value of 1, but additional constraints led to significant drops in model performance.

	Resident		Visitor	
	Business	Leisure	Business	Leisure
Observations	6,706	8,269	7,207	9,667
Parameters	74	76	73	82
Final LL	-14870.7	-17443.2	-15104.7	-20265.4
Adjusted $\rho^2(0)$	0.3441	0.3590	0.3615	0.3502

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\beta_{\text{air-fare}}$	-	-	-	-	-	-	-0.0021	-1.53
$\beta_{\text{LN(access cost)}}$	-1.1331	-7.04	-0.7850	-6.92	-0.8011	-8.12	-0.6316	-4.86
$\beta_{\text{LN(flight time)}}$	-2.3415	-3.44	-2.8067	-3.45	-2.1560	-2.14	-4.4597	-4.93
$\beta_{\text{LN(frequency)}}$	0.5716	2.50	0.8593	4.17	0.4707	1.86	0.6014	3.00
$\beta_{\text{LN(IVT}^\dagger)}$	-1.3946	-6.27	-1.4594	-10.04	-1.5659	-11.38	-1.1370	-6.50
$\delta_{\text{national carrier}}$	-	-	-	-	0.4328	1.52	-	-

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
π_{A1}	0.6113	1.36	0.9511	0.15	0.7126	1.32	0.9222	0.32
π_{A2}	1.0000	-	1.0000	-	1.0000	-	0.8736	0.21
π_{A3}	1.0000	-	0.2609	1.92	0.5595	1.45	1.0000	-
π_{A4}	0.8940	1.29	0.8333	2.17	1.0000	-	0.8259	1.35
π_{A5}	1.0000	-	1.0000	-	0.7268	1.90	0.6572	2.15
π_{A6}	0.7795	0.44	0.2099	1.52	1.0000	-	1.0000	-
π_{A7}	1.0000	-	0.9025	0.36	1.0000	-	0.8585	0.51
π_{A8}	0.6339	0.84	0.6199	1.04	0.3981	0.63	0.4927	1.97
π_{A9}	1.0000	-	0.4089	3.70	1.0000	-	1.0000	-
π_{A11}	0.7186	1.46	0.6923	2.38	0.5391	3.03	0.5260	2.34
π_{A12}	0.6185	2.02	0.9020	0.43	1.0000	-	0.2869	2.80
π_{A13}	1.0000	-	0.6950	1.84	1.0000	-	0.8392	0.63
π_{A14}	0.7758	0.39	1.0000	-	0.4072	2.01	0.7271	0.79
π_{A15}	0.5633	1.48	0.3234	3.57	0.4669	1.04	0.2337	3.09
π_{A16}	1.0000	-	1.0000	-	0.8942	1.33	0.8575	0.93
π_{A17}	0.8992	0.37	0.7721	0.60	0.8096	1.21	1.0000	-
π_{A18}	0.6091	0.93	1.0000	-	1.0000	-	1.0000	-
π_{A19}	0.7654	0.41	0.4931	1.53	0.7494	0.67	0.5711	1.60
π_{A20}	0.4341	2.50	1.0000	-	0.8362	0.58	0.8649	0.46
π_{A21}	0.4869	1.77	0.5568	1.61	0.1316	2.48	0.4243	2.34
π_{A22}	1.0000	-	1.0000	-	1.0000	-	0.7133	0.91
π_{A24}	1.0000	-	1.0000	-	0.8487	0.58	1.0000	-
π_{A25}	1.0000	-	1.0000	-	1.0000	-	0.8156	0.62
π_{A27}	0.7238	1.15	1.0000	-	1.0000	-	0.8578	0.66
π_{A28}	0.8700	0.76	1.0000	-	1.0000	-	1.0000	-
π_{A29}	1.0000	-	1.0000	-	1.0000	-	0.6307	1.85
π_{A30}	0.3878	2.67	0.5162	1.48	1.0000	-	0.7903	1.22
π_{A31}	0.6622	1.42	0.5528	2.40	0.4298	3.40	0.5109	3.05
π_{A32}	0.6874	0.97	0.8226	0.67	0.2730	3.14	0.6720	0.78
π_{A34}	1.0000	-	0.6940	2.28	0.6711	1.50	0.5483	2.30
π_{A37}	0.4285	0.85	0.7507	1.45	1.0000	-	0.8065	0.44

† IVT = in-vehicle access time

T-statistics calculated with respect to 0 for taste coefficients, and with respect to 1 for structural parameters.

Table 10.6: NL estimation results for London data, using nesting by airline

	Resident		Visitor	
	Business	Leisure	Business	Leisure
Observations	6,706	8,269	7,207	9,667
Parameters	60	61	61	62
Final LL	-14816.7	-17581.8	-15114.6	-20422.7
Adjusted $\rho^2(0)$	0.3470	0.3544	0.3616	0.3458

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\beta_{\text{air-fare}}$	-	-	-	-	-	-	-0.0021	-2.14
$\beta_{\text{LN(access cost)}}$	-1.0197	-7.23	-0.7841	-5.83	-0.8258	-7.61	-0.6838	-5.27
$\beta_{\text{LN(flight time)}}$	-1.4941	-2.95	-2.2619	-2.87	-1.3507	-1.72	-3.4752	-4.15
$\beta_{\text{LN(frequency)}}$	0.3196	1.96	0.8227	4.28	0.3746	1.93	0.5419	2.94
$\beta_{\text{LN(IVT}^\dagger)}$	-0.9553	-3.27	-1.5280	-8.95	-1.3575	-9.36	-1.1774	-6.99
$\delta_{\text{national carrier}}$	-	-	-	-	0.3869	1.89	-	-

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
Ψ_{Car}	0.6062	2.56	0.7815	2.62	0.7553	2.11	0.7145	3.10
Ψ_{Hire}	0.3700	2.40	0.9244	0.21	0.4598	3.66	1.00	-
Ψ_{LDC}	0.7635	0.69	1.00	-	1.00	-	0.5954	2.63
Ψ_{MC}	0.5778	2.15	0.9548	0.41	0.5081	4.59	0.7206	1.92
Ψ_{PT}	1.00	-	1.00	-	1.00	-	1.00	-
Ψ_{Taxi}	0.6356	2.53	0.7986	1.77	0.7156	2.44	0.6467	2.32

† IVT = in-vehicle access time

T-statistics calculated with respect to 0 for taste coefficients, and with respect to 1 for structural parameters.

Table 10.7: NL estimation results for London data, using nesting by access-mode

Discussion

The presentation of the NL results has shown that each of the three nesting approaches offers significant improvements in model fit over the corresponding MNL model, across the four population groups.

At the same time, the analysis has revealed important differences across population segments in terms of the optimal two-level nesting structures. As such, using the adjusted ρ^2 measure as a means of comparison[17], the model using nesting by access-mode leads to the best performance for resident business travellers, while the model using nesting by airline leads to the best performance in the two leisure groups. For visiting business travellers, the performance of the two models is indistinguishable, with the higher number of parameters for the model using nesting by airline nullifying its LL advantage. Finally, it is interesting to note that, for leisure travellers, the model using nesting by access-mode gives the poorest performance, while, for business travellers, this is the case with the model using nesting by airport.

[17]Simple LR-tests cannot be used in this case, given that the models are not nested. While non-nested tests could be used for model comparison in this case (cf. Ben-Akiva & Lerman 1985), it was decided that, for the present purpose, the use of the simple ρ^2 measure, which takes into account the number of parameters, was sufficient.

	Dimension	Resident Business	Resident Leisure	Visitor Business	Visitor Leisure
NL	airport	1	2	2	2
NL	airline	18	18	21	13
NL	access-mode	1	2	2	2
CNL	airport	1	2	1	2
CNL	airline	9	27	14	16
CNL	access-mode	2	2	1	1

Table 10.8: Structural constraints required in NL and CNL models

10.4.5 CNL models

We next turn our attention to the estimation of the CNL models, where the base specification of utility was again the same as that used for the MNL and NL models.

In the present context, a total of 48 nests were used in the CNL models (5 airports, 37 airlines, and 6 access-modes). Aside from leading to the use of 48 separate nesting parameters (to allow for differential levels of correlation in different nests), this leads, in the presence of a choice set of 324 combined alternatives, to a total of 972 allocation parameters (324 along each dimension). As each alternative is associated with exactly one airport, one airline, and one access-mode, only one allocation parameter along each of the three dimension is not constrained a priori to zero for a given alternative. From this, it can also be seen that, given the condition that the allocation parameters for each alternative sum to 1, a total of 648 can be identified. Although this number is reduced somewhat due to availability conditions, this still leads to a very expensive estimation process, and can be seen to result in an over-parameterised model, an issue that was already highlighted by Hess (2004) in the estimation of a CNL model with the SF-bay area data.

A preliminary analysis showed that, although the estimation of the allocation parameters leads to gains in model fit, these are not statistically significant, given the huge cost in terms of the number of parameters. Additionally, the estimation of the allocation parameters leads to major issues with parameter identification and very significant increases in computational cost. As such, the decision was taken to constrain all non-zero allocation parameters to a value of $\frac{1}{3}$, such that an alternative is associated in equal parts with an airport, an airline, and an access-mode. With the use of fixed allocation parameters, it is not immediately clear how the CNL model can reduce to one of the three NL models, although an approximation can be obtained when the structural parameters along two dimensions reduce to a value of 1. Given this complication, nested LR tests were once again replaced by the adjusted ρ^2 statistic.

The results of the CNL models are summarised in Table 10.9 (main results, plus structural parameters for airport and access-mode dimensions) and Table 10.10 (structural parameters for airline dimension).

It is of interest to first look at the constraints that are required to yield acceptable values for the structural parameters. As such, the number of required constraints, along each dimension, is given in Table 10.8 for each of the four models, along

	Resident		Visitor	
	Business	Leisure	Business	Leisure
Observations	6,706	8,269	7,207	9,667
Parameters	91	74	89	87
Final LL	-14603.3	-17437.6	-14988.2	-20142.9
Adjusted $\rho^2(0)$	0.3551	0.3592	0.3658	0.3540

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\beta_{\text{air-fare}}$	-	-	-	-	-	-	-0.0020	-1.39
$\beta_{\text{LN(access cost)}}$	-0.9911	-8.02	-0.7901	-5.87	-0.7975	-8.62	-0.6708	-6.92
$\beta_{\text{LN(flight time)}}$	-1.4201	-3.16	-1.8270	-2.50	-1.3789	-1.71	-3.5148	-3.21
$\beta_{\text{LN(frequency)}}$	0.2453	1.10	0.9027	4.60	0.5306	1.61	0.5523	2.56
$\beta_{\text{LN(IVT}^\dagger)}$	-1.0718	-9.19	-1.5515	-12.04	-1.4368	-11.14	-1.0552	-6.67
$\delta_{\text{national carrier}}$	-	-	-	-	0.4081	1.35	-	-

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
λ_{LCY}	0.5412	1.23	1.0000	-	0.8346	0.78	1.0000	-
λ_{LGW}	0.6177	2.63	0.1050	0.67	0.1000	0.60	0.0783	0.25
λ_{LHR}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
λ_{LTN}	0.2627	2.95	0.0961	4.53	0.2644	2.47	0.6873	1.74
λ_{STN}	0.2608	2.18	0.5311	2.04	0.1806	2.28	0.1603	0.30

	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
Ψ_{Car}	0.0402	-	0.7250	0.98	0.5778	1.20	0.4581	0.40
Ψ_{Hire}	0.3087	1.03	0.2124	-	0.1048	0.60	0.6127	1.20
Ψ_{LDC}	1.0000	-	1.0000	-	0.7936	0.25	0.2734	2.15
Ψ_{MC}	0.1493	2.58	0.7165	1.14	0.2924	2.09	0.4573	1.38
Ψ_{PT}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
Ψ_{Taxi}	0.4877	1.13	0.6670	1.32	0.6540	0.93	0.1109	2.82

† IVT = in-vehicle access time

T-statistics calculated with respect to 0 for taste coefficients, and with respect to 1 for structural parameters.

Table 10.9: CNL estimation results for London data, part I

with the constraints in the corresponding NL models[18]. The results show that, overall, fewer constraints are required in the CNL models. However, while there is a significant reduction in the constraints required in the two business models, there is a significant increase in constraints for the model for resident leisure travellers, and an increase by two constraints in the corresponding model for visitors. It can be seen that the differences arise primarily in the airline dimension.

In terms of model performance, it can be seen that the four CNL models give a better fit to the data (in terms of adjusted ρ^2) than the MNL model, or any of the three NL structures. Again, however, there are some significant differences across the four population segments. Indeed, for resident business travellers, the

[18] It is worth pointing out that these constraints reflect the correlation structure in the data, as retrieved by the model, and not theoretical identification requirements.

improvement in LL offered by the CNL model over the MNL model is 35% bigger than the combined improvement offered by the three NL models. In the remaining three models, the improvement is smaller than the combined gain in LL by the three NL models, but it is in each case still significantly larger than the average improvement offered by the three NL models (twice as large in the two models for visitors). In fact, it be seen that the only group where the performance of the CNL model is not convincing is that for resident leisure travellers. Here, the performance of the model is still better than that of the NL models using nesting by airport and by access-mode, but there is little gain in performance to be obtained when compared to the model using nesting by airline. This can be seen to be a direct result of the higher number of constraints required along the airline dimension for the CNL model in this population segment. Again, several parameters experience a drop in significance when compared to the MNL model, with the biggest drop occurring in the case of the frequency coefficient for resident business travellers, which is now only significant at the 73% level, but whose exclusion leads to considerable drops in model fit. A number of the estimated structural parameters are again not statistically different from 1, while others are very close to zero. Here, it should be noted that it was not possible to produce a reliable value for the standard error of π_{A10}[19] in the model for resident business travellers, despite the use of a robust estimator. This is an indication of the complexity of the model that was estimated here.

In closing, it can be seen that the CNL model does have the potential to offer gains in performance when compared to the three two-level NL structures. Furthermore, given the problems with using multi-level NL structures[20], the CNL model has clear conceptual advantages. Finally, further gains in performance can be expected with the use of a flexible formulation of the allocation parameters, for example with a parameterisation as a function of socio-demographics.

10.4.6 Comparison of substantive results

The final step of the analysis, before proceeding to model validation, is concerned with a comparison of the actual substantive results across population groups and across models. To allow for a consistent comparison, only coefficients estimated across all four population subgroups should be involved in these comparisons. In the presence of four such coefficients, namely the marginal utilities of changes in (the logarithms of) in-vehicle access time, access cost, flight frequency, and flight time, three trade-offs were used. These are the willingness to accept increases in access cost in return for decreases in access time (i.e. VTTS), the willingness to accept increases in access time in return for increases in flight frequency, and the willingness to accept increases in access time in return for decreases in flight time. Trade-offs involving fare and the allegiance to the national carrier would be of interest, but given that these two coefficients were each only estimated in one model, and not jointly, the calculation of such trade-offs was not possible in the present study, but is dealt with extensively in the SP case-study (cf. Chapter 11).

[19] The estimated value of π_{A10} is arbitrarily close to zero.

[20] No multi-level NL structures were estimated in this chapter, but the observations made in Section 8.4.3 apply.

	Resident				Visitor			
	Business		Leisure		Business		Leisure	
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
π_{A1}	0.1888	1.05	1.0000	-	0.4976	0.84	1.0000	-
π_{A2}	0.1726	1.88	1.0000	-	0.4936	2.39	0.1367	1.71
π_{A3}	1.0000	-	1.0000	-	0.1628	1.68	1.0000	-
π_{A4}	0.4727	2.01	1.0000	-	1.0000	-	0.7661	0.81
π_{A5}	0.1233	1.52	1.0000	-	0.4404	1.47	0.3054	-
π_{A6}	0.1016	0.14	1.0000	-	0.2101	1.10	0.1271	7.20
π_{A7}	1.0000	-	1.0000	-	1.0000	-	0.8871	0.25
π_{A8}	0.4773	0.40	0.6816	0.47	0.3354	0.33	0.5610	0.89
π_{A9}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
π_{A10}	0.0752	†	1.0000	-	1.0000	-	1.0000	-
π_{A11}	0.5512	1.00	0.6829	0.96	0.1196	0.61	0.0613	0.30
π_{A12}	0.2936	2.17	1.0000	-	0.7029	0.30	0.2192	2.52
π_{A13}	0.9049	0.26	0.4500	1.41	0.8986	0.20	0.8286	0.36
π_{A14}	0.0998	2.96	1.0000	-	0.2796	4.43	0.2348	3.41
π_{A15}	0.5752	0.82	0.1554	2.06	0.1083	3.94	0.1079	1.24
π_{A16}	0.4087	1.77	1.0000	-	0.6396	1.70	0.1000	2.03
π_{A17}	0.2772	1.75	0.0584	-	0.4494	1.32	1.0000	-
π_{A18}	0.3386	0.88	1.0000	-	0.9075	0.17	1.0000	-
π_{A19}	0.1051	-	0.0502	1.47	0.6685	0.24	0.2115	1.71
π_{A20}	0.1012	5.53	1.0000	-	1.0000	-	1.0000	-
π_{A21}	0.4790	1.28	0.0949	1.18	0.1072	3.04	1.0000	-
π_{A22}	1.0000	-	1.0000	-	1.0000	-	0.6505	0.37
π_{A23}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
π_{A24}	0.7482	0.70	1.0000	-	0.6214	0.50	1.0000	-
π_{A25}	1.0000	-	1.0000	-	0.8376	0.22	0.7344	0.52
π_{A26}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
π_{A27}	0.2320	2.21	1.0000	-	1.0000	-	0.8216	0.37
π_{A28}	0.4722	1.78	1.0000	-	1.0000	-	1.0000	-
π_{A29}	0.8907	0.20	1.0000	-	0.4366	-	0.2525	0.99
π_{A30}	0.2394	1.29	0.0510	1.39	1.0000	-	0.9219	0.28
π_{A31}	0.1197	5.56	0.1835	0.75	0.2278	2.07	0.5838	1.58
π_{A32}	0.5304	2.09	1.0000	-	0.2502	1.25	0.0475	2.85
π_{A33}	0.8717	0.26	1.0000	-	1.0000	-	1.0000	-
π_{A34}	1.0000	-	1.0000	-	0.5096	0.81	0.5889	0.91
π_{A35}	0.0993	5.83	1.0000	-	1.0000	-	1.0000	-
π_{A36}	1.0000	-	1.0000	-	1.0000	-	1.0000	-
π_{A37}	0.1233	0.83	0.8262	0.73	0.6993	0.68	1.0000	-

† Estimate of standard error not reliable

T-statistics calculated with respect to 1.

Table 10.10: CNL estimation results for London data, part II

		Resident		Visitor	
		Business	Leisure	Business	Leisure
MNL	mean	17.45	24.03	38.66	24.07
	std.dev.	24.83	37.60	45.69	35.93
NL by airport	mean	19.18	22.77	40.52	20.93
	std.dev.	27.30	35.62	47.88	31.25
NL by airline	mean	19.08	24.02	41.70	21.93
	std.dev.	27.15	37.57	49.28	32.74
NL by access-mode	mean	14.52	25.18	35.07	20.97
	std.dev.	20.67	39.38	41.44	31.31
CNL	mean	16.76	25.37	38.44	19.16
	std.dev.	23.86	39.69	45.42	28.60

Table 10.11: Trade-off between in-vehicle access time and access cost (£/hour)

All four attributes used in the trade-offs enter the utility under a log-transform, making the calculation of trade-offs slightly more complicated than in the case of linear specifications of utility, where the trade-offs are given by the simple ratio of the estimated coefficients. Indeed, the trade-off between two attributes is given by the ratio of the partial derivatives of the utility with respect to these attributes. Let $U = \ldots + \beta_1 \ln(z_1) + \beta_2 \ln(z_2) + \ldots$. The ratio of the partial derivatives of U with respect to z_1 and z_2 is then given by $\frac{\beta_1}{\beta_2}\frac{z_2}{z_1}$, as opposed to the simple $\frac{\beta_1}{\beta_2}$ ratio used in the case of a linear parameterisation. In the present study, for each trade-off, the actual values of z_1 and z_2 for the observed journeys were used to calculate $\frac{z_2}{z_1}$, and statistics were calculated for the distribution of the appropriate ratios across respondents. In the case of the ratio of two coefficients using a log-transform, and in the presence of non-perfectly correlated variations in z_1 and z_2, this approach is clearly preferable to the commonly adopted use of $\frac{\overline{z_2}}{\overline{z_1}}$ (with $\overline{z_1}$ and $\overline{z_2}$ giving the mean values of z_1 and z_2 respectively), as it avoids potentially significant levels of bias in the calculation of trade-offs. Furthermore, this approach yields respondent-specific trade-offs, allowing the calculation of a set of statistics for the distribution of the trade-offs, where it should be noted that these variations are an effect of the varying values for the concerned attributes, and do not as such give variations in tastes across respondents, but rather give an indication of the varying levels of trade-offs under different market conditions.

The results of the calculation of the trade-offs are summarised in Table 10.11 for the VTTS, Table 10.12 for the trade-off between flight time and in-vehicle access time and Table 10.13 for the trade-off between flight frequency and in-vehicle access time. In each case, the tables present the mean value of the respective trade-off in the different population groups, along with the associated standard deviation, and show the values across the five different models estimated.

The first observation that can be made is that there are some variations across the five models in the calculated values for the different trade-offs. In the first trade-off (VTTS), and the third trade-off (frequency *vs* IVT), the values in the MNL model and the two NL models using nesting by airport and by airline are generally roughly similar, while those produced with the NL model using nesting by access-mode and

		Resident		Visitor	
		Business	Leisure	Business	Leisure
MNL	mean	1.07	1.14	1.09	2.96
	std.dev.	0.70	0.90	0.98	2.79
NL by airport	mean	0.97	0.94	0.98	2.80
	std.dev.	0.63	0.74	0.88	2.63
NL by airline	mean	1.13	1.34	1.12	3.73
	std.dev.	0.74	1.06	1.01	3.51
NL by access-mode	mean	1.06	1.03	0.81	2.81
	std.dev.	0.69	0.81	0.73	2.64
CNL	mean	0.89	0.82	0.78	3.17
	std.dev.	0.58	0.65	0.70	2.98

Table 10.12: Trade-off between flight time and in-vehicle access time

		Resident		Visitor	
		Business	Leisure	Business	Leisure
MNL	mean	6.76	19.67	5.83	16.90
	std.dev.	10.02	29.36	8.14	22.64
NL by airport	mean	6.45	21.32	5.76	19.75
	std.dev.	9.56	31.84	8.05	26.46
NL by airline	mean	7.10	20.02	5.64	17.84
	std.dev.	10.51	29.88	7.88	23.91
NL by access-mode	mean	5.79	18.30	5.17	15.53
	std.dev.	8.58	27.33	7.23	20.81
CNL	mean	3.96	19.78	6.93	17.66
	std.dev.	5.87	29.53	9.68	23.66

Table 10.13: Trade-off between frequency and in-vehicle access time (min/flight)

the CNL model are quite different. The second trade-off, between flight time and in-vehicle access time, shows a different trend, where the biggest outliers are this time generated by the model using nesting by airline, in the two models for residents and in the model for visiting leisure travellers, while the observations for visiting business travellers in the model using nesting by access-mode and the CNL model need to be put into context by noting the lower significance of the flight time coefficient in these two models (cf. Table 10.7 and Table 10.9). The differences in the trade-offs across models are possibly more significant than could have been expected on the basis of the small differences in model fit, highlighting the flatness of the LL function, but also stressing the differences between model structures in terms of implied choice behaviour.

Aside from comparing the calculated trade-offs across the five model structures, it is of interest to compare their values across the four population subgroups. Here, a major issue arises. Indeed, the calculation of the trade-off between access time and access cost produces a counter-intuitive result, suggesting that, for residents, the VTTS is higher for leisure travellers than for business travellers, depending on the model considerably so (ranging from 19% to 73%). On the other hand, for visitors, the results consistently show higher VTTS for business travellers than for leisure travellers, with differences ranging from 61% to 101%.

Clearly, this issue needs to be addressed. There are two possible reasons for an underestimated VTTS; an underestimation of the marginal utility of travel time changes, and an overestimation of the marginal utility of access cost changes. In the present analysis, it seems likely that both factors play a role. This insight is partly gained from a separate analysis carried out to look at the choice-behaviour by respondents on the actual observed access-journey[21]. This analysis produced two main findings:

- Business travellers are more likely to use premium PT modes for their access-journey than leisure travellers, with, for example, a lower market share for the Tube.

- On access-journeys using combinations of modes, there is, on the non-final stages, a higher market share for taxis in the case of business travellers than in the case of leisure travellers, and a much lower market share for the Tube.

These two findings apply for residents as well as for visitors, but seem to play a bigger role in the former group. In combination, these two observations can be used to explain the counter-intuitive findings in the calculation of trade-offs. Indeed, it should be remembered that, for data reasons, the present study uses highly aggregate PT data, and does not differentiate between *standard* and *premium* modes. At the same time, again for data reasons, the chosen mode for a given traveller is defined to be the mode used in the final part of the access-journey. As such, it can be seen that, with the above two observations, the level-of-service data used in model estimation underestimates the access cost for journeys by business travellers, and also overestimates the access time. This clearly has an effect on the estimated coefficients, leading to an underestimation of the access time coefficient, and an overestimation of the access cost coefficient. Additionally, it should be remembered that the VTTS is in the present case calculated as $\frac{\beta_{AT}}{\beta_{AC}}\frac{AC}{AT}$. At the same time as the estimates lead to a lower than warranted ratio of $\frac{\beta_{AT}}{\beta_{AC}}$, the *biased* level-of-service data leads to an underestimation of the ratio $\frac{AC}{AT}$, which further underestimates the VTTS.

To illustrate the differences, the ratio between the access cost and the access time variables for the actual observed journey was calculated, using the available data (i.e. common PT data, and single mode only). This yielded mean values of 25.84 pence per minute for resident business travellers, with a corresponding value of 21.53 pence per minute for resident leisure travellers. Although this does suggest a slightly higher spending rate for business travellers, the differences are small. On the other hand, for visitors, the corresponding values are 35.56 pence per minute for business travellers, and 20.29 pence per minute for leisure travellers. This thus shows a much bigger difference between the spending rates for business and leisure travellers in the case of visitors. It is conceivable that the spending rate for resident business travellers is indeed lower than for visiting business travellers, for example due to a lower reliance on taxis (and a higher reliance on cheaper minicabs), but this

[21]For this, the market shares of different access-modes were calculated from the survey data, using a higher level of disaggregation than was possible in the actual modelling analysis.

is unlikely to be on the scale indicated by the data. This argument is supported by the findings for leisure travellers, where the spending rates for residents and visitors are very similar. As such, this brief analysis does indeed suggest some bias in the data, which could explain the counter-intuitive findings.

Clearly, this issue leads to unreliable estimates of the trade-offs, making them inapplicable for use in cost-benefit analysis and forecasting. At the same time, it is not clear what effect, if any, this bias in the access-journey level-of-service data has on the results in terms of model structure. This can only be addressed with the reanalysis of the models on more disaggregate access-journey level-of-service data, which is an important avenue for future research. It should also be noted that the incorporation of ad hoc correction factors, in the form of increased cost and decreased time attributes, being an unreliable approach in any case, is hampered by the lack of specific route choice information for the access-journeys.

Attempts were made to estimate models on subsets of the data which were less likely to contain biased access time and cost information. One such approach consisted of estimating models for those travellers who use only a single mode for their entire access-journey. The analysis was restricted to residents only, and made use of MNL structures. The results show a mean VTTS of £23.76 per minute for resident business travellers, and £26.91 per minute for resident leisure travellers. This shows that the VTTS for resident business travellers increases by 36%, while that for resident leisure travellers increases by just 12%. A higher VTTS would indeed be expected for travellers using single modes as opposed to combinations of modes, but the fact that the gap between the two groups decreases so significantly does suggest that part of the bias in the data is indeed caused by the level-of-service data for travellers who use multiple modes on their access-journey. Nevertheless, some bias clearly remains, such that the requirement for more detailed level-of-service data persists.

Another source of potential bias in the access cost attribute (and by extension the associated coefficient) is the exclusion of parking cost from the calculation of the access cost for car-journeys, where the use of a separate parking cost coefficient led to problems with parameter significance while the inclusion of parking cost into an overall car cost attribute led to a drop in model performance. An analysis of the data showed a higher preference for premium parking options (closer to the airport, but more expensive) for business travellers than for leisure travellers. This thus again leads to lower access time and higher access cost, with the above-discussed effects this has in terms of biased VTTS measures. The likelihood of this issue being a prime source of the bias in the results for residents is increased by the fact that the treatment of parking cost information has a much bigger impact for residents than for visitors. It can also be seen that for business travellers, cost-sensitivity is further decreased by the higher incidence of cases where costs are covered by the employer. In the present analysis, this could not be accounted for in the models in the absence of cost-bearing information, but it can be seen that, in cases where such a treatment is possible, it is important to somehow incorporate this into the calculation of the VTTS, as a VTTS estimate based solely on costs not covered by the employer will be biased.

To test the likely impact on the VTTS calculation of bias in the access cost attribute for car-users, a separate analysis was carried out for travellers using car as the access-mode, where again, the sub-analysis was limited to residents, and made use of MNL models only. This approach led to significant reductions in the sample-size, through the exclusion of respondents using access-modes other than car, and also led to a different model specification, as the access-journey choice-dimension is now obsolete. This approach clearly leads to a loss of information, as the only differences between alternatives in terms of access time and access cost are now a function of the distance between the ground-level origin and the various airports. Furthermore, in the absence of a treatment of parking cost, access cost and access time are now almost perfectly correlated. As expected, these effects lead to much lower VTTS than in the models accounting for the presence of different modes, and especially the inclusion of modes with much higher cost/time ratios, such as taxi. The two models produce VTTS of £4.65 and £6.90 respectively, which are lower than those reported in Table 10.11 by around 75%. More importantly however, the ratio between the values for leisure and business travellers increases from 1.38 to 1.48, which can serve as an indication of the bias caused by the absence of parking cost information from the models.

Despite the above discussion of the likely bias in the estimated trade-offs, several interesting observations can nevertheless be made on the basis of Tables 10.11, 10.12 and 10.13.

The first observation is that the estimated VTTS, even for leisure travellers (where the issue of biased data plays a lesser role), are lower than those reported in previous studies of airport choice behaviour, where, as an example, Pels et al. (2003) produce VTTS for business travellers between \$1.97 and \$2.90 per minute in the SF-bay area. While there could clearly be differences across regions[22], it seems more likely that the use of a non-linear specification is the main reason for the lower values; indeed, much higher values, together with a lower model fit, were obtained when using a linear specification. While previous research in airport choice modelling has generally made use of a log-transform for flight-frequency, access time and access cost have usually been treated in a linear fashion, which could have caused the high implied VTTS.

The second observation relates to the relative sensitivity to access time and flight time. Here, the findings for visitors can again be judged to be more reliable, and would indicate that, while business travellers are relatively equally sensitive to access time and flight time, leisure travellers are far more sensitive to flight time. Here, the correlation between flight time and aircraft-type plays an important role, and the lower objection to using turboprop flights by business travellers than by leisure travellers can help to explain the results[23]. Additionally, the average flight time is longer for leisure travellers than for business travellers, further reducing the appeal of turboprop flights. Here, there is little opportunity for comparing the results to those obtained in other studies, where this trade-off is often not available.

Finally, for the willingness to accept increases in access time in return for increases in flight frequency, the values are higher for leisure travellers than for business travellers, which is a reflection of lower VTTS for leisure travellers. This finding applies to residents as well as visitors (cf. Table 10.13), suggesting that the main source

[22] Here, no comparable values for other studies involving the London airports were available.

[23] There is also possibly a greater opportunity for working during a flight than during the access-journey.

of bias in the VTTS estimates could be access cost, rather than access time[24]. The low value for the trade-off for resident business travellers in the CNL model needs to be put into context by noting the high associated standard error (cf. Table 10.9). The actual implied values equate to between 10% and 30% of the average observed access time, and as such, are possibly on the low end of the scale. Nevertheless, they are generally higher than the values reported in the SF-bay study (cf. Section 9.4.6), where additionally, in the present study, there are major differences between the two purpose segments. In the SF-bay study, such differences were only observed in the models for visitors, where, in contrast with the London results, a higher willingness was observed for business travellers than for leisure travellers.

10.5 Model validation

The final part of the analysis is concerned with model validation. For this, the five different model structures were applied to the four validation samples[25], with the final parameter values produced in estimation. It should be noted that, although the differences in LL between the five model structures are significant from a statistical point of view, they are relatively modest, in terms of the difference in LL per observation. As such, little differences in performance can be expected between the five structures in the validation process, which should rather be seen as a process for establishing the overall performance of the models, and for comparing the performance across population groups, as well as for making sure that the models were not overfitted to the estimation data. For each observation, the estimation software generates a choice probability for each of the 324 elementary alternatives; these can be summed up appropriately to obtain the choice probability for the different airports (5), airlines (37) and access-modes (6). From this, the probability of correctly predicting a given respondent's choices along each of the three choice dimensions can be retrieved straightforwardly, and averaging over observations yields the *average probability of correct prediction* in each of the choice dimensions.

The results of the validation process are summarised in Table 10.14. It should first be noted that the low probabilities for the elementary alternatives must be put into context by remembering that the total number of such alternatives is 324, with an average of 30 available alternatives per individual in the validation sample. As expected, the results show little variation between the five model structures, where additionally, it is again not clear a priori what measure of error should be associated with these values, such that no inferences on differences between models should be drawn on the basis of these results. Even though the differences in performance are thus only minor, the more complex models should still be preferred, given their greater intuitive correctness. This is reinforced by the more significant differences in terms of trade-offs (cf. Section 10.4.6).

In terms of differences across population segments, the results suggest better

[24]This insight is gained by comparing the ratio between the trade-offs for business and leisure across residents and visitors, where the comparison is again not entirely reliable due to the potential differences between residents and visitors.

[25]The validation samples contained 353 observations for resident business travellers, 434 observations for resident leisure travellers, 379 observations for visiting business travellers, and 508 observations for visiting leisure travellers

		Resident		Visitor	
		Business	Leisure	Business	Leisure
MNL	Elementary alternative	16.01%	13.54%	18.54%	15.63%
	Airline	48.01%	41.87%	46.56%	40.48%
	Airport	61.47%	52.71%	60.28%	49.03%
	Access-mode	39.27%	38.68%	45.75%	51.44%
NL (airport)	Elementary alternative	16.03%	13.62%	18.43%	15.87%
	Airline	47.84%	41.57%	46.25%	40.41%
	Airport	61.19%	51.95%	59.77%	48.62%
	Access-mode	39.51%	39.28%	46.34%	52.00%
NL (airline)	Elementary alternative	16.17%	13.99%	18.24%	15.86%
	Airline	47.71%	41.39%	46.18%	40.17%
	Airport	61.34%	52.42%	60.07%	49.25%
	Access-mode	39.54%	39.81%	45.97%	51.99%
NL (access)	Elementary alternative	16.50%	13.71%	18.49%	15.83%
	Airline	48.62%	41.98%	47.10%	40.74%
	Airport	62.88%	53.05%	61.72%	49.93%
	Access-mode	38.58%	38.51%	45.24%	51.32%
CNL	Elementary alternative	16.43%	13.87%	18.38%	15.99%
	Airline	47.78%	41.82%	46.56%	40.59%
	Airport	62.47%	52.75%	60.67%	49.17%
	Access-mode	39.14%	39.32%	45.86%	51.80%

Table 10.14: Average correct prediction performance of London models on validation samples

performance for the business models than for the leisure models, except for the access-mode dimension in the models for visitors. In terms of differences between residents and visitors, the performance is very similar, except for the access-mode dimension, where better performance is obtained for visitors, despite the fact that the modal split for visitors is more diverse (lower market share for car).

The average probabilities of correct prediction obtained in the present study are well below those obtained in the SF-bay study, where rates of up to 85% were obtained for airport choice, with rates of up to 60% for airline choice, and rates of up to 85% for access-mode choice. This however needs to be put into context by noting that the choice set used in the SF-bay study was considerably smaller (3 airports, 8 airlines and 6 access-modes). Furthermore, the exceedingly high market share for car made the analysis of access-mode choice behaviour in the SF-bay area almost trivial. Finally, it seems that airport-captivity plays a much bigger role in the SF-bay area than in London, where the levels of competition are much higher. This suggests that the models estimated in this study yield very satisfactory performance, even though they should still only be seen as a first step in the search of an "optimal" specification.

The final point of the analysis consists of ensuring that the models have not been overfitted to the estimation data. For this, the models were applied to the estimation data, using the final parameter estimates. Given the small differences in performances between the different model structures, this comparison was limited

	Resident		Visitor	
	Business	Leisure	Business	Leisure
Elementary alternative	17.21%	14.33%	17.08%	16.20%
Airline	50.27%	44.15%	45.17%	41.16%
Airport	62.46%	52.82%	58.94%	50.33%
Access-mode	40.30%	39.38%	44.61%	52.59%

Table 10.15: Average correct prediction performance of London (MNL) models on estimation samples

to the four MNL models (one for each population subgroup). The results of this analysis are summarised in Table 10.15. They show very similar performance to that obtained on the validation sample in Table 10.14, with slightly better performance on the estimation sample in all groups except for visiting business travellers. The differences are too small to indicate any general trend, suggesting that the models have indeed not been overfitted on the estimation sample.

10.6 Summary and Conclusions

This chapter has described an analysis of the combined choice of airport, airline and access-mode for passengers departing from the London area, using three different types of GEV structure; MNL, two-level NL, and CNL.

In common with most previous studies, the analysis has shown that access time is a prime determining factor in travellers' choice of departure airport, while flight frequency, access cost and flight time also play a role. At this point, it should be noted again that the frequency variable can be seen as a proxy for visibility and scheduling convenience, while the flight time variable can also be seen as a proxy for aircraft-type, and for on-time performance, given that the block-time incorporates taxi time, and hence takes into account congestion. As in many previous studies, it was not possible to estimate a consistent significant effect of air-fare[26], nor of airline-allegiance[27], a fact that is down to the general low quality of the level-of-service data for the associated attributes.

In terms of model performance, all attempted nesting approaches lead to significant gains in fit. The use of two-level NL models, which allow for the treatment of correlation along a single dimension of choice, lead to improvements in performance over the MNL model, and show differences across population groups in terms of the optimal nesting structure. The theoretical discussions in Section 8.4.3 have highlighted the deficiencies of the NL model in the present context, being limited to accounting for correlation along at most two dimensions of choice, where only a limited treatment applies along the second dimension. This gives a clear advantage to the CNL model, which allows for the simultaneous analysis of correlation along all three dimensions of choice. These theoretical advantages of the CNL model are reflected in the estimation results, showing gains in model performance and insights

[26] This was only possible for visiting leisure travellers.

[27] This was only possible for visiting business travellers, in the form of a dummy variable showing allegiance to non-UK airlines.

into choice behaviour, suggesting that this model form can indeed serve as a valuable tool in the analysis of air-travel choice behaviour.

At this point, it must be stressed again that the study did produce a counter-intuitive result in the models for residents, showing lower VTTS for business travellers than for leisure travellers. While it was possible to identify the most likely source for this bias, the absence of more detailed access-journey level-of-service data prevented the estimation of more refined models. As such, it remains to be seen what effect, if any, the data problems had on the conclusions in terms of model structure. The relatively consistent results across the four population segments in terms of the advantages of advanced model structures however somewhat increase the confidence in these findings.

A number of avenues for future research can be identified, not least of which the use of more advanced model structures, allowing jointly for cross-nesting, continuous deterministic and random taste heterogeneity. Further refinement of the auxiliary datasets can also be expected to lead to gains in model performance. Finally, aside from accounting for correlation between alternatives sharing a given airport, airline or access-mode (or a combination thereof), it is also of interest to test for correlation between alternatives at different, yet comparable airports (e.g. STN & LTN), or different airlines and access-modes.

Chapter 11

Stated preference case-study of airport and airline choice

11.1 Introduction and context

The two RP case-studies presented in Chapter 9 (SF-bay) and Chapter 10 (Greater London) have shown the difficulties that can arise in the absence of detailed level-of-service information relating to the choices actually faced by respondents[1], leading to an inability to offer a reliable treatment of factors such as air-fares, flight availability and airline allegiance. The issues described in the London study in the context of VTTS estimation (Section 10.4.6) have highlighted that data problems can also lead to counterintuitive results along the access-mode dimension. The main aim of this chapter is to illustrate how SP data can be used to alleviate these problems.

The biggest advantage of SP data in the present context comes in the availability of exact data on the alternatives that respondents were actually faced with. Similarly, the issue of uncertainty with regards to flight availability does not come into play, as it is known exactly what alternatives were open to a given respondent. This is also strongly related to the issue of capacity. Even in the presence of an adequate weighting strategy when using RP data, the dummy variables associated with a given airline or a given aircraft type do capture effects of flight availability as a function of capacity. This problem of biased dummy variables does not arise in the case of SP data; a negative estimate for a given airline or aircraft dummy does indeed signal a negative effect on utility associated with that specific airline or aircraft type[2]. On the other hand, it should be noted that some of these problems are, in the case of SP data, simply postponed to the forecasting stage.

However, another major difference arises between the use of RP and SP data in air-travel research. As described in Section 8.3, one of the variables with the greatest explanatory power in RP case-studies of air-travel choice behaviour is flight frequency. As mentioned previously, it should be noted again that, with the possible exception of travellers on very flexible tickets, frequency is not taken into account by travellers in the way it is modelled. Rather, it captures a host of other factors,

[1]This, of course, is a problem with all RP data; it is just a bit more exaggerated in the case of air-travel.

[2]This reasoning is based on the assumption that any SP-design related factors are captured by an appropriate set of constants.

most notably visibility, capacity, and schedule delay between the actual and optimal departure time, on the basis of an assumption of a relatively even spread of departure times. In the case of SP data, visibility and capacity need not be taken into account, as described above. Furthermore, by presenting travellers with a set of actual disaggregate flight options, frequency does not play a role in the description of the alternatives, but a direct treatment of schedule delay becomes possible.

The study conducted by Adler et al. (2005) on the same data used here reveals significant effects for a range of variables that are generally not well estimated in RP studies, such as air-fares, schedule delay and airline and airport allegiance. Furthermore, the study shows important differences between business and leisure travellers, and the use of a MMNL model on the same data reveals the prevalence of significant random variations in tastes across respondents within these two groups.

Aside from using a further segmentation of the leisure segment into holiday and VFR travellers, the study presented in this chapter aims to expand on the work by Adler et al. (2005) in two main directions.

- Firstly, the main estimation work is preceded by a detailed investigation of the non-linearities in response to changes in explanatory variables, using a preliminary analysis based on Box-Cox transforms. The aim of this analysis is to explore the potential for using non-linear transforms for a host of attributes that are generally treated in a linear fashion.

- Secondly, the study aims to explore continuous interactions between taste coefficients and socio-demographic variables. This treatment of deterministic taste heterogeneity, which has clear conceptual advantages over more arbitrary segmentation approaches, does not seem to have found widespread application in air-travel research thus far. In fact, it can be argued that this also extends to other areas of transport research, where modellers still rely mainly on the use of segmentations or simple linear interactions in the analysis of deterministic taste heterogeneity. It should also be said that the rise in popularity of mixture models has contributed to this situation, with modellers increasingly relying purely on a random treatment of taste heterogeneity, despite the advantages of the other methods in terms of interpretation.

The remainder of this chapter is organised as follows. Section 11.2 presents a description of the SP data used in the analysis, and Section 11.3 discusses the specification of the utility functions in the various models. Sections 11.4 and 11.5 present the results of the MNL and MMNL models respectively, and Section 11.6 summarises the findings of the analysis.

11.2 Description of data

The survey data used in this analysis were collected via the internet in 2001 from a sample of around 600 individuals who had made a paid domestic air trip within the twelve months prior to the interview taking place (Resource Systems Group Inc. 2003).

The first stage of the survey was an RP exercise, collecting data on the most recent domestic air-trip by a respondent, along with socio-demographic information, and information on membership in frequent flier programmes. Besides actual level-of-service information for the observed trip, the survey also collected qualitative data, indicating the level of satisfaction with the observed trip, along the airport as well as airline dimension. On the basis of the observed trip, a number of alternative flight options, in terms of airports and airlines, were compiled, and the respondents were asked to rank them in order of preference. For the airline options, the ranking was performed under the assumption of equal fares, while the ranking of airports was performed independently of the differences in access time. The rankings of airlines and airports thus serve as proxy variables for service quality attributes not included directly in the later model specification[3].

The actual SP survey uses a binomial choice set, with ten choice situations per individual[4]. In each choice-situation, the respondent is faced with a choice between the current observed trip, and an alternative journey option, compiled on the basis of the information collected in the RP part of the survey. These two alternatives are hereafter referred to as the *RP alternative* and the *SP alternative* respectively. A fractional factorial experimental design was used in the generation of the choice situations, and the airports and airlines used in the choice sets for a given individual were selected on the basis of the ranking compiled in the RP survey.

Aside from the actual airline and airport names[5], the attributes used to describe the two alternatives in the SP survey include flight time, the number of connections, the air-fare, the arrival time[6], the aircraft type, and the on-time performance of the various flights. Access cost was not included in the surveys (in the absence of an actual specification of the mode choice dimension), such that a calculation of the VTTS on the access-journey is not possible, although there is the possibility of calculating a trade-off between access time and air-fare. No choice is given between different travel classes; this can be regarded as an upper-level choice.

As an illustration, Figures 11.1, 11.2 and 11.3 show screen-shots from the actual SP survey, where the third, ninth and tenth choice-situation for a given respondent are reproduced here, showing variation in all attributes of the *SP alternative* across observations, while the attributes for the *RP alternative* are kept constant across all the observations for the same respondent.

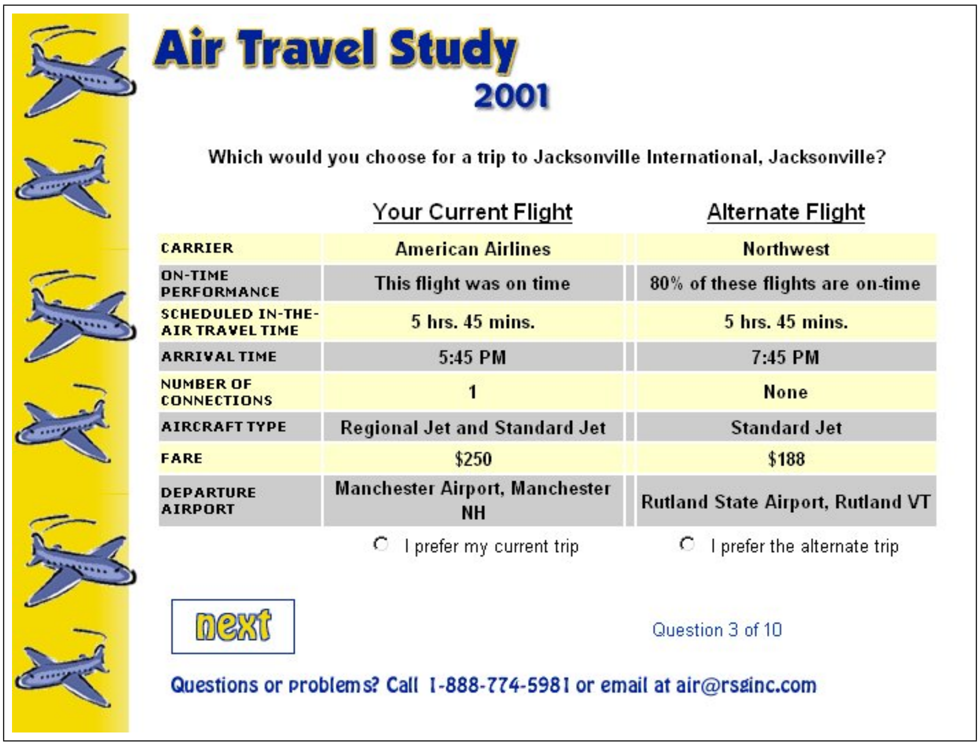
The final sample contains data collected from 589 respondents; with 10 choice-situations per respondent, a sample size of 5,890 observations is obtained, split into 1,190 business travellers, 1,840 holiday travellers and 2,860 VFR travellers. Further segmentations, for example by employment status, did not provide additional gains in performance. Given the small sample sizes, especially for the business segment, and the high number of explanatory variables, the decision was taken to include all observations in the estimation process, rather than *waste* some of them on a validation sample. This decision is also partly motivated by the relative lack of insight gained from using a validation sample in the two RP case-studies (Chapters

[3]It should be noted that, for airlines, there is potential correlation between the ranking of airlines and the membership in frequent flier programmes.

[4]The binomial nature of the data prevents the use of nesting structures. On the other hand, this reduces problems in terms of respondents being confronted with too much information (too many alternatives).

[5]From which access times can be inferred.

[6]From which schedule delays can be calculated, with the help of information on desired arrival times.



©2001, Resource Systems Group, Inc.

Figure 11.1: Example screen-shot for SP survey: third choice-situation

9 and 10), where the performance in the validation sample was very similar to that in the estimation sample, although there is of course a possibility of a different level of risk of overfitting when using SP data.

11.3 Model specification

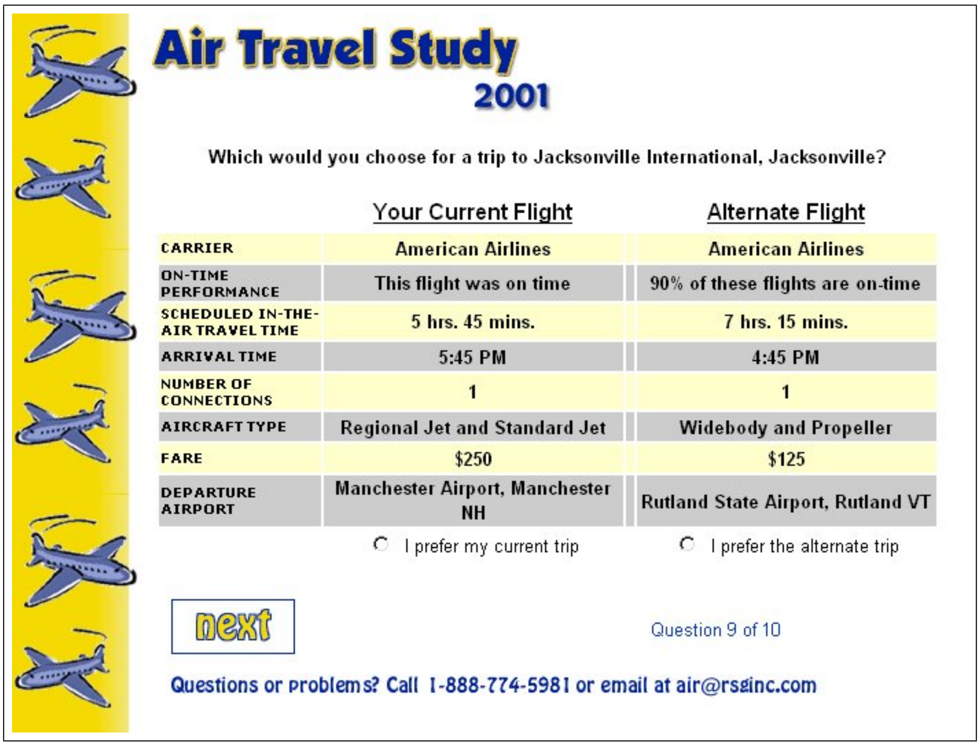
The description of model specification is split into three parts. We first look at the explanatory variables that were included in the specification search. We then describe the specification of the utility in terms of non-linearities in response rate, before turning our attention to the modelling of continuous interactions between explanatory attributes and socio-demographic (and/or trip-specific) characteristics.

11.3.1 Explanatory variables used in specification search

A high number of variables were included in the initial specification search. These are now looked at in turn.

- **Airport dummy variables**: On the basis of the ranking of airports provided by the respondent, dummy variables were associated with the different ranks. The number of airports included in the ranking was limited to four[7], and

[7]It should be noted that the set of airports, and the ranking therein, varies across respondents.



Air Travel Study
2001

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

	Your Current Flight	Alternate Flight
CARRIER	American Airlines	American Airlines
ON-TIME PERFORMANCE	This flight was on time	90% of these flights are on-time
SCHEDULED IN-THE-AIR TRAVEL TIME	5 hrs. 45 mins.	7 hrs. 15 mins.
ARRIVAL TIME	5:45 PM	4:45 PM
NUMBER OF CONNECTIONS	1	1
AIRCRAFT TYPE	Regional Jet and Standard Jet	Widebody and Propeller
FARE	$250	$125
DEPARTURE AIRPORT	Manchester Airport, Manchester NH	Rutland State Airport, Rutland VT
	I prefer my current trip	I prefer the alternate trip

next

Question 9 of 10

Questions or problems? Call 1-888-774-5981 or email at air@rsginc.com

©2001, Resource Systems Group, Inc.

Figure 11.2: Example screen-shot for SP survey: ninth choice-situation

the dummy variable for the lowest-ranked airport was normalised to zero. Attempts were also made to estimate an additional constant, associated with the airport closest to the passenger's ground-level origin, which is not necessarily the preferred airport.

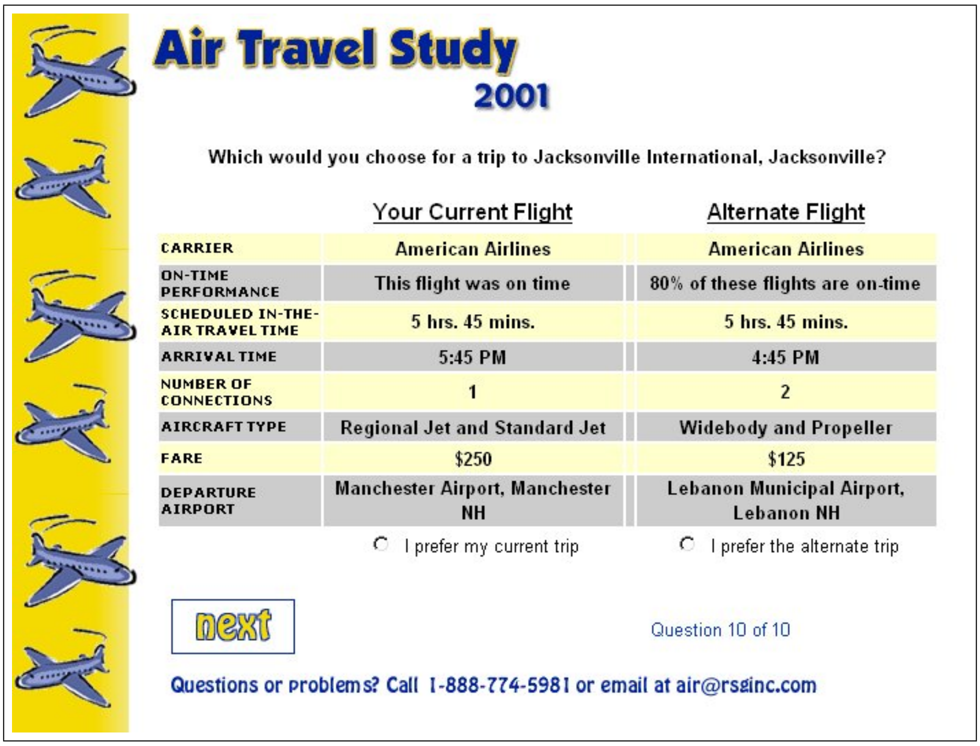
- **Airline dummy variables**: As was the case for the different airport options, four dummy variables were again specified, associated with the three top-ranked airlines and the lowest-ranked airline, where this final dummy variable was normalised to zero.

- **Frequent flier information**: Three dummy variables were included in the base specification, to account for the effects of frequent flier (FF) membership. These were associated with *standard* membership, *elite* membership, and *elite plus* membership.

- **Air-fare**: The fare of a given flight option, in US$.

- **Flight time**: The time from the departure-gate to the arrival-gate (in minutes), also referred to as block time, which includes taxi time.

- **Connections**: The number of connections for a given flight, with three possible levels, 0, 1 and 2. Instead of assuming a linear effect, two separate dummy variables were initially estimated, associated with single and double-connecting flights.


Air Travel Study
2001

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

	Your Current Flight	Alternate Flight
CARRIER	American Airlines	American Airlines
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	4:45 PM
NUMBER OF CONNECTIONS	1	2
AIRCRAFT TYPE	Regional Jet and Standard Jet	Widebody and Propeller
FARE	$250	$125
DEPARTURE AIRPORT	Manchester Airport, Manchester NH	Lebanon Municipal Airport, Lebanon NH
	I prefer my current trip	I prefer the alternate trip

next

Question 10 of 10

Questions or problems? Call 1-888-774-5981 or email at air@rsginc.com

©2001, Resource Systems Group, Inc.

Figure 11.3: Example screen-shot for SP survey: tenth choice-situation

- **Schedule delay**: Separate coefficients were estimated for the penalties associated with an earlier than desired arrival time (SDE), and a later than desired arrival time (SDL).

- **Aircraft-type**: Four different types of aircraft were used in the SP survey; turboprop, regional jet, single-aisle jet, and wide-body jet. Appropriate dummy variables were defined, with single-aisle jet used as the base.

- **On-time performance (OTP)**: For the RP alternative, information was collected on whether the flight was on time or not, while, for the SP alternative, five different levels were used, ranging from 50% to 90% probability of being on time. The high number of levels (7) of the attribute, in conjunction with the low number of observations for some of these levels, led to a decision not to use separate dummy variables for the different levels, but to use a marginal coefficient, in conjunction with appropriate non-linear transforms where applicable (cf. Section 11.3.2).

- **Access time**: The access times for the two alternatives, in minutes, on the mode chosen in the RP survey.

- **Inertia variables**: Attempts were made to account for respondent inertia with the help of a number of variables. Aside from an ASC for the RP alternative (which admittedly also captures other factors), airport and airline inertia constants were included in the utility of the SP alternative in the case where the RP airport or airline was reused in the SP alternative. Additionally, attempts were made to account for inertia effects by including a variable giving the number of flights in the past twelve months in the utility function of the RP alternative.

- **Qualitative variables**: Attempts were also made to include qualitative variables in the utility of the RP alternative, such as the level of satisfaction expressed by the respondent in relation to service. None of these variables was found to have a significant effect, such that quality of service factors are seemingly all captured in the ASCs.

There are differences across the three population segments in terms of which of these variables have a significant impact. These differences are described in detail in the presentation of the MNL results (Section 11.4).

11.3.2 Non-linearities

Except for those variables listed in Section 11.3.1 for which a separate coefficient was associated with each possible level, there are no a priori grounds for believing that a linear specification of utility is appropriate.

With this in mind, for each of the three population segments, an analysis was conducted to test for the presence of non-linear responses. In this analysis, Box-Cox transforms[8] were used for access time, air-fare, flight time, on-time performance, and the two schedule delay variables. On the basis of the results of this Box-Cox analysis, a choice was then made between a linear and a non-linear formulation, where, in the latter, a log-transform was used in the case of decreasing marginal returns, and a power-formulation was used in the case of increasing marginal returns. As such, the Box-Cox transforms are used only in an explanatory role, and are replaced by more *well-behaved* transforms in the final model, easing the numerical issues. The drops in flexibility resulting from this approach were minimal, with the original estimates for λ always being relatively close to values of 0 or 1.

Again, there were differences across the three population groups in terms of the optimal specification, and these are described in detail in the presentation of the MNL results. It should be noted that the use of different specifications in different groups complicates the comparisons across groups, given for example the big impact that the use of a log-transform can have on the estimated mean WTP (cf. Section 10.4.6). However, it was judged that, on the basis of the differences in performance across specifications, the use of a suboptimal specification in some segments with the sole aim of facilitating cross-segment comparisons was not warranted.

11.3.3 Continuous interactions

While the majority of modelling analyses allow for some interactions between estimated parameters and socio-demographic attributes, these generally come in the

[8] With taste coefficient β_x associated with attribute x, the utility component is given by $\beta_x \frac{x^{\lambda_x}-1}{\lambda_x}$. Values of λ_x close to 1 indicate a linear response, while values smaller and larger than 1 indicate a concave and convex response respectively, where, with $\lambda = 0$, the Box-Cox transform is replaced by the log-transform.

form of a segmentation using separate models, or the use of separate coefficients in the same model. The treatment of such interactions in a continuous fashion is relatively rare, with the same applying for interactions between multiple explanatory variables. However, it is clear that such continuous treatments of interactions have advantages in terms of flexibility when compared to the more assumption-bound segmentation approaches. On the other hand, they pose greater demands in terms of the quality of auxiliary data as well as computational cost.

In the SP case-study presented in this chapter, two groups of continuous interactions were included in the final models, after an extensive specification search. The first interaction looks at the impact of travel-distance (in the form of flight time for the RP alternative) on the marginal utilities of access time, air-fare, on-time performance, and early and late arrival. For a given attribute x, the utility was specified as

$$U = \ldots + \beta_x \left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,x}} x + \ldots, \tag{11.1}$$

where FD gives the RP flight time to the current destination, and serves as a proxy for flight-distance, such that the same value of FD is used for the utilities of the RP and SP alternative. The division by the mean observed flight time $\overline{FD}$ ensures that β_x gives the marginal utility of changes in attribute x at the mean flight-distance in the current population segment[9]. With negative values for $\lambda_{FD,x}$, the sensitivity decreases with increases in FD, with the opposite applying in the case of positive values for $\lambda_{FD,x}$. Finally, the rate of the interaction is determined by the absolute value of $\lambda_{FD,x}$, where a value of 0 indicates a lack of interaction.

The same approach was used to account for an interaction between income[10] and the sensitivity to various attributes such as air-fare and access time. As an example, in the case of fare sensitivity, we have:

$$U = \ldots + \beta_{fare} \left(\frac{i}{\bar{i}}\right)^{\lambda_{inc,fare}} fare + \ldots, \tag{11.2}$$

where i gives the household income for the current respondent, with $\bar{i}$ giving the mean income in the appropriate population segment. Here, a negative estimate would be expected for $\lambda_{inc,fare}$, indicating reduced fare-sensitivity with higher income. A problem with this approach in the present context is caused by the fact that income information is presented in the form of a set of separate income-classes, as opposed to absolute income information, leading to a requirement for using class-midpoints, with the obvious averaging error this involves[11]. In practice, no significant interaction with income was identified for any attribute, except air-fare.

Interactions with other factors, such as trip duration, or party size, were not

[9] It can be seen that the normalisation is arbitrary in terms of having no effect on the estimate of $\lambda_{FD,x}$, or indeed on the model fit. However, it does have an effect on the estimated value for β_x, which gives the marginal utility of changes in attribute x when $FD = \overline{FD}$.

[10] In the form of annual household income.

[11] Here, it should be noted that similar averaging errors occur in the case where different income classes are grouped together in a segmentation approach, while the use of separate coefficients in each group risks leading to problems with parameter significance.

Parameters:	17
Observations:	1,190
Final log-likelihood:	-395.14
Adjusted $\rho^2(0)$:	0.5003

	est.	t-stat.
$\beta_{LN(access\,time)}$	-0.3725	-3.20
$\beta_{LN(fare)}$	-3.5344	-13.80
$\beta_{LN(flight\,time)}$	-1.6279	-6.27
$\beta_{LN(SDL)}$	-0.1390	-2.74
β_{SDE}	-0.0019	-1.61
β_{OTP}	0.0088	3.66
$\delta_{current}$	0.5979	4.21
$\delta_{FF\,standard}$	0.4168	2.37
$\delta_{FF\,elite\,and\,elite-plus}$	1.0628	2.60
$\delta_{top-ranked\,airport}$	0.7062	3.68
$\delta_{2^{nd}-ranked\,airport}$	0.2593	1.31
$\delta_{connecting\,flight}$	-0.3747	-2.39
$\delta_{wide-body}$	-0.2534	-1.22
$\delta_{regional\,jet}$	-0.6748	-3.49
$\delta_{turboprop}$	-0.8227	-3.73
$\lambda_{distance,SDE}$	-1.5941	-3.09
$\lambda_{income,LN(fare)}$	-0.1456	-1.61

Table 11.1: Estimation results for MNL model for business travellers

found to be significant.

11.4 MNL models

This section describes the findings of the MNL model fitting analysis[12]. It first looks at the findings in terms of optimal specification in the three population subgroups, before proceeding to a comparison of the actual substantive results across the three groups.

11.4.1 MNL model for business travellers

The findings from the analysis using the $1,190$ observations collected from business travellers are summarised in Table 11.1. Only parameters estimated in the final model are shown here, with any normalised or excluded parameters not listed explicitly. The normalisations used for multi-level attributes are those described in Section 11.3.1.

The analysis revealed effects for all the main continuous variables, including access time, air-fare, flight time, and early and late arrival. Except for the early arrival penalty[13], the analysis showed that the use of a log-transform led to signif-

[12] All models presented in this chapter were estimated with BIOGEME.

[13] The coefficient β_{SDE} is only significant at the 89% level.

icant gains in model performance, suggesting decreasing marginal returns for the associated attributes. Overall, this conclusion is consistent with intuition, with the possible exception of the schedule delay variables, where it is striking that the effect for early arrival was found to be linear, with that for late arrival showing decreasing marginal returns.

The results further show positive effects of improvements in on-time performance. Initial results showed a reduced sensitivity to on-time performance on longer flights, but this resulted in problems with significance for the actual on-time performance coefficient. Efforts to use a power formulation for on-time performance attribute were unsuccessful[14], as were efforts to use separate coefficients for different levels of on-time performance, such that the effect was specified to be linear.

In terms of other interactions, the estimates additionally suggest a reduced sensitivity to early arrival on longer flights, as well as reduced fare-sensitivity with higher income, where the effect is significant at the 89% level. No other interactions were found to be significant in this population segment, such as for example the relationship between flight-distance or income and the sensitivity to access time.

The final part of this discussion looks at the findings in terms of dummy variables. Here, a significant positive ASC was found to be associated with the current alternative, capturing inertia as well as a host of other effects. The estimation further shows a strong effect of frequent flier membership on the utility of an alternative. Here, the effect for elite and elite plus membership was so similar that a common coefficient was used, where the estimates show this effect to be over twice as large as for standard frequent flier membership. The fact that none of the airline dummy variables (linked to ranking) was found to be significant suggests that, for business travellers, airline allegiance is primarily influenced by membership in frequent flier programmes. In terms of airport allegiance, where the dummy for the lowest-ranked airport was used as the base, a significant effect could only be associated with the second and top-ranked airports, where the former one is significant only at the 81% level. Nevertheless, the results do suggest some effects of airport allegiance, especially for the *most preferred* airport.

Interestingly, the estimated dummy variables for flights with one and two connections were indistinguishable, leading to the use of a common factor, which suggests that, for business travellers, flights with multiple connections are not seen as more inconvenient than flights with a single connection[15], ceteris paribus. The final set of dummy variables, associated with aircraft type, show that single-aisle jets are clearly preferred over turboprop planes and regional jets, while the negative effect associated with wide-body jets is not statistically significant above the 78% level of confidence.

[14]This approach would have allowed for a much stronger *dislike* of very late flights than of flights which offer average on-time performance.

[15]It should be noted that this can be in part be seen as a result of the low incidence of flights with double connections in the data, an option that was only made available for destinations more than four hours away.

Parameters:	21
Observations:	1,840
Final log-likelihood:	-532.42
Adjusted $\rho^2(0)$:	0.5661

	est.	t-stat.
$\beta_{LN(access\,time)}$	-0.3488	-3.49
$\beta_{LN(fare)}$	-5.0039	-19.25
$\beta_{LN(flight\,time)}$	-1.9602	-6.76
β_{SD}	-0.0008	-1.57
β_{OTP}	0.0122	5.86
$\delta_{current}$	0.9379	6.97
δ_{FF}	0.1983	0.99
$\delta_{top-ranked\,airport}$	0.9354	4.22
$\delta_{2^{nd}-ranked\,airport}$	0.7179	3.11
$\delta_{3^{rd}-ranked\,airport}$	0.3213	1.29
$\delta_{top-ranked\,airline}$	0.4346	2.51
$\delta_{2^{nd}-ranked\,airline}$	0.3148	1.69
$\delta_{3^{rd}-ranked\,airline}$	0.3482	1.88
$\delta_{single\,connection}$	-0.3398	-2.33
$\delta_{double\,connection}$	-1.0783	-4.18
$\delta_{wide-body}$	0.2330	1.36
$\delta_{regional\,jet}$	0.0228	0.14
$\delta_{turboprop}$	-0.0310	-0.14
$\lambda_{distance,LN(fare)}$	0.1431	2.27
$\lambda_{distance,OTP}$	0.2631	1.71
$\lambda_{income,LN(fare)}$	-0.0430	-0.75

Table 11.2: Estimation results for MNL model for holiday travellers

11.4.2 MNL model for holiday travellers

The findings from the analysis using the 1,840 observations collected from holiday travellers are summarised in Table 11.2, which again only shows parameters included in the final model.

Like in the case of business travellers, the analysis revealed significant effects of access time, air-fare and flight time, where a log-transform was again found to be appropriate for all three attributes. The first difference with the business models arises in the treatment for schedule delay, where the use of linear effects was found to be preferable, and where, given the small differences between the effects for early and late arrival, a common coefficient was used (significant at the 88% level).

The results again show positive effects of improvements in on-time performance, where the estimated coefficient is highly significant in this model. The associated interaction term suggests that holiday travellers' sensitivity to on-time performance increases with flight-distance, although the associated effect is significant only at the 91% level. This can be explained for example by the notion that holiday flights are often pushed to the edges of the off-peak periods, where sensitivity to on-time performance may indeed be larger, and especially so on very long flights.

Other interactions again show a reduced fare-sensitivity with higher income, though the confidence level of the associated term is very low. The interaction terms also show that, for holiday travellers, fare sensitivity increases with flight-distance. It is important to put this into context by remembering that a log-transform is also used on the fare attribute. As such, the results simply suggest that, at a given fare level, increases are valued more negatively in the case of longer flights. A possible explanation for this could be the higher secondary costs associated with longer flights in the case of holiday travellers; such trips are generally more costly overall (e.g. longer duration), leading to a greater desire for savings when it comes to air-fares.

Like in the model for business travellers, the ASC associated with the RP alternative is again positive, and highly significant. However, some important differences arise for the remaining dummy variables. The first observation that can be made is that, as expected, frequent flier benefits play a much smaller role in this segment of the population; as such, it was only possible to estimate a common dummy variable for all levels of membership, and this was significant only at the 68% level. On the other hand, a significant positive effect is associated with the top-ranked airline. Positive effects are also associated with the second and third-ranked airlines, where these are less important and also only significant at lower confidence levels. Furthermore, the difference between these two dummy variables is not significant, although the actual estimates would suggest a higher value for the third-ranked airline than for the second-ranked airline. Aside from these problems, the findings would suggest that, for holiday travellers, airline *preference* plays a bigger role in allegiance than membership in frequent flier programmes. Additionally, positive effects, of decreasing magnitude as well as statistical significance, are associated with the three top-ranked airports.

Unlike for business travellers, the effect associated with flights with two connections is significantly larger than for flights with a single connection, and the scale of the difference (factor of 3) supports the decision not to use a linear effect, but to use two separate dummy variables. Finally, for the aircraft-type dummies, the results suggest that holiday travellers do not distinguish between single-aisle jets, regional jets, and turboprop planes, with the only aircraft dummy with a modestly significant value being that for wide-body aircraft, which are seemingly given preference over single-aisle jets.

11.4.3 MNL model for VFR travellers

The findings from the analysis using the $2,860$ observations collected from VFR travellers are summarised in Table 11.3, which again only shows parameters included in the final model.

An important difference arises immediately when comparing the results for VFR travellers to those for business and holiday travellers. Indeed, while access time and air-fare again enter the utility function under a log-transform, the specification search indicated that it is preferable to treat flight time in a linear fashion. Early and late arrival penalties are treated separately in this model, and both enter the utility in a linear form, where the penalty associated with late arrival is lower, and significant only at the lowly 62% level.

Three non-linear interactions could be retrieved from the data. As in the case

Parameters:	21
Observations:	2,860
Final log-likelihood:	-829.732
Adjusted $\rho^2(0)$:	0.5709

	est.	t-stat.
$\beta_{LN(access\,time)}$	-0.3602	-3.96
$\beta_{LN(fare)}$	-4.7477	-23.84
$\beta_{flight\,time}$	-0.0086	-8.96
β_{SDE}	-0.0012	-3.25
β_{SDL}	-0.0007	-0.87
β_{OTP}	0.0105	6.09
$\delta_{current}$	0.4345	3.70
$\delta_{top-ranked\,airport}$	1.0506	4.93
$\delta_{2^{nd}-ranked\,airport}$	1.0299	5.32
$\delta_{3^{rd}-ranked\,airport}$	0.4880	2.39
$\delta_{closest\,to\,home}$	0.5281	3.10
$\delta_{top-ranked\,airline}$	0.3971	3.37
$\delta_{2^{nd}-ranked\,airline}$	0.2879	2.17
$\delta_{3^{rd}-ranked\,airline}$	0.0900	0.60
$\delta_{connecting\,flight}$	-0.3578	-3.15
$\delta_{wide-body}$	0.5248	3.28
$\delta_{regional\,jet}$	-0.1995	-1.50
$\delta_{turboprop}$	-0.3351	-2.03
$\lambda_{distance,LN(access\,time)}$	-0.4877	-1.91
$\lambda_{distance,LN(fare)}$	0.1915	3.44
$\lambda_{income,LN(fare)}$	-0.0531	-1.34

Table 11.3: Estimation results for MNL model for VFR travellers

of holiday travellers, these again show heightened fare sensitivity on longer flights, along with reduced fare sensitivity with higher income, where this is however only significant at the 82% level. Finally, unlike in the other two models, it was possible to retrieve a relationship between flight-distance and access time sensitivity, which is significant at the 94% level, and shows lower sensitivity to access time on longer flights, which would support a decision to shift long-haul flights to outlying airports, where the issue of point-to-point passengers on the required feeder-flights would however need to be addressed separately.

Like in the two other population segments, the ASC associated with the RP alternative is again positive and highly significant. However, in this segment, it was not possible to estimate a significant effect associated with frequent flier programmes, while the dummy variables associated with the two most preferred airlines are positive and significant at high levels of confidence. The results also indicate that airport allegiance plays a role, where there is however essentially no difference between the estimates of the dummies associated with the two top-ranked airports. Finally, unlike in the other two population segments, it was also possible to identify a significant positive effect associated with the airport closest to the passenger's ground-level origin.

For the same reasons as in the model for business travellers, a common effect was used for flights with single and double connections. In terms of aircraft-type, the difference between single-aisle jets and regional jets is significant only at the 87% level, while the results further indicate a significant dislike for turboprop flights, and a significant preference for wide-body jets over single-aisle jets.

11.4.4 Comparison of results across population segments

The description of the MNL model fitting exercises has already highlighted a number of differences between the specifications used in the three population segments. As such, it has been shown that frequent flier benefits matter more to business travellers, while simple airline preference plays a bigger role for leisure travellers. Other differences arise in the treatment of schedule delays; here, a common non-linear (decreasing) effect is used for holiday travellers, while for VFR travellers, the effect is linear, but the penalty associated with early arrival is larger than that associated with late arrival. For business travellers, SDL is treated in a non-linear fashion, while SDE is treated linearly, but the sensitivity to it decreases on longer flights. A difference also arises in the case of flight time, which is treated linearly for VFR travellers, while a log-transform is used for business and holiday travellers.

A number of other differences also arise in the treatment of interactions between attributes, where the results show higher fare sensitivity on longer flights for holiday and VFR travellers, with no interaction in the case of business travellers. Also, while holiday travellers are more sensitive to on-time performance on longer flights, there is no distance effect on the sensitivity to on-time performance for business and VFR travellers. In all segments, the results suggest reduced fare-sensitivity with higher income, although the interaction parameter never attains a high level of statistical significance. Finally, the results indicate decreased sensitivity to access time on longer flights only in the case of VFR travellers.

These differences in model specification need to be borne in mind when comparing the substantive results across the three population segments. The trade-off that is of the greatest interest is the computation of the willingness to pay measures for the different attributes. Additionally however, it is of interest to look at trade-offs involving the sensitivity to access time, for example to gauge the willingness to accept increases in access time in return for direct flights, or for flights on the preferred airline.

The calculation of the trade-offs from these models is made considerably more complicated than is usually the case, given the high number of non-linear terms in the utility functions. Indeed, there is no single case, among those trade-offs that are of interest, where the simple ratio between coefficients can be used. In those trade-offs involving attributes that enter the utility under a log-transform, the appropriate coefficient needs to be multiplied by the inverse of the associated attribute. In the case where both attributes involved in a trade-off enter under a log-transform, bias was avoided by using the mean value of the ratio between the two attributes across observations as the multiplier, as opposed to the ratio between the mean values of the two attributes.

The situation becomes more complicated again in the case of the fare coefficient, which interacts continuously with income. Here, the term in the utility function is given by $\beta\ \left(\frac{i_n}{\bar{i}}\right)^{\lambda_{inc,fare}}\ f\left(fare\right)$, where $f\left(fare\right)$ is in the present context always equal to $\ln\left(fare\right)$. Given that the derivative of $\left(\frac{i_n}{\bar{i}}\right)^{\lambda_{inc,fare}}$ with respect to $fare$ is 0, the partial derivative of the utility with respect to $fare$ is simply given by $\beta\ \left(\frac{i_n}{\bar{i}}\right)^{\lambda_{inc,fare}}\ f'\left(fare\right)$, where, given the use of a log-transform, we have that $f'\left(fare\right) = \frac{1}{fare}$. The inclusion of $\left(\frac{i_n}{\bar{i}}\right)^{\lambda_{inc,fare}}$ in the partial derivative shows that the actual value of the trade-off will vary across individuals as a function of income. A similar situation arises in the case of attributes where the associated coefficient interacts continuously with flight-distance, where, in the trade-offs, the coefficient is multiplied by $\left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,x}}$. Finally, in the case of holiday and VFR travellers, where the fare-coefficient interacts with income as well as flight-distance, a double multiplier needs to be used.

A special situation arises in the case of trade-offs involving flight time; here, major complications arise, given that the variable used as a proxy for flight-distance in the elasticity formulation is in fact the flight time variable collected in the RP survey (cf. Section 11.3.3). As such, it can be seen that this same variable FD, which is used in the elasticity specification, also enters the utility function of the RP alternative in the SP survey as the flight time attribute, FT. This means that any trade-off involving flight time needs to be calculated separately for the two alternatives. For the pure SP alternative, the partial derivative with respect to flight time is calculated in exactly the same way as for any other attribute, where any continuous interactions with income as well as non-linear transforms are taken into account separately.

However, for the RP alternative used inside the SP survey, the situation becomes more complicated. The actual term involving the flight time attribute will be treated in the same way as in the pure SP alternative. The difficulty arises with regards to any attributes whose marginal utility is defined to depend on flight-distance. As an example, let:

$$U_{RP} = \ldots + \beta_{FT}\ f_{FT}\left(FT_{RP}\right) + \beta_y\ \left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,y}}\ f_y\left(y_{RP}\right) + \ldots. \tag{11.3}$$

The fact that, for the RP alternative in the SP survey, the variables FD and FT are equivalent means that the derivative of $\beta_y\ \left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,y}}\ f_y\left(y_{RP}\right)$ with respect to FT is no longer zero, but is given by

$$\beta_y\ \lambda_{FD,y}\left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,y}-1}\left(\frac{\overline{FD}-\frac{FD}{N}}{\overline{FD}^2}\right)\ f_y\left(y_{RP}\right),$$

where N is the total number of observations, and where the dependency of $\overline{FD}$ on FD is taken into account. The extension of this approach to the case with multiple attributes that have an interaction with flight-distance is straightforward. As such, it can be seen that a trade-off involving flight time will, for the RP alternative, include an additional term for each attribute whose marginal utility has a non-zero flight-distance elasticity. The inclusion of this *correction* term is burdensome, but causes no complications otherwise.

In the present analysis, the comparison was limited to two mains sets of trade-offs, looking at the willingness to accept increases in fare and access time respectively, in return for *improvements* in other determinants of choice. All attributes were included in the calculation of trade-offs, with the exception of the flight time variable. This is partly motivated by the above discussion, in conjunction with the fact that the set of variables where distance elasticity was taken into account varied across the three population groups. This would have led to an unreliable comparison of the trade-off across models, given the differences in the *correction factor*[16]. Additionally however, there is little information to be gained from trade-offs involving flight time, when already looking at trade-offs involving the highly correlated connection and aircraft-type variables, as well as a separate treatment of on-time performance[17]. It should also be noted that trade-offs involving aircraft-type were only calculated in the case of willingness to pay indicators, where the benefits of looking at the willingness to accept access time increases in return for flying on a specific aircraft are limited. Finally, in each case, the trade-offs are presented for the average flight-distance and household income in that population-segment, such that $\left(\frac{i_n}{\bar{i}}\right)^{\lambda_{inc,fare}}$ and $\left(\frac{FD}{\overline{FD}}\right)^{\lambda_{FD,y}}$ become equal to 1. The effects of changes in flight-distance and income on coefficient values, and hence trade-offs, are discussed towards the end of this section.

The results are summarised in Table 11.4 for the willingness to pay indicators, and Table 11.5 for the willingness to accept increases in access time. In each case, several coefficients used in the trade-offs were not significant at the 95% level, as pointed out in Sections 11.4.1, 11.4.2 and 11.4.3, and this is indicated appropriately in the presentation of the trade-offs.

The results show important differences between the three model groups, and while there are strong similarities between the two non-business segments for several of the trade-offs[18], the use of separate models is justified by the differences in other trade-offs, and the differences in the optimal specification, as discussed in Sections 11.4.2 and 11.4.3.

Consistent with a priori expectations, the results show a much greater willingness to accept higher fares in return for shorter access times for business travellers than for holiday or VFR travellers, by a factor of just over 2. Given the use of an air-fare coefficient as opposed to an access cost coefficient in the calculation of the ratio, this trade-off does not correspond to a standard VTTS measure. Nevertheless, the estimates give an indication of the monetary values of reductions in access time. In fact, the high values, especially for business travellers, are broadly consistent with previous research which actually used an access cost coefficient in the calculation of the trade-off. For example, Pels et al. (2003) report values of between \$1.97/min and \$2.90/min for business travellers in the SF-bay area. Lower values were reported in older studies; for example, Harvey (1986) gives a value of \$0.69/min in the SF-bay area, while Furuichi & Koppelman (1994) give a value of \$1.21/min for air-travellers in Japan. These high values, when compared to other contexts, can be explained by

[16] There are additional differences, given the linear treatment of flight time in the VFR model.

[17] This point relates to the fact that the flight times are block times, which take into account taxi time and hence also airport-congestion.

[18] Most notably in the willingness to accept increases in air-fare in return for shorter access-journeys, and vice-versa.

	Business	Holiday	VFR$^{(\dagger)}$
Reduction in access time (1 hour)	75.40	35.80	35.48
Reduction in SDE (1 hour)	13.27$^{(*)}$	2.61$^{(*)}$	3.68
Reduction in SDL (1 hour)	11.08		2.25$^{(*)}$
On-time perf. (+10%)	10.39	7.02	5.57
FF elite or elite-plus *vs* none	125.24	11.44$^{(*)}$	-
FF standard *vs* none	49.12		-
Top airline *vs* worst	-	25.07	21.06
2^{nd} airline *vs* worst	-	18.16$^{(*)}$	15.27
3^{rd} airline *vs* worst	-	20.09$^{(*)}$	4.77$^{(*)}$
Top airport *vs* worst	83.22	53.97	55.73
2^{nd} airport *vs* worst	30.56$^{(*)}$	41.42	54.63
3^{rd} airport *vs* worst	-	18.54$^{(*)}$	25.89
Airport closest to home	-	-	28.02
No connection *vs* one connection	44.15	19.60	18.98
No connection *vs* two connections		62.21	
Jet *vs* wide-body	29.86$^{(*)}$	-	-
Jet *vs* regional jet	79.51	-	10.59$^{(*)}$
Jet *vs* turboprop	96.94	1.79$^{(*)}$	17.77
Wide-body *vs* jet	-	13.45$^{(*)}$	27.84
Regional jet *vs* jet	-	1.31$^{(*)}$	-

$^{(*)}$ Coefficient used in numerator of trade-off not significant at 95% level

$^{(\dagger)}$ Visiting friends or relatives

Table 11.4: MNL trade-offs, part 1: willingness to pay ($)

a variety of factors, including the lower rate of air-trips (as opposed to other travel, e.g. commuting), the greater inflexibility in terms of timing, and the severe financial penalty incurred by arriving at the airport late, and missing the flight[19].

The models also indicate a higher willingness by business travellers to pay for reductions in schedule delay and for improved on-time performance[20]. Interestingly, the models suggest that, except for holiday travellers, respondents are more sensitive to early than to late arrival, a finding that should however be put into context given the small differences, and high associated standard-errors.

Perhaps the most striking difference between population groups comes in the willingness of business travellers to pay $125 to fly on an airline where they hold an elite frequent-flier account. Even though this figure decreases to $49 in the case of

[19] As such, it can be argued that travellers associate a longer access-journey with a higher risk of missing their flight. The financial penalty does not apply for all types of passengers, as it depends on the ticket type. Nevertheless, the delay resulting from missing a specific flight can still cause severe inconvenience.

[20] It should be noted that, across purposes, the findings in relation to sensitivity to on-time performance as well as schedule delay are consistent with the observations of Bates et al. (2001), who, in the context of rail-travel, argue that, with limited influence on determining the actual departure time (when compared to car), travellers care about reliability in addition to schedule delay.

<table>
<tr><th></th><th>Business</th><th>Holiday</th><th>VFR(†)</th></tr>
<tr><td>Reductions in fare ($1)</td><td>2.14</td><td>4.61</td><td>4.57</td></tr>
<tr><td>Reduction in SDE (1 hour)</td><td>17.38(*)</td><td rowspan="2">8.25(*)</td><td>12.24</td></tr>
<tr><td>Reduction in SDL (1 hour)</td><td>17.00</td><td>7.49(*)</td></tr>
<tr><td>On-time perf. (+10%)</td><td>13.60</td><td>22.16</td><td>18.53</td></tr>
<tr><td>FF elite or elite-plus vs none</td><td>163.97</td><td rowspan="2">36.10(*)</td><td>-</td></tr>
<tr><td>FF standard vs none</td><td>64.31</td><td>-</td></tr>
<tr><td>Top airline vs worst</td><td>-</td><td>79.11</td><td>70.08</td></tr>
<tr><td>2^{nd} airline vs worst</td><td>-</td><td>57.31(*)</td><td>50.81</td></tr>
<tr><td>3^{rd} airline vs worst</td><td>-</td><td>63.40(*)</td><td>15.88(*)</td></tr>
<tr><td>Top airport vs worst</td><td>108.96</td><td>170.29</td><td>185.43</td></tr>
<tr><td>2^{nd} airport vs worst</td><td>40.01(*)</td><td>130.69</td><td>181.78</td></tr>
<tr><td>3^{rd} airport vs worst</td><td>-</td><td>58.49(*)</td><td>86.14</td></tr>
<tr><td>No connection vs one connection</td><td rowspan="2">57.81</td><td>61.86</td><td rowspan="2">63.15</td></tr>
<tr><td>No connection vs two connections</td><td>196.29</td></tr>
</table>

(*) Coefficient used in numerator of trade-off not significant at 95% level

(†) Visiting friends or relatives

Table 11.5: MNL trade-offs, part 2: willingness to accept increases in access time (min)

standard membership, the figures are still much higher than for holiday travellers, while no such effects could be identified for VFR travellers[21]. In these latter two groups, the results however show a certain willingness to pay a premium for flying on either of the top-ranked airlines. The results also show a willingness to pay higher fares for flying out of one of the top-ranked airports, where this willingness is especially high for the top-ranked airport in the case of business travellers, while VFR travellers are also willing to pay an additional premium of $28 for flying out of the airport closest to their home. In terms of paying a premium for direct flights, the results again suggest a higher willingness for business travellers, although the different treatment in the case of holiday travellers results in a higher value for the trade-off in the case of flights with 2 connections in this group. A difference arises between the three population groups in the trade-offs looking at the willingness to pay for flying on a specific type of aircraft. Here, the differences in the *most-valued* type of aircraft led to a different base-type. The actual results suggest a high willingness by business travellers to pay for flying on single-aisle jets, while VFR travellers are willing to pay a premium for wide-body jets over single-aisle jets, in addition to a premium for avoiding turboprop flights.

The findings for the trade-offs looking at the willingness to accept increases in

[21] The results are broadly consistent (albeit showing slightly higher values) with those of Proussaloglou & Koppelman (1999), who show a higher willingness to pay such a premium in the case of business travellers than in the case of leisure travellers. As such, the premium for standard membership is $21 in the case of business travellers, compared to $7 in the case of leisure travellers. These values increase in the case of the programme in which respondents participate most actively, with valuations between $52 and $72 for business travellers, compared to between $18 and $26 for leisure travellers.

Distance + 50%	β_{SDE}	$\beta_{LN(fare)}$	β_{OTP}	$\beta_{LN(access\,time)}$
Business	-47.60%	-	-	-
Holiday	-	5.97%	11.26%	-
VFR	-	-	8.07%	-17.94%

Distance - 50%	β_{SDE}	$\beta_{LN(fare)}$	β_{OTP}	$\beta_{LN(access\,time)}$
Business	201.91%	-	-	-
Holiday	-	-9.44%	-16.67%	-
VFR	-	-	-12.43%	40.22%

	Income + 50%	Income - 50%
	$\beta_{LN(fare)}$	$\beta_{LN(fare)}$
Business	-5.73%	10.62%
Holiday	-1.73%	3.02%
VFR	-2.13%	3.75%

Table 11.6: Effects of interaction terms on coefficient-values in MNL models

access time do, overall, show a lower willingness for business travellers than for holiday and VFR travellers, which is to be expected. The main exception again comes in the case of frequent-flier benefits, where the results suggest that business travellers are willing to fly out of more distant airports in return for flying on an airline whose frequent-flier programme they are a member of. Some of the findings, especially in the two leisure groups, show very high values for the trade-offs. Here, the limitations of an approach looking at simple ratios between coefficients should be kept in mind, while also noting that real-world choice set formation would not allow for the inclusion of airports located more than a few hours from a respondent's home[22].

However, one trade-off involving access time is of major interest, especially in the context of the increased use by low cost carriers of outlying airports, namely the willingness to accept increases in access time in return for reductions in air-fares[23]. Here the high willingness, especially in the two leisure groups, can help to at least partly explain the success of such operators in being able to draw travellers away from network carriers and centrally-located airports to more regional bases, with often poor ground-level access facilities.

Here, it should also be noted that some of the trade-offs presented in this section

[22] The average access time in the RP survey was just under one hour.

[23] This trade-off shows the importance of using the *correct* calculation for the multiplier inside the trade-off. Indeed, the non-linearities in the ratio between the access time and fare attributes mean that the willingness to accept increases in fare in return for reductions in access time is in this case not the simple counter-part of the willingness to accept increases in access time in return for reductions in fare. The use of the ratio of means instead of the mean of ratios in the calculation of the multiplier would falsely indicate that the one trade-off is simply the inverse of the other.

are very high, which could potentially be a reflection of the well-established notion that in SP studies, there is a tendency for respondents to exaggerate their responsiveness to changes in attributes (cf. Louviere et al. 2000, Ortúzar 2000). As such, the findings presented here are potentially vulnerable to such exaggeration. In this context, the use of a combined RP/SP approach is of interest.

The final part of the MNL analysis looks at the findings in terms of interactions with socio-demographic and trip-related attributes. For this, the change in the concerned coefficients (with the obvious impact on trade-offs) was calculated in the case of increases and decreases by 50% in flight-distance[24] and in household income, with results summarised in Table 11.6. Comparisons between models are only possible in two cases. As such, the results show that the effect of changes in flight-distance on the sensitivity to on-time performance is more significant for holiday than for VFR travellers. More interestingly however, the results show that the relationship between income and fare-sensitivity is stronger in the case of business travellers than in the two leisure groups, although the lower level of significance for the associated elasticity parameter needs to be borne in mind (cf. Tables 11.1, 11.2 and 11.3).

11.5 Extension to MMNL models

In this section, we extend the analysis from Section 11.4 to the use of MMNL models, with the aim of accommodating any variations in tastes across respondents that could not be explained in a deterministic fashion.

The MMNL models estimated in this section are based on the final MNL specifications presented in Tables 11.1, 11.2 and 11.3, using the same specification in terms of response-rate to changes in the various attributes (linear *vs* non-linear). A major issue arises here due to the non-linear specifications of utility, and specifically, the continuous interactions with flight-distance and income. These interactions lead to a more complicated form for the derivatives of the log-likelihood function, and hence higher estimation cost[25]. While, in the MNL models, this increase in computational cost is acceptable, the fact that mixture models require the computation of the choice probabilities for a high number of draws at each iteration of the optimisation algorithm makes the approach very expensive in the case of MMNL models. Initial estimation results (using a low number of draws) showed that the actual values of the interaction parameters remained largely unchanged when compared to the MNL models, such that, in the final MMNL models, the interaction parameters were kept fixed at the MNL estimates[26], allowing for the use of a high number of draws, leading to a high level of precision in simulation[27]. The fact that the interaction parameters are thus kept fixed in the MMNL models means that nested LR-tests cannot be

[24] In the form of flight time for the RP alternative.

[25] The estimation software, BIOGEME, uses numerical derivatives for the non-linear part of utility functions. The other non-linearities used in the present model, in the form of log-transforms, can be applied at the data-level.

[26] An additional extension, allowing the interaction parameters to vary randomly across individuals, was not explored in this context. In any case, it can be argued that the effects of such a variation can also be captured by allowing for variation in the associated taste coefficient.

[27] In this analysis, 10,000 MLHS draws were used per individual and per dimension, ensuring very low simulation error.

used to compare the performance to that of the MNL models; here, the adjusted ρ^2 measure can give an indication of the gains in performance obtained by allowing for random variations in tastes across respondents.

The next issue that needs to be addressed is the choice of distribution to be used for any coefficients specified to vary randomly across respondents. In the context of the discussions in Chapter 4, the use of flexible distributions, such as the S_B, has clear advantages over the commonly used Normal distributions. However, their use leads to a very significant increase in computational cost, especially when additionally allowing for a treatment of the repeated choice nature of the dataset[28]. On these grounds, the estimation results presented in this section are based on the use of the Normal distribution for all randomly distributed coefficients. This leads to less reliable results, especially in terms of behaviour in the tails of the population, a fact that needs to be borne in mind in model interpretation. Given these limitations, the scope of this analysis focusses less on producing accurate results in terms of randomly distributed willingness to pay indicators, but rather aims to give an account of the extent of random variations in tastes across air-travellers, in addition to the results produced in Section 11.4 in terms of deterministic variations.

11.5.1 MMNL model for business travellers

The estimation results for the MMNL model for business travellers are summarised in Table 11.7. The estimation revealed significant levels of random taste heterogeneity in three taste coefficients, namely those associated with changes in access time[29], fare and on-time performance, as well as in three dummy variables, namely those associated with the second-ranked airport, connecting flights, and regional jets. With these six randomly distributed parameters, and despite the drop in flexibility resulting from using fixed interaction parameters, the MMNL model offers an improvement in the adjusted ρ^2 measure by 12.79%. These results suggest significant levels of random variation across respondents in the business group, in addition to the deterministic variation accounted for by the two interaction parameters. The results however also show the effects of using the Normal distribution. While there is no a priori sign assumption for the dummy variable associated with regional jets, a positive coefficient would be expected for the dummy variable associated with the second-ranked airport (given that the worst-ranked airport is used as the base), with a negative effect for the dummy variable associated with connecting flights. However, with the use of the Normal distribution, probabilities of counter-intuitively signed coefficients of 35% and 24% are obtained for the airport and connection dummies respectively. Similar issues arise in the case of the three randomly distributed β coefficients, with probabilities of incorrectly signed coefficient-values of 10%, 2% and 18% for the access time, air-fare and on-time performance coefficients respectively. Given the discussion in Chapter 4, it is important to stress that these results should primarily be seen as an artefact of using the Normal distribution, and not as evidence of individuals with non-rational behaviour.

[28] In this study, we accommodate this under the assumption that tastes vary across respondents, but not across observations for the same respondent.

[29] Dispersion parameter significant at the 93% level.

Parameters:	21
Observations:	1,190
Final LL (MMNL):	-338.35
Adjusted $\rho^2(0)$ (MMNL):	0.5643
Final LL (MNL):	-395.14
Adjusted $\rho^2(0)$ (MNL):	0.5003

	est.	t-stat.
$\beta_{LN(access\ time)}$ (μ)	-0.8173	-3.47
$\beta_{LN(access\ time)}$ (σ)	0.6464	1.79
$\beta_{LN(fare)}$ (μ)	-6.4991	-7.97
$\beta_{LN(fare)}$ (σ)	3.0004	4.59
$\beta_{LN(flight\ time)}$	-2.7906	-5.40
β_{SDE}	-0.0025	-2.21
$\beta_{LN(SDL)}$	-0.1507	-1.51
β_{OTP} (μ)	0.0295	3.29
β_{OTP} (σ)	0.0316	3.26
$\delta_{current}$	0.7304	1.99
$\delta_{FF\ standard}$	0.9452	2.45
$\delta_{FF\ elite\ and\ elite-plus}$	1.2312	2.00
$\delta_{top-ranked\ airport}$	1.6639	4.51
$\delta_{2^{nd}-ranked\ airport}$ (μ)	0.5233	1.30
$\delta_{2^{nd}-ranked\ airport}$ (σ)	1.3649	2.33
$\delta_{connecting\ flight}$ (μ)	-1.2430	-3.30
$\delta_{connecting\ flight}$ (σ)	1.7559	3.59
$\delta_{wide-body}$	-0.5105	-1.12
$\delta_{regional\ jet}$ (μ)	-0.7415	-2.04
$\delta_{regional\ jet}$ (σ)	2.0589	3.64
$\delta_{turboprop}$	-0.9488	-2.55
$\lambda_{distance,SDE}$	-1.5941	-
$\lambda_{income,LN(fare)}$	-0.1456	-

Table 11.7: Estimation results for MMNL model for business travellers

11.5.2 MMNL model for holiday travellers

The estimation results for the MMNL model for holiday travellers are summarised in Table 11.8. The analysis revealed significant levels of random taste heterogeneity for four marginal utility coefficients, namely those associated with access time, air-fare, flight time and on-time performance[30], as well as three dummy variables, namely those associated with membership in a frequent-flier programme, flights with a single connection, and turboprop flights. With these seven randomly distributed parameters, the MMNL model obtains an improvement in the adjusted ρ^2 measure by 8.23% when compared to the MNL model, where the scope for improvement was reduced by the higher MNL ρ^2 measure when compared to the model for business travellers. While the results suggesting a relatively even spread of people with a

[30]Dispersion parameter significant at the 94% level.

Parameters:	25
Observations:	1,840
Final log-likelihood (MMNL):	-468.97
Adjusted $\rho^2(0)$ (MMNL):	0.6127
Final log-likelihood (MNL):	-532.42
Adjusted $\rho^2(0)$ (MNL):	0.5661

	est.	t-stat.
$\beta_{LN(access\,time)}$ (μ)	-0.7479	-2.61
$\beta_{LN(access\,time)}$ (σ)	1.3649	3.31
$\beta_{LN(fare)}$ (μ)	-10.1548	-7.22
$\beta_{LN(fare)}$ (σ)	4.2677	5.37
$\beta_{LN(flight\,time)}$ (μ)	-3.9619	-4.96
$\beta_{LN(flight\,time)}$ (σ)	3.0112	3.11
β_{SD}	-0.0017	-2.20
β_{OTP} (μ)	0.0226	4.05
β_{OTP} (σ)	0.0177	1.90
$\delta_{current}$	1.3714	4.31
δ_{FF} (μ)	0.6280	1.32
δ_{FF} (σ)	2.4540	3.78
$\delta_{top-ranked\,airline}$	0.8742	2.43
$\delta_{2^{nd}-ranked\,airline}$	0.8859	2.22
$\delta_{3^{rd}-ranked\,airline}$	0.5743	1.57
$\delta_{top-ranked\,airport}$	2.1411	3.98
$\delta_{2^{nd}-ranked\,airport}$	1.6695	3.13
$\delta_{3^{rd}-ranked\,airport}$	0.8284	1.64
$\delta_{single\,connection}$ (μ)	-0.6381	-2.21
$\delta_{single\,connection}$ (σ)	1.1748	2.50
$\delta_{double\,connection}$	-1.7875	-3.76
$\delta_{wide-body}$	0.5024	1.46
$\delta_{regional\,jet}$	0.3635	1.26
$\delta_{turboprop}$ (μ)	-0.5851	-0.95
$\delta_{turboprop}$ (σ)	2.4627	3.28
$\lambda_{distance,LN(fare)}$	0.1431	-
$\lambda_{distance,OTP}$	0.2631	-
$\lambda_{income,LN(fare)}$	-0.0430	-

Table 11.8: Estimation results for MMNL model for holiday travellers

negative or positive attitude towards turboprop flights (59%-41%) are not a priori counter-intuitive, the distributional effects on the sign conclusions need to be borne in mind in the interpretation of results showing probabilities of 29% and 9% for a wrongly-signed access time and flight time coefficient respectively, along with a 10% and 40% probability of a wrongly-signed effect for on-time performance and frequent-flier membership. Again however, the results do suggest the presence of significant levels of random variations in taste across holiday travellers.

Parameters:	25
Observations:	2,860
Final LL (MMNL):	-757.57
Adjusted $\rho^2(0)$ (MMNL):	0.6052
Final LL (MNL):	-829.732
Adjusted $\rho^2(0)$ (MNL):	0.5709

	est.	t-stat.
$\beta_{LN(access\,time)}$ (μ)	-0.7475	-3.92
$\beta_{LN(access\,time)}$ (σ)	1.2134	5.00
$\beta_{LN(fare)}$ (μ)	-8.2048	-9.97
$\beta_{LN(fare)}$ (σ)	3.7696	6.38
$\beta_{flight\,time}$ (μ)	-0.0150	-7.15
$\beta_{flight\,time}$ (σ)	0.0073	1.77
β_{SDE}	-0.0019	-2.65
β_{SDL} (μ)	-0.0026	-1.46
β_{SDL} (σ)	0.0084	2.12
β_{OTP} (μ)	0.0221	5.32
β_{OTP} (σ)	0.0230	3.28
$\delta_{current}$	0.3957	1.92
$\delta_{top-ranked\,airline}$	0.7192	3.21
$\delta_{2^{nd}-ranked\,airline}$	0.6297	2.45
$\delta_{3^{rd}-ranked\,airline}$	0.3868	1.52
$\delta_{top-ranked\,airport}$	1.7623	4.45
$\delta_{2^{nd}-ranked\,airport}$	1.8251	5.05
$\delta_{3^{rd}-ranked\,airport}$	0.9143	2.50
$\delta_{closest\,to\,home}$	0.9847	2.86
$\delta_{connecting\,flight}$ (μ)	-0.5829	-2.78
$\delta_{connecting\,flight}$ (σ)	0.6266	1.23
$\delta_{wide-body}$	0.8675	2.87
$\delta_{regional\,jet}$ (μ)	-0.2086	-0.80
$\delta_{regional\,jet}$ (σ)	1.2241	3.67
$\delta_{turboprop}$	-0.5571	-1.97
$\lambda_{distance,LN(access\,time)}$	-0.4877	-
$\lambda_{distance,LN(fare)}$	0.1915	-
$\lambda_{income,LN(fare)}$	-0.0531	-

Table 11.9: Estimation results for MMNL model for VFR travellers

11.5.3 MMNL model for VFR travellers

The estimation results for the MMNL model for VFR travellers are summarised in Table 11.9. The analysis revealed significant variations for all β coefficients except that associated with early departures, where the dispersion parameter for the flight time coefficient is additionally only significant at the 92% level. The results also show variation in the dummy variables associated with connecting flights[31], and

[31]Dispersion parameter significant at the 78% level.

regional jets. With these seven random parameters, the MMNL model offers an improvement in the adjusted ρ^2 measure by 6% over the MNL model. In this model, the use of the Normal distribution again leads to problems of interpretation, with probabilities of counter-intuitive values of 27% for the access time coefficient, 38% for the SDL coefficient, 17% for the on-time coefficient, and 18% for the dummy variable associated with connecting flights.

11.5.4 Conclusions for MMNL analysis

The three MMNL analyses described in this section have revealed that, in addition to the taste heterogeneity accounted for through a segmentation into three separate population groups, and the use of continuous interactions with socio-demographic variables, there is additional, non-quantifiable variation in a large number of the parameters used in model specification. This gives the MMNL model a clear advantage in terms of flexibility over its closed form counterpart. Here, it should however also be noted that part of the reason for the improvements in model fit offered by the MMNL models is the treatment of the repeated choice nature of the data, something that was not possible in the MNL models.

The MMNL estimation processes conducted as part of this study were limited to the use of the Normal distribution, opening the door to potentially misleading results in terms of the behaviour in the tails of the population. Although, for what is arguably the most important coefficient, namely that associated with air-fare, the probability of a counter-intuitively signed coefficient is in the present study always negligible, this is not the case for other coefficients, notably that associated with access time. Especially in the context of air-travel, it is very difficult to make a case for the presence of individuals with negative valuations of access time reductions, such that these results should indeed be seen as an artefact of the distributional assumptions.

These issues with counter-intuitive signs lead to problems in the computation of trade-offs on the basis of simulation. As such, to avoid issues with cancelling out in the numerator[32], and problems caused by the inclusion in the denominator of values straddling zero, it would again be necessary to remove the upper and lower few percentiles of the distribution of some of the coefficients. In cases where the share of counter-intuitively signed coefficient-values is large, as with the present estimates, the number of percentile points that would need to be removed is however so high that the resulting distribution would have a much reduced variance when compared to the original distribution. As such, this not only leads to very unreliable results in terms of the variation in the trade-offs[33], but in fact leads to an approximation of a simple ratio of means approach, which has its own limitations, in terms of producing biased trade-offs.

Given the above discussion, the computation of trade-offs in this chapter is limited to the case of MNL models, as described in Section 11.4.4. Even though, given the results in terms of the extent of random variations in tastes, it is likely that the trade-offs from the MNL models are themselves biased, there is no a priori reason for

[32] Caused by the presence of positive as well as negative values.

[33] Here, the bias is now caused by the censoring, as opposed to the issue of counter-intuitively signed coefficient-values.

believing that this bias is any larger than would be the case with trade-offs based on MMNL estimates produced with the help of a Normal distribution, given the issues discussed above. The results presented in this section merely serve as an indication of the potential of MMNL structures in the analysis of air-travel choice behaviour. They also highlight the need for future work, using more appropriate distributional assumptions. Here however, additional work is required to provide more powerful implementations of flexible distributions such as the S_B, easing the computational as well as numerical problems faced when using existing code.

11.6 Summary and Conclusions

This chapter has described a study of air-travel choice behaviour making use of SP data collected in the US in 2001.

In common with the results from the RP studies described in Chapters 9 and 10, the analysis presented in this chapter has highlighted the important role that ground-level distance plays in airport choice behaviour. However, while, in the two RP studies, it was not generally possible to retrieve a significant and meaningful effect of changes in air-fares, the results from this SP study have shown air-fare to be the variable with the most explanatory power, across the three population segments used in the analysis. This result is consistent with intuition, and highlights a certain advantage of SP data in this context, given that reliable information is available on the choices that respondents were actually faced with. Additionally, in the context of SP data, data protection issues do not apply, where, aside from air-fares, this also applies in the case of airline allegiance. As such, while no effects of airline allegiance could be identified in the RP case-studies[34], the SP analysis presented in this chapter has revealed significant effects in response to membership of frequent-flier programmes, as well as general airline-preference. Although these results do suggest a certain advantage for SP data in the analysis of air-travel choice behaviour, these advantages need to be put into context by remembering the usual limitations affecting this type of data. This in turn suggests that an important avenue for future research in air-travel comes in the use of a combination of RP and SP data, as discussed by Algers & Beser (2001).

Aside from illustrating the potential advantages of SP data, the study described in this chapter has also achieved several other aims. One of the main innovations in the context of air-travel is the use of a continuous treatment of the interactions between socio-demographic attributes and the sensitivity to travel-attributes. The improvements in performance obtained with this approach were significant, and the approach has clear theoretical advantages in terms of flexibility over more basic methods, such as a simple segmentation into different income-classes.

Another important topic addressed in this chapter is the way in which attributes enter the utility function. Although the use of log-transforms for some of the attributes, such as flight frequency, has now become commonplace, other attributes, such as air-fare and access time, are in general still treated in a linear fashion in aviation research. The estimation work conducted in this chapter has shown this to

[34]With the exception of allegiance to the national carrier by visiting business travellers in the London study (cf. Section 10.4.2).

be inappropriate in many cases, consistent with the results from the London case-study (Chapter 10). Instead of simply comparing the use of a log-transform to a linear approach, the work described in this chapter made use of Box-Cox transforms in a preliminary analysis. Although no incidence of such cases was discovered in the present analysis[35], the use of this approach can also alert the modeller to the presence of variables with increasing marginal returns, something that is not possible when simply comparing the results of a linear and a log-linear approach.

Here, it should also be noted that the analysis has revealed important differences in the optimal specification across the three population segments, in terms of the variables included in the utility function, the way these variables enter the utility function, and the interaction with socio-demographic and trip-related attributes.

Finally, the analysis has also highlighted the gains in model performance and insights into travel-behaviour that can be obtained with the help of model structures allowing for random variations in tastes, such as MMNL. The gains in the present case were larger than those reported in the SF-bay case-study (Chapter 9), which can be explained on the grounds of more detailed level-of-service data, as well as the presence of repeated observations for the same individual. However, the study has also highlighted some of the complications that still limit the widespread applicability of mixture models in large-scale modelling analyses. The main issue identified in the present context is the requirement, imposed by computational issues, to rely on the use of the Normal distribution, leading to counter-intuitive results in terms of behaviour in the tails of the population, and an inability to produce adequate trade-offs based on random coefficients. Aside from addressing issues of implementation, an important avenue for future research in this context is the analysis of the potential advantages of using a direct specification of such willingness to pay indicators (cf. Fosgerau 2004), as opposed to basing them on ratios of randomly distributed coefficients.

[35]Initial results showing increasing marginal returns for improvements in on-time performance came at the cost of a severe drop in significance for the associated taste coefficient.

Chapter 12

Summary, conclusions and directions for future research

This chapter provides a summary of the work described in this thesis, suggests avenues for future research, and presents the conclusions of the thesis.

12.1 Summary

Given the layout of the thesis, it makes sense to structure the summary in the same way, looking first at the work presented in the theoretical part, before proceeding to the findings of the case-studies of air-travel choice behaviour conducted in the applied part.

12.1.1 Theoretical part

The theoretical part of the thesis can again be divided into three sub-parts, discussing issues related to the cost of estimation and application of mixture models, the specification of random taste heterogeneity, and issues related to the error-structure of advanced discrete choice models.

Simulation processes

The work described in Chapter 3 looks at ways of addressing the cost of estimation and application of discrete choice models whose choice probabilities do not have a closed form expression, such as MMNL.

The discussion in this thesis focusses on the use of alternatives to PMC draws in the simulation of these choice probabilities, in the form of *cleverly-crafted* sequences of draws which provide a higher uniformity of coverage of the area of integration. The review of existing work highlights the major differences between the various available methods, where there is a strong relationship between the performance of a method and the cost of implementation and generation of the draws. Here, it can be argued that, for the purpose of the estimation and application of advanced discrete choice models, the more basic approaches have a certain advantage, allowing for easy implementation and generalisation, while still offering significant improvements in performance over PMC draws.

However, as discussed in detail in Chapter 3, important issues arise with the commonly used Halton sequences, where problems may in fact occur in relatively low-dimensional applications, unlike generally assumed in the existing literature. Additionally, there are also major shortcomings with the adapted versions of the Halton sequence. This discussion forms the motivation for the development of the Modified Latin Hypercube Sampling (MLHS) approach. Although the results in a large-scale, sixteen-dimensional application, are not entirely conclusive, calling for more research, the findings from a less complex four-dimensional application show that important reductions in computational cost can be obtained with the use of the MLHS approach as an alternative to PMC draws. Additionally, the approach has major advantages in terms of implementation when compared to more flexible methods, and in terms of generalisation to high-dimensional problems, when compared to Halton-based approaches.

Specification of random taste heterogeneity

The work presented in Chapters 4 and 5 is concerned with the representation of random taste heterogeneity across respondents, a very timely issue in the face of the mounting popularity of the MMNL model.

In the context of continuous distributions (Chapter 4), the discussion highlights the important risks of misleading results when relying on the commonly used Normal distribution. Here, the unbounded nature of the distribution, together with its strict symmetry assumption, can lead to results showing non-trivial shares of respondents with counter-intuitively signed coefficients, even in the case where such coefficient values are not actually *revealed* by the data. This notion is supported by the findings of a case-study that makes use of eleven different continuous distributions, in what is one of the most extensive such comparisons to date. The findings not only highlight the risk of counterintuitive results when using the Normal distribution[1], but also show that more flexible distributions, including some that had not previously received widespread exposure in the field of discrete choice modelling, such as the Johnson S_B or S_U, are much less likely to produce misleading results. Additionally however, the analysis shows that model fit on its own is not the best indicator of performance, especially when interested in behaviour in the tails of the population.

While the use of flexible continuous distributions can reduce the risk of biased results, there are situations in which even the most flexible distributions are not able to reproduce the distribution of tastes in the population. In this context, chapter 5 discusses an alternative approach for the analysis of random taste heterogeneity, based on using discrete mixtures of an underlying GEV model across a finite set of support points. While such approaches have been used previously in the analysis of travel behaviour (Gopinath 1995, Dong & Koppelman 2003), they have received comparatively limited exposure, despite their appealing characteristics in terms of being free from any assumptions in relation to the shape of the distribution. In the present context, the discrete mixture approach is used in a novel way, namely for accommodating a share of the population that is indifferent to changes in a given attribute. The analysis in this chapter shows that the presence of such zero

[1] Or other symmetrical distributions.

valuations, if left untreated, as is commonly the case, can result in significant bias in the estimates, as well as leading to lower model fit. The use of a discrete mixture approach can help to significantly reduce this risk.

Error-structure

Although looking at quite separate issues, Chapters 6 and 7 have a common factor in terms of being concerned with the representation of inter-alternative correlation.

The work in Chapter 6 discusses an important issue of specification and interpretation, related to the potential risk of confounding between random taste heterogeneity and substitution patterns between alternatives. The discussion shows that, if modellers allow for the effects of only either of the two phenomena, they are at risk of producing biased results, with the findings in relation to the modelled phenomenon, say random taste heterogeneity, being masked by the effects of the unmodelled phenomenon, say correlation between the unobserved part of utility of different alternatives. These theoretical claims are supported by results from six different case-studies. As such, the advice offered in this chapter is that researchers should always strive to jointly allow for random taste heterogeneity and correlated error-terms, with the help of a GEV mixture model, or an appropriately specified MMNL model[2]. Although some risk of confounding still persists even with such advanced models, this is much reduced when compared to the more basic approaches.

The work presented in Chapter 7 discusses the development of a novel structure, which allows for random variations in the substitution patterns between alternatives across respondents, leading to increased model flexibility. Such a Mixed Covariance model can be specified either with purely random variation or with a mixture between random and deterministic variation, for example as an extension of the COVNL model of Bhat (1997). Additionally, the model can be based on an underlying GEV or ECL structure. Finally, the model can be specified as a continuous mixture or as a discrete mixture. The discussion in Chapter 7 is mainly concerned with the development of the model structure, and there is still a need for large-scale testing. Nevertheless, the basic example conducted in Section 7.3 suggests that the model structure is able to retrieve variations in the error-structure across respondents, hence avoiding a source of bias in forecasting applications.

12.1.2 Applied part

The applied part of the thesis discusses the findings of three case-studies of air-travel choice-behaviour, using RP data collected in the SF-bay area (Chapter 9) and Greater London (Chapter 10), and SP data collected in the US (Chapter 11). We will now briefly summarise the findings of the three studies, before discussing the overall findings of the applied part of the thesis.

SF-bay area case-study

The analysis using the RP survey data collected in the SF-bay area suggests that the main attributes affecting choice-behaviour across all population segments are

[2] I.e. using a combination of the RCL and ECL approaches.

access time and flight-frequency. Effects for other variables, such as air-fare, flight time, access cost, and aircraft-size could only be retrieved in some of the population segments. Additionally however, habit formation seems to play a consistent role across population segments, with travellers more likely to fly out a given airport again if they have done so in the past.

In terms of model structure, the findings from this analysis show that all three NL models, using nesting by airport, airline or access-mode, lead to increases in model-fit over the corresponding MNL model[3], with differences in the optimal nesting structure across the six population segments used. Additionally, gains in model fit are obtained by using a MMNL model, allowing for random variations across respondents in the sensitivity to factors such as access time.

Finally, an important contribution of this chapter is the development of a framework for the simultaneous modelling of choices by air-travellers along the airport, airline and access-mode dimension. Aside from having conceptual advantages, the analysis also confirms that this approach leads to better prediction performance.

Greater London case-study

The analysis using the RP survey data collected in Greater London reveals significant effects of changes in access time, access cost, flight time and frequency across all four population segments, while air-fare is only found to have a significant effect for visiting leisure travellers, with visiting business travellers being the only segment of the population that shows any effects of airline allegiance, in terms of a preference for non-UK carriers.

In terms of model structure, the results from the London study again show that all three two-level NL models lead to improvements in model performance over the corresponding MNL models, where there are again differences across population segments in the optimal nesting structure.

The main contribution in this chapter is the use of a CNL structure for the simultaneous analysis of correlation along all three dimensions of choice. The analysis shows that this approach leads to better model fit than any of the three two-level NL structures. Additionally, as discussed in Chapter 8, this approach has important advantages by not relying on a specific ordering of the different levels of nesting.

From a topical point of view, this chapter also has the merit of being the first study of its type in the London area, making use of advanced discrete choice structures for the joint analysis of choices by air-travellers along multiple dimensions of choice.

SP case-study

The results produced by the study using the SP data for airport and airline choice are quite different from those produced in the RP studies. As such, while access time is still found to have a major impact on choice-behaviour, the most important factor is now air-fare, where it was not possible to estimate a significant effect for this attribute across all segments in the two RP studies. Additionally, with the SP

[3]With the exception of the models using nesting by airport and airline for visiting VFR travellers.

data, it is now possible to retrieve effects of airline allegiance, schedule delay, and on-time performance.

From a methodological point of view, the SP study again shows an advantage for the more flexible model structures, with MMNL obtaining significant improvements in model fit over MNL, where these improvements are larger than in the case of the SF-bay study[4]. Additionally, in this study, the quality of the data is sufficiently high to allow for an analysis of the continuous interactions between taste coefficients and socio-demographic attributes, showing, amongst others, a decreasing fare sensitivity with higher income. Such approaches, despite their obvious benefits in terms of flexibility as well as interpretation, have thus far only received very limited exposure, especially in the field of air-travel behaviour research.

Summary

The results of the three studies are not directly comparable, given the geographical differences, the differences in the age of the data[5], and the differences in survey design as well as data type (RP *vs* SP). Nevertheless, some conclusions can be reached.

From a model structure point of view, all three case-studies have shown that the use of more advanced model structures can lead to improvements in model fit. However, although the improvements are statistically significant, they are too small to lead to any major differences in model performance. Nevertheless, the advanced model structures provide further insights into choice behaviour, and there are also differences in the substantive results between the various models.

The main observation that can be made in the comparison of the results across the three studies is the greater ability of the SP models to retrieve significant effects for a range of variables that are generally not well estimated in RP studies, such as air-fares, schedule delay and airline and airport allegiance. This is an illustration of the complications that arise with the use of RP survey data in the analysis of air-travel choice behaviour, where there are issues of data quality in relation to air-fares and availabilities, while information on a number of other attributes, notably the membership in frequent flier programmes, is not generally available in such datasets[6]. Additionally, it should be noted that problems in RP studies can also arise in terms of the measurement of attributes along secondary dimensions of choice, as in the case of access cost, where bias in the data led to an underestimation of the VTTS for resident business travellers in the London study in Chapter 10.

The one common observation that can be made from the three case-studies is that the results do suggest that access time plays a major role in the choice process, with passengers being highly captive to their local airport. As such, the attractiveness of outlying airports depends heavily on good access-connections, unless there are other incentives, such as low air-fares. This is reflected in the fact that only low-cost carriers find it relatively easy to attract passengers to outlying airports that are not served by convenient and fast ground-level services. It is conceivable that the

[4]It is likely that this is at least partly due to the presence of multiple observations per individual, which facilitates the analysis of taste heterogeneity, and allows for correlation across replications.

[5]Especially when comparing the SP data to the two sets of RP data.

[6]This also generally applies in the case of other RP data sources, such as bookings data.

sensitivity to access time decreases with flight time[7], such that moving long-haul services to outlying airports would seem wise; this however causes problems as the associated (and necessary) short-haul feeder flights will also carry point-to-point passengers, who will again have a preference for more centrally-located airports.

12.2 Directions for future research

A number of directions for future research can be identified, where these are again divided into two parts, relating to the theoretical and applied parts of the thesis respectively.

12.2.1 Discrete choice theory

In the context of the discussions in the theoretical part of the thesis, there is clearly some scope for further testing, with the aim of establishing whether the results produced in this work extend to other datasets and choice-scenarios. This applies especially to the findings in terms of zero VTTS in Chapter 5, and the findings in terms of confounding between different components of the error-structure, as discussed in Chapter 6. Although more research on the specification of continuous taste heterogeneity would also be of interest, this should apply especially to the exploration of further alternative distributions, and the development of efficient implementations for the estimation of models based on flexible distributions. Finally, more testing is necessary both for the Mixed Covariance structure discussed in Chapter 7, and the MLHS approach discussed in Chapter 3, where, in the context of simulation efficiency, more work is also required in terms of the applicability to discrete choice modelling of approaches producing multi-dimensional draws directly, as opposed to using combinations of one-dimensional sequences.

12.2.2 Air-travel behaviour research

A number of important avenues for future research can also be identified in the context of the applied part of the research. Given the conclusions in terms of the complications that arise in the case of RP data, and the risk of bias when relying solely on the use of SP data, one example is the combined use of RP and SP data, as discussed by Algers & Beser (2001). In this context however, important issues need to be addressed in terms of the compatibility between datasets, and it is likely that tailor-made surveys need to be used (as opposed to relying on existing RP and SP surveys). Here, collaborations between researchers and industry players such as airlines and airports are of interest. Additionally, it remains to be seen how fruitful an approach this is in the context of RP survey data, as opposed to RP bookings data, given the issues of availabilities arising with the former.

From a methodological point of view, it is important to further explore the use of mixture models, but with the use of more flexible distributions than the Normal. Here, important issues of computational cost however need to be addressed. Also, the more in-depth exploration of the use of the CNL model remains an important

[7] As suggested by the results for VFR travellers in Section 11.4.3

avenue for future work. Additionally, as mentioned before, it is of interest to explore correlations between the separate choice-dimensions[8]. Finally, the separate findings with regards to taste heterogeneity and complex substitution patterns highlight the need for a study using mixture models capable of jointly accommodating these two phenomena, especially given the discussion in Chapter 6 in relation to the risk of confounding between the two phenomena. Again, important computational issues arise in this case.

12.3 Conclusions

The research presented in the theoretical parts of this thesis highlights the problems of specification and interpretation that arise with the use of advanced model structures, for example in terms of assumptions relating to the distribution of taste coefficients, and the error-structure of the *true* model, where significant risks of misspecification exist, potentially leading to misinterpretation and misguided policy implications.

In this context, an important point needs to be addressed. Indeed, advanced models clearly have the potential to offer improvements in performance and accuracy in cases where the assumptions made by less flexible models, such as MNL and NL, cannot be justified. On the other hand, there are evidently also cases in which the use of the more basic models is acceptable, and where the additional gains in performance obtained with the advanced models is negligible in the face of the associated increases in computational cost. However, with the gain in popularity of models offering a flexible treatment of the error-structure, modellers are more and more relying on such structures to explain processes that could otherwise be accommodated in the observed part of utility, which, for interpretation purposes, is clearly preferable. This applies especially in the context of the analysis of variations in tastes across respondents, where there is a trend for modellers to increasingly rely purely on a random treatment.

The applied part of the thesis has shown that, in the context of air-travel behaviour research, advanced model structures do indeed have the potential to lead to better performance than the more basic approaches, while also providing further insights, and potentially different substantive results. Overall however, the results show that the gains in performance are relatively modest, especially when compared to the very significant increases in the cost of estimation and application[9]. In this context, issues arise not just with the use of mixture models, but also in the case of advanced GEV structures, where, in the London case-study (Chapter 10), major numerical complications arose in the estimation of the CNL models. Additionally, the use of such models can lead to increases in data requirements, which can further reduce their appeal, despite their conceptual advantages. These problems are an illustration of the difficulties of moving a *theoretical* model into the *real-world*, and are at least one reason for the prevailing gap between the state-of-the-art and the state-of-practice.

[8] As opposed to correlation along individual choice-dimensions.

[9] Which plays a crucial role in practice, given that a high number of forecasting applications may be needed in policy analysis.

Additionally, although the theoretical discussions in the first part of this thesis show the importance of the assumptions made in model specification, and offer some guidelines for good practice, the applied part of the thesis shows that, for practical modelling purposes, these guidelines often need to be violated. This relates mainly to the distributional assumptions in the case of random coefficients models, but also applies to the recommendation to always jointly allow for the effects of random taste heterogeneity across individuals and substitution patterns between alternatives. This is again an illustration of the difficulties involved in using advanced model structures in large-scale analyses.

This discussion calls for greater cooperation between *researchers* and *practitioners*. More effort needs to go into improving the transferability of model structures from theory to practice, and work is also needed to reduce the cost of application of such advanced structures in large-scale policy-oriented analyses. The people who develop the more advanced model structures need to do a better job at selling their advantages to those people that are actually supposed to use them, rather than just their peers. This extends particularly to showing that the differences between structures go beyond model fit[10], in that the use of the advanced approaches leads to a reduction in the risk of bias, in willingness to pay indicators as well as substitution patterns, two components that are of crucial interest to policy-makers.

[10]Where, as shown in the case-studies presented in this work, the differences may be relatively modest.

References

Abay, G. (1999), *Nachfrageabschaetzung Swissmetro: Eine stated-preference Analyse (in German)*, Berichte des Nationalen Forschungsprogrammes 41: Verkehr und Umwelt, Bericht F1, EDMZ, Bern.

Adler, T., Falzarano, C. S. & Spitz, G. (2005), *Modeling Service Trade-offs in Air Itinerary Choices*, paper presented at the 84^{th} Annual Meeting of the Transportation Research Board, Washington, DC.

Algers, S., Bergstroem, P., Dahlberg, M. & Lindqvist Dillen, J. (1998), *Mixed Logit Estimation of the Value of Travel Time*, Working Paper, Department of Economics, Uppsala University, Uppsala, Sweden.

Algers, S. & Beser, M. (2001), 'Modelling choice of flight and booking class - a study using stated preference and revealed preference data', *International Journal of Services Technology and Management* **2**(1-2), 28–45.

Andrews, R. L. & Srinivasan, S. (1995), 'Studying consideration effects in empirical choice models using scanner panel data', *Journal of Marketing Research* **32**(February), 30–41.

Armstrong, P., Garrido, R. A. & Ortúzar, J. de D. (2001), 'Confidence interval to bound the value of time', *Transportation Research Part E: Logistics and Transportation Review* **37**(1), 143–161.

Ashford, N. & Bencheman, M. (1987), 'Passengers' choice of airport: an application of the Multinomial Logit model', *Transportation Research Record* **1147**, 1–5.

Ashley, D., Hanson, P. & Veldhuis, J. (1995), 'A policy-sensitive traffic forecasting model for Schiphol airport', *Journal of Air Transport Management* **2**(2), 89–97.

Axhausen, K. W., Koenig, A., Abay, G., Bates, J. & Bierlaire, M. (2004), *Swiss Value of Travel Time Savings*, paper presented at the European Transport Conference, Strasbourg, France.

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. & Haupt, T. (2002), 'Observing the rhythms of daily life: A six-week travel diary', *Transportation* **29**(2), 95–124.

Baret, M.-P. (2001), *L'aéroport Marseille-Provence révise ses objectifs de fréquentation à la baisse*, Le Monde, June 9^{th} 2001, Paris.

Barret, S. D. (2000), 'Airport competition in the deregulated European aviation market', *Journal of Air Transport Management* **6**, 13–27.

Basar, G. & Bhat, C. R. (2004), 'A Parameterized Consideration Set model for airport choice: an application to the San Francisco Bay area', *Transportation Research Part B: Methodological* **38**(10), 889–904.

Bastin, F. (2004), Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models, PhD thesis, Namur, Belgium.

Bastin, F., Cirillo, C. & Hess, S. (2005), *Evaluation of optimisation methods for estimating mixed logit models*, Transportation Research Record, forthcoming.

Bastin, F., Cirillo, C. & Toint, P. L. (2003), *Numerical Experiments with AMLET, a New Monte-Carlo Algorithm for Estimating Mixed Logit Models*, paper presented at the 10^{th} International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Bates, J. (1987), 'Measuring travel time values with a discrete choice model: a note', *The Economic Journal* **97**, 493–498.

Bates, J., Ashley, D. & Hyman, G. (1987), *The Nested Incremental Logit Model: Theory and Application to Mode Choice*, Proceedings of the PTRC summer annual meeting, Volume 290, PTRC, London.

Bates, J., Polak, J. W., Jones, P. & Cook, A. (2001), 'The valuation of reliability for personal travel', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 191–229.

Batley, R. P. & Daly, A. (2003), *Establishing equivalence between nested logit and hierarchical elimination-by-aspects choice models*, paper presented at the European Transport Conference, Strasbourg, France.

Battersby, B. (2004), *Conclusions on Capacity, Utility and Air-Travel Demand in Regional Australia*, paper presented at the Annual Air Transport Research Society conference, Istanbul, Turkey.

Becker, S. G. (1965), 'A theory of the allocation of time', *Economic Journal* **75**(299), 493–517.

Bell, J. C., Walls, R., Allan, J. R., Watola, G., Burton, N. H. K., Musgrove, A. J. & Rehfish, M. M. (2003), *Study on the potential safety risks from birds at and around a potential new airport at Cliffe marshes and measures for mitigating those risks*, study commissioned by the Department for Transport, London.

Ben-Akiva, M. (1973), Structure of passenger travel demand models, PhD thesis, Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. (1974*a*), *Note on the specification of a Logit Model with Utility Functions that Include Attributes of Competing Alternatives*, Working Paper, Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. (1974*b*), 'Structure of passenger travel demand models', *Transportation Research Record* **526**, 26–41.

Ben-Akiva, M. (1977), *Choice Models with simple choice set generating process*, Working Paper, Dept. of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. & Bierlaire, M. (1999), Discrete choice methods and their applications to short term travel decisions, *in* R. Hall, ed., 'Handbook of Transportation Science', Kluwer Academic Publishers, Dordrecht, The Netherlands, chapter 2, pp. 5–34.

Ben-Akiva, M. & Boccara, B. (1995), 'Discrete choice models with latent choice sets', *International Journal of Research in Marketing* **12**(1), 9–24.

Ben-Akiva, M. & Francois, B. (1983), *μ-homogenous generalized extreme value model*, Working Paper, Department of Civil Engineering, MIT, Cambridge, MA.

Ben-Akiva, M. & Lerman, S. R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.

Bhat, C. R. (1995), 'A heteroscedastic extreme value model of intercity travel mode choice', *Transportation Research Part B: Methodological* **29**(6), 471–483.

Bhat, C. R. (1997), 'Covariance Heterogeneity in Nested Logit models: economic structure and application to inter-city travel', *Transportation Research Part B: Methodological* **31**(1), 11–21.

Bhat, C. R. (1998*a*), 'Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling', *Transportation Research Part A: Policy and Practice* **32**(7), 495–507.

Bhat, C. R. (1998*b*), 'Analysis of travel mode and departure time choice for urban shopping trips', *Transportation Research Part B: Methodological* **32**(6), 361–371.

Bhat, C. R. (2000*a*), 'A multi-level cross-classified model for discrete response variables', *Transportation Research Part B: Methodological* **34**(7), 567–582.

Bhat, C. R. (2000*b*), 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation Science* **34**(2), 228–238.

Bhat, C. R. (2001), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model', *Transportation Research Part B: Methodological* **35**(7), 677–693.

Bhat, C. R. (2003), 'Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences', *Transportation Research Part B: Methodological* **37**(9), 837–855.

Bhat, C. R. & Guo, J. (2004), 'A mixed spatially correlated Logit model: formulation and application to residential choice modeling', *Transportation Research Part B: Methodological* **38**(2), 147–168.

Bierlaire, M. (2002), *The Network GEV model*, Proceedings of the 2^{nd} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Bierlaire, M. (2003), *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3^{rd} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Bierlaire, M. (2005), *A theoretical analysis of the cross-nested Logit model*, Annals of Operations Research, forthcoming.

Bierlaire, M., Axhausen, K. W. & Abay, G. (2001), *The acceptance of modal innovation : The case of Swissmetro*, Proceedings of the 1^{st} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Bierlaire, M., Bolduc, D. & McFadden, D. (2003), *Characteristics of Generalized Extreme Value Distributions*, Technical report, Econometrics Laboratory, University of California, Berkeley, CA.

Bierlaire, M., Lotan, M. & Toint, P. L. (1997), 'On the overspecification of multinomial and Nested Logit models due to alternative specific constants', *Transportation Science* **31**(4), 363–371.

Boeing (2004), *Current Market Outlook*, Market Analysis, Boeing Commercial Airplanes, Seattle, WA.

Bondzio, L. (1996), Modelle für den Zugang von Passagieren zu Flughäfen (Models for the Passengers' Access to Airports), PhD thesis, Ruhr-University, Bochum.

Bowman, J. L. (2004), *Logit kernel (or mixed logit) models for large multidimensional choice problems: identification and estimation*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Boyd, J. & Mellman, J. (1980), 'The effect of fuel economy standards on the US automotive market: a hedonic demand analysis', *Transportation Research Part A: General* **14**(5-6), 367–378.

Braaten, E. & Weller, G. (1979), 'An improved low-discrepancy sequence for multidimensional Quasi-Monte Carlo integration', *Journal of Computational Physics* **33**, 249–258.

Bradley, M. A. (1998), Behavioural models of airport choice and air route choice, *in* J. de D. Ortúzar, D. Hensher & S. R. Jara-Diaz, eds, 'Travel behaviour research: updating the state of play (IATBR 94)', Elsevier, Oxford, pp. 141–159.

Bradley, M. A. & Daly, A. (1996), Estimation of logit choice models using mixed stated-preference and revealed-preference information, *in* P. R. Stopher & M. Lee-Gosselin, eds, 'Understanding Travel Behaviour in an Era of Change', Elsevier, Oxford, chapter 9, pp. 209–231.

Bratley, P. & Fox, B. L. (1988), 'Implementing Sobol's quasi-random sequence generator', *ACM Transactions on Mathematical Software* **14**, 88–100.

Bratley, P., Fox, B. L. & Niederreiter, H. (1992), 'Implementation and tests of low-discrepancy sequences', *ACM Transactions on Modeling and Computer Simulation* **2**, 195–213.

Breffle, W. S., Morey, E. & Waldman, D. (2005), Gaussian Quadrature versus Simulation for the Estimation of Random Parameters: Some Evidence from Stated-Preference Choice Data, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 17, pp. 341–354.

Bresnahan, T., Stern, S. & Trajtenberg, M. (1997), 'Market Segmentation and the Sources of Rents from Innovation : Personal Computers in the late 1980's', *Rand Journal of Economics* **28**, 17–44.

Brooke, A. S., Caves, R. & Pitfield, D. E. (1994), 'Methodology for predicting European Short-Haul Air Transport Demand from Regional Airports: An Application to East Midlands International Airport', *Journal of Air Transport Management* **1**(1), 37–46.

Bruning, E. R. & Prentice, B. E. (2002), *A comparison of country-of-origin loyalties for Canadian, Mexican, and American consumers*, paper presented at the International Federation of Scholarly Associations of Management (IFSAM) 2002 Conference, Gold Coast, Queensland, Australia.

Burge, P. & Rohr, C. (2004), *DATIV: SP Design: Proposed approach for pilot survey*, Tetra-Plan in cooperation with RAND Europe and Gallup A/S.

CAA (1996), *UK airports Passenger survey, CAP 677 - Passengers at Birmingham, Gatwick, Heathrow, London City, Luton, Manchester and Stansted in 1996*, The Civil Aviation Authority, London.

Cabret, N. (2004), *La décision d'Air France de supprimer la ligne Grenoble-Orly menace l'avenir de l'aéroport isérois*, Le Monde, June 30^{th} 2004, Paris.

Cantillo, V. & Ortúzar, J. de D. (2005), 'A semi-compensatory discrete choice model with explicit attribute thresholds of perception', *Transportation Research Part B: Methodological* **39**(7), 641–657.

Cardell, S. & Dunbar, F. (1980), 'Measuring the societal impacts of automobile downsizing', *Transportation Research Part A: General* **14**(5-6), 423–434.

Carrasco, J. A. & Ortúzar, J. de D. (2002), 'Review and assessment of the nested logit model', *Transport Reviews* **22**(2), 197–218.

Cascetta, E., Nuzzolo, A., Russo, F. & Vitetta, A. (1996), *A modified logit route choice model overcoming path overlapping problems, specification and some calibration results for interurban networks*, Proceedings of the 13th International Symposium on the Theory of Road Traffic Flow, Lyon, France.

Caussade, S., Ortúzar, J. de D., Rizzi, L. I. & Hensher, D. (2005), 'Assessing the influence of design dimensions on stated choice experiment estimates', *Transportation Research Part B: Methodological* **39**(7), 621–640.

Caves, R., Ndoh, N. N. & Pitfield, D. E. (1991), *Route Choice Modeling applied to the Choice between Mature Airports and Emergent Airports in their Shadow*, paper presented at the 31^{th} RSA European Congress, Lisbon, Portugal.

Cherchi, E. & Ortúzar, J. de D. (2004), *Random effect vs. error component in the Mixed Logit model: correlation, heteroscedasticity, preference and response unobserved heterogeneity*, paper presented at the 10^{th} triennial World Conference on Transport Research, Istanbul, Turkey.

Cherchi, E. & Polak, J. W. (2005), *The Assessment of User Benefits Using Discrete Choice Models: Implications of Specification Errors Under Random Taste Heterogeneity*, paper presented at the 84^{th} Annual Meeting of the Transportation Research Board, Washington, DC.

Chernew, M., Gowrisankaran, G. & Scanlon, D. (2002), *Learning and the Value of Information: The Case of Health Plan Report Cards*, National Bureau of Economic Research, Working paper 8589.

Chin, A. T. H. (2002), 'Impact of frequent flyer programs on the demand for air travel', *Journal of Air Transportation* **7**(2), 53–86.

Chintagunta, P., Jain, D. & Vilcassim, N. (1991), 'Investigating heterogeneity in brand preference in logit models for panel data', *Journal of Marketing Research* **28**, 417–428.

Chu, C. A. (1989), *Paired Combinatorial Logit Model for Travel Demand Analysis*, Proceedings of the fifth World Conference on Transportation Research, Yokahama, Japan.

Cirillo, C. & Axhausen, K. W. (2004), *Evidence on the distribution of values of travel time savings from a six-week diary*, Arbeitsbericht Verkehrs- und Raumplanung 212, IVT, ETH Zurich.

Coldren, G. M. & Koppelman, F. S. (2005), 'Modeling the competition among air-travel itinerary shares: GEV model development', *Transportation Research Part A: Policy and Practice* **39**(4), 345–365.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K. & Mukherjee, A. (2003), 'Modeling aggregate air-travel itinerary shares: logit model development at a major US airline', *Journal of Air Transport Management* **9**(6), 361–369.

Daganzo, C. F. (1979), *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York.

Dalal, S. R. & Klein, R. W. (1988), 'A flexible class of discrete choice models', *Marketing Science* **7**(3), 232–251.

Daly, A. (1987), 'Estimating Tree Logit models', *Transportation Research Part B: Methodological* **21**(4), 251–267.

Daly, A. (2001*a*), 'Alternative tree Logit models: comments on a paper of Koppelman and Wen', *Transportation Research Part B: Methodological* **35**(8), 717–724.

Daly, A. (2001*b*), *Recursive Nested Extreme Value Model*, Working Paper 559, ITS, University of Leeds, Leeds, U.K.

Daly, A. & Bierlaire, M. (2005), *A general and operational representation of GEV models*, Transportation Research Part B: Methodological, forthcoming.

Daly, A., Hess, S., Hyman, G., Polak, J. W. & Rohr, C. (2005), *Modelling Time Period Choice: Experience from the UK and the Netherlands*, paper presented at the European Transport Conference, Strasbourg, France.

Daly, A. & Zachary, S. (1978), Improved multiple choice models, *in* D. Hensher & Q. Dalvi, eds, 'Identifying and Measuring the Determinants of Mode Choice', Teakfields, London.

de Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M. A. & Hofman, F. (2003), 'A model for time of day and mode choice using error components logit', *Transportation Research Part E: Logistics and Transportation Review* **39**(3), 245–268.

de Neufville, R. (1995), 'Management of multi-airport systems: A development strategy', *Journal of Air Transport Management* **2**(2), 99–110.

De Serpa, A. C. (1971), 'A theory of the economics of time', *The Economic Journal* **81**(324), 828–846.

Debreu, G. (1960), 'Review of R.D. Luce, Individual choice behavior: A theoretical analysis', *American Economic Review* **50**, 186–188.

Dennis, N. (2004), *Running out of steam? - the growth prospects for European low-cost airlines*, paper presented at the Air Travel Research Society Annual Meeting, Istanbul, Turkey.

Dennis, N. (2005), 'Industry consolidation and future airline network structures in Europe', *Journal of Air Transport Management* **11**(3), 175–183.

DfT (1997), *Air traffic forecasts for the United Kingdom 1997*, Department for Transport, London.

DfT (2000), *Air traffic forecasts for the United Kingdom 2000*, Department for Transport, London.

DfT (2003*a*), *The future development of air transport in the UK: South East (2nd edition)*, Department for Transport, London.

DfT (2003*b*), *The Future of Air Transport: Government White Paper*, The Department for Transport, London.

Domencich, T. A. & McFadden, D. (1975), *Urban Travel Demand : A behavioural analysis*, North Holland, Amsterdam.

Dong, X. & Koppelman, F. S. (2003), *Mass Point Mixed Logit Model: Development and Application*, paper presented at the 10^{th} International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Doornik, J. A. (2001), *Ox: An Object-Oriented Matrix Language*, Timberlake Consultants Press, London.

EPRI (1997), *Methodology for Predicting the Demand for New Electricity-Using Goods*, EA-593, Project 488-1, Final Report, Electric Power Research Institute, Palo Alto, CA.

Evans, A. (1972), 'On the theory of the valuation and allocation of time', *Scottish Journal of Political Economy, February* pp. 1–17.

Evans, M., Hastings, N. & Peacock, B. (2000), *Statistical Distributions*, Wiley series in probability and statistics, third edn, John Wiley & sons, inc., New York.

Fosgerau, M. (2004), *Investigating the distribution of the value of travel time savings*, Working Paper, Danish Transport Research Institute, Copenhagen, Denmark.

Franz, J. D. (1996), *Metropolitan Transportation Commission Airline Passenger Survey: Final Report*, prepared by J.D. Franz Research for the Metropolitan Transport Commission, Sacramento, CA.

Fry, T. R. L. & Harris, M. N. (2002), *The DOGEV Model*, Working Paper, Department of Econometrics and Business Statistics, Monash University, Monash, Australia.

Furuichi, M. & Koppelman, F. S. (1994), 'An analysis of air traveler's departure airport and destination choice behaviour', *Transportation Research Part A: Policy and Practice* **28**(3), 187–195.

Garrido, R. A. (2003), *Estimation Performance of Low Discrepancy Sequences in Stated Preferences*, paper presented at the 10^{th} International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Garrow, L. (2004), *Comparison of models representing correlation and random taste variation: An application to airline passengers' rescheduling behavior*, Published doctoral dissertation, Northwestern University, Evanston, Illinois.

Garrow, L. & Koppelman, F. S. (2004*a*), 'Multinomial and nested logit models of airlines passengers no-show and standby behaviour', *Journal of Revenue and Pricing Management* **3**(3), 237–253.

Garrow, L. & Koppelman, F. S. (2004*b*), 'Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information', *Journal of Air Transport Management* **10**(6), 401–411.

Gaudry, M. & Degenais, M. (1978), 'The Dogit model', *Transportation Research Part B: Methodological* **13**(2), 105–111.

Gaudry, M., Jara-Diaz, S. R. & Ortúzar, J. de D. (1989), 'Value of time sensitivity to model specification', *Transportation Research Part B: Methodological* **23**(2), 151–158.

Gaudry, M. & Wills, M. J. (1978*a*), 'Estimating the functional form of travel demand models', *Transportation Research* **12**, 257–289.

Gaudry, M. & Wills, M. J. (1978*b*), 'Testing the Dogit Model with aggregate time-series and cross-sectional travel data', *Transportation Research Part B: Methodological* **13**(2), 155–166.

Geman, S. & Geman, D. (1984), 'Stochastic relaxation gibbs distributions and the bayesian restoration of images', *IEEE Transactionson Pattern Analysis and Machine Intelligence* **6**, 721–741.

González-Savignat, M. (2004), 'Competition in air transport: The case of high speed train', *Journal of Transport Economics and Policy* **38**(1), 77–108.

Gopinath, D. (1995), Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand, PhD thesis, MIT, Cambridge, MA.

Greene, W. & Hensher, D. (2003), 'A latent class model for discrete choice analysis: contrasts with mixed logit', *Transportation Research Part B: Methodological* **37**(8), 681–698.

Greene, W., Hensher, D. & Rose, J. (2005), *Accounting for heterogeneity in the variance of unobserved effects in mixed logit models*, Transportation Research Part B: Methodological, forthcoming.

Hajivassiliou, V. A. & McFadden, D. (1998), 'The method of simulated scores for the estimation of LDV models', *Econometrica* **66**, 863–896.

Halton, J. (1960), 'On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals', *Numerische Mathematik* **2**, 84–90.

Hammersley, J. M. & Morton, K. (1956), 'A new monte carlo technique: Antithetic variates', *Proceedings of the Cambridge Philosophical Society* **52**, 449–474.

Han, B., Algers, S. & Engelson, L. (2001), *Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by Using the Mixed Logit Model*, paper presented at the 80^{th} Annual Meeting of the Transportation Research Board, Washington, DC.

Harvey, G. (1986), 'Study of airport access mode choice', *Journal of Transportation Engineering* **112**(5), 525–545.

Harvey, G. (1987), 'Airport choice in a multiple airport region', *Transportation Research Part A: General* **21**(6), 439–449.

Hastings, W. (1970), 'Monte carlo sampling methods using markov chains and their applications', *Biometrika* **57**, 97–109.

Hausman, J. & Wise, D. (1978), 'A conditional Probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences', *Econometrica* **46**(2), 403–426.

Heckman, J. L. (1981), Statistical Models for Discrete Panel Data, *in* C. Manski & D. McFadden, eds, 'Structural Analysis of Discrete Data with Econometric Applications', MIT Press, Cambridge, MA, chapter 3, pp. 114–178.

Heiss, F. (2002), 'Structural Choice Analysis with Nested Logit models', *The Stata Journal* **2**(3), 227–252.

Hellekalek, P. (1984), 'Regularities in the distribution of special sequences', *Journal of Number Theory* **18**, 41–55.

Hensher, D. (1999), 'HEV Choice Models as a Search Engine for the Specification of Nested Logit Tree Structures', *Marketing Letters* **10**(4), 339–349.

Hensher, D. (2001*a*), 'Measurement of the valuation of travel time savings', *Journal of Transport Economics and Policy* **35**(1), 71–98.

Hensher, D. (2001*b*), 'The sensitivity of the valuation of travel time savings to the specification of unobserved effects', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 129–142.

Hensher, D. (2001*c*), 'The valuation of commuter travel time savings for car drivers: Evaluating alternative model specifications', *Transportation* **28**, 101–118.

Hensher, D. (2004), 'Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings', *Journal of Transport Economics and Policy* **38**(3), 425–446.

Hensher, D. & Greene, W. (2002), 'Specification and estimation of the nested logit model: alternative normalisations', *Transportation Research Part B: Methodological* **36**(1), 1–17.

Hensher, D. & Greene, W. (2003), 'The Mixed Logit Model: The State of Practice', *Transportation* **30**(2), 133–176.

Hensher, D., Stopher, P. R. & Louviere, J. (2001), 'An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application', *Journal of Air Transport Management* **7**(6), 373–379.

Hess, S. (2004), *A model for the joint analysis of airport, airline, and access-mode choice for passengers departing from the San Francisco Bay area*, paper presented at the European Transport Conference, Strasbourg, France.

Hess, S. (2005), *Reducing the cost of estimating mixture models: adaptive drawing techniques and other useful approaches*, CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. & Axhausen, K. W. (2005), *Distributional assumptions in the representation of random taste heterogeneity*, paper presented at the 5^{th} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Hess, S., Bierlaire, M. & Polak, J. W. (2004), *A note on the effects of unobserved travel-experience factors on the estimation of value of travel-time savings*, CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*a*), Capturing taste heterogeneity and correlation structure with Mixed GEV models, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 4, pp. 55–76.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*b*), *Discrete mixture models with applications to the estimation of value of travel-time savings*, CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*c*), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research Part A: Policy and Practice* **39**(2-3), 221–236.

Hess, S., Bolduc, D. & Polak, J. W. (2005), *Modélisation des choix discrets avec hétérogénéité aléatoire dans la structure de covariance des erreurs*, (Modelling of discrete choice processes with random variations across respondents in the covariance structure), paper presented at the 45^{th} annual congress of the Canadian Economic society, Manoir Richelieu, Charlevoix, Québec, Canada.

Hess, S. & Polak, J. W. (2003*a*), *A comparison of Scrambled and Shuffled Halton sequences for Simulation Based Estimation*, CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. & Polak, J. W. (2003*b*), *An alternative method to the scrambled Halton sequence for removing correlation between standard Halton sequences in high dimensions*, paper presented at the 42^{nd} Meeting of the European Regional Science Association, Jyväskylä, Finland.

Hess, S. & Polak, J. W. (2004*a*), *Mixed Logit estimation of parking type choice*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. & Polak, J. W. (2004*b*), *On the use of Discrete Choice Models for Airport Competition with Applications to the San Francisco Bay area Airports*, paper presented at the 10^{th} triennial *WCTR* conference, Istanbul, Turkey.

Hess, S. & Polak, J. W. (2005*a*), *Accounting for random taste heterogeneity in airport-choice modelling*, Transportation Research Record, forthcoming.

Hess, S. & Polak, J. W. (2005*b*), 'Mixed logit modelling of airport choice in multi-airport regions', *Journal of Air Transport Management* **11**(2), 59–68.

Hess, S., Polak, J. W. & Bierlaire, M. (2005), Functional approximations to alternative-specific constants in time-period choice-modelling, *in* H. Mahmassani, ed., 'Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Proceedings of the 16^{th} International Symposium on Transportation and Traffic Theory', Elsevier, Amsterdam, chapter 28, pp. 545–564.

Hess, S., Polak, J. W. & Daly, A. (2003), *On the performance of shuffled Halton sequences in the estimation of discrete choice models*, paper presented at the European Transport Conference, Strasbourg, France.

Hess, S., Polak, J. W., Daly, A. & Hyman, G. (2004), *Flexible Substitution Patterns in Models of Mode and Time of Day Choice: New evidence from the UK and the Netherlands*, paper presented at the European Transport Conference, Strasbourg, France, October 2004.

Hess, S., Train, K. & Polak, J. W. (2005), *On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice*, Transportation Research Part B: Methodological, forthcoming.

Horowitz, J. (1991), 'Reconsidering the multinomial Probit model', *Transportation Research Part B: Methodological* **25**(6), 433–438.

Innes, J. D. & Doucet, D. H. (1990), 'Effects of access distance and level of service on airport choice', *Journal of Transportation Engineering* **116**(4), 507–516.

Jara-Diaz, S. R. (2000), Allocation and valuation of travel time savings, *in* D. Hensher & K. Button, eds, 'Handbooks in Transport, Vol. 1: Transport Modelling', Pergamon Press, Oxford, pp. 303–319.

Jara-Diaz, S. R. (2003), 'On the goods-activities technical relations in time allocation theory', *Transportation* **30**, 245–260.

Jara-Diaz, S. R. & Guevara, C. A. (2003), 'Behind the subjective value of travel time savings - the perception of work, leisure, and travel from a joint mode choice activity model', *Journal of Transport Economics and Policy* **37**, 29–46.

Kamakura, K. W. & Russell, G. (1989), 'A probabilistic choice model for market segmentation and elasticity structure', *Journal of Marketing Research* **26**, 379–390.

Kanafani, A. & Sadoulet, E. (1977), 'The partitioning of long haul air traffic - a study in multinomial choice', *Transportation Research* **11**(1), 1–8.

Kerlouegan, P. & Gelie, P. (2001), *Air France réorganise ses fréquences et prépare la saison estivale*, Le Figaro, March 6^{th} 2001, Paris.

Kitamura, R. & Bunch, D. S. (1990), Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components, *in* M. Koshi, ed., 'Transportation and Traffic Theory', Elsevier Science Publishing Co., Amsterdam, pp. 477–496.

Kocis, L. & Whiten, W. J. (1997), 'Computational investigations of low-discrepancy sequences', *ACM Transactions on Mathematical Software* **23**(2), 266–294.

Koenig, A., Abay, G. & Axhausen, K. W. (2004), *Zeitkostenansaetze im Personenverkehr*, Schriftenreihe 1065, final report for SVI 2001/534, Bundesamt für Strassen, UVEK, Bern, Switzerland.

Koppelman, F. S. (1975), Travel Prediction with Models of Individualistic Choice Behaviour, PhD thesis, MIT, Cambridge, MA.

Koppelman, F. S. & Sethi, V. (2000), Closed form discrete choice models, *in* D. Hensher & K. Button, eds, 'Handbook of Transport Modeling', Pergamon Press, Oxford, pp. 211–228.

Koppelman, F. S. & Sethi, V. (2005), 'Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior', *Transportation Research Part B: Methodological* **39**(9), 825–853.

Koppelman, F. S. & Wen, C.-H. (1998), 'Alternative Nested Logit Models: structure, properties and estimation', *Transportation Research Part B: Methodological* **32**(5), 289–298.

Koppelman, F. S. & Wen, C.-H. (2000), 'The Paired Combinatorial Logit model: properties, estimation and application', *Transportation Research Part B: Methodological* **34**(2), 75–89.

Koppelman, F. S. & Wen, C.-H. (2001), 'Alternative Nested Logits: a response to comments by Andrew Daly on an earlier paper by Frank Koppelman and Chieh-Hua Wen', *Transportation Research Part B: Methodological* **35**(8), 725–729.

Kroes, E., Bradley, M. A. & Veldhuis, J. (1994), *Integrated analysis of air travel demand in Europe*, PTRC 22^{nd} Summer Annual Meeting, PTRC, London.

Krommer, A. R. & Ueberhuber, C. W. (1994), *Numerical Integration on Advanced Computer Systems*, Springer, Berlin.

Lafferiere, R. & Gaudry, M. (1993), *Testing the linear inverse power transformation logit mode choice model*, Working paper CRT 892, Universit de Montral, Montreal, Canada.

Lapparent, M. D. & de Palma, A. (2002), *Non-linearities in the valuations of time estimates*, Proceedings of the PTRC Annual Meeting, 2002, PTRC, London.

Lee, B. J., Fujiwara, A., Zhang, J. & Sugie, Y. (2003), *Analysis of Mode Choice Behaviours based on Latent Class Models*, paper presented at the 10^{th} International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Lee, L.-F. (1992), 'On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete choice models', *Econometric Theory* **8**, 518–552.

Lee, L.-F. (1995), 'Asymptotic bias in simulated maximum likelihood estimation of discrete choice models', *Econometric Theory* **11**, 437–483.

Lin, F.-B. (1977), 'Airport choice in low demand region', *Transportation Engineering Journal* **103**(6), 711–727.

Louviere, J., Hensher, D. & Swait, J. (2000), *Stated Choice Models: Analysis and Application*, Camrbridge University Press, Cambridge.

Luce, R. (1959), *Individual choice behavior: a theoretical analysis*, J.Wiley and Sons, New York.

Mackie, P. J., Jara-Diaz, S. R. & Fowkes, A. S. (2001), 'The value of travel time savings in evaluation', *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3), 91–106.

Mandel, B. (1998), Measuring competition in air transport, *in* W. Pfähler, H.-M. Niemeier & O. G. Mayer, eds, 'Airports and Air Traffic: Regulation, Privatisation and Competition', Peter Lang Press, Hamburg, Germany.

Mandel, B., Gaudry, M. & Rothengatter, W. (1997), 'A disaggregate Box-Cox Logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand forecasts', *The Annals of Regional Science* **31**, 99–120.

Manski, C. (1977), 'The structure of random utility models', *Theory and Decision* **8**, 229–254.

Marschak, J. (1960), Binary choice constraints on random utility indications, *in* K. Arrow, ed., 'Stanford Symposium on Mathematical Methods in the Social Science', Stanford University Press, Stanford, CA, pp. 312–329.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–142.

McFadden, D. (1975), *On Independence, Structure, and Simultaneity in Transportation Demand Analysis*, Working Paper No. 7511, Urban Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley, CA.

McFadden, D. (1978), Modelling the choice of residential location, *in* A. Karlquist, ed., 'Spatial Interaction Theory and Planning Models', North Holland, Amsterdam, chapter 25, pp. 75–96.

McFadden, D. (1981), Econometric models of probabilistic choice, *in* C. Manski & D. McFadden, eds, 'Structural Analysis of Discrete Data with Econometric Applications', MIT Press, Cambridge, MA, USA, chapter 5, pp. 198–272.

McFadden, D. (1989), 'A method of simulated moments for estimation of the multinomial Probit without numerical integration', *Econometrica* **57**(5), 995–1026.

McFadden, D. (2000), *Disaggregate Behavioral Travel Demand's RUM Side: A 30-Year Retrospective*, Proceedings of the 9^{th} IATBR Conference, Gold Coast, Australia.

McFadden, D., Talvitie, A., Cosslett, S., Hasan, I., Johnson, M., Reid, F. & Train, K. (1977), *Demand Model Estimation and Validation*, Urban Travel Demand Forecasting Project, The institute of transportation studies at the University of California, Berkeley and Irvine, CA.

McFadden, D. & Train, K. (2000), 'Mixed MNL Models for discrete response', *Journal of Applied Econometrics* **15**, 447–470.

McFadden, D., Train, K. & Tye, W. B. (1978), 'An application of diagnostic tests for the Independence from Irrelevant Alternatives property of the multinomial Logit model', *Transportation Research Record* **637**, 39–45.

McGrath, E. (1970), *Fundamentals of Operations Research*, West Coast University Press, San Francisco.

McKay, M. D., Conover, W. J. & Beckman, R. J. (1979), 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', *Technometrics* **21**(2), 239–245.

Metropolis, N. C., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1092.

Mokhtarian, P. L. & Salomon, I. (2001), 'How derived is the demand for travel? some conceptual and measurement considerations', *Transportation Research Part A: Policy and Practice* **35**(8), 695–719.

Monteiro, A. B. F. & Hansen, M. (1997), 'Improvements to airport ground access and behavior of multiple airport system: BART extension to San Francisco International Airport', *Transportation Research Record* **1562**, 38–47.

Morikawa, T. (1989), Incorporating Stated Preference Data in Travel Demand Analysis, PhD thesis, MIT, Cambridge, MA.

Morokoff, W. J. & Caflisch, R. E. (1994), 'Quasi random sequences and their discrepancies', *SIAM Journal on Scientific Computing* **15**(6), 1251–1279.

MTC (2000), *Aviation Demands Forecast: Executive Summary*, MTC Metropolitan Transportation Commission, Oakland, CA.

Munizaga, M. A. & Alvarez-Daziano, R. (2002), *Evaluation of Mixed Logit as a practical modelling alternative*, paper presented at the European Transport Conference, Cambridge, UK.

Munizaga, M. A., Heydecker, B. & Ortúzar, J. de D. (2000), 'Representation of heteroskedasticity in discrete choice models', *Transportation Research Part B: Methodological* **34**(3), 219–240.

Nason, S. D. (1980), 'Analyzing ticket-choice decisions of air travellers', *Transportation Research Record* **768**, 20–24.

NATS (1998), *Second Passenger Allocation Model (SPAM) : Pen Portrait Version 1*, National Air Traffic Services, London.

Ndoh, N. N., Pitfield, D. E. & Caves, R. (1990), Air transportation passenger route choice: a Nested Multinomial Logit analysis, *in* M. M. Fisher, P. Nijkamp & Y. Y. Papageorgiou, eds, 'Spatial Choices and Processes', Elsevier Science Publishers B.V., North Holland, pp. 349–365.

Niederreiter, H. & Xing, C. (1998), Nets, (t,s)-sequences, and algebraic geometry, *in* P. Hellekalek & G. Larcher, eds, 'Random and Quasi-Random Point Sets, Lecture Notes in Statistics, 138', Springer, New York, pp. 267–302.

OAG (1996), *World Airways Guide Fares Supplement 1996*, Reed Elsevier, Dunstable, United Kingdom.

OEF (1999), *The Contribution of the Aviation Industry to the UK Economy*, Oxford Economic Forecasting, Oxford.

ONS (1996), *International Passenger Survey*, The Office for National Statistics, London.

Oort, O. (1969), 'The evaluation of travelling time', *Journal of Transport Economics and Policy* **3**, 279–286.

Ortúzar, J. de D. (2000), *Stated Preference Modelling Techniques: PTRC Perspectives 4*, PTRC Education and Research Services Ltd, London.

Ortúzar, J. de D. (2001), 'On the development of the nested logit model', *Transportation Research Part B: Methodological* **35**(2), 213–216.

Owen, A. B. (1995), Randomly permuted (t,m,s)-nets and (t,s)-sequences, *in* H. Niederreiter & J.-S. Shiue, eds, 'Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing', Springer, New York, pp. 351–368.

Ozoka, A. I. & Ashford, N. (1989), 'Application of disaggregate modelling in aviation systems planning in Nigeria: a case study', *Transportation Research Record* **1214**, 10–20.

Papola, A. (2004), 'Some developments on the cross-nested Logit model', *Transportation Research Part B: Methodological* **38**(9), 833–854.

Pathomsiri, S., Mahmassani, H. & Haghani, A. (2004), *Airport Choice Model for Baltimore-Washington Region*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Pels, E., Nijkamp, P. & Rietveld, P. (2001), 'Airport and airline choice in a multi-airport region: an empirical analysis for the San Francisco bay area', *Regional Studies* **35**(1), 1–9.

Pels, E., Nijkamp, P. & Rietveld, P. (2003), 'Access to and competition between airports: a case study for the San Francisco Bay area', *Transportation Research Part A: Policy and Practice* **37**(1), 71–83.

Pels, E. & Rietveld, P. (2004), 'Airline pricing behaviour in the London-Paris market', *Journal of Air Transport Management* **10**(4), 277–281.

Proussaloglou, K. & Koppelman, F. S. (1995), 'Air carrier demand: an analysis of market share determinants', *Transportation* **22**(4), 371–388.

Proussaloglou, K. & Koppelman, F. S. (1999), 'The choice of air carrier, flight, and fare class', *Journal of Air Transport Management* **5**(4), 193–201.

Psaraki, V. & Abacoumkin, C. (2002), 'Access mode choice for relocated airports: the new Athens International Airport', *Journal of Air Transport Management* **8**(2), 89–98.

Rajaonarison, D., Bolduc, D. & Jayet, H. (2005), 'The k-deformed multinomial logit model', *Economics Letters* **86**(1), 13–20.

RAPC (2000), *Regional Airport System Plan Update 2000*, Regional Airport Planning Committee, Oakland, CA.

Redmond, L. S. & Mokhtarian, P. L. (2001), 'The positive utility of the commute: modeling ideal commute time and relative desired commute amount', *Transportation* **28**, 179–205.

Resource Systems Group Inc. (2003), *Air Travelers 2003: The New Realities?*, annual Air Survey project report.

Revelt, D. & Train, K. (1998), 'Mixed Logit with repeated choices: households' choices of appliance efficiency level', *Review of Economics and Statistics* **80**(4), 647–657.

Revelt, D. & Train, K. (1999), *Customer-specific taste parameters and Mixed Logit: Households' choice of electricity supplier*, Working Paper No. E00-274, Department of Economics, University of California, Berkeley, CA.

Richardson, A. J. (2003), 'Some evidence of travelers with zero value-of-time', *Transportation Research Record* **1854**, 107–113.

Salomon, I. & Mokhtarian, P. L. (1998), 'What happens when mobility-inclined market segments face accessibility-enhancing policies?', *Transportation Research Part D: Transport and Environment* **3**(3), 129–140.

Sándor, Z. & Train, K. (2004), 'Quasi-random simulation of discrete choice models', *Transportation Research Part B: Methodological* **38**(4), 313–327.

Scott Wilson Kirkpatrick (1999), *National airport accessibility model: model status report*, prepared for the Department for Transport by Scott Wilson Kirkpatrick & Co Ltd, Basingstoke, United Kingdom.

Scott Wilson Kirkpatrick (2004), *Rules and Modelling: a User's Guide to SPASM, Edition 2*, prepared for the Department for Transport by Scott Wilson Kirkpatrick & Co Ltd, Basingstoke, United Kingdom.

Shaw, J. E. H. (1988), 'A quasirandom approach to integration in Bayesian statistics', *The Annals of Statistics* **16**(2), 895–914.

Sillano, M. & Ortúzar, J. de D. (2004), 'Willingness-to-pay estimation from mixed logit models: some new evidence', *Environment & Planning A* **37**(3), 525–550.

Sipress, A. (1999), *Not all commuters driven crazy*, pages A1 and A12, Washington Post, October 18.

Skinner, R. E. J. (1976), 'Airport choice: An empirical study', *Transportation Engineering Journal* **102**(4), 871–883.

Small, K. (1987), 'A discrete choice model for ordered alternatives', *Econometrica* **55**(2), 409–424.

Sørensen, M. (2003), Discrete Choice Models: Estimation of Passenger Traffic, PhD thesis, Centre for Traffic and Transport, Technical University of Denmark, Lyngby, Denmark.

Spanier, J. & Li, L. (1997), Quasi-monte carlo methods for integral equations, *in* 'Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Proceedings of a Conference at the University of Salzburg', Springer Lecture Notes on Statistics 127, Springer Verlag, Salzburg, Austria.

Steckel, J. H. & Vanhonacker, W. R. (1988), 'A heterogeneous conditional logit model of choice', *Journal of Business and Economic Statistics* **6**(3), 391–398.

Suzuki, Y. (2003), 'Airline frequent flyer programs: equity and attractiveness', *Transportation Research Part E: Logistics and Transportation Review* **39**(4), 289–304.

Suzuki, Y., Crum, M. R. & Audino, M. J. (2003), 'Airport choice, leakage, and experience in single-airport regions', *Journal of Transportation Engineering* **129**(2), 212–218.

Swait, J. (2001*a*), 'Choice set generation within the Generalized Extreme Value family of discrete choice models', *Transportation Research Part B: Methodological* **35**(7), 643–666.

Swait, J. (2001*b*), 'A non-compensatory choice model incorporating attribute cutoffs', *Transportation Research Part B: Methodological* **35**(10), 903–928.

Swait, J. & Adamowicz, W. (1996), *The Effect of Choice Environment and Task Demands on Consumer Behavior: Discriminating between Contribution and Confusion*, Working Paper, Department of Rural Economy, University of Alberta, Edmonton, Canada.

Swait, J. & Adamowicz, W. (2001), 'Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice', *Organizational Behavior and Human Decision Processes* **86**(2), 141–167.

Swait, J. & Ben-Akiva, M. (1985), 'Incorporating random constraints in discrete models of choice set generation', *Transportation Research Part B: Methodological* **21**(2), 91–102.

Swait, J. & Ben-Akiva, M. (1987), 'Empirical test of a constrained choice discrete model: mode choice in Sao Paulo, Brazil', *Transportation Research Part B: Methodological* **21**(2), 103–115.

Swait, J. & Bernardino, A. (2000), 'Distinguishing taste variation from error structure in discrete choice data', *Transportation Research Part B: Methodological* **34**(1), 1–15.

Thompson, A. & Caves, R. (1993), 'The projected market share for a new small airport in the south of England', *Regional Studies* **27**(2), 137–147.

Thurstone, L. (1927), 'A law of comparative judgement', *Psychological Review* **34**, 273–286.

Timmermans, H., Borgers, A. & van der Waerden, P. (1991), 'Mother logit analysis of substitution effects in consumer shopping destination choice', *Journal of Business Research* **23**(4), 311–323.

Train, K. (1999), *Halton sequences for Mixed Logit*, Working Paper, Department of Economics, University of California, Berkeley, California.

Train, K. (2001), *A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit*, Working Paper, Department of Economics, University of California, Berkeley, CA.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.

Train, K. & Hudson, K. (2000), *The impact of information in vehicle choice and the demand for electric vehicles in California*, project report, National Economic Research Associates, San Francisco.

Train, K., McFadden, D. & Ben-Akiva, M. (1987), 'The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices', *Rand Journal of Economics* **18**(1), 109–123.

Train, K. & Sonnier, G. (2005), Mixed logit with bounded distributions of correlated partworths, *in* R. Scarpa & A. Alberini, eds, 'Applications of Simulation Methods in Environmental and Resource Economics', Springer Publisher, Dordrecht, The Netherlands, chapter 7, pp. 117–134.

Truong, P. & Hensher, D. (1985), 'Measurement of travel time savings and opportunity cost from a discrete choice model', *The Economic Journal* **95**, 438–451.

Tuffin, B. (1996), 'On the use of low discrepancy sequences in Monte Carlo methods', *Monte Carlo Methods and Applications* **2**, 295–320.

Tuffin, B. (1998), A new permutation choice in Halton sequences, *in* H. Niederreiter, P. Hellekalek, G. Larcher & P. Zinterhof, eds, 'Monte-Carlo and quasi-Monte-Carlo methods 1996, Lecture notes in statistics 127', Springer, New York, pp. 427–435.

Tversky, A. (1972), 'Elimination by aspects: a theory of choice', *Psychological Review* **79**, 281–299.

Veldhuis, J., Essers, I., Bakker, D., Cohn, N. & Kroes, E. (1999), 'The Integrated Airport Competition Model, 1998', *Journal of Air Transportation World Wide* **4**(2).

Viton, P. A. (2004), 'Will mixed logit change urban transport policies?', *Journal of Transport Economics and Policy* **38**(3), 403–423.

Vovsha, P. (1997), 'Application of a Cross-Nested Logit model to mode choice in Tel Aviv, Israel, Metropolitan Area', *Transportation Research Record* **1607**, 6–15.

Vovsha, P. (1999), *Comparative Analysis of Different Spatial Interaction Models*, paper presented at the 78^{th} Annual Meeting of the Transportation Research Board, Washington, DC.

Vovsha, P. & Bekhor, S. (1998), 'The Link-Nested Logit model of route choice: overcoming the route overlapping problem', *Transportation Research Record* **1645**, 133–142.

Walker, J. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, PhD thesis, MIT, Cambridge, MA.

Walker, J. (2002), 'Mixed logit (or logit kernel) model: Dispelling misconceptions of identification', *Transportation Research Record* **1805**, 86–98.

Walker, J., Ben-Akiva, M. & Bolduc, D. (2003), *Identification of the Logit Kernel (or Mixed Logit) Model*, Working Paper, Department of Civil Engineering, MIT, Cambridge, MA.

Wang, X. & Hickernell, F. J. (2002), 'Randomized Halton sequences', *Mathematical and Computer Modelling* **32**, 887–899.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S. & Poulsen, C. S. (1999), 'Discrete and continuous representations of unobserved heterogeneity in choice modeling', *Marketing Letters* **10**(3), 219–232.

Weeks, M. (1997), 'The multinomial probit model revisited: A discussion of parameter estimability, identification and specification testing', *Journal of Economic Surveys* **11**(3), 296–320.

Wei, W. & Hansen, M. (2005), 'Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets', *Transportation Research Part E: Logistics and Transportation Review* **41**(4), 315–327.

Wen, C.-H. & Koppelman, F. S. (2001), 'The Generalized Nested Logit Model', *Transportation Research Part B: Methodological* **35**(7), 627–641.

Williams, H. C. W. L. (1977), 'On the Formulation of Travel Demand Models and Economic Evaluation Measures of User Benefit', *Environment & Planning A* **9**(3), 285–344.

Williams, H. C. W. L. & Ortúzar, J. de D. (1982), 'Behavioural theories of dispersion and the mis-specification of travel demand models', *Transportation Research Part B: Methodological* **16**(3), 167–219.

Windle, R. & Dresner, M. (1995), 'Airport choice in multi-airport regions', *Journal of Transportation Engineering* **121**(4), 332–337.

Yoo, K. E. & Ashford, N. (1997), 'Carrier choices of air passengers in pacific rim: Using comparative analysis and complementary interpretation of revealed preference and stated preference data', *Transportation Research Record* **1562**, 1–7.

Young, W. & Morris, J. (1981), 'Evaluation by individuals of their travel time to work', *Transportation Research Record* **794**, 51–59.

Appendix A

Confounding between substitution patterns and random taste heterogeneity: ECL results

A.1 Introduction

This appendix presents a summary of the estimation results obtained with ECL models as part of the case-studies described in Chapter 6. Here, we look specifically at the models estimated on the three datasets used in the forecasting applications presented in Section 6.3, namely the data generated by the two-level NL model, the data generated by the RCL model with a randomly distributed travel cost coefficient, and the data generated by the NL mixture model.

A.2 Estimation results

The estimation results of the three ECL models are summarised in Table A.1. In each case, the final ECL model acts as an approximation to a NL mixture structure, with a single nest shared by two alternatives. To ensure that, prior to adding in random taste heterogeneity, the ECL model structure is homoscedastic (as in the case of a GEV model), an additional error-component with equal variance is added to the utility of the alternative nested on its own (cf. Garrow 2004). As such, in the case where car and rail are nested together, the three utility functions are specified as:

$$\begin{aligned} U(car) &= V(car) + \sigma\xi_1 + \varepsilon_1 \\ U(rail) &= V(rail) + \sigma\xi_1 + \varepsilon_2 \\ U(SM) &= V(SM) + \sigma\xi_2 + \varepsilon_3, \end{aligned} \tag{A.1}$$

where $V(car)$, $V(rail)$ and $V(SM)$ give the remaining part of utility for car, rail and SM respectively, potentially including randomly distributed taste coefficients, and ε_1, ε_2 and ε_3 are *iid* type I extreme value variates. The correlation between the car and rail alternatives is accommodated through sharing the same error-component ξ_1, which follows a standard Normal distribution. The utility of the SM alternative contains an independent error-component $\xi_2 \sim N(0, 1)$, where the multiplication

True model	NL		RCL		NL mixture	
Final LL	-900.83		-1321.77		-1142.96	
adj. $\rho^2(0)$	0.7227		0.5962		0.6493	
	est.	t-stat.	est.	t-stat.	est.	t-stat.
δ_{car}	-8.465	-6.04	-4.319	-11.80	-7.679	-9.31
δ_{SM}	-6.682	-6.35	-3.642	-9.42	-5.577	-10.80
β_{TC} (μ)	-0.222	-6.47	-0.107	-14.79	-0.197	-9.24
β_{TC} (σ)	0.024	1.54	0.044	14.62	0.072	9.29
β_{HW} (μ)	-0.046	-6.60	-0.021	-9.12	-0.039	-10.48
β_{HW} (σ)	0.000	0.35	-	-	0.003	0.64
$\beta_{TT,car}$ (μ)	-0.072	-6.40	-0.036	-11.39	-0.060	-9.23
$\beta_{TT,car}$ (σ)	0.007	1.56	-	-	0.006	1.11
$\beta_{TT,rail}$ (μ)	-0.090	-6.29	-0.047	-13.39	-0.079	-10.72
$\beta_{TT,rail}$ (σ)	0.003	0.30	-	-	0.000	0.10
$\beta_{TT,SM}$ (μ)	-0.081	-6.48	-0.040	-10.71	-0.073	-9.53
$\beta_{TT,SM}$ (σ)	0.004	0.78	-	-	0.005	0.87
σ (rail-SM)	2.176	5.01	-	-	2.326	6.58
σ (rail-car)	-	-	0.804	3.54	-	-
	CHF/hour		CHF/hour		CHF/hour	
VTTS (car) (μ)	19.83		24.68		20.85	
VTTS (car) (σ)	3.06		14.45		9.62	
VTTS (rail) (μ)	24.69		31.88		27.7	
VTTS (rail) (σ)	2.9		18.66		12.41	
VTTS (SM) (μ)	22.1		27.53		25.38	
VTTS (SM) (σ)	2.7		16.12		11.55	

Table A.1: Estimation results for ECL models

by a common σ ensures equal variance across alternatives, while still maintaining the correlation between the unobserved utility terms for car and rail. Here, the correlation between the unobserved utility components for car and rail is given by $\frac{\sigma^2}{\frac{\pi^2}{6}+\sigma^2}$, with $\frac{\pi^2}{6}$ being the variance of the standard type I extreme value term. The use of the above *correction-approach* ensures that no *uncontrolled* heteroscedasticity acts on the utilities. *Controlled* heteroscedasticity could in this case be allowed by introducing additional error-components, with a new standard deviation ($\neq \sigma$), although issues of identification would need to be dealt with in that case (cf. Walker 2001).

We will now look at the estimation results for the three models in turn. The ECL model estimated on the data generated by the two-level NL model (cf. Section 6.2.1) obtains very similar model fit to the corresponding NL and NL mixture models presented in Table 6.1 (-900.2 and -899.15 respectively, compared to -900.83). The estimated VTTS measures are also very similar to those reported for the NL and NL mixture models, albeit with marginally higher standard deviations (when compared to the NL mixture model). Finally, the ECL model shows a correlation of 0.74 between the unobserved utilities for rail and SM, which is essentially identical to that of the *true* model from Section 6.2.1, where a value of 0.75 was used.

The ECL model estimated on the data generated by the RCL model with a

Representative individual

Original choice probabilities

	Car	Rail	SM
True model	46.74%	21.63%	31.63%
ECL	49.98%	19.49%	30.53%

Forecasted choice probabilities

	Car	Rail	SM
True model	50.73%	8.99%	40.28%
ECL	53.55%	7.47%	38.98%

Relative change in choice probabilities

	Car	Rail	SM
True model	8.54%	-58.45%	27.35%
ECL	7.14%	-61.67%	27.69%

Bias in predicted change

	Car	Rail	SM
ECL	-16.38%	+5.51%	+1.23%

Population-level

Original market shares

	Car	Rail	SM
True model	55.69%	23.69%	20.62%
ECL	55.82%	23.65%	20.99%

Forecasted market shares

	Car	Rail	SM
True model	58.46%	13.22%	28.32%
ECL	58.58%	13.14%	28.76%

Relative change in market shares

	Car	Rail	SM
True model	4.97%	-44.19%	37.36%
ECL	4.95%	-44.45%	37.02%

Bias in predicted change

	Car	Rail	SM
ECL	-0.27%	+0.60%	-0.91%

Table A.2: Forecasting exercise using data generated by two-level NL model: ECL results

randomly distributed travel cost coefficient (cf. Section 6.2.4) obtains very similar model fit to the corresponding RCL and NL mixture models presented in Table 6.4 (-1322.96 and -1320.93 respectively, compared to -1321.77). Again, the mean VTTS measures are very similar, although the ECL model again leads to higher estimates for the standard deviations. Finally, the ECL model shows a correlation of 0.28 between the unobserved utility components for rail and car, compared to 0.33 in the corresponding NL mixture model. This suggests that, with this data, the ECL model, like the NL mixture model, is also vulnerable to confounding, given that, in the true model, no explicit correlation was assumed between the unobserved utilities for the car and rail alternatives. Again however, the risk of confounding is significantly lower than in the closed form NL model.

With the final dataset, generated by the NL mixture model (cf. Section 6.2.6), the ECL model obtains slightly better model fit than the corresponding NL mixture model presented in Table 6.6 (-1142.96 compared to -1147.78). The mean VTTS measures are again very similar, but this time, the ECL model leads to a less severe overestimation of the standard deviations than the corresponding NL mixture models. Finally, in terms of correlation, the ECL model shows a level of 0.77 between the unobserved utility terms for rail and SM, which is identical to that obtained with the NL mixture model.

A.3 Forecasting analysis

To complete the analysis, we use the three ECL models presented in Table A.1 in repeating the forecasting analysis described in Section 6.3. The results of these

Representative individual

Original choice probabilities

	Car	Rail	SM
True model	38.64%	27.65%	33.71%
ECL	40.15%	26.76%	33.09%

Forecasted choice probabilities

	Car	Rail	SM
True model	43.79%	18.32%	37.89%
ECL	46.02%	17.67%	36.32%

Relative change in choice probabilities

	Car	Rail	SM
True model	13.31%	-33.72%	12.39%
ECL	14.62%	-33.98%	9.74%

Bias in predicted change

	Car	Rail	SM
ECL	+9.83%	+0.76%	-21.45%

Population-level

Original market shares

	Car	Rail	SM
True model	53.94%	23.78%	22.28%
ECL	53.51%	24.26%	22.23%

Forecasted market shares

	Car	Rail	SM
True model	56.68%	15.02%	28.30%
ECL	56.37%	15.89%	27.73%

Relative change in market shares

	Car	Rail	SM
True model	5.08%	-36.82%	26.99%
ECL	5.34%	-34.49%	24.78%

Bias in predicted change

	Car	Rail	SM
ECL	+5.23%	-6.33%	-8.20%

Table A.3: Forecasting exercise using data generated by RCL model with randomly distributed travel cost coefficient: ECL results

experiments are summarised in Table A.2 for the data generated by the two-level NL model, Table A.3 for the data generated by the RCL model, and Table A.4 for the data generated by the NL mixture model. In each case, the results are shown for the representative individual and at the population-level, and the results for the true model are repeated for reference purposes.

Aside from some slight differences, mostly in the case of the representative individual, the results are very similar to those obtained with the corresponding NL mixture models in Section 6.3, with comparable levels of bias. This suggests that ECL models can be used in a similar fashion to non-MMNL GEV mixture models in the joint analysis of inter-alternative correlation and random taste heterogeneity, with similar performance in terms of avoiding confounding between the two phenomena. The choice of model structure needs to be made on a case-by-case basis, as a function of the differences in estimation cost[1], as well as the relative severity of the structure-specific issues, such as the identification conditions in the ECL model (cf. Walker 2001).

[1]This depends on the complexity of the nesting structure, as discussed in Section 2.9.2

Representative individual

Original choice probabilities

	Car	Rail	SM
True model	46.79%	21.37%	31.84%
ECL	46.54%	19.31%	34.15%

Forecasted choice probabilities

	Car	Rail	SM
True model	50.65%	9.49%	39.86%
ECL	49.51%	8.70%	41.79%

Relative change in choice probabilities

	Car	Rail	SM
True model	8.25%	-55.58%	25.18%
ECL	6.38%	-54.95%	22.38%

Bias in predicted change

	Car	Rail	SM
ECL	-22.65%	-1.12%	-11.09%

Population-level

Original market shares

	Car	Rail	SM
True model	55.12%	23.65%	21.23%
ECL	55.81%	22.68%	21.51%

Forecasted market shares

	Car	Rail	SM
True model	57.82%	13.45%	28.73%
ECL	58.42%	12.75%	28.83%

Relative change in market shares

	Car	Rail	SM
True model	4.91%	-43.13%	35.31%
ECL	4.68%	-43.79%	34.05%

Bias in predicted change

	Car	Rail	SM
ECL	-4.67%	+1.54%	-3.56%

Table A.4: Forecasting exercise using data generated by two-level NL mixture model: ECL results

Appendix B

Covariance heterogeneity: Development of ECL approach

We now describe how the ECL formulation of the MMNL model can be adapted to allow for covariance heterogeneity. We first review the basic theory behind the ECL model (Section B.1) and show how it can be used to approximate the COVNL model (Section B.2). We then proceed to the case where the covariance heterogeneity is purely random (Section B.3), and to the case where part of the variation is deterministic with a remaining random part (Section B.4).

B.1 General ECL formulation

As described in Section 2.9.1, in the ECL model, correlation across alternatives is introduced through the use of error-components that are shared between alternatives that are closer substitutes for each other. The error-components take on the form of normally distributed random variables with a mean of zero, and a standard deviation of σ, where the estimate for σ is related to the correlation between the alternatives.

Ignoring for the moment the issues of identification discussed by Walker (2001), and the question of homoscedasticity[1], the utilities of two alternatives that have some correlation in the unobserved part of utility would be written as:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n} + \zeta_1 \tag{B.1}$$

and

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \zeta_1, \tag{B.2}$$

where $V_{i,n}$ and $V_{j,n}$ give the observed part of utility for alternatives i and j and respondent n, and $\varepsilon_{i,n}$ and $\varepsilon_{j,n}$ are *iid* type I extreme-value terms. The additional error-term ζ_1 is distributed $N(0, \sigma_1)$. With this, the covariance between the two alternatives is given by σ_1^2, while the variance for the individual utilities is given by

[1]Basic ECL approximations to GEV models are heteroscedastic, while GEV models are homoscedastic, an issue that can be addressed by *cancelling* out the heteroscedasticity in ECL models through the use of additional error-components, as shown in Appendix A.

$\sigma_1^2 + \frac{\pi^2}{6}$, leading to a correlation of:

$$Corr\left(U_{i,n}, U_{j,n}\right) = \frac{\sigma_1^2}{\sigma_1^2 + \frac{\pi^2}{6}}. \tag{B.3}$$

It is easy to see that it is possible to rewrite the utility of alternative j as:

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sigma_1 \xi_1, \tag{B.4}$$

where $\xi_1 \sim N\left(0,1\right)$, and where the subscript on ξ remains in use to guarantee that individual draws are taken for each error-component (with the same draws taken for the same error-component across alternatives).

For the choice probabilities, integration over the $N\left(0,1\right)$ draws for the error-components is required. Let Ψ_j define the set of error-components included in the utility function of alternative j, such that:

$$U_{j,n} = V_{j,n} + \varepsilon_{j,n} + \sum_{k \in \Psi_j} \sigma_k \xi_k \tag{B.5}$$

This notation allows for any structure for the error-components, including homoscedastic as well as heteroscedastic ones. The choice probability for alternative i and individual n is now given by:

$$P_n\left(i \mid \boldsymbol{\sigma}\right) = \int_{\xi_1} \ldots \int_{\xi_K} \left[\frac{exp\left(V_{i,n} + \sum_{k \in \Psi_i} \sigma_k \, \xi_k\right)}{\sum_{j \in C_n} exp\left(V_{j,n} + \sum_{l \in \Psi_j} \sigma_l \, \xi_l\right)} \prod_{k=1}^{K} \phi\left(\xi_k\right) \right] \mathrm{d}\xi_K \ldots \mathrm{d}\xi_1, \tag{B.6}$$

where K gives the total number of error-components used, and $\phi\left(\right)$ is the standard Normal density function.

B.2 Deterministic covariance heterogeneity in ECL models

The ECL formulation can be extended straightforwardly to allow for deterministic covariance heterogeneity by parameterising σ_k, for example by setting $\sigma_k = f\left(\boldsymbol{\theta}, \boldsymbol{z_n}\right)$, where $\boldsymbol{\theta}$ is a vector of parameters, and where $\boldsymbol{z_n}$ is defined as before. The only condition applying to $f\left(\right)$ is that it yields positive values for the standard deviations[2]; equation (B.3) guarantees that the resulting correlation falls between 0 and 1.

[2]This merits some clarification. Estimation code can deal with negative values for standard deviation parameters in the case where they are only used in the form of variances as opposed to standard deviations; in fact, in unconstrained estimation, it can often be observed that estimation packages produce *negative* estimates for the standard deviations. The problems arise in the case where $f\left(\right)$ allows for positive as well as negative values for σ, leading to an underestimated mean level of correlation.

B.3 Purely random covariance heterogeneity in ECL models

In the standard ECL formulation of the MMNL model, the choice probabilities are obtained by integration over the distribution of the error-components, with additional integration over the distribution of random taste-coefficients in the case of added random taste heterogeneity. Focussing for now on the case of error-components for correlation only (as opposed to additional taste heterogeneity), random covariance heterogeneity can be introduced by additional integration over the distribution of the variances of the error-components.

The choice probability is in this case given by:

$$P_n(i) = \int_{\sigma_1} \dots \int_{\sigma_K} \left[P_n(i \mid \boldsymbol{\sigma}) \prod_{k=1}^{K} g(\sigma_k \mid \boldsymbol{\theta_k}) \right] \mathrm{d}\sigma_K \dots \mathrm{d}\sigma_1, \tag{B.7}$$

where $P_n(i \mid \boldsymbol{\sigma})$ is the choice probability for alternative i, conditional on the vector of standard deviations $\boldsymbol{\sigma}$, as in equation (B.6), and where $g(\sigma_1 \mid \boldsymbol{\theta_1})$ is the density function for σ_1, with parameters given by the vector $\boldsymbol{\theta_1}$. Here, an appropriate choice of distribution for the standard deviations is of crucial importance, given that they need to take on positive values[3]. An alternative to the use of bounded distributions comes in the use of a transform mapping monotonically from the real domain to the space of positive numbers. The adaptation of equation (B.7) to this case is straightforward.

B.4 Deterministic and random covariance heterogeneity in ECL models

The extension of the approach described in Section B.3 to the case allowing jointly for deterministic and random covariance heterogeneity is straightforward. We reuse the formulation from Section B.2, where $\sigma = f(\boldsymbol{\theta}, \boldsymbol{z_n})$. This time however, we allow some of the elements of $\boldsymbol{\theta}$ to be randomly distributed across individuals. The choice probability for alternative i and decision-maker n is now rewritten as:

$$P_n(i) = \int_{\boldsymbol{\theta_1}} \dots \int_{\boldsymbol{\theta_K}} \left[P_n(i \mid \sigma_k = f(\boldsymbol{\theta_k}, \boldsymbol{z_n}) \forall k) \prod_{k=1}^{K} g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k}) \right] \mathrm{d}\boldsymbol{\theta_K} \dots \mathrm{d}\boldsymbol{\theta_1}, \tag{B.8}$$

where $\boldsymbol{\theta_k}$ is distributed according to $g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k})$, and where the notation allows for correlation between individual elements in $\boldsymbol{\theta_k}$. It can easily be seen that this approach reduces to the purely random formulation in Section B.3 if those parameters associated with z_n are zero[4], and the purely deterministic formulation in Section B.2, in the case where $g(\boldsymbol{\theta_k} \mid \boldsymbol{\Omega_k})$ produces only a single (fixed) value for the vector $\boldsymbol{\theta_k}$.

[3]Again, this requirement is used solely to avoid an underestimation of the mean level of correlation in the case where the distribution yields positive as well as negative estimates for σ.

[4]I.e., only a constant is estimated, which is distributed randomly across respondents.

B.5 Discussion

The discussion presented here has shown how the ECL framework can be adapted to allow for deterministic as well as random covariance heterogeneity. In practice, it should be said that, due to the additional dimensions of integration, the mixed covariance ECL approach is generally more expensive in estimation and application than its GEV based counterparts described in Chapter 7, albeit that it has the advantage of a simpler form for the integrand (MNL *vs* more general GEV). An additional issue however arises with regards to identification, where appropriate conditions for identifiability need to be worked out on a case-by-case basis.

Appendix C

Frequently Used Acronyms

CAA: Civil Aviation Authority (UK)

CNL: Cross-Nested Logit

COVNL: Covariance Nested Logit

DfT: Department for Transport (UK)

ECL: Error-Components Logit

GEV: Generalised Extreme Value

GNL: Generalised Nested Logit

IIA: Independence from Irrelevant Alternatives

***iid*:** Independently and Identically Distributed

IVT: In-vehicle time

LCY: London City airport

LGW: Gatwick airport (London)

LHR: Heathrow airport (London)

LL: Log-likelihood

LR: Likelihood-ratio

LTN: Luton airport (London)

MLHS: Modified Latin Hypercube Sampling

MMNL: Mixed Multinomial Logit

mppa: million passengers per annum

MNL: Multinomial Logit

MTC: Metropolitan Transport Commission (California)

NL: Nested Logit

NNNL: Non-normalised Nested Logit

OAK: Metropolitan Oakland International airport (SF-bay)

***pdf*:** Probability Density Function

PMC: Pseudo Monte Carlo

QMC: Quasi Monte Carlo

RCL: Random Coefficients Logit

RP: Revealed Preference

RUM: Random Utility Model

SF-bay: San Francisco Bay

SFO: San Francisco International airport

SJC: Mineta San José International airport (SF-bay)

SM: Swiss-Metro

SP: Stated Preference

STN: Stansted airport (London)

UMNL: Utility Maximising Nested Logit

VFR: Visiting friends or relatives

VTTS: Value of travel-time savings

Appendix D

Publications

- **Hess, S.** & Polak, J.W. (2003), An alternative method to the scrambled Halton sequence for removing correlation between standard Halton sequences in high dimensions, paper presented in the R-sessions at the European Regional Science Conference, Jyväskylä, Finland.

- **Hess, S.**, Polak, J.W., Daly, A. (2003), On the performance of shuffled Halton sequences in the estimation of discrete choice models, paper presented at the European Transport Conference, Strasbourg, France.

- **Hess, S.**, Polak, J.W. (2004), Mixed Logit estimation of parking type choice, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

- **Hess, S.** & Polak, J.W. (2004), On the use of Discrete Choice Models for Airport Competition with Applications to the San Francisco Bay area Airports, paper presented at the 10^{th} triennial *WCTR* conference, Istanbul, Turkey.

- **Hess, S.**, Polak, J.W. Daly, A. & Hyman, G. (2004), Flexible Substitution Patterns in Models of Mode and Time of Day Choice: New evidence from the UK and the Netherlands, paper presented at the European Transport Conference, Strasbourg.

- Bastin, F., Cirillo, C. & **Hess, S.** (2005), Evaluation of optimisation methods for estimating Mixed Logit models, Transportation Research Record, forthcoming.

- Daly, A., **Hess, S.**, Hyman, G., Polak, J. W. & Rohr, C. (2005), Modelling Time Period Choice: Experience from the UK and the Netherlands, paper presented at the European Transport Conference, Strasbourg, France.

- **Hess, S.** (2005), A model for the joint analysis of airport, airline, and access-mode choice for passengers departing from the San Francisco Bay area, paper presented at the European Transport Conference, Strasbourg, France.

- **Hess, S.** & Axhausen, K.W. (2005), Distributional assumptions in the representation of random taste heterogeneity, paper presented at the 5^{th} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

- **Hess, S.**, Bierlaire, M. & Polak, J.W. (2005), Capturing taste heterogeneity and correlation structure with Mixed GEV models, in Scarpa, R. and Alberini, A. (eds.), Applications of Simulation Methods in Environmental and Resource Economics, Springer Publisher, Dordrecht, The Netherlands, chapter 4, pp. 55-76.

- **Hess, S.**, Bierlaire, M. & Polak, J.W. (2005), Discrete mixtures of GEV models, paper presented at the 5^{th} Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

- **Hess, S.**, Bierlaire, M. & Polak, J.W. (2005), Estimation of value of travel-time savings using Mixed Logit models, Transportation Research A, 39(2-3), pp. 221-236.

- **Hess, S.** & Polak, J.W. (2005), Accounting for random taste heterogeneity in airport-choice modelling, Transportation Research Record, forthcoming.

- **Hess, S.** & Polak, J.W. (2005), Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area, Transportation Research Part E: Logistics and Transportation Review, forthcoming.

- **Hess, S.** & Polak J.W. (2005), Mixed Logit modelling of airport choice in multi-airport regions, Journal of Air Transport Management, 11(2), pp.59-68.

- **Hess, S.**, Polak, J.W. and Bierlaire, (2005), Functional approximations to alternative-specific constants in time-period choice-modelling, Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Proceedings of the 16th International Symposium on Transportation and Traffic Theory, H.S.Mahmassani (ed.), Elsevier, Amsterdam, chapter 28, pp. 545-564.

- **Hess, S.**, Train, K. & Polak, J.W. (2005), On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice, Transportation Research B, forthcoming.