Theory and practice in modelling air travel choice behaviour

Stephane Hess

Institute for Transport Studies
University of Leeds

February 17, 2009

Abstract

There is growing interested in modelling the choices made by air travellers. However, these choices are complex and many studies don't do that complexity justice. In this chapter, we take another look at the different choices made by an air passenger for a single journey and discuss ways of modelling these processes. We make a number of recommendations for good practice and also highlight a number of issues that need to be dealt with by the analysts. Finally, we present an application making use of state-of-the-art modelling techniques in an air travel behaviour context.

1 Introduction

An increasing number of studies are looking at modelling air travel choice behaviour, mainly with the help of discrete choice models. Existing applications range from the choice of air as a mode of travel [e.g. 1], to the choice of airport in multi airport regions [e.g. 2, 3, 4, 5, 6, 7], the choice of airline or fare classes [e.g. 8, 9], and the choice of access mode [e.g. 10, 11]. Some applications look jointly at multiple travel dimensions, e.g. the joint choice of airport and airline [e.g. 12] or even the choice of an airport, airline and access mode triplet [e.g. 13]. Finally, there is also an increasing reliance on advanced model structures available, used for example in the representation of random taste heterogeneity across travellers [e.g. 6, 7, 14], or the multi-dimensional correlation between alternatives sharing sub-choices along some of the travel dimensions [15].

Despite the recent progress in the area, a lot of work remains to be done. Indeed, the choice processes undertaken by air travellers are arguably more complex than those taken with other modes. While authors are gradually acknowledging this in their work, and while the use of advanced model structures has allowed for more realism, many studies still make strong assumptions that unduly simplify the choice processes. Crucially, there is also still a general lack of understanding of the actual choice processes undertaken by air travellers, a fact that is not helped by the dynamic nature of the problem at hand, as witnessed for example with the advance of low cost carriers.

This chapter aims to take an overview look at the modelling issues and makes some suggestions for good practice. We first give an overview of the choice processes undertaken by air travellers (Section 2). This is followed in Section 3 by a discussion setting targets for practical research in this area while also acknowledging certain simplifications that are necessary in such analyses. Section 4 is concerned with data issues, while Section 5 looks at the question of model structure. Finally, Section 6 presents an empirical example, and Section 7 provides a brief summary of the chapter.

2   Air travel choice behaviour

The choices made by an air passenger for a single journey are complex and involve decisions along a number of different dimensions, some of which are strongly interrelated. In simple terms, the choices made by an air traveller can be divided into three main subcategories, namely those at the origin side, those at the destination side, and those concerning the actual air journey. It is the latter ones and to some extend the origin side decisions that have received the most attention in the existing literature.

At the origin side of an air journey, a passenger chooses a departure airport and also makes a number of choices relating to the ground level journey to this departure airport. In many ways, the destination side choices are the mirror image of those made at the origin side.

This leaves us with choices relating to the actual air journey. Here, the first choice is that of an airline operating a route to the chosen destination. While most passengers will travel on a single airline for the duration of their journey, there are some routings on which passengers will have to rely on a combination of airlines, a situation that has in recent years increased in complexity given that a large number of routes are now operated under *code share* agreements. The next level of choice is that of a routing, which

looks first at the choice between direct and connecting flights, before, when choosing a connecting flight, looking at the choice between different routes in terms of 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.

The above discussion has already shown that air journeys involve decisions along a multitude of dimensions. What makes the analysis of these choices even more complicated are the complex interdependencies between the various dimensions of choice, both in terms of interactions as well as ordering of priorities. To illustrate the latter point, it should be obvious that, for most passengers, the choice of destination will influence the choice set in terms of airlines or departure airports. However, there may also be passengers where the decision to travel on a specific airline (e.g. low-cost airline) will determine the choice set of possible destinations. From this perspective, a simultaneous modelling approach is clearly preferable to a potentially misguided sequential approach.

Finally, it should also be acknowledged that passengers actually also make a decision to use air in the first place, as opposed to travelling on a different mode. For most destinations, this choice is a direct result of distance, making all other modes either impractical or impossible. However, for a number of short-haul destinations, there is increasing competition with high speed rail, and the analysis of the choices in this context are an interesting area for further work. However, mainly due to data requirements, most applications looking at detailed air travel choice dimensions have to rely on the assumption that the choice of air as a mode of travel has been made at a higher level, prior to making choices relating to the actual air journey.

3   Guidance for good modelling practice and contrast with status quo

The above discussion has highlighted the complexity of the choice processes undertaken by air travellers. This section sets out some guidelines for good practice and contrasts them with the status quo in modelling work.

3.1 Guidelines

3.1.1 Recognise the multi dimensional nature of the choice process

Air travellers make choices among a multitude of travel dimensions and not recognising this in practical work can potentially lead to biased results. As an example, let us look at the case of a study that looks solely at the choice of airport but not the choice of airline. If a traveller in this study has a strong allegiance to an airline flying from a smaller airport with low overall frequency, then a simple airport choice model will be confused by the fact that this passenger actually flies from an airport with lower frequency to the desired destination. From this point of view, it is important to work on the basis of frequencies specific to airport-airline pairings, and not airport-specific frequency, as has commonly been the case in existing work. The same reasoning applies to other combinations of choice dimensions. In practice, this issue can be addressed by dividing the multi-dimensional alternative into a combination of elementary alternatives, with a passenger simultaneously choosing one alternative along each dimension of choice.

3.1.2 Account for correlation along different choice dimensions

By understanding the multidimensional 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 (e.g. same airline, same access mode). This clearly creates correlation between alternatives, and in the likely case where not all characteristics are observed by the analyst, this will result in correlated error terms. Here, it is important to account for this correlation, as described in Section 5.2.

3.1.3 Use highly disaggregate level-of-service data

Many existing studies have produced poor results partly as a result of using an insufficient level of disaggregation in the level-of-service data, often as an effect of using simplifications along a number of choice dimensions. While some aggregation is almost inevitable, the use of an excessive amount of aggregation can lead to biased results. This can for example arise in the case of studies making use of weekly (or even monthly) data instead of daily data, hence ignoring the often significant variations in level-of-service attributes between different days of the week, especially in terms of flight frequencies.

3.1.4 Account for differences in behaviour across travellers

As in most contexts, there are significant differences between individual air travellers in their sensitivities to attributes affecting their journey, e.g. air fares and departure times. As a result, it is crucial to take such taste heterogeneity into account at the modelling stage, an issue that is addressed in more detail in Section 5.3.

3.2 Limitations of practical research

Without exemption, existing studies of air travel choice behaviour use a number of simplifications of the choice processes, often due to data issues. These simplifications can be looked at in turn for each of the various dimensions of choice:

- The **choice of destination** and the actual **decision to travel** are not generally modelled, where a reservation applies for the latter in the case of SC surveys giving respondents the option *not to travel*. It is thus normally assumed that these decisions are taken at an upper level, prior to the air-journey specific choices.

- As already mentioned above, the same reasoning applies to the **decision to travel by air**. Here, it is thus important to acknowledge that the estimates obtained from such models relate to the part of the population that has decided to travel by air, and are not representative of the overall population.

- The majority of studies of air travel choice behaviour look solely at the choice of **departure airport** and ignore the choice of **arrival airport**, a simplification primarily resulting from data issue.

- Just as for the choice of airport, the analysis of **ground level decisions** is generally limited to the origin end, again primarily due to data reasons. Additionally, most studies are only able to look at the choice of main mode, ignoring the possibility of trip-chaining, as well as the choice of different routes. The effects of these restrictions are dependent on the geographical context.

- In RP models, **trip timing** cannot generally be modelled due to a relative lack of information on preferred departure time and flight availabilities. These issues do not apply in the case of SC models.

5

- The issue of **_flight routing_** is generally left untreated in RP studies, while in SC studies, this is often limited to a choice between direct and connecting flights.

- Given the major role that airline allegiance plays for some travellers, advanced studies increasingly model the **_choice of airline_** in addition to the choice of airport. However, combinations of airlines are generally not allowed in such models, and additional issues can arise in the case of code share flights.

4   Data issues

On one of the main problems that need to be faced in the modelling of air travel choice behaviour is that of data quality. This is the topic of the present section, where the discussion is divided into two parts, looking first at RP data before turning our attention to SC data.

4.1   Revealed Preference data

4.1.1   Availability and attributes of unchosen alternatives

With RP data, the main issue that needs to be faced is the relative lack of information on the choice set that the traveller was faced with. While detailed information is generally available on the chosen flight, this does not extend to unchosen alternatives. In fact, in many cases, it is not even just the attributes of these alternatives that are unknown, but also their availability to a given traveller at a given time. Generally, an assumption of availability is required, alongside significant aggregation of level of service attributes for the unchosen flight options. Similar problems also arise along other dimensions, such as the access journey related choices, where again information on the availability and attributes of unchosen modes is often not available.

4.1.2   Fare data

Air fares should naturally be expected to play a major role in air travel choice behaviour. Despite this, the majority of RP studies of air travel choice behaviour have struggled or been unable to retrieve meaningful marginal fare effects. This is almost certainly a direct effect of the poor quality of the fare data, characterised by a high level of aggregation. Indeed, it is in most studies only possible to obtain information on the average fare charged

by a given airline on a specific route. This clearly involves a great deal of aggregation, as no distinction is made between the fares paid across different travellers (i.e. in terms of travel classes as well as booking classes). With the dynamic nature of air fares, the levels used in modelling thus often bear little or no resemblance to those actually faced in the choice set, with the unavoidable implication of poor modelling results. Even though some progress can be made with the help of bookings data, issues of aggregation do remain. In fact, it can be seen that, in RP studies, disaggregate choice data is used in conjunction with aggregate level-of-service data, for at least some of the attributes. While, for some characteristics, this may be acceptable, it does, as described above, create significant problems in the treatment of air fares, and flight availability by extension.

4.1.3 Frequent flier information

While it is well known that airline allegiance as a result of membership in frequent flier programmes plays an important role in air travel choice behaviour, such information is often not collected in passenger surveys. As a result, this potentially crucial influence on choice behaviour cannot usually be taken into account in RP case studies.

4.1.4 Influence of other attributes

While most applications generally only look at a limited set of attributes, such as access time, flight time, frequency and fare, various other factors, such as on-time performance and in-flight entertainment conceivably also have an influence on travellers' choices. Often, information on such attributes is however not available, potentially significantly increasing the error terms in the models, although some effects may be captured in airline specific constants.

4.1.5 Survey design issues and inter-dataset compatibility

The main input into RP air travel behaviour studies comes in the form of data collected at departure airports, generally by airport operators or civil aviation authorities. The factors of interest in these questionnaires often differ from those relevant to a modelling analysis (e.g. the lack of information on frequent flier programme membership), with the obvious impacts this has on any analysis. Additionally, compatible sources of level-of-service data need to be found.

7

4.2   Stated Choice data

Some of the problems discussed above in the context of RP data can be alleviated by making use of SC data collected on the basis of tailor-made surveys, designed specifically for the use in advanced modelling analyses. Here, the main advantage is the fact that complete and accurate information is available on all alternatives faced by the respondent. From this point of view, it should come as no surprise that studies making use of SC data have been much more successful in retrieving significant effects for factors such as air fares and airline allegiance [cf. 16, 17]. However, it is important to remember that the use of SC data does pose some philosophical problems, in terms of how the behaviour differs from that observed in RP data [see e.g. 18]. Additionally, issues of survey complexity need to be addressed. To allow respondents to better relate to the presented choices, surveys now often include a real-world reference alternative in the choice set; this however poses some additional issues, as discussed recently by Hess [19]. Both RP and SC data have advantages in their own right, and to a large extent, the choice of the optimal approach depends on the issues to be investigated, as well as the quality of the level-of-service data in the RP context. An interesting approach in this context is to combine RP and SC data, as done by Algers and Beser [20], hence correcting for the bias inherent to models estimated on SC data. The problem in this case however is one of obtaining compatible RP and SC datasets.

5   Model structure

Probably the most important question to address at the modelling stage is the choice of a mathematical structure. Given the important differences across travellers both in terms of behaviour as well as choice context, the use of a disaggregate modelling approach is clearly preferable to an aggregate one, and here, discrete choice structures belonging to the class of random utility models (RUM) have established themselves as the preferred approach. For an in-depth discussion of these modelling structures, see Train [21]. Here, we look solely at two main issues of great relevance in the analysis of air travel behaviour, namely that of inter-alternative correlation in the error terms (Section 5.2) and that of taste heterogeneity (Section 5.3) across travellers. This is preceded by a brief introduction of some common notation.

8

5.1 Basic concepts

A discrete choice model looks at the choices made by a *decision maker n* amongst a set of mutually exclusive alternatives contained in a choice set C_n. Each alternative $i = 1, \ldots, I$ in the choice set has an associated 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. Under the assumption of *utility maximising behaviour*, respondent n will choose alternative i if and only if $U_{i,n} > U_{j,n}$ $\forall j \neq i$, with $i, j \in C_n$.

The utility of an alternative is a function of its attributes and the tastes of a decision maker. Given inherent randomness in behaviour as well as data limitations, only part of the utility can be observed, such that we rewrite:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n}, \tag{1}$$

with $V_{i,n}$ and $\varepsilon_{i,n}$ 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 as faced by decision-maker n[1], and β_n is a vector of parameters representing the tastes of decision-maker n, which is to be estimated from the data.

Due to the presence of the unobserved utility term $\varepsilon_{i,n}$, 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, where the individual probabilities are given by:

$$P_n(i) = P\left(\varepsilon_{j,n} - \varepsilon_{i,n} < V_{i,n} - V_{j,n} \ \forall j \neq i\right). \tag{2}$$

After noting that the mean of the unobserved utility terms can be added to the observed part of utility in the form of an alternative-specific constant (ASC), the vector $\varepsilon_n = \{\varepsilon_{1,n}, \ldots, \varepsilon_{I,n}\}$ is now defined to be a random vector with joint density $f(\varepsilon_n)$, zero mean and covariance matrix Σ. From this, we can rewrite (2) as:

$$P_n(i) = \int_{\varepsilon_n} I\left(\varepsilon_{j,n} - \varepsilon_{i,n} < V_{i,n} - V_{j,n} \ \forall j \neq i\right) f(\varepsilon_n) \, d\varepsilon_n, \tag{3}$$

where $I(\cdot)$ is the indicator function which equals 1 if the term inside brackets is true and 0 otherwise. Different assumptions on the distribution of the error terms lead to different forms for the choice probabilities, and the

---

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

multidimensional integral in Equation 3 will only takes a closed form for certain choices of distribution for ε_n. In the most basic model, the Multinomial Logit (MNL) model, the error terms are assumed to be distributed identically and independently (*iid*), with more advanced models allowing for complex interdependencies.

5.2   Correlation between alternatives

As mentioned in Section 3.1.2, 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. Indeed, it cannot generally be expected that any commonalities between two alternatives sharing a common component along one or more of the choice dimensions would need to be explained in the observed part of utility. This makes basic models that assume independence of error terms, such as MNL, inappropriate for such modelling purposes.

The typical departure from the MNL model in the context of air travel choice behaviour has been to make use Nested Logit (NL) models which form the most basic nesting structure within the generalised Extreme Value (GEV) family of models, introduced by McFadden [22]. This set of models are all based on the use of the extreme-value distribution and allow for various levels of correlation among the unobserved part of utility across alternatives. This is accomplished by dividing the choice set into nests of alternatives, with increased correlation, and thus higher cross-elasticities, between alternatives sharing a nest. Alternatives sharing a nest are more likely substitutes for each other. These structures are generally represented by an upside-down tree , with the root at the top, elementary alternatives at the bottom, and composite alternatives, or nests, in between.

In a two level NL model. In this model, alternatives are grouped into mutually exclusive nests, where, for each nest, a structural (nesting) parameter is estimated that relates to the level of correlation in the error terms of alternatives sharing that nest. One example of such a structure is to nest the alternatives by airport, i.e. allowing for heightened correlation between alternatives sharing the same departure airport. This structure is illustrated in Figure 1, with K mutually exclusive nests, one for each airport, and where each nest has its own nesting parameter, λ_k, thus allowing for different correlation levels with different airports.

The example in Figure 1 can be replicated for the case of correlation along the airline dimension as well as correlation along the access mode dimension (or other dimensions of choice). However, the structures discussed
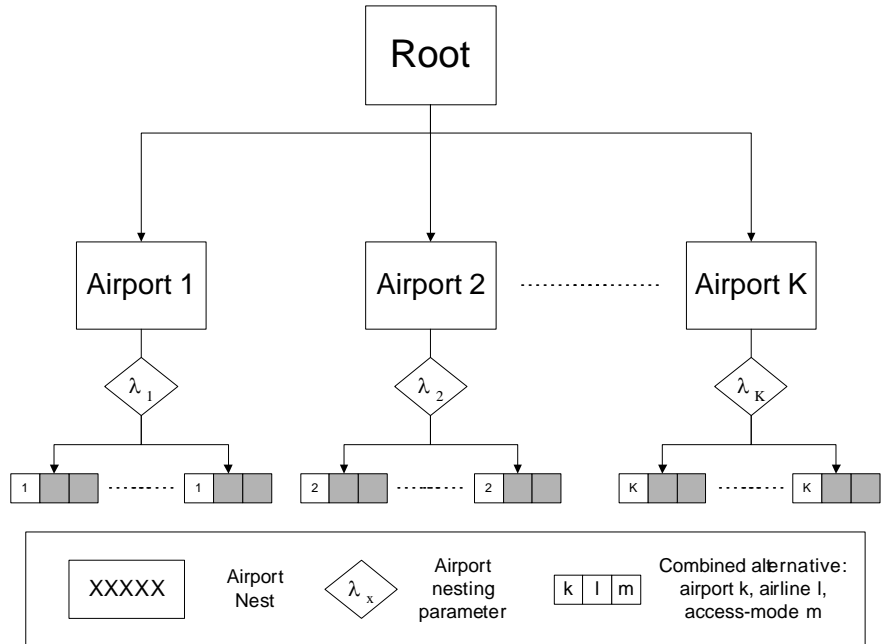
Figure 1: Structure of two level NL model, using nesting along airport-dimension

so far only allow for correlation along one dimension of choice at a time. While it is possible to extend the NL structure to multiple dimensions of nesting, as shown in Figure 2 for the case of airport and airline choice[2], where π_l is used as the nesting parameter for airline nests, this is not optimal. Indeed, the ordering of choice dimensions plays a role and the full level of correlation is only allowed for along the highest level of nesting. 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

---

[2]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.
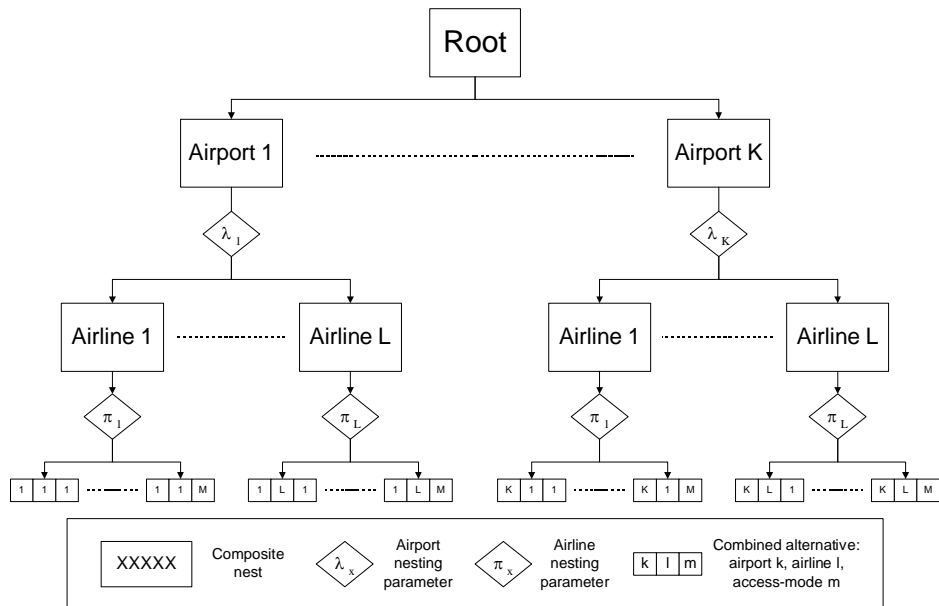
Figure 2: Structure of three level NL model, using nesting along airport-dimension and airline-dimension

on airline l for that airport k. The model is thus 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. Finally, it can also be noted that this structure can only accommodate correlation along all but one of the dimensions of choice. Indeed, using the example shown in Figure 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 contain a single alternative, as the airline nest preceding the access mode nest would contain exactly one alternative for each access mode. As such, the lower level of nesting becomes obsolete.

Both of these problems can be addressed by using a Cross-Nested Logit (CNL) model, as discussed in the application in Section 6. This structure would be 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. The resulting structure
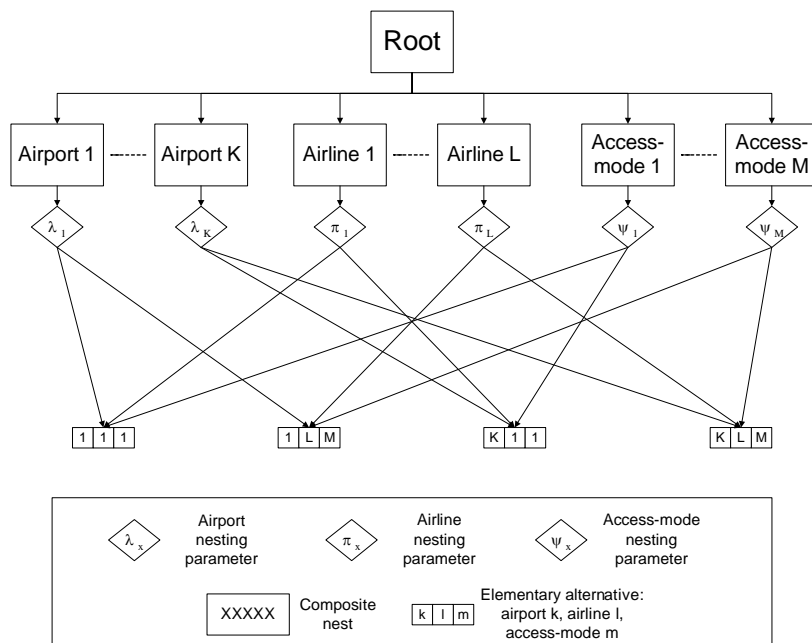
Figure 3: Structure of CNL model for the joint analysis of correlation along the airport, airline and access mode dimensions

allows for correlation along all three dimensions of choice and does so in a simultaneous rather than sequential fashion, i.e. not giving priority to one dimension of choice.

An example of such a model is shown in Figure 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 triplets of alternatives is shown. Additionally, the allocation parameters [cf. 21, chapter 4], governing the proportion by which an alternative belongs to each of the three nests, are not shown in Figure 3.

5.3 Variation in behaviour across respondents

Another important issue that needs to be addressed during model specification are potential differences between travellers in their sensitivities to changes in attributes defining the alternatives. Such differences can clearly

be seen to be likely to apply to air fare and travel sensitivities, but may also extend to other factors, such as the willingness to pay for flying on a certain airline or on a certain type of aircraft. From this point of view, it is clearly a major assumption to make an assumption of taste homogeneity.

Various possible approaches exist to allow for taste heterogeneity. The most basic approach is to use discrete segmentations of the sample population, estimating separate models for different subsets of the sample, or in a less extreme case, separate coefficients within the same model. There are also many situations in which tastes can be expected to be related to socio-demographic attributes, such as for example lower cost sensitivity with rising income. While it would be possible to segment the sample population into different income classes, it may, in many cases, be preferable to allow for continuous interactions between tastes and sensitivities. Here, the modeller needs to additionally make a choice between linear or non-linear interactions.

Due to data limitations as well as inherent randomness in sensitivities, not all taste heterogeneity can be explained in a deterministic fashion, and analysts are increasingly turning to models such as Mixed Multinomial Logit (MMNL) that allow for a representation of random (as opposed to deterministic) taste heterogeneity. In a MMNL model, the choice probabilities are given by integrals of MNL probabilities over the assumed distribution of taste coefficients. Although incredibly flexible, the MMNL model comes with a high computational cost due to the reliance on simulation in estimation and application.

The various methods for representing taste heterogeneity are characterised by differences in terms of flexibility, ease of implementation/estimation, and ease of interpretation. These factors are strongly correlated, with more advanced specifications offering gains in flexibility at the expense of higher computational cost as well as issues in interpretation. Nevertheless, whichever approach is used, it is important to at least challenge the assumption of taste homogeneity.

5.4   Discussion

The modelling approaches described in Section 5.2 and Section 5.3 have quite separate aims; the analysis of inter-alternative correlation, along multiple dimensions, and the representation of deterministic and random variations in choice behaviour. When both phenomena play a role simultaneously, analysts may need to rely on even more advanced model structures that allow for the joint representation of inter-alternative correlation and random

14

taste heterogeneity [cf. 23, 24].

6 Empirical example

For the empirical example used in this section, we summarise part of the results described in the study by Hess and Polak [15]. Particularly, we look at the combined choice of airport, airline and access mode for air passengers departing from Greater London.

6.1 Description of data

The application described here made use of RP data, with information on actual trips taken from the 1996 passenger survey conducted by the Civil Aviation Authority [25]. The final sample used for the analysis reported here contained data from 8,704 respondents travelling for business reasons with the London departure being the outbound leg of the journey. A total of 31 destinations used in the analysis, all served by a single airport, and spread across Great Britain, Europe, the Middle East and North America. For short haul destinations an assumption was made that respondents made an a priori assumption to travel by air. Air-side level-of-service data were obtained from BACK aviation, with information on fares compiled from the International Passenger Survey [26] and the fare supplement of the Official Airways Guide for 1996 [27]. As is the case with most RP studies, the resulting dataset is of highly aggregate nature, leading to the previously discussed problems in the estimation of the marginal utility of air fares (cf. Section 4.1.2.) Additionally, no information was available on frequent flier programmes. Finally, for the analysis of the ground-level choice dimension, data from the National Airport Access Model (NAAM) were obtained for the base year 1999 [28], and 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.

6.2 Model structure and specification

The final choice contained 5 departure airports, namely Heathrow (LHR), Gatwick (LGW), Stansted (STN), Luton (LTN) and London City (LCY). This mix of major airports, outlying airports and a small city centre airport means the area if of special interest in the context of the present book. Along the airline dimension, there were 37 options, airlines, with 6 modes

available along the access journey dimension[3]. This leads to a total of $1,110$ combinations of airports, airlines and access modes arise. However, with not all airlines operating from all airports, the total number of airport-airline pairs is actually 54, which reduces the number of alternatives (airport, airline, access mode triplets) to 324[4].

A large number of different attributes were used in the initial modelling analysis, including attributes relating to the air journey, such as frequency, fare, flight time, aircraft type, and seat capacity, and attributes relating to the access journey, such as access cost, in-vehicle access time, out-of-vehicle access time, wait time, number of interchanges, and parking cost. Different model structures were investigated, namely MNL, two-level NL, and CNL, where in all models, weights were used in the specification of the log-likelihood function to account for the quota used in data collection, which are not representative of the population level.

6.3 Model results

Despite an extensive specification search, only a limited number of attributes were observed to have a significant impact on behaviour, namely access cost, in-vehicle access time (IVT), flight frequency, and flight time, where a log-transform was used for all four attributes. Crucially, no effect could be identified for air fare, where this is at least partly due to the poor quality of the data as discussed in Section 4.1.2. The list of significant attributes was observed to stay identical across model structures, i.e. MNL, NL and CNL.

Table 1 presents a summary of the mathematical performance of the different models. Here we can see that all three NL models outperform the MNL model, with the best performance offered by the model using nesting by access mode. In turn, the CNL model outperforms not only MNL but also all three NL structures. Finally, the the total improvement of the CNL model over the MNL model is bigger than the combined improvements in the adjusted ρ^2 measure for the three NL models. However, while the results show that the CNL model offers significant improvements over the more simple NL models, these come at the expense of a very significant increase in estimation cost.

We next turn our attention to substantive results, which are summarised

---

[3]The options were private car, rental car, public transport (rail, bus, local transport), long distance coach, taxi and minicab (MC), where, for data reasons, no combinations of modes were considered in the present analysis.

[4]The number of available alternatives for specific individuals in the estimation sample ranges from 6 to 58, with a mean of 31.

Table 1: Model performance on London data

Model	Final LL	Parameters	Adjusted ρ^2
MNL	-14,945.3	55	0.3445
NL by airport	-14,896.1	59	0.3465
NL by airline	-14,870.7	74	0.3469
NL by access mode	-14,816.7	60	0.3499
CNL	-14,603.9	91	0.3578

in Table 2. Given the use of the log-transform in nominators as well as denominators, the various trade-offs were calculated separately for each individual[5], and summary statistics were then calculated across respondents. We can see that the first three models produce roughly similar results, while those produced by the CNL model and the NL model using nesting access mode are more extreme (when compared to the three first models), especially when looking at the value of travel time savings (VTTS) measures for the model using nesting by access mode.

Another observation that can be made for the trade-offs is that the VTTS measures are markedly lower than those reported for example by Pels et al. [3], although they are still higher than in other contexts, which can be explained partly by concepts of risk-averseness, as discussed for example by Hess and Polak [7]. Travellers are willing to pay for a reduction in the risk of missing their flight, where this risk clearly increases with access time. The still high values should also be put into context by noting that the average access journey in this population segment was measured as 57 minutes.

Finally, a detailed analysis of the correlation structure in the different models [cf. 15] highlights the presence of high levels of correlation between the errors for alternatives sharing the same airport, airline or access mode. Here, the CNL model was more successful at retrieving these correlation patterns than was the case for the different NL structures.

7 Summary and Conclusions

The discussion in this chapter has highlighted the complexity of the choice processes undertaken by air travellers, with decisions being taken along a multitude of dimensions. In practice however, it is almost inevitable to use

---

[5]With $U = \ldots + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \ldots$, the ratio of the partial derivatives of U with respect to x_1 and x_2 is given by $\frac{\beta_1}{\beta_2} \frac{x_2}{x_1}$, as opposed to the simple $\frac{\beta_1}{\beta_2}$ ratio used in the case of a linear parameterisation.

Table 2: Model results for London data

	IVT vs. access cost (£/hour)	Freq. vs. access cost (£/flight)	Freq. vs. IVT (hours/flight)	Flight time vs. IVT
MNL				
Minimum	1.18	0.02	0.01	0.04
Mean	16.24	1.56	0.11	1.07
Maximum	143.38	231.05	4.06	7.43
Standard deviation	25.44	4.85	0.18	0.7
NL by airport				
Minimum	1.3	0.02	0.01	0.04
Mean	17.85	1.63	0.11	0.97
Maximum	157.65	242.38	3.87	6.72
Standard deviation	27.98	5.09	0.17	0.63
NL by airline				
Minimum	1.29	0.03	0.01	0.04
Mean	17.76	1.79	0.12	1.13
Maximum	156.8	265.11	4.26	7.85
Standard deviation	27.83	5.57	0.19	0.74
NL by access mode				
Minimum	0.99	0.02	0.01	0.04
Mean	13.52	1.11	0.1	1.05
Maximum	119.35	164.74	3.47	7.31
Standard deviation	21.18	3.46	0.15	0.69
CNL				
Minimum	1.16	0.01	0	0.04
Mean	15.96	0.9	0.07	0.95
Maximum	140.89	132.96	2.38	6.59
Standard deviation	25	2.79	0.1	0.62

some simplifications of the choice process, partly because of modelling complexity, but mainly because of data issues. In this context, the *advantages* of SC data make the use of such datasets an important avenue for further research, potentially in conjunction with compatible RP data.

The problems that need to be faced when making use of RP data were also highlighted in the empirical example in Section 6 where it was not

possible to retrieve significant effects for a range of important variables that included air fares, schedule delay and airline and airport allegiance. From a model structure point of view, the application has 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 one common observation from this application and that of other studies is that the results do suggest that access time plays a major role in the choice process, with passengers having a strong preference for 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 sensitivity to access time decreases with flight time [cf. 17], 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.

Acknowledgements

References

[1] Mar González-Savignat. Competition in air transport: The case of high speed train. *Journal of Transport Economics and Policy*, 38(1):77–108, 2004.

[2] E. Pels, P. Nijkamp, and P. Rietveld. Airport and airline choice in a multi-airport region: an empirical analysis for the San Francisco bay area. *Regional Studies*, 35(1):1–9, 2001.

[3] E. Pels, P. Nijkamp, and P. Rietveld. 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, 2003.

[4] S. Pathomsiri, H. Mahmassani, and A. Haghani. *Airport Choice Model for Baltimore-Washington Region.* paper presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, 2004.

[5] G. Basar and C. R. Bhat. 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, 2004.

[6] S. Hess and J. W. Polak. Accounting for random taste heterogeneity in airport-choice modelling. *Transportation Research Record*, 1915:36–43, 2005.

[7] S. Hess and J. W. Polak. Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management*, 11(2): 59–68, 2005.

[8] Kimon Proussaloglou and F. S. Koppelman. Air carrier demand: an analysis of market share determinants. *Transportation*, 22(4):371–388, 1995.

[9] A. T. H. Chin. Impact of frequent flyer programs on the demand for air travel. *Journal of Air Transportation*, 7(2):53–86, 2002.

[10] A. B. F. Monteiro and M. Hansen. Improvements to airport ground access and behavior of multiple airport system: BART extension to San Francisco International Airport. *Transportation Research Record*, 1562:38–47, 1997.

[11] V. Psaraki and C. Abacoumkin. Access mode choice for relocated airports: the new Athens International Airport. *Journal of Air Transport Management*, 8(2):89–98, 2002.

[12] L. Bondzio. *Modelle für den Zugang von Passagieren zu Flughäfen (Models for the Passengers' Access to Airports).* PhD thesis, Ruhr-University, Bochum, 1996.

[13] S. Hess and J. W. Polak. Airport, airline and access mode choice in the san francisco bay area. *Papers in Regional Science*, 85(4):543–567, 2006.

[14] S. Hess. Posterior analysis of random taste coefficients in air travel choice behaviour modelling. *Journal of Air Transport Management*, 13 (4):203–212, 2007.

[15] S. Hess and J. W. Polak. 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*, 42(2):63–81, 2006.

[16] T. Adler, C. S. Falzarano, and G. Spitz. *Modeling Service Trade-offs in Air Itinerary Choices*. paper presented at the 84th Annual Meeting of the Transportation Research Board, Washington, DC, 2005.

[17] S. Hess, T. Adler, and J. W. Polak. Modelling airport and airline choice behavior with stated-preference survey data. *Transportation Research Part E: Logistics and Transportation Review*, 43:221–233, 2007.

[18] J. J. Louviere, D. A. Hensher, and J. Swait. *Stated Choice Models: Analysis and Application*. Cambridge University Press, Cambridge, 2000.

[19] S. Hess. *Treatment of reference alternatives in stated choice surveys for air travel choice behaviour*. Journal of Air Transport Management, forthcoming, 2008.

[20] S. Algers and M. Beser. 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, 2001.

[21] K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, MA, 2003.

[22] D. McFadden. Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. W. Weibull, editors, *Spatial Interaction Theory and Planning Models*, chapter 25, pages 75–96. North Holland, Amsterdam, 1978.

[23] J. Walker. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. PhD thesis, MIT, Cambridge, MA, 2001.

[24] S. Hess, M. Bierlaire, and J. W. Polak. Capturing taste heterogeneity and correlation structure with Mixed GEV models. In R. Scarpa and A. Alberini, editors, *Applications of Simulation Methods in Environmental and Resource Economics*, chapter 4, pages 55–76. Springer Publisher, Dordrecht, The Netherlands, 2005.

[25] CAA. *UK airports Passenger survey, CAP 677 - Passengers at Birmingham, Gatwick, Heathrow, London City, Luton, Manchester and Stansted in 1996.* The Civil Aviation Authority, London, 1996.

[26] ONS. *International Passenger Survey.* The Office for National Statistics, London, 1996.

[27] OAG. *World Airways Guide Fares Supplement 1996.* Reed Elsevier, Dunstable, 1996.

[28] Scott Wilson Kirkpatrick. *National airport accessibility model: model status report.* prepared for the Department for Transport by Scott Wilson Kirkpatrick & Co Ltd, Basingstoke, 1999.