



Consumer choice behaviour and strategies of air transportation service providers

Michael A.B. van Eggermond

ETH Zurich and Delft University of Technology

Nadine Schuessler

ETH Zurich

Kay W. Axhausen

ETH Zurich

Conference paper STRC 2007

STRC

7th Swiss Transport Research Conference
Monte Verità / Ascona, September 12. – 14. 2007

Consumer choice behaviour and strategies of air transportation service providers

Michael A.B. van Eggermond	Nadine Schuessler	Kay W. Axhausen
IVT	IVT	IVT
ETH Zürich	ETH Zürich	ETH Zürich
TU Delft		
CH-8093 Zürich	CH-8093 Zürich	CH-8093 Zürich
phone: +41-44-633 67 37	phone: +41-44-633 30 85	phone: +41-44-633 39 43
fax: +41-44-633 10 57	fax: +41-44-633 10 57	fax: +41-44-633 10 57
michael@eggermond.com	schuessler@ivt.baug.ethz.ch	axhausen@ivt.baug.ethz.ch

September 2007

Abstract

This paper presents an analysis of air traveller's itinerary choice in Europe. The analysis is based on three datasets, which cover altogether 70 origin-destination pairs in Europe. These OD-pairs differ in many aspects such as geographic coverage, number of available connections, number of transfers and travel time.

The first dataset contains itinerary fares collected from the Web in the period September-November 2006 for flights departing in November 2006. The second dataset contains the booked travel itineraries in November 2006 through customer reservation systems. Finally, detailed itinerary information is obtained from the Official Airline Guide (OAG) By matching these three datasets, a data set has been derived that will allow to model the influence of several variables, e.g. departure time, travel time, code shares, number of transfers and fare, on the itinerary choice in air transport.

With the derived data set, it is possible to construct a choice set with the alternatives an air traveller faced at the booking day. Estimated models include models with the sensitivity of air travellers for fare over time and per duration of stay. Also, a continuous departure time function is implemented in the choice model.

Keywords

Aviation, Discrete choice modeling, Airlines, Departure time modeling, Pricing

1. Introduction

Compared to public and private transportation, the number of studies using discrete choice models to represent passenger behaviour in the aviation sector is fairly limited. Potential application areas for discrete choice models are airport choice modelling and air connection choice (itinerary) modelling, or combinations of both. They form a challenging research area as they include complex choices across multiple dimensions. Itinerary choice modelling can aid airlines with their medium and long-term planning as they provide carriers with an understanding of the relative importance of different service factors. This understanding can also aid airlines and online travel portals with the listings of different air connections on their websites.

Several datasets have been necessary for this research: a dataset that contains tickets bookings through computer reservation systems (CRS) for November 2006, a dataset with fares observed in the period September 2006 – November 2006 for flights departing in November 2006 on 70 origin-destination pairs and the official airline guide database (OAG, 2006). They have been combined to form a comprehensive dataset for the analysis of itinerary choice.

The remainder of the paper is structured as follows: First, an overview will be given of choice modelling in the aviation sector. This overview is structured from an airline and airport point of view and help to identify unaddressed and potential research and application areas of discrete choice models. Following this overview, the modelling framework will be presented. Special attention is paid to choice set formulation. The available datasets and descriptive statistics are then discussed in section 4. Before presenting and discussing model estimation results in section 5, the specification of the utility function is presented. The paper concludes with a summary of the main findings.

2. Literature Overview

As discussed in the previous section, this literature overview is structured by the various planning levels that airline managers face. After the overview regarding the airlines, a brief overview of airport choice literature will be given, together with a potential application area of choice models in this setting.

Ordered by long term to short term decision (or strategic to tactical), the following decision stages can be recognized (Belobaba 2006):

- *Fleet planning*, which regards the number and type of aircraft to acquire or retire. Criteria for aircraft evaluation include technical performance and characteristics, economics of operation and revenue generation, marketing and environmental issues and political and international trade concerns;
- *Route evaluation*, which regards what network structure to operate and which city-pairs to serve. Considerations include forecasts of potential demand and revenues, airline's market share of total demand and network implications for costs and revenues;
- *Schedule development*, which regards frequency planning, timetable development, fleet assignment and aircraft rotation planning. In this stage the demand per itinerary is necessary and the response of demand to a decrease or increase in service level per time period (Lohatepanont & Barnhart 2004);
- *Pricing*, which regards the products, fares and restriction for each origin-destination market. Current challenges lie within the field of price elasticity estimation and willingness-to-pay;
- *Revenue management*, how many bookings should be accepted, by type of fare to maximize the revenue of each flight and over the network. This can than also be seen as inventory control for airlines. It is estimated that revenue management systems increase revenues by 4-6% (Talluri & van Ryzin 2005).

In all planning stages, forecasting demand plays an important role. The level of detail varies from aggregate (i.e. development of air demand, in general, origin-destination market) to disaggregate (i.e. origin-destination pair, leg).

Coldren *et al* (2003) argue that disaggregate demand models can be used to support long and intermediate decision-making, as current studies of air-travel market allocation do not give an airline's management enough planning information due to its lack of detail on carrier service attributes in different markets. Studies discussed in their literature overview are either based on a high level of geographic aggregation or limited to a small number of city-pairs. However, they do not discuss how a disaggregate demand model may be applied to fleet planning or route evaluation.

Parker (2007) discusses several potential applications of discrete choice models with regard to airline planning. One application is the incorporation of a passenger choice behavior in a market simulator, named the Universal Market Simulator (UMS). This is a discrete event simulator, in which airlines and passengers act as agents. After running a number of simulations, demand is assigned to airlines and the network. He mentions some features still lacking in the UMS, such as models for airport choice and more specific choice models of passenger behaviour. Parker addresses the application of the notion of consumer surplus, coupled with discrete choice models in order to evaluate a network change. This network change can follow from the introduction of new equipment to the impact of a low cost carrier.

Both studies highlight the usage of discrete choice models in the context of strategic and tactical planning as they address fleet planning and route evaluation.

Carrier (2006) argues that previous studies have not included fare and schedule convenience on a detailed level, which ultimately influences passenger choice and sees as a potential application area pricing policy and revenue management. Such a level of detail might, however, be unnecessary for strategic and tactical planning, as also argued by Grammig *et al* (2005). They argue that fare is an outcome of the revenue management in place, and not necessary for network planning. Boeing for instance, offers a high and low resolution discrete choice model (Parker 2007), and apply them for different purposes and planning levels. Carrier (2006) analyzes the joint choice of an itinerary and a fare product based on past booking data. Talluri and van Ryzin (2004) step into more detail and apply a simple discrete choice model to revenue management and compare it to a current revenue management method. The incorporation of a discrete choice model in revenue management algorithms lead to an increase in revenue. They consider an individual making a choice for a fare product on an itinerary and not a choice between itineraries, as earlier discussed studies (Coldren, *et al.* 2003, Coldren & Koppelman 2005, Garrow, *et al.* 2007, Parker 2007) do. A considerable part of revenue management literature covers standby and overbooking forecasting. Discrete choice modelling is applied here by Garrow and Koppelman (2004a, 2004b). These studies offer a more detailed description of standby and no-show behaviour, as they use disaggregated data and offer an analysis of rescheduling behaviour.

Several studies have been carried into airport choice behaviour. Bondzio (1996) conducted a study regarding airport choice in Southern-Germany and showed that travel time to the airport played an important role and that access time was more important for business passengers than for leisure travellers. Pels *et al* (2001, 2003) analyzed the combined choice of airport and airline in the San Francisco area and the combined choice of access-mode and airport. In the first study they find that airline choice is nested within airport choice, i.e. the competition between airlines departing from the same airport is more severe between airlines departing from different airports. In their second study they analyze the joint choice of access-mode and

airport, showing high sensitivity to access time, especially for business travellers. Business passengers also consider frequency of the flight to be important. Leisure travellers consider access cost and itinerary fare more important. A case study of the London area is presented by Hess and Polak (2006). Their study reveals that business travellers are very reluctant to accept increases in access journeys; outlying airports depend heavily on good-access connections and/or low air fares. The results of these studies show that strong differences exist between preferences of leisure and business travellers.

An aspect not directly related to the interaction between travellers and airports is demand management. Demand management refers to any set of administrative or economic measures and regulations aimed at constraining the demand for access to a busy airfield and/or modifying the temporal characteristics (de Neufville & Odoni 2003). Three approaches are available: purely administrative, purely economic and hybrid approaches, which combine the previous two. The fundament of the administrative approach is the slot: an interval of time reserved for the arrival or departure of the flight. Airlines do not necessarily have to use assigned slots. Economic approaches utilize congestion pricing (Brueckner 2002, Pels & Verhoef 2004), which has goal to internalize external costs. A hybrid approach would consist of the assignment of slots and congestion pricing, where the landing fees would be published prior to the slot assignment. These slots could then be auctioned. Without carrying into much detail, a potential application of the itinerary choice models of Coldren *et al* (Coldren, *et al.* 2003, Coldren & Koppelman 2005) can be seen here. Coldren showed that the choice probabilities of different itineraries, holding all other attributes constant, differed per hour-of-day. For airlines such models can help in determining their willingness-to-pay for a certain slot; such models can help airports with the differentiation of their slot and landing fees.

From the preceding discussion, two largely unaddressed and interesting research themes can be distilled. First, this is the willingness-to-pay of a traveller for several aspects of an itinerary, such as departure time and service level are not incorporated in a discrete choice model framework, despite the potential of doing so. Second, the incorporation of discrete choice models in revenue management applications deserves more attention. Without carrying into detail on architecture of such systems, this can be either in the demand forecasting stage in such a system, but this can also be on the moment a potential customer requests a fare. This fare can then be adjusted to the willingness-to-pay of a customer based on the limited information a traveller enters on a booking site.

3. Modelling Framework

3.1 MNL-Model

With discrete choice models, a decision-makers' choice is described; any choice is made, by definition, from a non-empty set of alternatives. The utility U_{iq} of an alternative i for a decision-maker q is defined by:

$$U_{iq} = V_{iq} + \varepsilon_{iq} = f(\beta, x_{iq}) + \varepsilon_{iq} \quad (1)$$

with a deterministic part V_{iq} that consists of a function $f(\beta, x_{iq})$ of the vector β of taste parameters and the vector x_{iq} of attributes of the alternative, the decision-maker and the choice situation. In addition, socio-demographic attributes of decision-maker q can be included in the deterministic part of the utility function. The non-deterministic, non-observable part of the utility function is captured by ε_{iq} .

Decision-maker n will chose the alternative from set C with the highest utility:

$$P(i | C_q) = P[U_{iq} \geq U_{jq} \forall j \in C_q] = P[U_{iq} \max_{j \in C_q} U_{jq}] \quad (2)$$

The most widely applied discrete choice model is the Multinomial Logit (MNL) model. With this model, the choice probability of each alternative i can be calculated as:

$$P(i | C_q) = \frac{e^{V_{iq}}}{\sum_j e^{V_{jq}}} \quad (3)$$

In the next paragraph, a closer look will be given to notion of the choice set and the determination of the choice set.

3.2 Choice Set Formulation

The environment of the decision maker determines the universal set of alternatives. Any single decision maker considers a subset of this universal set of alternatives, the choice set or consideration set (Ben-Akiva & Lerman 1985, p. 33). The identification of the list of alternatives is usually referred to as choice set generation or choice set formation. It is however important to make a clear distinction between choice set generation and choice set formation (Bovy 2007). In the ensuing, it is assumed that choice set generation is a process performed by the analyst. Choice set formation is the result of a behavioural process of an individual and thus equals the consideration set. Bovy (1990, 2007) gives a thorough of the

behavioural process and makes a distinction between different stages in the choice set formation process.

At least four dimensions can be recognized when an individual chooses for an itinerary and which influence the composition of the consideration set:

1. The first dimension is the booking period dimension: an individual can choose to book his ticket any time in the period between the decision to make a trip and the preferred departure time.
2. The second dimension is the dimension that includes the choice of air transportation service provider. An individual may consider all possible transportation service providers for his journey, but it is also very well possible that an individual is bounded to a carrier through a loyalty scheme or shows a preference for low-cost carriers;
3. The third dimension is the departure time choice: a traveller may make a trade-off between her preferred departure and thus preferred arrival time and attributes of other known alternatives.
4. The fourth dimension concerns the fare of an itinerary. A traveller might not consider all the fares or fare products offered by an airline.

Several types of consideration sets can be distinguished when modelling itinerary choice. The consideration set will contain different alternatives, based on the ranges a traveller considers of each dimension. In Table 1 two levels are assigned to the booking time dimension (single booking consideration/multiple booking consideration), the information acquisition dimension (one carrier/multiple carriers) and preferred arrival time dimension (low preference/high preference). This leads to eight types of choice sets.

Two of choice set types are less likely to occur, namely choice set V and VI. Travellers' having a high arrival time preference are more likely to immediately book their ticket, either because they perceive a sense of urgency or because they consider price less important as arrival time. Choice sets based on multiple considerations, are likely to become very large. Booking sites for instance offer 50 itineraries per time. If all possible itineraries would be included for each time a traveller acquires information, the number of alternatives in the choice set would become very large. An approach as proposed by Chorus(2007) might offer a possibility to reduce the choice set. Chorus presents a discrete choice modelling approach which describes the full sequence of possibly multiple information acquisitions, followed by a travel choice. In all cases, it can prove hard to determine the traveler's choice set, especially if revealed preference data is used. Even if a choice set is observed, it may difficult to determine the consideration set of a traveler.

Table 1 Combination of ranges of dimensions

Choice Set Type	Booking time dimension	Information acquisition dimension	Preferred arrival time dimension
I	Single booking consideration	One carrier	High preference
II	Single booking consideration	Multiple carriers	High preference
III	Single booking consideration	One carrier	Low preference
IV	Single booking consideration	Multiple carriers	Low preference
V	Sequence of booking considerations	One carrier	High preference
VI	Sequence of booking considerations	Multiple carriers	High preference
VII	Sequence of booking considerations	One carrier	Low preference
VIII	Sequence of booking considerations	Multiple carriers	Low preference

4. Data

4.1 Available Datasets

Three revealed preference datasets are available for the purpose of this research:

- tickets bookings through computer reservation systems (CRS) in November 2006;
- fares observed in the period September – November 2006 for departures in November 2006 on 70 origin-destination pairs;
- the official airline guide (OAG, 2006).

The first dataset contains detailed records of passenger bookings in November 2006 through Computer Reservation Systems (CRS). CRS systems included in the dataset are Amadeus, Abacus, Galileo, Worldspan and Apollo. A rough comparison with Eurostat figures has led us to believe that the CRS data cover between the 40% and 90% of the passenger bookings on any one route. Variables included in the CRS dataset are: booking date, trip origin, trip destination, leg origin, leg destination, departure date, return date, departure and arrival times, carrier abbreviation, and flight number per leg.

The second dataset was obtained by webbots querying Expedia (<http://www.expedia.de>) on a nearly daily basis in the period September – November 2006 for flights departing in November 2006 on 70 origin-destination pairs in Europe. Three durations of stay were queried: a trip returning on the same day, a trip returning on the next day and a trip returning in two weeks time. Variables obtained from Expedia include query date, trip origin, trip destination, departure date, return date, departure and arrival times, carrier name(s), flight number(s) and most notably fare.

Third, detailed information of carrier schedules was obtained from the Official Airline Guide (OAG, 2006). This dataset contains variables such as operating day, operating airline, code share airlines, departure and arrival time and type of aircraft operated.

In order to use these datasets, two steps had to be taken. First, the datasets were matched to obtain complete air connection information. Second, choice sets were extracted.

Air connection fares were added to the passenger bookings by matching the CRS dataset and Expedia dataset on query date, departure date, duration of stay and outbound and inbound flight number combination. OAG information was added by matching flight numbers and carrier abbreviations, taking into account code share. In the end, nearly 19,000 choices with

fare information are observed. Unfortunately, no characteristics of the decision-makers could be included. Therefore, no trip purposes are known and no fare parameter per user category and parameters for deviation of preferred departure time (scheduled delay) could be estimated, as done by Garrow, Jones and Parker (2006).

4.2 Choice Set Comparison and Descriptive Statistics

4.2.1 Days in Advance of Booking & Duration of Stay

First, two aspects of the choice sets will be discussed that are independent of the type of choice set, namely the number of days in advance a ticket is booked and the duration of stay.

Figure 1 shows the percentage of tickets booked per day in advance, i.e. 3 days before departure 15% of the tickets were booked. 50% of the tickets are booked up to 8 days before departure; 95% of the tickets in 36 days in advance.

Figure 1 Days in Advance of Booking

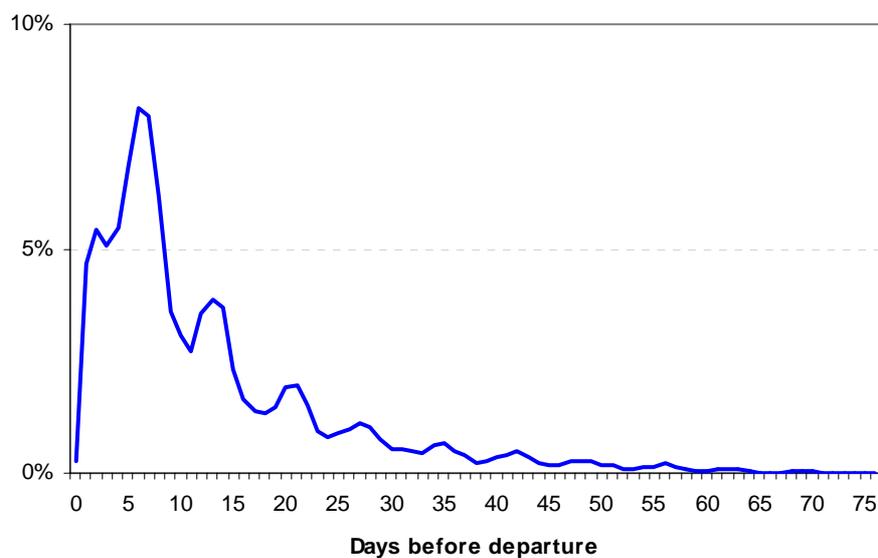
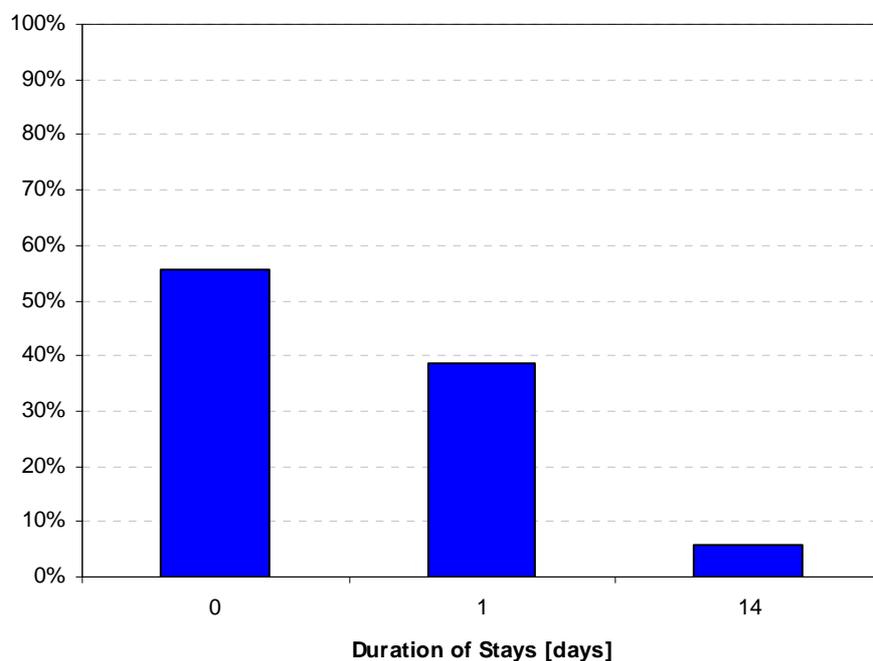


Figure 2 shows the duration of stay in the chosen alternatives after the matching of the datasets. 55% of the passengers return from their destination the same day, almost 40% of the tickets are for a duration of stay of 1 day. The remaining 5% stays 14 days at their destination.

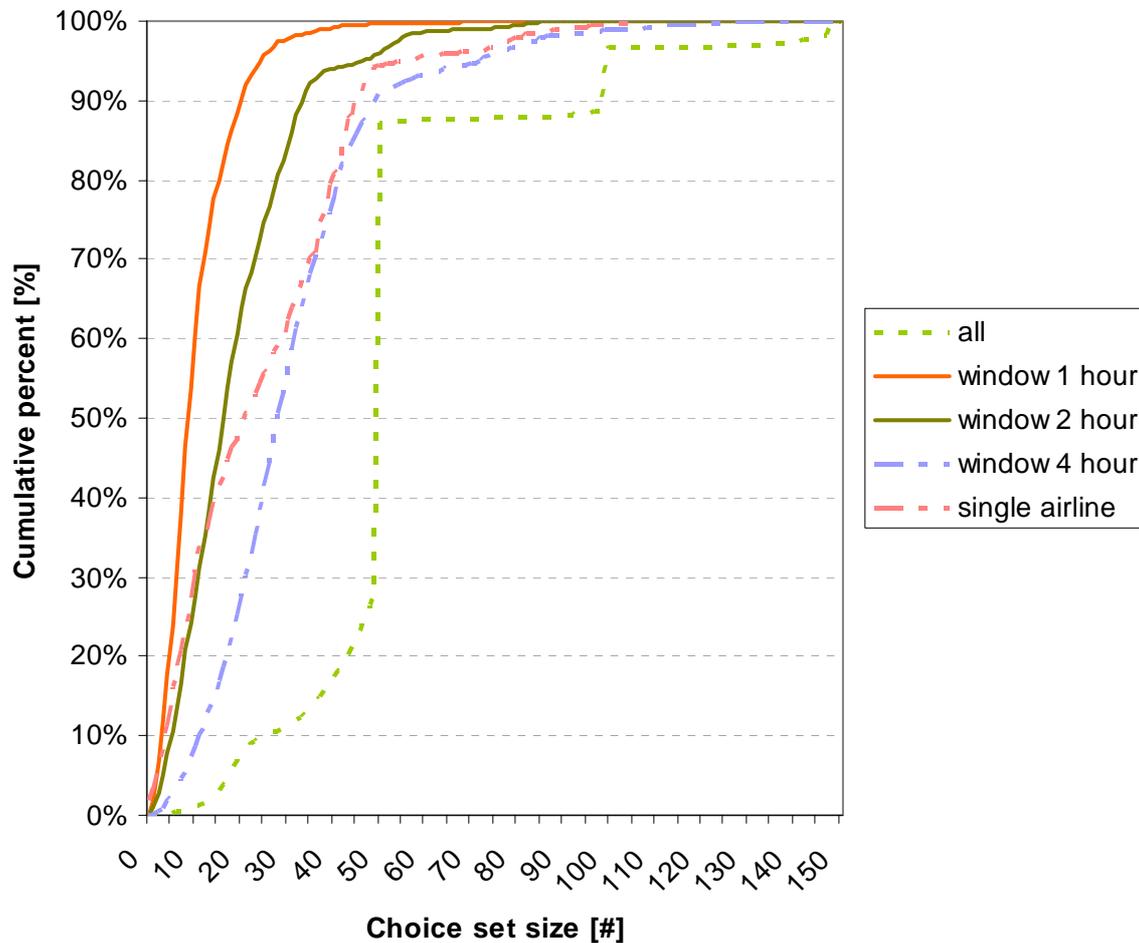
Figure 2 Duration of stay



4.2.2 Choice set specific descriptive statistics

As discussed in section 3.2, it is possible to formulate different types of choice sets. A closer look is given to the arrival time dimension and to the carrier dimension. The arrival time dimension is defined as a window: around each chosen itinerary an arrival time window is defined, which includes all itineraries arriving up to n hours earlier or later.

Figure 3 Choice set size



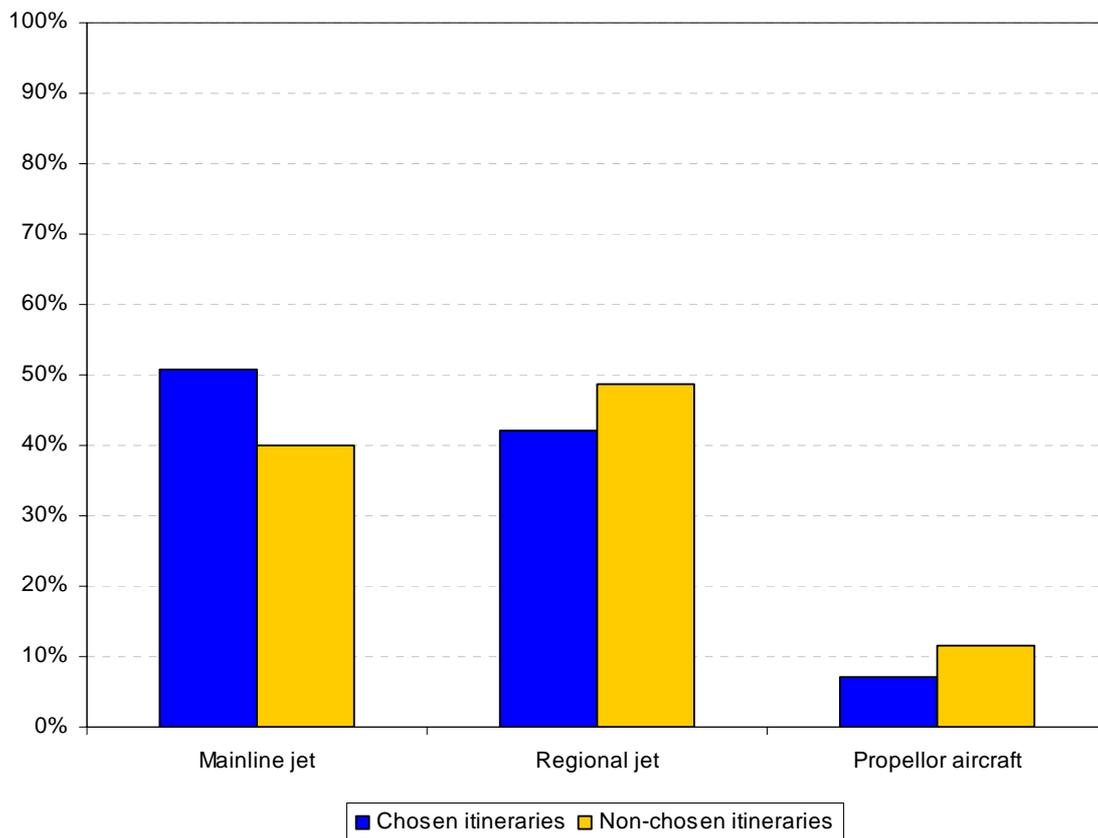
In the figure the choice set size is depicted. It can be seen how the choice set size steadily increases if the window is enlarged. If window is set to 1 hour, 90% of the choice sets contain 20 alternatives or less, if the window is set to 2 hours this number becomes 30. A window of 4 hours leads to even larger choice sets. If a passenger considers all flights on arriving on the same day, a choice set can contain up to 150 flights, almost 90% of the choice sets will contain 50 alternatives. Choice sets containing a single airline (the airline equal to the chosen airline) are also fairly large. This indicates that on a number of origin-destination pairs, only a few carriers serve the route.

In the ensuing, the descriptive statistics of choice sets containing all itineraries available on day of booking for the chosen departure day and chosen origin-destination pair on Expedia.

A distinction can be made between three types of aircraft. These are the mainline jet, regional aircraft and propeller aircraft. The latter aircraft clearly forms a distinctive category. The first two are less clear when for instance looking at number of seats. However, aircraft manufacturers make a clear distinction on their websites. The Airbus 320-series and the Boeing 737-series are considered to be mainline jets; Embraers are considered to be regional jets.

A preference structure can be recognized: mainline jets are chosen more often than regional jets; regional jets are chosen much more often as propeller aircraft. The non-chosen itineraries do not follow this preference structure: itineraries served by regional jets are offered more as mainline jets.

Figure 4 Type of aircraft in chosen and non-chosen itineraries



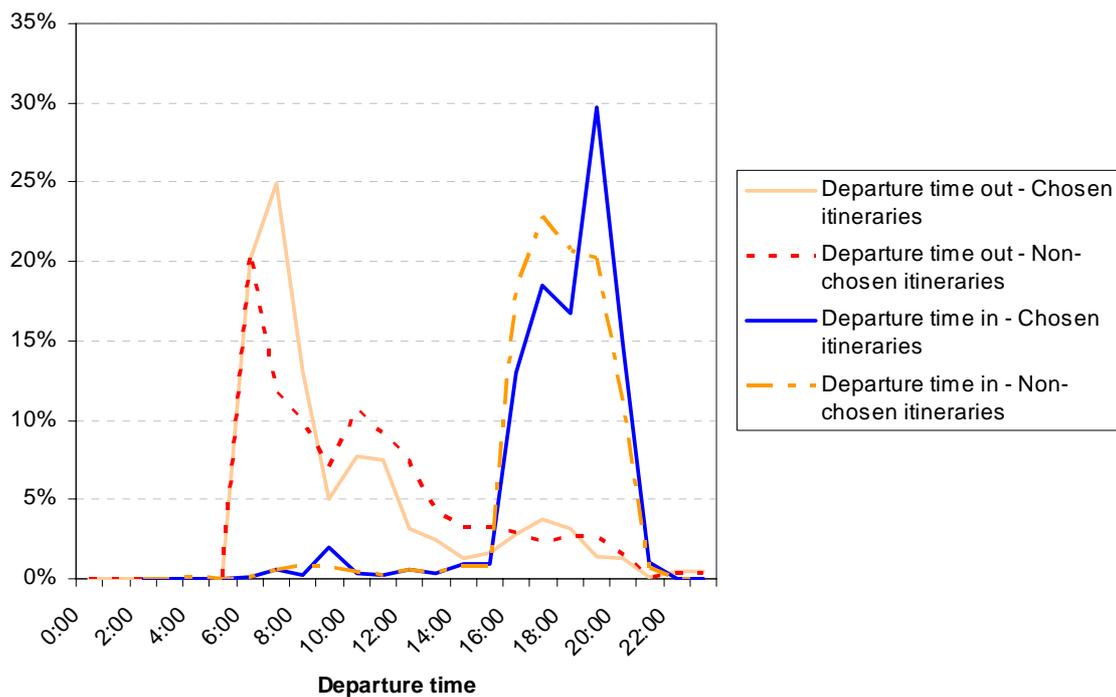
A further distinction between itineraries is their departure time. It is chosen to first aggregate itineraries per hour, i.e. 5:00 – 5:59, 6:00 – 6:59. A higher level aggregation can then be made in the next steps.

In Figure 5 the results can be seen. Most chosen itineraries depart in the period 6:00 – 9:00; a slight increase can be observed in the periods 10:00 and 12:00 and the 16:00 – 19:00. The departure time of the non-chosen itineraries is distributed somewhat more evenly across the day, but follows the same trend.

An opposite trend can be observed for the inbound part of the itinerary; most chosen and non-chosen itineraries depart between 16:00 and 21:00. Again, the non-chosen itineraries are distributed more evenly in this period.

Both the trends for the outbound and inbound part of the itinerary can be expected when combined with the earlier findings with regard to the duration of stay. As a large part of the observed booking fall in the first stay category (duration of stay of 0 days) passengers will have to leave in the morning and return in the evening.

Figure 5 Departure time of chosen and non-chosen itineraries



5. Specification and Results

5.1 Specification

Following travellers' preferences and the available data, several itinerary characteristics are assumed to influence passenger utility:

- The carrier *marketing* the itinerary. This is accounted for by including a dummy variable $D_{carrier_i}$ identifying the carriers of interest. As reference an airline is chosen that occurs frequently in the chosen alternatives;
- The itinerary being a *code-share*. Code sharing is an agreement between two airlines, under which an airline operating a service allows another airline to offer that service to the travelling public under its own flight designator. The practice is most often used to show connecting flights as being on one airline. This is accounted for by including a dummy variable $D_{code-share}$ which indicates if the itinerary contains a code-share;
- The *type of aircraft* operated on the itinerary. A distinction is made between a propeller aircraft, $D_{propeller-aircraft}$, regional jet, $D_{regional-jet}$, and a mainline jet. The latter is the reference category;
- The *total travel time* of the itinerary. The total travel time TT is calculated as flight time plus waiting time at a transfer airport;
- The *number of transfers* of the itinerary. Despite the two transfers being offered on some itineraries, it is never chosen. Therefore, the variable TR indicates if the itinerary contains a transfer;
- The *fare* of the itinerary. This is the fare, $Fare$, of the itinerary as listed on Expedia, taxes not included;
- The *booking time* of the itinerary and *stay duration*. It is hypothesized, that the latter two influence the sensitivity to fare of a traveller. Passengers booking farther in advance will have a higher sensitivity to fare as compared to traveller booking close to their departure date. Also, passengers staying short at their destination are likely to have a lower sensitivity for fare, as these will probably travel for business purposes. A distinction is made between three stay durations $D_{stay_category_k}$ and five booking periods $D_{booking_period_m}$.
- The *departure time* of the itinerary. On the one hand, it is hypothesized that passengers will prefer departing in the morning and returning in the evening. These preferences will

most likely vary per duration of stay. On the other hand, it is important for both airlines and airports to know these preferences for schedule (re)design and slot auctions. As an aggregation level an hourly basis is chosen, which is indicated by $D_{time_period_j}$. For example, itineraries departing between 8:00 and 8:59 fall in category 8, where an itinerary departing at 9:00 falls in category 9.

More specifically, the utility function of a traveller is defined as:

$$U = \sum_{i=1}^{I-1} \beta_{carrier_i} D_{carrier_i} + \beta_{code_share} D_{code_share} + \beta_{regional-aircraft} D_{regional-aircraft} + \beta_{propellor-aircraft} D_{propellor-aircraft} \\ + \beta_{total-travel-time} TT + \beta_{transfer} TR + \beta_{fare} Fare \\ + \sum_{k=1}^K \sum_{j=1}^{J-1} \beta_{time_period_j} D_{time_period_j} D_{stay_category_k}$$

In addition, several variants of this function will be estimated. First, as already discussed, it is hypothesized that passengers staying longer at their destination will have a higher sensitivity to fare (and a lower sensitivity to departure time). In stead of estimating a single β_{fare} , the parameter for fare is replaced by $\sum_{n=1}^N \beta_{fare_n} D_{stay_category_n} Fare$, where $D_{stay_category_n}$ indicates if the itinerary is for period n . The same approach is followed for booking period, where the parameter β_{fare} is replaced by $\sum_{m=1}^M \beta_{fare_m} D_{booking_period_m} Fare$ and $D_{booking_period_m}$ indicates if the itinerary is booked in period m .

The discretization of departure time might give strange changes in choice probabilities. Koppelman *et al* (2007) propose an approach which is adopted from Zeid *et al* (2006) to overcome this problem. Zeid *et al* (2006) propose a trigonometric function to replace dummy variables. The partial utility of departure time then becomes:

$$U(t) = \beta_{\sin 2} \sin\left(\frac{2\pi t}{1440}\right) + \beta_{\sin 4} \sin\left(\frac{4\pi t}{1440}\right) + \beta_{\sin 6} \sin\left(\frac{6\pi t}{1440}\right) \\ + \beta_{\cos 2} \cos\left(\frac{2\pi t}{1440}\right) + \beta_{\cos 4} \cos\left(\frac{4\pi t}{1440}\right) + \beta_{\cos 6} \cos\left(\frac{6\pi t}{1440}\right) \quad (4)$$

Where t is the departure time in minutes and 1440 the number of minutes per day. As can be seen, this a Fourier series approach. Gramming *et al* (2005) model departure time preference with a similar approach, namely $\sum_{q=1}^3 \gamma_q \sin\left(\frac{2\pi q}{T} t_j + \phi_q\right)$, and estimate the parameters $\gamma_1, \gamma_2, \gamma_3, \phi_1, \phi_2, \phi_3$. As can be seen, this is equal to $\sum_{q=1}^3 \gamma_q \sin\left(\frac{2\pi q}{T} t_j\right) \cos \phi_q + \sum_{q=1}^3 \gamma_q \cos\left(\frac{2\pi q}{T} t_j\right) \sin \phi_q$ and therefore equal to formula (4). Formula (4) is more convenient to implement in a software

package such as BIOGEME (Bierlaire 2003). Both approaches should yield the same results and require six parameters to be estimated.

5.2 Results

In Table 2, the reference model is presented. Model estimation is carried out with BIOGEME (Bierlaire 2003). For space reasons, not all departure time parameters are shown. Model performance indicators are shown in Table 4.

All parameters the flight attributes carry the expected signs. The estimated parameters for aircraft attributes follow the anticipated preference structure. Thereby, the dummy variables for individual carriers resulted in significantly better model results than those capturing specific carrier attributes. Obviously the perception of carriers is more complex than indicated by these attributes. However, from the modeller's point of view it would be interesting to classify the carriers in a next step with respect to their attributes.

Regarding the treatment of the relationship between journey time and transfers, those models accounting for the total travel time and the number of transfers delivered better results. The other variant would be more intuitive for European flights, for which the waiting time represents a significant part of the overall journey time if a transfer is necessary. However, the number of chosen alternatives with a transfer is rather low in this sample. This leads to a better performance of models accounting for the number of transfers and not the actual waiting time.

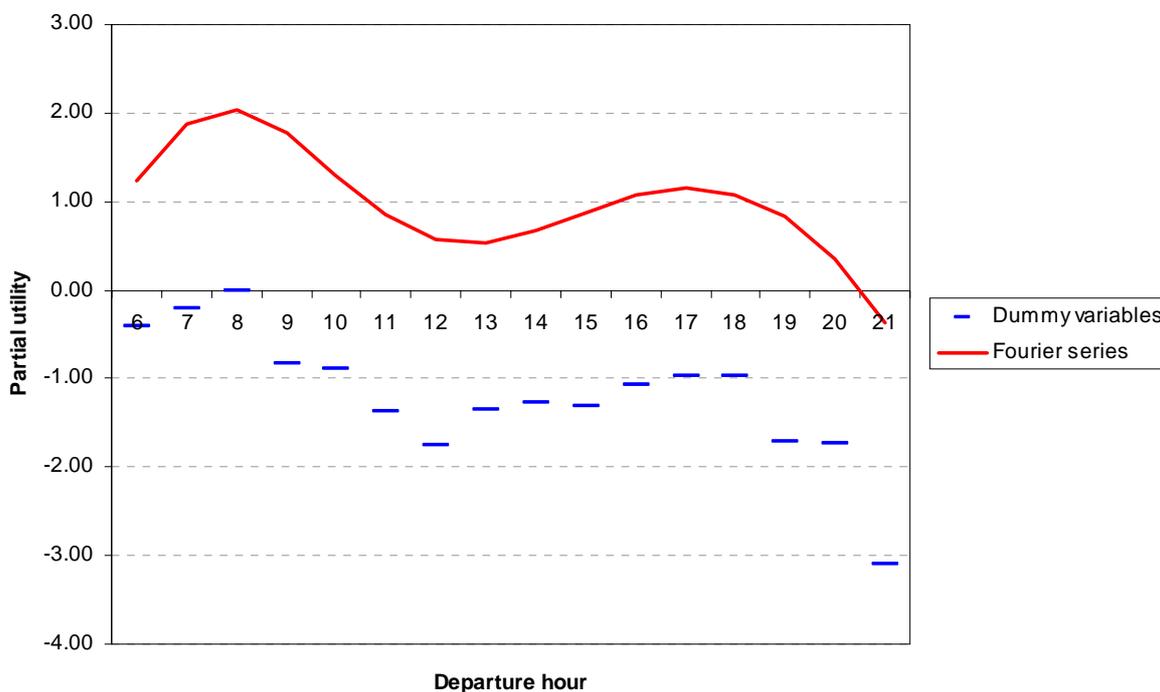
Adding fare to the models increased their explanatory power significantly, as did the inclusion of outbound departure hour, which show an expected preference structure. Passengers who opted for a stay of a fortnight do not show a clear preference for the departure period of their inbound flight. Departing in the morning is preferred by all types of passengers. However, the inclusion of the latter variables influenced the sign and value of carrier variables. Seen in the light of passenger preferences for a certain departure time and the high number of observations of passengers staying at their destination for a short period of time, this is reasonable.

As discussed in section 5.1, a second approach to model departure time preferences is experimented with. The results of this approach are shown on the estimated parameters for departure time of passengers staying 1 night at their destination is shown in Figure 6. It can be seen that the Fourier series approach yields the same results as the dummy variables. An advantage of the Fourier series approach is that a continuous approximation of departure time preferences is made. Also, no reference category is necessary and the number of parameters to be estimated is far less.

Table 2 Example estimated model

Attributes	Estimated parameter	Robust t-test
<i>Carrier attributes</i>		
Carrier 1 n	Not presented	
<i>Flight attributes</i>		
Non-code share	0.0000	
Code share	-0.9300	-12.7956
Total travel time out	-0.0158	-6.6836
Number of transfers	-4.1352	-11.0242
<i>Aircraft attributes</i>		
Mainline jet	0.0000	
Regional aircraft	-0.1199	-4.3033
Propeller aircraft	-1.6111	-15.3361
<i>Fare attribute</i>		
Fare	-0.0067	-79.9771
<i>Departure time out- duration of stay 0 days</i>		
6:00 - 6:59	-0.3489	-10.1552
7:00 - 7:59	0.2357	5.4939
8:00 - 8:59	0.0000	0.0000
9:00 - 9:59	-1.1259	-18.2996
10:00 - 10:59	-1.4389	-31.0608
11:00 - 11:59	-1.9421	-33.6937
12:00 - 12:59	-2.7223	-37.4082
13:00 - 13:59	-4.6270	-16.4480
14:00 - 14:59	-4.2553	-19.7668
15:00 - 15:59	-5.3732	-5.3017
<i>Departure time out- duration of stay 1-14 days</i>	Not presented	

Figure 6 Fourier series approach versus dummy variables for departure hour



In Table 3 the estimated parameters for several specifications of the fare parameter are shown. Passengers staying at their destination 2 weeks are more sensitive to fare as passengers staying at their destination one night or returning on the same day. Passengers staying at their destination one day are the least sensitive to fare. It is hypothesized, that this is because these passengers already have opted for an overnight stay.

Six booking periods are identified based on the number of forecasts that are made prior to departure of an itinerary. Booking period 0 is up to 3 days before departure, booking period 1 indicates between 4 and 7 days before departure, booking period 2 indicates between 7 and 14 days before departure, booking period 3 indicates between 14 and 21 days before departure, booking period 4 indicates between 21 and 28 days before departure and booking period 5 indicates longer as 28 days before departure. The estimated parameters for fare follow the anticipated preference structure: travellers booking further in advance are more sensitive to fare as travellers booking close to departure.

Model performance (Table 4) does not vary much per estimated model. Models including a further specification of fare perform slightly better as models that do not include a further specification of fare.

Table 3 Different specifications of fare attribute

	Base model		Fare parameter per stay category		Fare parameter per booking period	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Fare	-0.0067	-79.9771	-	-	-	-
Fare duration of stay 0 days	-	-	-0.0075	-72.6337	-	-
Fare duration of stay 1 day	-	-	-0.0054	-43.4048	-	-
Fare duration of stay 2 days	-	-	-0.0085	-9.4202	-	-
Fare booking period 0	-	-	-	-	-0.0063	-27.4071
Fare booking period 1	-	-	-	-	-0.0065	-43.4222
Fare booking period 2	-	-	-	-	-0.0065	-49.8324
Fare booking period 3	-	-	-	-	-0.0069	-33.0146
Fare booking period 4	-	-	-	-	-0.0074	-23.7068
Fare booking period 5	-	-	-	-	-0.0077	-27.0626

Table 4 Model performance

	Base model	Model with fare per stay category	Model with fare per booking period	Model with Fourier series approach
Number of estimated parameters	66	68	71	43
Number of observations	18416	18416	18416	18416
Number of individuals	18416	18416	18416	18416
Null log-likelihood	-68956.1	-68956.1	-68956.1	-68956.1
Init log-likelihood	-68956.1	-68956.1	-68956.1	-68956.1
Final log-likelihood	-46798.4	-46714.8	-46784.5	-46892.5
Likelihood ratio test	44315.6	44482.7	44343.3	44127.3
Rho-square	0.3213	0.3225	0.3215	0.3200
Adjusted rho-square	0.3204	0.3216	0.3205	0.3193

6. Conclusions and Outlook

In this paper, several relatively new and interesting areas application areas of discrete choice models in the aviation sector have been identified. The further incorporation of discrete choice models in revenue management systems is an example of this. More interesting may be however the willingness-to-pay of a traveller for certain itinerary characteristics. Potential application areas here include demand management of airports and real-time willingness-to-pay estimation of travellers through online booking channels.

For this purpose, an MNL model for itineraries in Europe is estimated. The model includes carrier attributes, aircraft attributes and a fare attribute and is based on revealed preference data, such as actually booked tickets through CRS systems and fares as observed on the web-based booking system Expedia.

Estimated models show that travellers have a different sensitivity for fare over time and per duration of stay. Passengers staying at their destination for only a short period of time prefer itineraries leaving in the morning and returning in the morning. Despite this is not being new, it is now possible to couple a monetary value to such parameters. A Fourier series approach of the modelling of departure time is a possible replacement of dummy variables for departure time and requires less parameters to be estimated. Different sensitivities for departure day are also imaginable and remain a part of future research.

The research presented here only forms a part of a larger study towards itinerary choice behaviour. Seen in the light of this study, further attention should be paid to overcoming the IIA-assumption of the MNL-model, either by more complex model structures or, preferably, by the incorporation of a similarity factor.

Acknowledgements

The authors would like to thank Karl Isler and Henrik Imhof (Swiss International Airlines) for providing the data set containing the ticket booking through CRS system and the webbots to observe the fares. Their contribution was central to the success of this study. In addition, we would like to thank Michel Bierlaire for providing Biogeme for the logit estimations. Also, we would like to thank Claude Weis and Alexander Erath for numerous discussions regarding model estimation, variables to incorporate and their help with Biogeme. Finally, we would like to thank Professor Bovy and Dr. Rob van Nes for their valuable input.

References

- Belobaba, P. (2006) *Airline Management*, MIT Open courseware, Massachusetts Institute of Technology, Boston.
- Ben-Akiva, M. and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Massachusetts.
- Bierlaire, M. (2003) *BIOGEME: a free package for the estimation of discrete choice models*, 3rd Swiss Transport Research Conference, Ascona.
- Bonzio, L. (1996) Modelle für den Zugang von Passagieren zu Flughäfen, *PhD Thesis*, Ruhr-Universität Bochum, Bochum.
- Bovy, P. H. L. (2007) *Modeling route choice sets in transportation networks: a preliminary synthesis*, Tristan V Conference, Phuket (Thailand).
- Bovy, P. H. L. and E. Stern (1990) *Route choice and way finding in transport networks*, Kluwer Academic Publishers, Dordrecht.
- Brueckner, J. K. (2002) Airport Congestion When Carriers Have Market Power, *The American Economic Review* **92** (5) 1357-1375.
- Carrier, E. (2006) *Modeling the Joint Choice of an Airline Itinerary and Fare Product: Implications for Airline Pricing Strategies*, 86th Annual Meeting of the Transportation Research Board, Washington D.C.
- Chorus, C. G., J. L. Walker and M. E. Ben-Akiva (2007) A model of travel information response: describing the full sequence of information acquisition and travel choice, *Working paper*.
- Coldren, G. and F. Koppelman (2005) Modeling the competition among air-travel itinerary shares: GEV model development, *Transportation Research A* **39** (4) 345-365.
- Coldren, G., F. Koppelman, K. Kasturirangan and A. Mukherjee (2003) Modeling aggregate air-travel itinerary shares: logit model development at a major US airline, *Journal of Air Transport Management* **9** (6) 361-369.
- de Neufville, R. and A. Odoni (2003) *Airport Systems: Planning, Design and Management*, McGraw-Hill, New York.
- Garrow, L. A., S. P. Jones and R. A. Parker (2007) How much airline customers are willing to pay: An analysis of price sensitivity in online distribution channels, *Journal of Revenue and Pricing Management* **5** (4) 271-290.
- Garrow, L. A. and F. S. Koppelman (2004a) Multinomial and nested logit models of airline passengers' no-show and standby behaviour, *Journal of Revenue and Pricing Management* **3** 237-253.
- Garrow, L. A. and F. S. Koppelman (2004b) Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information, *Journal of Air Transport Management* **10** (6) 401-411.
- Grammig, J., R. Hujer and M. Scheidler (2005) Discrete choice modelling in airline network management, *Journal of Applied Econometrics* **20** (4) 467-486.

- Hess, S. and J. W. Polak (2006) 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.
- Koppelman, F., G. Coldren and R. A. Parker (2007) Schedule Delay Impacts on Air-Travel Itinerary Demand, *Submitted January, 2007*.
- Lohatepanont, M. and C. Barnhart (2004) Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment, *Transportation Science* **38** (1) 19-32.
- Parker, R. A. (2007) *What Passengers Will Pay To Fly: A Review of Passenger Choice Modeling at Boeing*, 11th World Congress of Transportation Research (WCTR), Berkeley, California.
- Pels, E., P. Nijkamp and P. Rietveld (2001) Airport and Airline Choice in a Multiple Airport Region: An Empirical Analysis for the San Francisco Bay Area, *Regional Studies* **35** (1) 1 - 9.
- Pels, E., P. Nijkamp and P. Rietveld (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. and E. T. Verhoef (2004) The economics of airport congestion pricing, *Journal of Urban Economics* **55** (2) 257-277.
- Talluri, K. and G. van Ryzin (2004) Revenue Management Under a General Discrete Choice Model of Consumer Behavior, *Management Science* **50** (1) 15-33.
- Talluri, K. T. and G. J. van Ryzin (2005) *The Theory and Practice of Revenue Management*, Springer, New York.
- Zeid, M. A., T. F. Rossi and B. Gardner (2006) *Modeling time Of Day Choice in the Context of Tour and Activity Based Models*, 85th Annual Meeting of the Transportation Research Board, Washington D.C.