The research put forward in this paper was carried out during a stay at the IVT (Institute for Transport Planning and Systems) of the ETH Zurich.
USING DISCRETE CHOICE MODELS TO SUPPORT STRATEGIC DECISION-MAKING OF AIR TRANSPORTATION SERVICE PROVIDERS.

ABSTRACT

The focus of this paper is on understanding the relative valuation of non-monetary and monetary characteristics of an itinerary based on revealed preference data. A discrete choice modeling approach is followed.

The research presented makes use of a number of datasets: 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 (OAG). They have been combined to form a comprehensive dataset for the analysis of itinerary choice.

From this research it is concluded that the contribution to utility of fare is large: in a direct itinerary fare yields the largest contribution to utility, together with departure time. In an average non-chosen itinerary, the largest contribution to utility is given by a transfer.

1 INTRODUCTION

In recent years air traveler behavior research has seen a strong increase in activity, stimulated by the aviation industry who perceive an increased need for an understanding of traveler behavior on a micro-level. Such an understanding can aid airlines with decisions involving network and schedule planning, pricing strategies and improvement of revenue management controls. Furthermore, an increased understanding of traveler needs can aid online travel agents with their competitive advantage over carrier websites by offering a complete, tailor-made overview of available itineraries. Finally, an increased understanding of traveler behavior can aid airports by providing insight in access mode and airport choice.
Studies on traveler behavior address choices over the entire decision-making spectrum, starting with the choice for air as mode of traveler (González-Savignat 2004), access mode and airport (Bondzio 1996, Tron, et al. 2007) and access mode, airport and airline (Hess & Polak 2006b, Hess & Polak 2006a, Hess, et al. 2007) and itinerary choice (Coldren, et al. 2003, Coldren & Koppelman 2005, Grammig, et al. 2005, Theis, et al. 2006). The results of the latter studies address the needs of airlines with regard to network planning and scheduling, as they reveal the relative valuation of service characteristics by travelers.

The increasing competition between flag and low-cost carriers, simplified fare structures and the widespread availability of fare information through on-line distribution channels has led to an increased need in the understanding of the role of fare and fare product in decision-making, in addition to the understanding of service characteristics provided by the earlier mentioned studies. Practitioners point out the necessity of joint pricing and revenue management (Ratliff & Vinod 2005) and real-time, dynamic fares, where dynamic is defined as a fare influenced by seat availability, expectation of competing demand, prices of competitors and alternatives for the consumer (Westermann 2006).

Research has considered the choice of fare product (Proussaloglou & Koppelman 1999, Carrier 2006), the willingness-to-pay of passengers (Theis, et al. 2006, Garrow, et al. 2007) and the incorporation of fare product choice in revenue management (RM) systems (Talluri & van Ryzin 2004). Both revealed preference (RP) and stated preference data (SP) has been used for studies concerning itinerary choice modeling. Earlier mentioned studies focusing on the willingness-to-pay of passengers for service characteristics have used a stated-preference data.

The majority of the studies concerning willingness-to-pay have used SP data, having the advantage that the exact information (e.g. alternatives, attributes) presented to the respondent is known. While offering these advantages, SP data represents choices made in a hypothetical context and not, as is the case with RP data, choices made in real-life situations. However, in a
typical revealed preferences dataset, cross sectional dataset, the challenge for the researcher is to
determine which alternatives are available to an individual (Ortuzar & Willumsen 2001).
The focus of this study is on understanding the relative valuation of non-monetary and monetary
characteristics of an itinerary based on revealed preference data, contributing to current literature
by using revealed preference data and the incorporation of fare. The fact that actual behavior is
analyzed can prove both important for convincing both practitioners and researchers of the
findings. In addition, special attention will be paid to the incorporation of information a passenger
presents at the moment of requesting an itinerary, such as duration of stay and booking period,
thereby specifically addressing the possibilities of dynamic pricing and listing.
The research presented in this paper makes use of a number of datasets: 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 (OAG). They have been
combined to form a comprehensive dataset for the analysis of itinerary choice.

2 CONCEPTUAL FRAMEWORK

2.1 Traveler decision-making and traveler segmentation

Prior to making an itinerary choice, a traveler makes a series of decisions. Common in
transportation modeling are the following choices: destination choice, mode choice, departure
time choice and route choice. In airline choice modeling, several more choices can be recognized,
such as origin airport choice, access mode choice and egress mode choice. The outcome of all
these choices is dependent on the characteristics of the journey at hand (supply) the
characteristics of the traveler or decision-maker (demand) and the perceived utility of the activity
at the destination end.
The motivation of a traveler to undertake the trip stems from the fact that the perceived utility at
the activity end of the trip (e.g. leisure, business) minus the disutility of the trip, resulting in the
net-utility, is higher than the net-utility of not taking part in the activity or the net-utility of other
possible destinations and corresponding activities.

If we focus on the characteristics of the traveler, several characteristics can be observed at the
moment of booking, with proxy variables such as departure day and duration of stay. Other
factors influencing the decision of the traveler can only be influenced by third parties, such as
network knowledge (is there an airport at my preferred destination?) and trip purpose.

Journey characteristics on the other hand, consist of the characteristics of the access mode (e.g.
fuel costs, public transport costs), origin airport (e.g. security, parking costs, comfort), itinerary
(e.g. fuel surcharge, fare, transfers) and egress mode.

Without carrying into too much detail, it is important to recognize that airlines, airports, travel
agents and other stakeholder posses instruments to influence these decisions, such as departure
time, network knowledge and fare setting.

2.2 Discrete Choice Models

To analyze air traveler behavior, discrete choice models are applied, a widely accepted modeling
approach in the transport modeling community (e.g. Ben-Akiva & Lerman 1985, Train 2003).

The underlying hypothesis is that the decision-maker is rational and has full information on all
available alternatives.

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 $U_{iq}$, or value, of an alternative $i$ to
the traveler or decision-maker $q$ is defined by:

$$U_{iq} = V_{iq} + \epsilon_{iq} = f(\beta, x_{iq}) + \epsilon_{iq}$$
with a deterministic part \( V_q \) that consists of a function \( f(\beta, x_q) \) of the vector \( \beta \) of taste parameters and the vector \( x_q \) 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 \( \varepsilon_q \), the error or stochastic term. This error term captures unobserved alternative attributes, unobserved individual characteristics, measurement errors and proxy variables (Ben-Akiva & Bierlaire 1999).

The most commonly used discrete choice model is the Multinomial Logit (MNL) Model due to its ease of estimation and simple mathematical structure (McFadden 1974). It is based on the assumption that the random terms, often called error terms or disturbances, are identically and independently (i.i.d.) Gumbel distributed. The choice probability of each alternative \( i \) can be calculated as:

\[
P(i \mid C_q) = \frac{e^{V_i}}{\sum_j e^{V_j}}
\]

Despite known deficiencies of the MNL-model, such as the Independence of Irrelevant Alternatives (IIA) property, it is chosen to use this model as a starting point for this research, as it can give valuable insight in the perception and the relative valuation of itinerary service characteristics and the magnitude of the willingness-to-pay of air travelers. Other research is being conducted in the field of the IIA property and accounting for similarity between alternatives in general and itineraries specific by the authors.

2.3 Choice Set Formation and Terminology

Each choice is made from a set of alternatives or the choice set. The environment of the decision-maker determines the composition of the choice set. Several approaches are mentioned in
literature to determine the choice set which contains the alternatives which were available to the
decision maker (e.g. Manski 1977, Bovy & Stern 1990, Swait 2001).

If the theoretical and behavioral concepts of choice set formation in the case of route choice (e.g. Bovy & Stern 1990, Hoogendoorn-Lanser 2005) would be translated to itinerary choice several decision dimensions influencing the composition of the choice set can be recognized. Without getting into detail, it is important to realize that a traveler can inform himself multiple times and on websites of carriers and/or travel portals. Furthermore, a traveler will filter itineraries based on arrival time and itinerary fare.

In addition, attention should be paid to choice set notions from a traveler’s perspective and an analyst’s perspective. An analyst is not aware of the itineraries considered by the traveler, and can only approximate the composition of the traveler’s choice set. An in-depth discussion of these choice set notions is given by van Eggermond (2007). The necessity to make this distinction stems from the fact that the model estimates are dependent on the composition of the choice set.

3 DATA

3.1 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;
- Official Airline Guide.

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 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 Worldwide Limited 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.

Itinerary 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 shares. In the end, nearly 19 000 choices with fare information are observed. An extensive discussion of the matching of the datasets can be found in van Eggermond (2007), extracted attributes and attribute levels are presented in
Table 1.
Table 1: Description of attributes and attribute levels

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level and definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier</td>
<td>First carrier listed in itinerary</td>
</tr>
<tr>
<td>Propeller aircraft</td>
<td>Dummy variable indicating if any leg of the outbound itinerary is operated by a propeller aircraft.</td>
</tr>
<tr>
<td>Regional jet</td>
<td>Dummy variable indicating if any leg of the outbound itinerary is operated by a regional jet.</td>
</tr>
<tr>
<td>Mainline jet</td>
<td>Dummy variable indicating if any leg of the outbound itinerary is operated by a mainline jet.</td>
</tr>
<tr>
<td>Code-share</td>
<td>Dummy variable indicating if any leg of the outbound itinerary contains a code share.</td>
</tr>
<tr>
<td>Departure hour outbound itinerary</td>
<td>Departure hour of the outbound itinerary (local time)</td>
</tr>
<tr>
<td>Departure hour inbound itinerary</td>
<td>Departure hour of the inbound itinerary (local time)</td>
</tr>
<tr>
<td>Fare</td>
<td>Fare of the itinerary as listed on Expedia on the booking day for the same duration of stay and same departure day.</td>
</tr>
<tr>
<td>Transfer</td>
<td>Indicates if the itinerary contains a transfer</td>
</tr>
<tr>
<td>Total travel time</td>
<td>Sum of in vehicle time and transfer time in minutes</td>
</tr>
</tbody>
</table>

3.2 Relation between traveler segmentation and datasets

It was previously argued that trip purpose and traveler type influenced the outcome of the itinerary choice. However, no information is known on trip purpose and traveler type in these datasets in general and at the moment of booking in particular. However, information is available on duration of stay, booking time and departure day. With these variables, it is possible to proxy the type of traveler and the trip purpose by using a segmentation that includes these variables.

4 Empirical Results

4.1 Data analysis

In this section, the most important results of the data analysis will be presented. First, general remarks with regard to the datasets will be made. Second, several aspects of the data analysis will be highlighted.

Two main deficiencies of the bookings in the MIDT data can be recognized. First, the number of chosen itineraries containing a transfer is low. Only 0.5% of the travelers opt for an itinerary for a transfer, opposed to a transfer being offered in 42% of the cases. In the context of the observed
durations of stay, this is not strange: passengers returning the same or next day will prefer not to transfer, as a transfer would take up much of their valuable time at their destination. This is also reflected in the duration of stay in minutes of the passengers and is shown in **Error! Reference source not found.**. It can be seen that passengers opt for a duration of stay between the 300 minutes and 720 minutes, whereas the offered itineraries have shorter duration of stay on average.

With regard to the duration of stay, 55.77% of the passengers return the same day, 38.63% of the passengers return the next day and 5.60% returns after 6 days. Travelers returning the same day depart on weekdays, traveler returning the next day depart from Monday to Thursday. Travelers returning after 6 days do not show a clear preference for departure day.

![Cumulative percentage of duration of stay](image)

\[N_{obs} = 10,537, N_{non-chosen} = 546,939\]

**Figure 1**: Difference arrival time outbound itinerary–departure time inbound itineraries returning the same day
The second main deficiency is that the number of low cost carriers in both the MIDT and Expedia dataset is low. For instance, in the US, Southwest Airlines fares cannot be obtained from Expedia or MIDT. The same holds for some low cost carriers in Europe. 4.0% of the booked itineraries contain a flight operated by a low cost carrier, opposed to a low cost carrier being offered in 3.8% of the itineraries. The remaining itineraries either contain a regional carrier (13.8% of the cases) or a flag carrier (82.2% of the cases). Omitting itineraries from low-cost carriers could limit the number of alternatives. It is recommended for follow-up research to look into possibilities to overcome these deficiencies, such as compiling data sets from different sources. However, the authors believe the dataset used this paper can serve as a starting point for further research towards the usage of revealed preference data in the case of itinerary choice.

5% of the travelers book their ticket up to 36 days in advance, 50% of the tickets are booked 8 days before departure and 85% of the tickets are booked up to 3 days in advance.

Combined with the fact that a low number of departures in the weekend are observed and the MIDT dataset is for November 2006, it is thought that the dataset contains a large number of business travelers, thus offering the opportunity to analyze the choice behavior of a less diffuse group of travelers.

It should be noted, that the MIDT dataset contains a large number of bookings for longer than 30 days. This leads to believe that travelers exploit irrationalities in revenue management systems, where a combination of two return tickets may be cheaper than two one-way tickets.

A further distinction between itineraries is their departure time. In this study, itineraries are aggregated by hour and per stay category, i.e. 5:00 – 5:59, 6:00 – 6:59. Most chosen itineraries returning on the same day depart in the period 6:00 – 9:00 and return between 16:00 and 22:00. Most chosen itineraries returning on the next day depart in the period 6:00 – 9:00. A second peak
can be observed during observed in the period 16:00 – 19:00. Again, the non-chosen itineraries are distributed somewhat more evenly than the chosen itineraries.

Passengers staying at their destination longer as six days do not show clear preference at first sight for the departure time of the outbound itinerary. However, a morning peak can be observed for the outbound itineraries and a peak in the late afternoon and early afternoon can be observed for the inbound itineraries.

It was argued in the previous section that it is necessary to make a distinction between different types of choice sets. A choice set containing all itineraries departing on the same day, available on the day of booking and for the same duration of stay is coined the objective choice set. By adding constraints to the choice set generation algorithm, such as a time window around the arrival time or the number of airlines in the choice set, the contents of the subjective choice set can be approximated. It should be noted though, that as no information on the passengers is known, the constraints remain fairly arbitrary.
It can be seen that the choice set size steadily increases if the window is enlarged. If the 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 departing on the same day, a choice set can contain up to 150 flights, 60% of the choice sets will contain 50 alternatives. A jump can be observed in the choice set size of latter category. This because in some cases, Expedia returns more itineraries than the 50 it usually does.
An analysis of the fare in the choice sets is presented in Error! Reference source not found. From each choice set, a subset containing the fares lower than the chosen fare and a subset containing the fares higher than the chosen fare is extracted. Each chosen fare is added to a bin, which holds the chosen fares of a certain interval (e.g. €100 - €109). This offers the opportunity to analyze the composition of fare in the choice set and the role of a time window.

It can be seen that the average fare in the lower subset indeed equals the chosen fare in the case of the lower chosen fares. As the chosen fare increases however, it can be seen that lower fares are available. Even if the window is limited, chosen fares do not equal the average of the lowest fare. Knowing this, it can be said that fare is not always the decisive criterion for an individual and that other service attributes play a role in decision-making.

Figure 3: Fare in choice set dependent on time window

\(N_{\text{chosen}} = 18\ 895, N_{\text{non-chosen 1 hour}} = 184\ 252 \text{ and } N_{\text{non-chosen same day}} = 968\ 352\)

An analysis of the fare in the choice sets is presented in Error! Reference source not found.. From each choice set, a subset containing the fares lower than the chosen fare and a subset containing the fares higher than the chosen fare is extracted. Each chosen fare is added to a bin, which holds the chosen fares of a certain interval (e.g. €100 - €109). This offers the opportunity to analyze the composition of fare in the choice set and the role of a time window.

It can be seen that the average fare in the lower subset indeed equals the chosen fare in the case of the lower chosen fares. As the chosen fare increases however, it can be seen that lower fares are available. Even if the window is limited, chosen fares do not equal the average of the lowest fare. Knowing this, it can be said that fare is not always the decisive criterion for an individual and that other service attributes play a role in decision-making.
4.2 Modeling results

The definitive model presented in Table 2 contains carrier constants, a dummy variable representing if the itinerary contains a code-share, the total travel time, a variable representing a transfer, variables representing the type of aircraft. Furthermore, departure hour variables and a fare variable are included per stay category. With this approach, an explicit choice is made for a segmentation of passengers only with regard to fare and departure time preferences. Model estimation is carried out with BIOGEME (Bierlaire 2003).

The inclusion of fare and departure time of outbound and inbound itinerary led to a significant increase of explanatory power, starting from an adjusted rho-square of 0.20 for models excluding fare, increasing to 0.25 for models with fare and 0.33 for the model including outbound and inbound departure time. In addition, models including a Fourier series approach representing departure time were estimated (e.g. Grammig, et al. 2005, Zeid, et al. 2006, Koppelman, et al. 2007). Despite a lower number of parameters to be estimated in the case of a Fourier series, it was chosen to continue with dummy variables representing departure time, as a continuous series of observations is necessary for the Fourier series approach. Finally, models with different fare parameters were estimated, revealing different sensitivities for passengers departing on weekdays and weekends and a higher sensitivity of fare for passengers booking long before departure, as compared to passengers booking close to departure.

Two remarks should be made with regard to the terminology of the estimated parameters:

- The parameter estimates for carrier attributes, flight attributes and aircraft attributes are generic parameter estimates, assuming that the taste parameters for these attributes are similar across all segments;

- The parameter estimates for departure time and fare are segment specific attributes. The parameter estimates can be compared within a segment and between segments. For instance, passengers returning the same day perceive departing between 12:00-12:59 1.65
times worse than passengers returning the next day perceive departing in the same period, as the ratio of these parameter estimates is -2.76/-1.68.

Table 2: Parameter estimates of the best MNL-model (\(N_{\text{obs}} = 18,416\) and \(N_{\text{cases}} = 800,897\))

<table>
<thead>
<tr>
<th></th>
<th>Estimated parameter</th>
<th>Robust t-test</th>
<th>Estimated parameter</th>
<th>Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Carrier constants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Not presented -</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flight attributes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-code share</td>
<td>0.0000</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code share</td>
<td>-0.9215</td>
<td>-12.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total travel time out</td>
<td>-0.0116</td>
<td>-5.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td>-4.6511</td>
<td>-12.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aircraft attribute</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainline jet</td>
<td>0</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional aircraft</td>
<td>-0.1530</td>
<td>-5.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propeller aircraft</td>
<td>-1.5518</td>
<td>-14.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Departure times – return same day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Outbound</strong></td>
<td><strong>Inbound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:00 - 6:59</td>
<td>-0.3543</td>
<td>-10.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7:00 - 7:59</td>
<td>0.3012</td>
<td>6.99</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8:00 - 8:59</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9:00 - 9:59</td>
<td>-1.0473</td>
<td>-17.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10:00 - 10:59</td>
<td>-1.4841</td>
<td>-31.06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11:00 - 11:59</td>
<td>-2.0104</td>
<td>-34.11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12:00 - 12:59</td>
<td>-2.7596</td>
<td>-37.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13:00 - 13:59</td>
<td>-4.8408</td>
<td>-17.27</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14:00 - 14:59</td>
<td>-4.4808</td>
<td>-20.74</td>
<td>-1.6323</td>
<td>-5.08</td>
</tr>
<tr>
<td>15:00 - 15:59</td>
<td>-5.5495</td>
<td>-5.45</td>
<td>-0.7659</td>
<td>-2.99</td>
</tr>
<tr>
<td>16:00 - 16:59</td>
<td>0</td>
<td>-</td>
<td>0.5991</td>
<td>14.42</td>
</tr>
<tr>
<td>17:00 - 17:59</td>
<td>-</td>
<td>-</td>
<td>0.8246</td>
<td>19.54</td>
</tr>
<tr>
<td>18:00 - 18:59</td>
<td>-</td>
<td>-</td>
<td>0.8816</td>
<td>22.03</td>
</tr>
<tr>
<td>19:00 - 19:59</td>
<td>-</td>
<td>-</td>
<td>0.3325</td>
<td>7.55</td>
</tr>
<tr>
<td>20:00 - 20:59</td>
<td>-</td>
<td>-</td>
<td>0.4206</td>
<td>3.09</td>
</tr>
<tr>
<td>21:00 - 21:59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22:00 - 22:59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Departure time – return next day</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Outbound</strong></td>
<td><strong>Inbound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:00 - 6:59</td>
<td>-0.3794</td>
<td>-6.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7:00 - 7:59</td>
<td>-0.2656</td>
<td>-3.97</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8:00 - 8:59</td>
<td>0</td>
<td>-</td>
<td>-1.2016</td>
<td>-5.34</td>
</tr>
<tr>
<td>9:00 - 9:59</td>
<td>-0.8122</td>
<td>-9.64</td>
<td>-0.7210</td>
<td>-2.29</td>
</tr>
<tr>
<td>10:00 - 10:59</td>
<td>-0.8318</td>
<td>-11.89</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11:00 - 11:59</td>
<td>-1.3406</td>
<td>-16.28</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12:00 - 12:59</td>
<td>-1.6706</td>
<td>-19.41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13:00 - 13:59</td>
<td>-1.3736</td>
<td>-16.25</td>
<td>0.5032</td>
<td>1.73*</td>
</tr>
<tr>
<td></td>
<td>Estimated parameter</td>
<td>Robust t-test</td>
<td>Estimated parameter</td>
<td>Robust t-test</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>14:00 - 14:59</td>
<td>-1.1646</td>
<td>-14.88</td>
<td>-0.0533</td>
<td>0.29**</td>
</tr>
<tr>
<td>15:00 - 15:59</td>
<td>-1.1883</td>
<td>-12.13</td>
<td>0.2588</td>
<td>1.27**</td>
</tr>
<tr>
<td>16:00 - 16:59</td>
<td>-1.0239</td>
<td>-14.26</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>17:00 - 17:59</td>
<td>-0.9566</td>
<td>-14.63</td>
<td>0.5907</td>
<td>11.21</td>
</tr>
<tr>
<td>18:00 - 18:59</td>
<td>-0.9299</td>
<td>-12.82</td>
<td>0.7830</td>
<td>14.02</td>
</tr>
<tr>
<td>19:00 - 19:59</td>
<td>-1.6873</td>
<td>-19.47</td>
<td>1.0273</td>
<td>20.76</td>
</tr>
<tr>
<td>20:00 - 20:59</td>
<td>-1.6964</td>
<td>-16.29</td>
<td>0.5686</td>
<td>9.96</td>
</tr>
<tr>
<td>21:00 - 21:59</td>
<td>-2.9532</td>
<td>-11.47</td>
<td>0.4250</td>
<td>3.02</td>
</tr>
</tbody>
</table>

**Departure time - return after 6 days**

**Outbound**

<table>
<thead>
<tr>
<th></th>
<th>Estimated parameter</th>
<th>Robust t-test</th>
<th>Estimated parameter</th>
<th>Robust t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:00 - 6:59</td>
<td>-0.9844</td>
<td>-2.11</td>
<td>0.3294</td>
<td>0.90**</td>
</tr>
<tr>
<td>7:00 - 7:59</td>
<td>-0.3082</td>
<td>0.72**</td>
<td>0.8557</td>
<td>3.34</td>
</tr>
<tr>
<td>8:00 - 8:59</td>
<td>0</td>
<td>-</td>
<td>-0.2100</td>
<td>0.58**</td>
</tr>
<tr>
<td>9:00 - 9:59</td>
<td>-1.2215</td>
<td>-2.63</td>
<td>0.8315</td>
<td>2.97</td>
</tr>
<tr>
<td>10:00 - 10:59</td>
<td>-0.1765</td>
<td>0.41**</td>
<td>1.0259</td>
<td>4.02</td>
</tr>
<tr>
<td>11:00 - 11:59</td>
<td>-0.1006</td>
<td>0.23**</td>
<td>1.2060</td>
<td>3.78</td>
</tr>
<tr>
<td>12:00 - 12:59</td>
<td>-0.7790</td>
<td>-1.74*</td>
<td>0.7797</td>
<td>3.44</td>
</tr>
<tr>
<td>13:00 - 13:59</td>
<td>-1.5377</td>
<td>-3.39</td>
<td>1.2317</td>
<td>4.49</td>
</tr>
<tr>
<td>14:00 - 14:59</td>
<td>-0.8276</td>
<td>-1.74*</td>
<td>1.2646</td>
<td>4.62</td>
</tr>
<tr>
<td>15:00 - 15:59</td>
<td>-0.7482</td>
<td>-1.34*</td>
<td>1.5728</td>
<td>6.05</td>
</tr>
<tr>
<td>16:00 - 16:59</td>
<td>-0.6851</td>
<td>-1.46*</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>17:00 - 17:59</td>
<td>-0.5866</td>
<td>1.37**</td>
<td>1.0198</td>
<td>4.75</td>
</tr>
<tr>
<td>18:00 - 18:59</td>
<td>-0.4758</td>
<td>1.08**</td>
<td>1.6875</td>
<td>6.57</td>
</tr>
<tr>
<td>19:00 - 19:59</td>
<td>-1.7218</td>
<td>-3.11</td>
<td>1.1499</td>
<td>4.26</td>
</tr>
<tr>
<td>20:00 - 20:59</td>
<td>-1.2242</td>
<td>-2.39</td>
<td>1.2464</td>
<td>4.17</td>
</tr>
<tr>
<td>21:00 - 21:59</td>
<td>-2.5105</td>
<td>-2.33</td>
<td>0.1766</td>
<td>0.22**</td>
</tr>
</tbody>
</table>

**Fare**

- Fare return same day: -0.0077, -68.50
- Fare return next day: -0.0056, -41.96
- Fare return after 6 days: -0.0083, -8.83

- Number of estimated parameters: 100
- Number of observations: 18416
- Init log-likelihood: -69032.6
- Final log-likelihood: -46021.7
- Likelihood ratio test: 46021.9
- Rho-square: 0.333
- Adjusted rho-square: 0.332

* significant at a 85% level
** insignificant at a 85% level
5 FINDINGS

5.1 General findings

Almost all estimated parameters are significant at a 95% confidence level (t-value > 1.95). Among the non-significant parameters are variables representing carrier and departure time for travelers returning after six days.

The sign and magnitude of parameter estimates is in line with expected traveler behavior. Parameter estimates for fare, travel time and transferring are negative. Also, code-share itineraries have a lower value than non-code share itineraries and mainline jets are preferred above regional jets and propeller aircraft, as also found by Coldren et al. (2003).

5.2 Valuation of time and transfer

With the parameter estimates it is possible to estimate several ratios, such as the value of time of a traveler and the value of a transfer. As these parameter estimates are estimated with the MNL-model, the values presented in Error! Reference source not found. are obtained by simply dividing a parameter estimate by another parameter estimate. The estimate value of time and the monetary value of a transfer vary per duration of stay, as a separate fare parameter is estimated per duration of stay. The estimated value of a transfer expressed in minutes remains constant, as only a single parameter is estimated for both duration of stay and a transfer.
The estimated value of time varies from 84 €/h to 125 €/h, whereas the estimated value of a transfer varies from €559 to €832. A transfer is valued at 400 minutes. These estimates are in the expected range for air travelers.

Table 3: Estimated value of time and transfers per duration of stay

<table>
<thead>
<tr>
<th></th>
<th>0 days</th>
<th>1 day</th>
<th>&gt; 6 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of time [€/h]</td>
<td>90.50</td>
<td>125.09</td>
<td>84.02</td>
</tr>
<tr>
<td>Transfer [€]</td>
<td>602.56</td>
<td>832.92</td>
<td>559.41</td>
</tr>
<tr>
<td>Transfer [min]</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
</tbody>
</table>

5.3 Airline constants

The definitive MNL-model contained dummy variables representing carrier opposed to dummy variables representing type of carrier. Carrier preferences are more complex than a simple variable: parameter estimates vary strongly per carrier for carrier type. The relative valuation of carrier remains fairly small, compared to other service characteristics. This is in line with research carried out by Garrow et al. (2007). In their case, carrier variables yielded insignificant parameter estimates.

5.4 Fare

In an average chosen itinerary, fare yields the largest contribution to utility. Over 50% of the utility is contributed by fare. In itineraries containing a transfer the relative contribution of fare drops to approximately 20%. It should be noted however, that departure time also contributes significantly to utility.

Passengers returning the same day or after six days reveal a similar preference to fare. Passengers returning the next day perceive fare as less important.

It is hypothesized, that passengers returning the next day are less sensitive as fare only makes up a part of the total costs, which include an overnight stay. A second explanation could be that the fare differences for itineraries returning the same day are larger, as compared to itineraries...
returning the next day. Finally, an explanation could be found in the used segmentation: it is thought, that in stay category 2 passengers are included, who stay shorter at their destination than 6 weeks.

5.5 Departure time

In Figure 4(top) the estimated parameters for outbound and inbound departure hour dummy variables are shown. The reference for the outbound departure time is 8:00, as reference for the inbound departure time 16:00 is chosen.

The estimated values for the outbound and inbound departure time dummy variables for passengers returning the same day are all significant at the 95% level. Estimated parameters for the inbound departure time are all significant at the 85% level.

It can be seen that travelers prefer departing at 7:00, as compared to 8:00. Departing earlier is perceived as negative. Departing later is perceived more negative hour. For instance departing at 9:00 is perceived 3 times as negative as departing at 6:00. Departing at 15:00 is considered to be as negative as a transfer. Returning after 16:00 is preferred, whereas returning before 16:00 is perceived as negative.

In Figure 4(bottom) the estimated parameters for the outbound and inbound departure time dummy variables for passengers opting for an overnight stay are presented. The estimated parameters for the outbound and inbound departure time dummy variables are all significant at the 95% level. Estimated parameters (except one) for the inbound departure time are significant at the 95% level.

Passengers returning the next day prefer departing at 8:00. Departing after 19:00 is perceived as very negative. More or less indifference can be observed between 14:00 and 18:00, the difference in utility is -0.25, which is valued at €37, and is somewhat less as 10% of the average chosen ticket price.
Passengers prefer returning after 16:00, an exception being departing between 13:00 and 13:59 and perceive returning in the early morning as negative. The preference for 13:00 is thought to be because passengers still can use the morning for other activities, such as meetings.

Important to notice is that a clear difference exists in preference structure and a difference in the magnitude of the preferences. Passengers returning the same day perceive outbound departure time more important than passengers returning the next day, indicating the importance of segmenting passengers.
Figure 4 (top) Departure time preferences return the same day & (bottom) Departure time preference return next day.
5.6 Transfer

The definitive MNL-model of itinerary choice includes a dummy variable for a transfer and the total travel time, since these attributes yielded better results than a model with in-vehicle time and waiting time.

In addition, recent research suggests that waiting time is more complex than assumed up to now. Theis et al. (2006) conducted a stated preference experiment and showed that passengers actually have an aversion against short waiting times. It is hypothesized that this because passengers are afraid to miss their transfer flight or encounter luggage problems. The number of transfers observed in the chosen itineraries is too low to allow for such a differentiation.

Passengers are willing to pay € 685,- to avoid a transfer and values a transfer at 400 minutes. The question arises if this is a high figure, taking into consideration the used choice sets, which contain choices of traveler returning the same or next day. It depends: due to low number of transfer flights in the chosen flights, it can be argued that these are actually not considered by the traveler, as a transfer will take up most of the time.

6 Implications and Conclusions

The research put forward in this paper has addressed a number of topics, most notably the incorporation of fare using revealed preference data and segmenting passengers according to their duration of stay.

It is concluded that it is possible to compile a dataset suitable for logit model estimation from several data sources, in order to gain insight in traveler choice behavior on a micro-level. Used data sources include MIDT data, data collected from Expedia in the period September - November 2006. At the moment, generated objective choice sets are used for MNL-model estimation. These choice sets include itineraries available on the day of booking for the same departure day and the same duration of stay. The validity of this approach is confirmed by the
parameter estimates, which are in line with expected traveler behavior and current state-of-the-art itinerary choice research.

Travelers are willing-to-pay a premium for itineraries departing in the early morning and returning in the early evening. Yet, revenue management systems remain important, as does the estimation of the willingness-to-pay of a passenger. A simplified one-way fare structure, however, may be the best direction for the future, with a differentiation of fares per weekday, departure time and booking period. It is thought by the authors that closing low-fare high restriction fare classes leads to bookings of itineraries where the low-fares are available.

Furthermore, a mismatch between presented supply, itineraries, and observed demand is present on current website offering travel information. For instance, itineraries are offered returning almost instantly. In addition, transfer itineraries are offered but not booked. Focusing on better quality of the results instead of quantity may make booking through travel portals more attractive.

For medium term planning, scheduling is of interest. It is shown that is important to offer itineraries on the right time, arguably more important than price. Passengers choose for itineraries departing in the morning and are willing to pay a premium for these flights.

Airlines not focusing on business passengers but on leisure traffic can avoid airports during congested moments: passengers staying longer at their destination have a low preference for departure time, judged by the small and sometimes insignificant parameter estimates.

The results presented advocate strongly against hub-and-spoke systems, at least on intra European flights. Again, this can be due to the used data: most bookings used in the estimation process were on specific city-pairs and for passengers returning the same or next day. This group does however represent 60% of the bookings in Europe. Also, if it is economical to operate smaller aircraft than mainline jets, it is better to operate regional jets than propeller aircraft. With this, a general trend in aviation is followed which predicts an increase usage of regional jets due to their economic characteristics. Finally, based on the high penalty for a transfer, it is recommended to market direct connections, opposed to marketing the airline itself.
ACKNOWLEDGEMENTS

The research put forward in this paper was carried out during a stay by the author at the IVT, ETH Zürich. 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 Nadine Schüssler 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


