Modeling the Choice of an Airline Itinerary and Fare Product Using Booking and Seat Availability Data

Emmanuel Carrier

ecarrier@mit.edu

Department of Civil and Environmental Engineering

Massachusetts Institute of Technology

77 Massachusetts Avenue, Cambridge, MA 02139

June 9, 2008

Abstract:

In this research, we develop a methodology to analyze the choice of an airline itinerary and fare product based on booking data. Since non-booked travel alternatives are not recorded in airline bookings, booking data was combined with fare rules and seat availability data to incorporate the impact of airline pricing and revenue management and reconstitute the choice set of each booking. In addition, characteristics of the traveler and the trip were retrieved from the booking records and replaced trip purpose that is traditionally used to segment demand in airline markets. They were included as explanatory variables of a latent class choice model in which several factors can be used simultaneously to segment the market without necessarily dividing the bookings into many small sub-segments. In addition to an improvement in fit over previous models based on a deterministic segmentation of the market, the latent class choice model was found to provide a more intuitive segmentation of the demand between a core of time-sensitive business travelers and a mixed class of price-conscious leisure and business travelers. This model extends the range of applications of passenger choice models to additional airline planning decisions such as pricing and revenue management.

1. Introduction

1.1. Motivation and Research Objectives

Understanding passenger choice behavior is crucial to support a range of airline planning decisions such as network structure, schedule planning, pricing, revenue management. In particular, a better analysis of the choice of a fare product is now needed to support new pricing strategies based on offering travelers the choice between several branded product packages differentiated by features such as complementary advance seat selection or the flexibility to change travel plans. In addition, newly proposed revenue management models designed to maximize revenues under less restricted fare structures also require an analysis of the choice of a fare product and an estimate of the passenger willingness to purchase a higher fare when a lower-priced product is no longer available.

Previous studies of airline passenger choice have been based on the analysis of two types of data: Booking data (Coldren, 2005; Grammig et al., 2005) or stated preference data (Prossaloglou and Koppelman, 1999; Adler et al., 2005; Theis et al., 2006; Garrow et al., 2007;). Stated preference data collected through surveys of airline passengers provide the flexibility to model new non-existing alternatives as well as the opportunity to collect detailed information on the decision-maker (income, age, gender, travel patterns), the characteristics of the trip (trip purpose), even attitudes and perceptions regarding elements of the travel experience such as comfort or risk of misconnection. However, these studies suffer from the hypothetical nature of the experiment and are subject to a risk of response bias. This risk is particularly high for pricing applications when they are used to investigate the trade-off between price and other elements of airline service such as schedule or amenities.

In addition, the design of stated preference experiments cannot adequately represent the complexity of the product structure in airline markets and the impact of airline pricing and revenue management decisions on the passenger choice set. Respondents are typically asked to choose between a limited set of two to three alternatives that are designed to be representative of the products offered in the marketplace. However, such a simplified choice set does not reproduce in a sufficiently realistic manner the large number of travel alternatives viewed by prospective travelers in a real booking search, especially given the large selection provided by the increasingly sophisticated booking engines that power the large online travel retailers such as Travelocity, Orbitz or Expedia.

While not subject to a risk of response bias, previous studies based on booking data provide only a partial view of passenger choice behavior as they have been restricted to the sole choice of an airline itinerary and used primarily for schedule planning applications. This is due to the properties of booking data: Only the booked alternative is currently recorded in airline bookings and no information is available on other travel alternatives viewed by the passenger at the time of the booking. While the schedule of other travel alternatives can be obtained relatively easily from other sources such as the Official Airline Guide (OAG), information on other attributes such as the fare are difficult to collect as they depend on the state of the airline inventory at the time of the booking. The airline inventory is changing constantly based on the booking activity and the booking limits set by the airline revenue management system.

In addition, previous studies based on booking data did not take into account the heterogeneity of behavior across bookings, a major characteristic of airline markets. The conventional wisdom in the industry is to segment airline markets by trip purpose: Leisure travelers are considered to be very price-sensitive while business travelers place more emphasis on schedule convenience and service quality. However, trip purpose is not recorded in airline bookings. As a result, previous studies of airline passenger choice based on booking data did not segment the market between different categories of bookings and failed to test for heterogeneity of behavior.

The objective of this research is then to develop a passenger choice model that investigates the trade-off between the major dimensions of airline passenger choice such as schedule and price for different segments of airline travelers. In order to avoid the potentially high risk of response bias associated with pricing experiments in stated preference data, we will focus on analyzing actual choice behavior as reflected in past booking records. Booking data will be combined with other data sources such as fare rules and seat availability data to incorporate the impact of airline pricing and revenue management and reconstitute the passenger choice set at the time of the booking. In addition, elements of the booking records such as the characteristics of the trip and the profile of the traveler will be used to segment demand and search for differences in choice behavior across bookings.

1.2. Contributions

The contributions of this research can be divided into three categories: A choice set generation process that better reflects the characteristics of airline markets, advancements in passenger choice models and applications to airline planning decisions.

In this research, we developed a methodology to incorporate the impact of airline pricing and revenue management on the choice set of each booking by combining booking, fare rules and seat availability data. This choice set generation process reflects much more realistically the range of travel options effectively available to prospective travelers at the time of the booking than previous studies of airline passenger choice. It provides the foundation for the development of a model of the choice of an airline itinerary and fare product based on booking data.

This research also provides two important contributions to the development of passenger choice models. First, we developed an alternative to the segmentation of airline demand by trip purpose. Since it is unavailable in booking records, trip purpose was replaced by other elements found in airline bookings such as the characteristics of the trip (distribution channel, dates of travel) and the profile of the traveler (frequent flyer membership). These elements were used to estimate a latent class model of airline passenger choice that allows segmenting the market using multiple factors without dividing the bookings into a large number of small sub-segments. The latent class structure of the model was found to improve the fit of the model compared to previous specifications based on a deterministic segmentation scheme. It also leads to a more intuitive segmentation of the market between a core of time-sensitive business travelers and a mixed class of leisure and price-conscious business travelers.

Second, we improved the measurement of the time-of-day preferences of airline travelers by using a continuous function of time instead of a set of discrete time-period dummies. We generalized the formulation of a trigonometric function of time to better represent the time-of-day preferences of airline travelers in short-haul markets. Instead of being set to a full 24 hours, the duration of the daily cycle was estimated to take into account the low attractiveness of nighttime flights in short-haul markets. This approach also provided a flexible framework to model the time-of-day preferences of specific categories of travelers such as day trippers that travel exclusively on outbound morning flights due to the short duration of their trip.

This model of the choice of an airline itinerary and fare product extends the range of potential applications of passenger choice models to new areas of airline planning decisions such as pricing and revenue management. In particular, the parameter estimates of the model were used to estimate the sell-up behavior of airline passengers, a key input to all newly proposed revenue management algorithms designed to maximize revenues under less restricted fare structures.

1.3. Outline

The remainder of this paper is organized as follows: Section 2 describes the latent class model of airline passenger choice developed in this research. In Section 3, we propose a new formulation of a continuous function of time to model the time-of-day preferences of airline passengers. Section 4 describes the data collected for this research and the choice set generation process. Section 5 presents the estimation results of the model and discusses the benefits of the latent class approach relative to previous models based on a deterministic segmentation of the demand. In Section 6, we discuss the potential applications of the model focusing on how to use the parameter estimates to forecast the sell-up behavior of airline passengers. Finally, Section 7 summarizes the findings of this research and provides some concluding remarks.

2. Modeling Framework

The objective of this research is to develop a model of the major dimensions of airline passenger choice. As reported by Smith (2006), on-going research on the choice behavior of Travelocity customers shows that price is the most important factor in the selection of a travel alternative, followed by flight schedule and to a much lower extent, the carrier providing the service. In this research, we focus on how an airline can exploit its own data such as its booking and inventory databases to better understand the choice behavior of its passengers and develop choice-based decision support tools for a range of airline planning applications. As a result, this research is subject to the data limitations an airline is likely to face. Unlike a travel agent such as Travelocity that markets the inventory of several carriers, an airline does not have access to the bookings of its competitors. However, an airline has a record of the full set of bookings on its own network and is not limited to a specific and potentially biased subgroup of the market such as Travelocity customers. We will then assume the choice of a particular airline as given and study the choice behavior of airline passengers along the remaining two major dimensions, price and itinerary.

The dependent variable of the model in this single-airline framework is then set as the combination of an itinerary and a fare product. An itinerary or path is defined here as a sequence of flights between an origin and a destination point with specific departure and arrival times. A fare product is characterized by its price and a set of fare rules that define both its features such as the flexibility to change flights or cancel the trip and its conditions such as minimum stay or advance purchase requirements. For instance, a trip on a non-stop flight departing from the origin city at 7:30 a.m. and arriving at the destination city at 9:00 a.m. at a non-refundable fare of 200 EUR will be considered as one potential travel alternative in the market. A trip on the same flight but at a higher refundable fare of 400 EUR will be considered as another travel alternative.

The maximum number of alternatives in the universal choice set in an origin-destination market is then equal to the number of daily itineraries in the market multiplied by the number of fare products offered on each of these itineraries. However, the actual choice set varies for each booking based on the interaction between two sets of elements, airline planning decisions such as pricing and revenue management and passenger decisions such as the date of the booking and dates of travel. As a result, booking data was combined with fare rules and seat availability data to reconstruct with a high level of accuracy the choice set of each booking. A detailed description of the choice set generation process is provided in Section 4.

Discrete choice models are used to represent the choice of an individual decision-maker among a finite number of mutually exclusive and collectively exhaustive alternatives. Since the choice set of airline travelers is composed of a varying but finite number of potential travel alternatives, discrete choice models based on random utility theory were used as the basic tool for the model developed in this research. For a comprehensive description of random utility theory and discrete choice analysis, the reader is referred to Ben-Akiva and Lerman (1985).

The other major objective of this research is to investigate the impact of heterogeneity of behavior on airline passenger choice. Heterogeneity of behavior is a key characteristic of airline markets and the conventional wisdom in the industry is to segment airline bookings by trip purpose. Since trip purpose is not recorded in airline bookings, previous studies of airline passenger choice based on booking data have ignored heterogeneity of behavior. However, while trip purpose remains unobserved in airline bookings, booking data contain other elements that are correlated with trip purpose and can provide valuable information to segment airline bookings.

Two types of data available in booking records may prove useful to segment the demand: The traveler's profile and the characteristics of the trip. The traveler's profile does not depend on a single trip and includes both socio-economic characteristics of the traveler such as gender and travel-related characteristics such as frequent flyer membership and status. In addition to the traveler's profile, characteristics specific to the trip may also be useful to segment the market. In particular, the dates of outbound and inbound travel can be used to determine whether the trip included or not a stay at the destination over the weekend. Travel within a week is expected to be strongly correlated with business travel: Business trips are usually relatively short and business travelers tend to return home before the weekend, especially in short-haul markets. The distribution channel of the ticket also provides a fairly strong indicator of trip purpose. While many non-business travelers have shifted to online and direct channels of distribution, especially for simple travel needs such as a roundtrip ticket in a short-haul market, many business travelers still rely on traditional travel agents that provide a range of services such as billing, enforcement of company travel policies or access to discounted corporate fares.

While all these factors may be useful to segment the demand and identify differences in choice behavior across bookings, in practice, only a subset of them can be used under the conventional deterministic approach found in the literature. Otherwise, the number of segments will increase rapidly and each segment may be too small to identify differences in choice behavior across bookings and difficult to interpret. For instance, if three factors are used such as week travel, distribution channel of the ticket and frequent flyer membership, the bookings need to be divided into eight different segments.

A latent class choice model provides an alternative approach that has several advantages over previous models based on a deterministic segmentation of the demand. Latent classes are unobserved segments that allow the segmentation of the market using multiple factors without dividing the bookings into a large number of small sub-segments. Since we cannot directly identify to which latent class a particular booking belongs, a probabilistic assignment process is used also called the class membership model. These factors are specified as explanatory variables of the class membership model and parameter estimates are obtained that provide insight on their respective weight to the segmentation of the market.

In addition to the difference in choice behavior between different segments of air travel demand, such as business and leisure-style travelers that relate to the trade-off between price and schedule and the entire utility function, heterogeneity of behavior may also be associated with a specific part of the utility function. For instance, day trippers are likely to have specific time-of-day preferences due to the short duration of their trip. For day trip bookings, the choice of an itinerary is restricted to morning flight departures for the outbound part of the trip, so that the traveler has sufficient time to conduct his activities at the destination and travel back at the end of the day to the origin city. All other itineraries are excluded from the passenger choice set. Even after eliminating afternoon flights from the choice set, the time-of-day preferences of day trippers may still differ from the rest of the travelers. Among morning flights, day trippers are expected to have a stronger preference for very early departures in order to have as much time as possible during their short stay in the destination city.

Unlike trip purpose, which is unobserved in airline bookings, duration of stay can be inferred from the dates of inbound and outbound travel and day trippers identified for roundtrip tickets. In order to capture the preferences of day trippers, a specific set of variables will then be introduced in the class-specific choice models for the part of the utility function related to the choice of an itinerary. They will be used to determine whether the time-of-day preferences of day trippers differ from overnight bookings within each latent class of the model.

Figure 2-1 below summarizes the structure of the latent class model of airline passenger choice developed in this research.

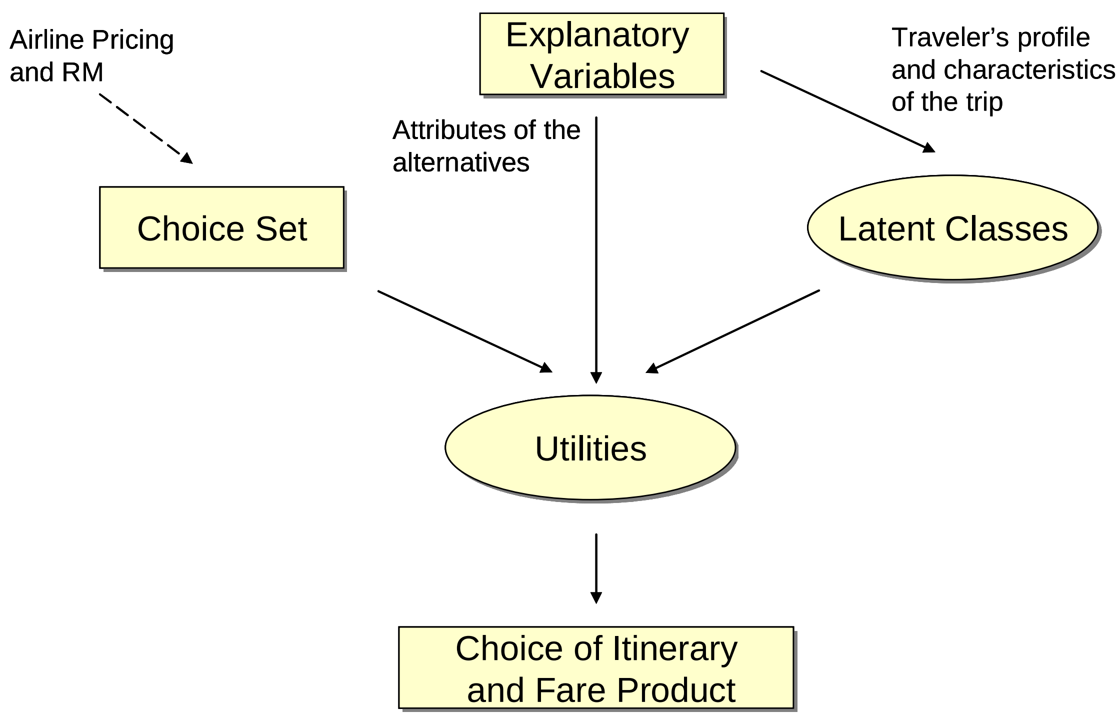


Figure 2-1: The Latent Class Model of Airline Passenger Choice

For a given booking and a travel alternative i, i = 1, 2, …, J_b where J_b is the number of alternatives in the choice set C of booking b, the basic form of the model can be written as follows:

$$P(i/X_M, X_C) = \sum_{s=1}^{S} \underbrace{P(s/X_M)}_{\textbf{Class Membership}} \underbrace{P(i/X_C, s)}_{\textbf{Choice}} \quad \forall i \in C \tag{2.1}$$

Where s = 1, 2, …, S are latent classes of bookings

X_M is a vector of explanatory variables of the class membership model

X_C is a vector of explanatory variables of the class-specific choice models

Given membership in class s, the class-specific choice model is written as follows:

$$y_{ib} = \begin{cases} 1 \text{ if } U_{ib} \geq U_{jb} \text{ for } j = 1, 2, \ldots, J_b \\ 0 \text{ otherwise} \end{cases}$$

$$U_{ib} = X_{Cib}\beta_C + \varepsilon_{ib} \tag{2.2}$$

Where y_{ib} indicates the chosen travel alternative and U_{ib} is the utility of travel alternative i for booking b. X_{Cib} is a (1xK) vector of the explanatory variables of the choice model, β_C is a (Kx1) vector of parameters and ε_{ib} is a random disturbance. The assumption that the disturbances are i.i.d. extreme value leads to the logit model specification. The class-specific choice probability of travel option i can then be expressed as follows:

$$P(i/X_{Cib}, s) = \frac{e^{X_{Cib,s}\beta_{C,s}}}{\sum_{j=1}^{J_b} e^{X_{Cjb,s}\beta_{C,s}}} \qquad \forall \ s \in S \quad \forall \ i \in C \tag{2.3}$$

where $\beta_{C,s}$ are the unknown parameters of the class-specific choice models

A multinomial logit model (MNL) specification is also used for the class membership model. However, it should be noted that, unlike for class-specific choice models, the MNL-type class membership model cannot be interpreted as derived from random utility theory. The probability of belonging to latent class s is then written as follows:

$$P(s / X_{Mb}) = \frac{e^{X_{Mb,s}\beta_M}}{\sum_{t=1}^{S} e^{X_{Mb,t}\beta_M}} \tag{2.4}$$

where β_M are the unknown parameters of the class membership model

The Latent Gold Choice software by Statistical Innovations (Vermunt and Magdison, 2005) that is specifically designed for the estimation of latent class choice models was used to estimate the parameters of the model.

3. Time-of-Day Preferences of Airline Travelers

While we discussed in the previous section the overall structure of a latent class model of airline passenger choice, we focus in this section on how to model a specific part of the passenger utility function, the time-of-day preferences of airline travelers.

The conventional approach to represent the time-of-day preferences of airline passengers is to divide the day into a finite number of time periods and specify a dummy variable for each period. For instance, Coldren (2005) divided the day into one-hour periods, except for night departures that were grouped into two longer periods, a 10 p.m. to midnight and a midnight to 5 a.m. period. However, since time is a continuous variable, the effect of any time-related variable should also be continuous. As time of day is cyclic with a cycle length of 24 hours, this function should be periodic so that the utility function takes the same value at time h and time h + 24 hours. Ben-Akiva and Abou Zeid (2007) proposed to take advantage of the properties of the trigonometric operators and to use a function of the following form:

$$U(h) = \beta_1 \sin(\frac{2\pi h}{24}) + \beta_2 \sin(\frac{4\pi h}{24}) + \beta_3 \sin(\frac{6\pi h}{24})$$
$$+ \gamma_1 \cos(\frac{2\pi h}{24}) + \gamma_2 \cos(\frac{4\pi h}{24}) + \gamma_3 \cos(\frac{6\pi h}{24}) + ... \quad (3.1)$$

Where β_1, β_2, …, γ_3 are unknown parameters to be estimated and h is the flight departure time. It can be easily verified that such a function satisfies the property U(h)=U(h + 24) for $0 \leq h \leq 24$. In particular the utility function takes the same value at the beginning and end of the daily cycle ensuring its continuity. The number of estimated parameters is determined empirically based on the resulting profile of the utility function and the statistical significance of the parameters.

However, there may be little demand for flights departing during some unattractive periods of the day and this could distort the parameter estimates of a continuous function of time defined over a full 24-hour daily cycle. For instance, in short-haul markets, very few passengers are expected to want to travel during nighttime. In a recent study of the choice of an airline itinerary based on a stated preference survey of prospective travelers at a travel website, Garrow et al. (2007) collected data on passenger ideal departure time for outbound travel in U.S. domestic markets. As shown in Figure 3-1 below, very few passengers stated that they wanted to depart between midnight and 6 a.m. in U.S. domestic North-South/South-North markets.

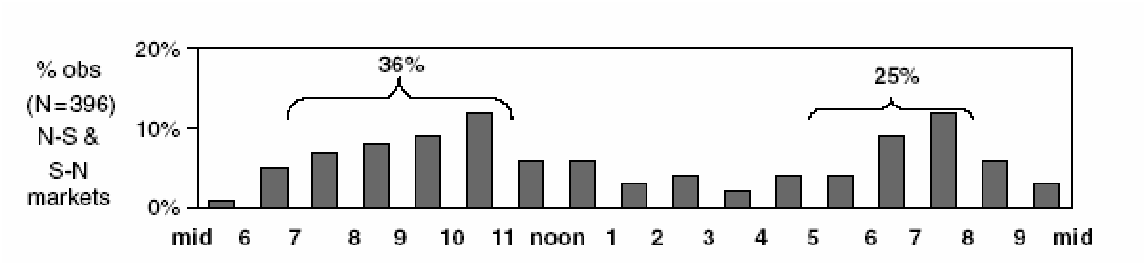


Figure 3-1: Passenger Ideal Departure Time in U.S. North-South markets (Source: Garrow, Parker and Jones, 2007)

As a result, we propose to adjust the duration of the cycle to an "effective" travel period d, equal to or less than 24 hours and starting at time s. Then, Equation (3.1) becomes:

$$U(h) = \beta_1 \sin(\frac{2\pi(h-s)}{d}) + \beta_2 \sin(\frac{4\pi(h-s)}{d}) + \beta_3 \sin(\frac{6\pi(h-s)}{d})$$
$$+ \gamma_1 \cos(\frac{2\pi(h-s)}{d}) + \gamma_2 \cos(\frac{4\pi(h-s)}{d}) + \gamma_3 \cos(\frac{6\pi(h-s)}{d}) + \ldots \quad (3.2)$$

Where $l - e \leq d \leq 24$

$0 \leq s \leq e$

With e and l the departure times in hours of respectively the earliest and latest itineraries in the market.

It can be verified that the value of the utility function is equal at the beginning and end of the cycle, U(s) = U (s+d). While a continuous function of time could potentially take different values at the beginning and end of the cycle when the duration of the cycle is less than 24 hours, this ensures that this property remains true as required if the duration of the cycle is estimated to cover a full 24 hour period.

Since the duration d and the start time s of the cycle are included inside the sinus and co-sinus functions in Equation (3.2), the passenger utility function is not linear in the parameters any more. As Latent Gold Choice does not support this type of utility function, s and d cannot be estimated directly using this software. However, it was found that the start time of the cycle s has no impact on the time-of-day preferences of airline travelers. This means that the value of the continuous function for a specific flight departure time such as a 7 a.m. flight departure is the same whether the cycle started at 5:30 or 6 a.m. Since it cannot be identified, we selected the start time of the cycle at a local minimum so that the continuous function starts and ends at a local low point. In addition, a trial and error method was used to search for the cycle duration d that maximizes the log-likelihood of the model.

Estimating the duration of the cycle also provides a flexible approach to model the time-of-day preferences of specific segments of airline travelers such as day trippers. As mentioned earlier, interaction variables will be used to capture the specific time-of-day preferences of day trip bookings. A continuous function specific to each type of booking is then included in the utility function and the duration and start time of the cycle is estimated for each category of bookings. As outbound day trippers are observed to book only morning flight departures, the estimated duration of the cycle is expected to be much shorter for day trip than for overnight bookings.

4. Data and the Choice Set Generation Process

As mentioned in Section 2, the choice set of each booking depends on a series of airline decisions such as pricing and revenue management. Since only the booked alternative is recorded in airline bookings and no information is available on non-booked alternatives, a data-intensive methodology was developed to reconstruct with a high level of accuracy the choice set of each booking along both dimensions of airline passenger choice considered in this research, itinerary and fare product. In order to incorporate the impact of airline pricing and revenue management on the passenger choice set, three types of data were combined: booking, fare rules and seat availability data.

Booking data was obtained from Amadeus, one of the major global distribution systems (GDS) used by travel agents to book airline tickets. In addition to the distribution of airline tickets through its network of affiliated travel agents, Amadeus also offers airlines the opportunity to outsource all their ticketing activities in its Amadeus System User (ASU) program also called Altea Sell. This means that, for ASU airlines, Amadeus has a record of all booking activity including bookings initiated by travel agents affiliated with a competing GDS or made through direct distribution channels such as the airline's own website. In order to have a complete set of booking records made through any distribution channel, data was collected for a few European short-haul markets out of Paris, in which major airlines offering non-stop service participate in the ASU program. In this study, we will focus on booking data collected for a major European airline in three markets connecting Paris with three major German business destinations: Dusseldorf (DUS), Frankfurt (FRA) and Stuttgart (STR).

Data was collected for outbound trips from Paris for two periods at the end of May 2005 (May 26 - May 31) and at the beginning of July 2005 (July 1 – July 7). A total of 2015 bookings are included in the dataset. In order to reduce the number of alternatives to a tractable size, only trips on non-stop flights were considered. Although adding connecting itineraries would make the study more realistic, this limitation has little impact in such short-haul markets as travel time on connecting itineraries is very unattractive compared to non-stop flights.

In addition to booking data, fare rules were obtained to incorporate the impact of airline pricing on the passenger choice set. The airline was using a similar fare structure in many European short-haul markets, including the three markets selected for this research. This fare structure is a mix of the traditional pricing strategy used by network airlines with a set of fare products that require a weekend or Saturday night stay and a low-cost airline pricing strategy with a set of non-flexible discounted fare products that can be used for travel within a week with several price points associated with different levels of required advance purchase.

Fare products are grouped into four categories as follows:

- A weekend fare product that is restricted to a departure on Friday or Saturday with a return on the following Sunday or Monday but requires only a very short one-day advance purchase.
- Traditional discounted fare products requiring a Saturday night stay with five price levels depending on the advance purchase requirement.
- Discounted fares valid for travel within a week but that are non-flexible and can neither be changed nor cancelled. Several price points are offered depending on the advance purchase requirement. This new set of products was introduced in March 2004 in European short-haul markets in response to growing low-cost competition.
- Fully flexible fares, either published (S) or available at a discount for eligible travelers through corporate contracts (BFIRME & SFIRME).

Table 4-1 below shows the list of all fare products with their equivalent one-way fare in the Paris-Frankfurt market.

Product	Fare Product Category	Fare Code	OW Fare (FRA)	AP (Days)	Maximum Stay	Cancellation Fee	Change Fee
1	Weekend	NWKEND	102 €	1	2 days	Not Permitted	Not Permitted
2	SN	NAP30	107 €	30	1 month	Not Permitted	Not Permitted
3	SN	EAP21	138 €	21	1 month	120 €	60 €
4	SN	WAP14	167 €	14	1 month	120 €	60 €
5	SN	QAP7	201 €	7	1 month	120 €	60 €
6	SN	MSX0	234 €	0	1 month	120 €	60 €
7	Week	AWEEK21	161 €	21	12 months	Not Permitted	Not Permitted
8	Week	UWEEK14	213 €	14	12 months	Not Permitted	Not Permitted
9	Week	UWEEK7	281 €	7	12 months	Not Permitted	Not Permitted
10	Week	RWEEK	325 €	0	12 months	Not Permitted	60 €
11	Flex	BFIRME	276 €	0	12 months	None	None
12	Flex	SFIRME	310 €	0	12 months	None	None
13	Flex	S	362 €	0	12 months	None	None

Table 4-1: Fare Structure (Source: Sabre)

Fare rules were applied based on the characteristics of each booking. For instance, a fare product that requires a Saturday night stay is eliminated from the passenger choice set if the dates of travel of the booking show that the traveler did not spend the Saturday night in the destination city. Similarly, a fare product that requires a 21-day advance purchase is eliminated if the booking occurred less than 21 days before departure.

In order to incorporate the impact of airline revenue management decisions, seat availability data was collected daily by Amadeus over a 3 month period prior to the booking data collection. For instance, for bookings collected for travel in the July 1-7, 2005 period, seat availability data was collected every day from April 1, to July 7, 2005 for all non-stop flights departing between July 1 and July 7. The booking and seat availability data were then matched to determine which fare products were available on the date of the booking on all non-stop itineraries offered for the same day of departure.

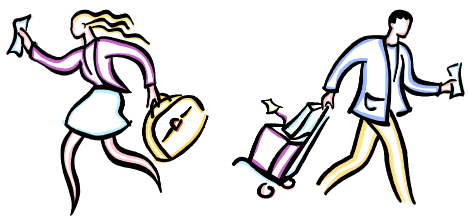
Additional steps were then used to further restrict the choice set. Dominance rules observed in the data were applied and, for instance, only the cheapest available fare product within a category was included in the choice set. In addition, since no data was available on which travelers were eligible for corporate contract discounts, it was assumed that all eligible travelers would select that product to take advantage of the discount and the full flexibility to change travel plans. As a result, for travelers that used a corporate discount, the model gets reduced to the sole choice of an itinerary. Finally, afternoon flights were removed from the choice set of day trip bookings as outbound day trippers were always observed to select morning flight departures due to the short duration of their trip.

Figure 4-1 below illustrates how the data is processed to generate the choice set of each booking.

Booking Date:
June 22, 2005 (AP=15)

Outbound Departure Date:
Thursday, July 7, 2005

Inbound Departure Date:
Saturday, July 9, 2005

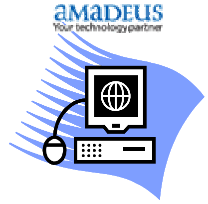


Booking Date:
June 22, 2005 (AP=15)

Outbound Departure Date:
Thursday, July 7, 2005

Inbound Departure Date:
Monday, July 11, 2005

aMaDEUS
Your technology partner



	FLIGHT 1	FLIGHT 2	FLIGHT 3
NWKEND	N	N	N
NAP30	N	N	N
EAP21	N	N	N
WAP14	N	N	N
QAP7	N	N	N
MSX0	N	N	N
AWEEK21	N	N	N
UWEEK14	Y	N	N
UWEEK7	N	N	N
RWEEK	N	N	N
BFIRME	N	N	N
SFIRME	N	N	N
S	Y	Y	Y

Flight 1, 6:40 a.m.:
Y9S9B9K9H9R9M9T9Q9V9L9X9U9W9E9A0N0I0

Flight 2, 7:45 a.m. :
Y9S9B9K4H4R0M0T0Q0V0L0X0U0W0E0A0N0I0

Flight 3, 10 a.m. :
Y9S9B4K3H3R0M0T0Q0V0L0X0U0W0E0A0N0I0

	FLIGHT 1	FLIGHT 2	FLIGHT 3
NWKEND	N	N	N
NAP30	N	N	N
EAP21	N	N	N
WAP14	Y	N	N
QAP7	N	N	N
MSX0	N	N	N
AWEEK21	N	N	N
UWEEK14	N	N	N
UWEEK7	N	N	N
RWEEK	N	N	N
BFIRME	N	N	N
SFIRME	N	N	N
S	N	Y	Y

Choice: Flight1, UWEEK14, 178.58 EUR OW

Choice: Flight1, WAP14, 116.08 EUR OW

Figure 4-1: An Example of the Choice Set Generation Process

This example, extracted from a similar dataset, shows the case of two passengers that both departed on July 7, 2005 and booked their trip 15 days in advance on June 22, 2005. Passenger 1 returned on Saturday, July 9, 2005 and her trip did not include a Saturday night stay while Passenger 2 came back on July 11, 2005 and stayed at the destination on the Saturday night.

For both passengers, only fare products that require less than 15 days of advance purchase were included in the choice set. However, fare products that require a Saturday night stay were excluded from the choice set of Passenger 1 and she chose the earliest flight and the UWEEK14 fare product that provides no flexibility to change travel plans and costs 178.58 EUR one-way. Cheaper fare products that require a Saturday night stay were included in the choice set of Passenger 2 and his final choice was for the same flight and a WAP14 product at a price of 116.58 EUR one-way. Please note that these fare products on the following two flights departing later in the morning were excluded from the choice set for both passengers due to the airline revenue management controls. For instance, the fare class availability display in the lower middle part of the figure indicates that fare class U is not available on flights 2 and 3 (the display shows U0) meaning that fare products UWEEK14 and UWEEK7 were not available for sale on those flights at the time of the booking.

5. Estimation Results

In this section, we present the estimation results of the latent class model of airline passenger choice proposed in this research. We first introduce in more detail the explanatory variables of the model. We then analyze the estimation results of a two-class latent class choice model with a continuous function of time. We finally discuss why such a model specification is preferred over a deterministic segmentation of airline bookings between week and non-week travelers as well as the benefits of using a continuous function of time over time-period dummies.

5.1. Explanatory Variables

The explanatory variables of the choice model include the attributes of the travel alternatives along the two dimensions considered in this research, itinerary and fare product. As only non-stop flights are considered and elapsed times are similar for all flights in a market, the attributes of an itinerary is reduced to the flight departure time. A continuous function of flight departure time with a variable duration of the cycle is used to model the time-of-day preferences of the passengers as described in Equation (3.2). A trial and error method with half-hour increments is used to estimate the duration of the cycle that maximizes the log-likelihood of the model. As mentioned earlier, a separate function is used for overnight and day trip bookings to model the specific time-of-day preferences of day trippers.

Regarding the choice of a fare product, attributes of the alternative include the fare paid by the traveler and the fare rules set by the airline pricing department. There are two types of fare rules. Some of them are conditions that need to be satisfied to book a specific product such as minimum and maximum stay or advance purchase requirements. These fare rules are not included as explanatory variables as their impact is incorporated in the choice set generation process. The other type of fare rules includes features of the product such as the flexibility to modify the booked itinerary and/or cancel the trip. The flexibility to change travel plans is used by the airline to differentiate fare products and segment demand. In particular, fare products from the Week category are highly restricted with no changes or cancellations permitted in most instances while the alternative full fare is refundable and allows for unlimited changes to the passenger’s itinerary.

In order to capture the disutility associated with the lack of flexibility to change travel plans, one option is to include a dummy variable for each fare product in the utility function. Since some fare products carry the same set of rules, an alternative approach is to group fare products by fare rule. The advantage of consolidating products by fare rule is that an estimate of the disutility value of a specific rule is then obtained. These estimates provide a metric to evaluate whether the current airline pricing strategy is effective or should be adjusted.

The rules regarding flexibility in travel plans were combined with the advance purchase requirement of the fare product as the uncertainty about a future trip and hence the need to change the itinerary or cancel the trip is expected to increase with how long in advance the traveler needs to anticipate its travel plans. Grouping the fare products is equivalent to constraining the value of the product-specific dummy variables to be equal across all products in a category. A likelihood ratio test can then be used to determine whether such a restriction is valid from a statistical point of view. Based on the estimation results of an extensive set of potential groupings and likelihood ratio tests, fare products were classified into the following four categories and three dummy variables were included in the utility function to capture the impact of fare rules on the choice behavior of airline passengers:

- Fare products that are not fully flexible and require at least 21 days of advance purchase (NON-FLEX & 21AP).
- Fare products that are not fully flexible and require either 7 or 14 days of advance purchase (NON-FLEX & 7-14AP).
- Fare products that are not fully flexible but do not require any advance purchase (NON-FLEX & 0AP).
- Fully flexible fare product.

Since the fully flexible fare is used as the base, the coefficients of these three dummy variables are expected to be negative reflecting the disutility associated with the lack of flexibility to change travel plans without penalty relative to the unrestricted fully flexible fare.

In addition to the attributes of the alternatives, characteristics of the trip and the traveler were extracted from the bookings and used to segment demand and capture heterogeneity of behavior. They were used as explanatory variables of the class membership model and include:

- Frequent flyer membership (FFP MEMBER): A dummy variable is used to indicate whether the traveler belongs to the airline's loyalty program.
- Distribution channel of the ticket (OFFLINE TA): A dummy variable is added to indicate if the ticket was booked through an offline travel agent.
- Week travel (MON to FRI): If outbound travel started on or after Monday and inbound travel occurred on or before Friday of the same week, this dummy variable will be set to 1.

5.2. Two-Class Latent Class Model of Airline Passenger Choice

The estimation results of a two-class latent class model of airline passenger choice are presented in Table 5-1 below.

		LATENT CLASS 1 75.3%			LATENT CLASS 2 24.7%		
		Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Class Membership	INTERCEPT	-4.60	3.66	-1.3	4.60	3.66	1.3
	FFP MEMBER	1.00	0.50	2.0	-1.00	0.50	-2.0
	MON to FRI	3.34	3.67	0.9	-3.34	3.67	-0.9
	OFFLINE TA	3.64	3.59	1.0	-3.64	3.59	-1.0
Choice Model	FARE	-0.0125	0.0052	-2.4	-0.0273	0.0075	-3.7
	NON-FLEX & 21AP	-2.46	1.07	-2.3	-3.71	1.25	-3.0
	NON-FLEX & 7-14AP	-2.23	0.67	-3.3	-1.52	0.84	-1.8
	NON-FLEX & 0AP	-1.33	0.25	-5.3	-0.01	0.47	0.0
	DUR-OV = 16 HOURS						
	SIN2PI-OV	0.33	0.10	3.3	0.18	0.15	1.2
	SIN4PI-OV	0.55	0.12	4.6	0.10	0.22	0.5
	SIN6PI-OV	0.09	0.10	1.0	-0.54	0.17	-3.3
	SIN8PI-OV	0.24	0.08	3.0	0.24	0.13	1.9
	COS2PI-OV	-0.15	0.05	-2.8	0.13	0.09	1.5
	COS4PI-OV	0.68	0.10	6.9	0.04	0.14	0.3
	COS6PI-OV	-0.33	0.08	-4.3	-0.29	0.13	-2.3
	COS8PI-OV	-0.03	0.13	-0.2	-0.17	0.23	-0.8
	DUR-DT = 9 HOURS						
	SIN2PI-DT	-1.51	0.12	13.1	-1.16	0.52	-2.2
	COS2PI-DT	1.75	0.34	5.1	5.19	3.27	1.6
Summary Statistics	Log L (0)			-3516.81			
	Log L			-3142.41			
	Number of parameters			34			
	Rho-bar-squared AIC			-3176.41			
	Rho-bar-squared BIC			-3271.75			

Table 5-1: Two-Class Latent Class Model of Airline Passenger Choice

5.2.1. Class Membership Model

Given that three dummy variables are included in the class membership model, they define a set of eight underlying categories of bookings also called covariate patterns. For each covariate pattern, the likelihood to belong to each latent class can be calculated given the parameter estimates of the class membership model. Figure 5-1 below shows the probability of belonging to latent class 1 for the different covariate patterns. Latent class 1 can be considered to reflect primarily business-type travel as business travelers tend to travel within a week and rely more frequently on traditional offline travel agents to book their tickets while latent class 2 appears to be primarily oriented toward leisure travelers.

Latent class 1 groups slightly over 75% of the observations as over 70% of the bookings belong to the two covariate patterns that combine an offline travel agent distribution channel and week travel and are estimated to belong almost entirely to class 1. This is consistent with a priori expectations as these markets are strongly oriented toward business travel. If we analyze further the distribution of bookings by covariate patterns and fare product category, over 80% of the travelers that purchased a Flex fare belong to these two covariate patterns versus only about 50% of the travelers who purchased a non-flexible fare from the Week category. Since they differ in their characteristics and belong in many instances to different covariate patterns, about 35% of the bookings for a product from the Week category are estimated to belong to latent class 2, compared to only 10% for bookings made by travelers that purchased the Flex fare product.

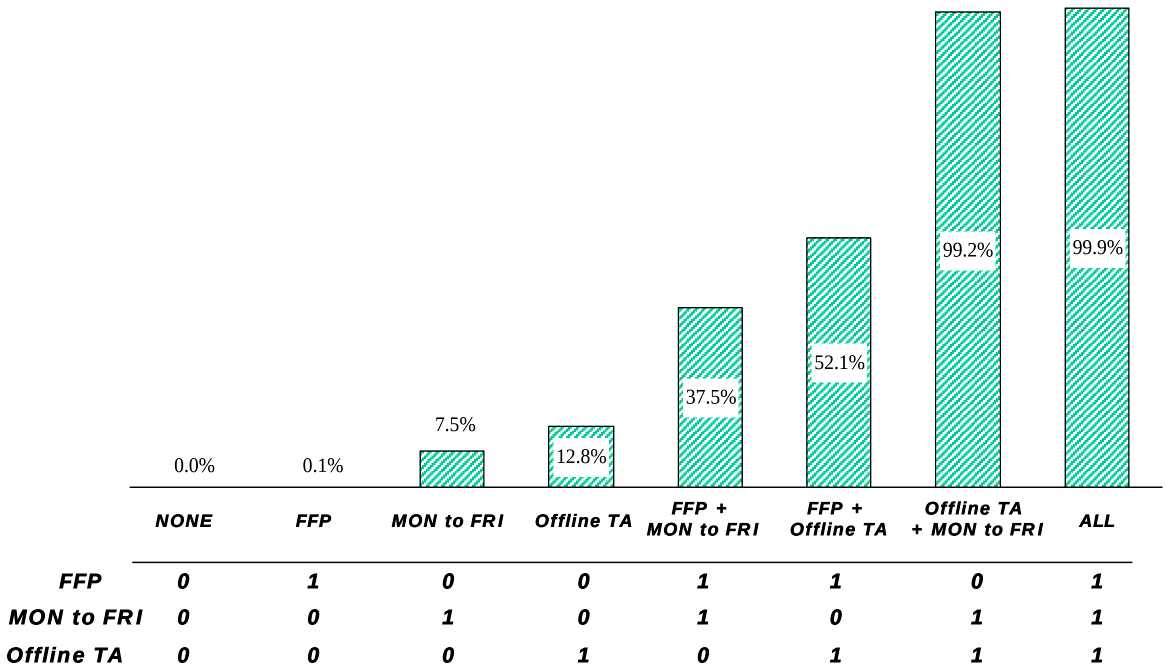


Figure 5-1: Latent Class 1 Membership Probabilities

The quasi-deterministic class membership probabilities found for half of the covariate patterns reflect the strong segmentation of the demand typically observed in airline markets. This trend is reinforced here by the airline pricing strategy since the underlying fare structure requires staying over for the weekend at the destination to book the most heavily discounted airfares.

This pattern is reflected in the high value of the standard errors obtained for two explanatory variables of the class membership model, the offline travel agent and the MON to FRI dummies. As a result, the value of the t-statistic for these two variables is under 1.65, meaning that they are not significantly different from zero, even at a 90% confidence level. As discussed in Galindo-Garre et al. (2004) and Vermunt et al. (2006), the maximum likelihood estimates of the parameters lie on the boundary of the parameter space when estimated model probabilities are equal to 0 or 1. Occurrence of boundary estimates causes the standard errors of the parameters to go toward infinity and the confidence intervals and significance tests become meaningless. These corner solutions occur frequently in latent class models, especially if the sample is relatively small and some covariate patterns are relatively sparsely populated as is here the case.

As a result, the estimation results of the class membership model should be interpreted primarily based on the relative magnitude of the coefficients rather than on their t-statistic. The offline travel agent and the MON to FRI dummies have coefficient estimates of similar magnitude while the frequent flyer member dummy has the same sign but a much lower estimated value. This is line with our expectations as business travelers tend to travel mostly within a week and use more frequently the services provided by a traditional travel agent to book their tickets than the rest of the travelers. They are also expected to be more likely to belong to the airline’s loyalty program as business travelers tend to travel more frequently and should be able to benefit more from the rewards of an airline frequent flyer program. However, this effect is not expected to be as strong as membership in airline loyalty programs is complementary and leisure travelers may also be able to accumulate enough credits to obtain a complimentary award ticket.

The number of latent classes used in the model was determined based on two elements: Model selection criteria such as the Akaike Information Criterion (AIC; Akaike, 1973 & 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978) and the interpretation of the class membership parameter estimates. Both the AIC and BIC indicated that a model with several latent classes was preferred over a model with a single class. In addition, the BIC that is often used to determine the number of classes suggested that the two-class model was superior while the less parsimonious AIC favored a model with three classes. However, the three-class model lead to a split of business-style latent class 1 bookings into two sub-segments that could not be easily interpreted and included similar proportions of bookings for the full flex product and non-flexible products from the Week category. As a result, a model specification with two latent classes was preferred.

5.2.2. Class-Specific Choice Models

All the parameter estimates relative to the choice of a fare product are of the expected sign. The fare coefficient is negative and significant for both latent classes. The parameter estimates of the dummy variables used to represent the lack of flexibility to change travel plans are also negative as the fully flexible but more expensive fare product is used as the base. In addition, the magnitude of these coefficients increases with the required level of advance purchase. This is also in line with our expectations meaning that the disutility associated with the lack of flexibility to change travel plans increases as travelers are required to further anticipate their travel plans and are likely to face a higher degree of uncertainty about their future schedule.

In order to compare how behavior regarding the lack of flexibility to change travel plans varies across latent classes, the ratio of these coefficients to the fare coefficient was calculated and is shown in Figure 5-2 below. Its absolute value can be interpreted as how much airline travelers are willing to pay for the opportunity to change their travel plans without penalty. The values obtained are largely higher for latent class 1 than for latent class 2 for all levels of required advance purchase meaning that business-style travelers are always willing to pay more for the flexibility to change their travel plans without penalty than leisure-style or price-conscious business travelers. This is in line with our expectations as business travelers tend to have less control over their schedule and are more likely to need to either cancel their trip or modify their itinerary.

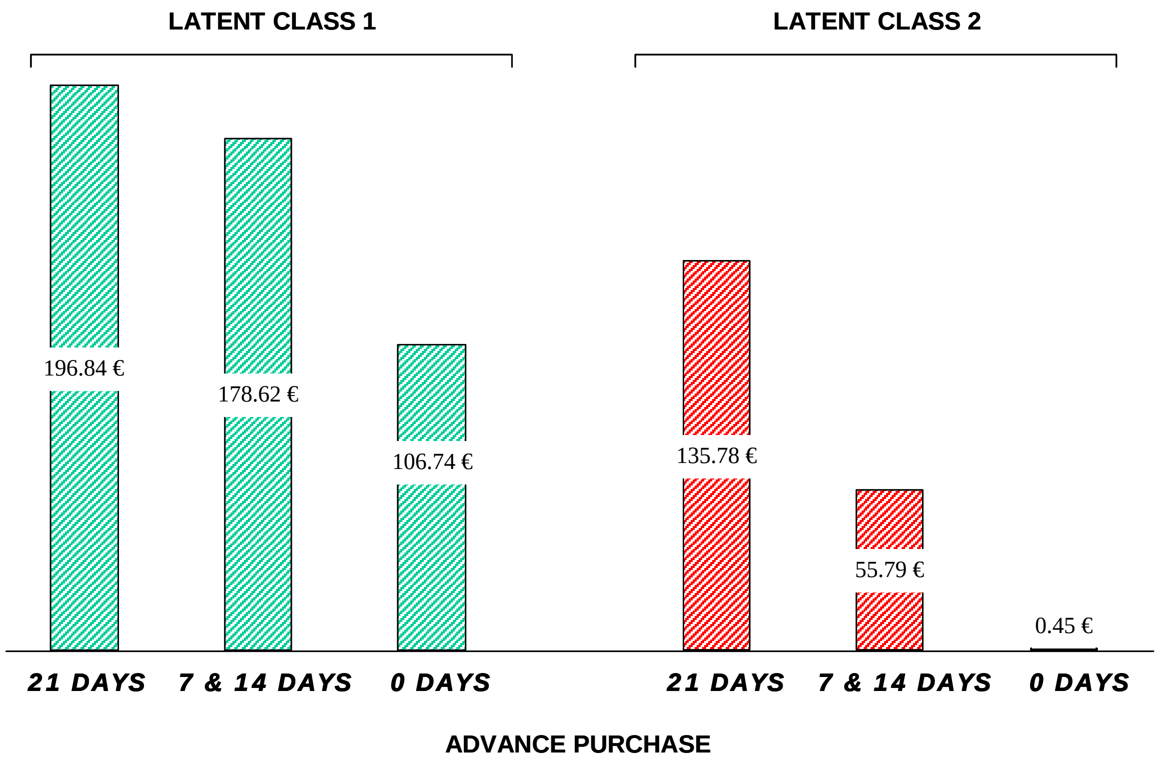


Figure 5-2: Willingness to Pay for the Flexibility to Change Travel Plans without Penalty

We now consider the parameters relative to the continuous function of time used to represent the choice of an airline itinerary or flight departure time. For both overnight and day trip bookings, the number of parameters of the trigonometric formulation was determined empirically based on the resulting profile of the function and the statistical significance of the parameters. A trigonometric formulation with eight parameters was preferred for overnight bookings while only two parameters could be identified for day trip bookings as the number of observations is reduced to the relatively few morning flight departures. A trial and error method with half-hour increments was used to determine the duration of the cycle that maximizes the log-likelihood of the model. A 16-hour duration was found for overnight bookings while a shorter 9-hour cycle is used for day trip bookings.

The parameter estimates are then used to calculate the fitted value of the function for any flight departure time. For latent class 1 and overnight bookings, the value of the continuous function of time is calculated as follows:

$$U(h) = 0.33 * \sin(\frac{2\pi h}{16}) + 0.55 * \sin(\frac{4\pi h}{16}) + 0.09 * \sin(\frac{6\pi h}{16}) + 0.24 * \sin(\frac{8\pi h}{16})$$
$$- 0.15 * \cos(\frac{2\pi h}{16}) + 0.68 * \cos(\frac{4\pi h}{16}) - 0.33 * \cos(\frac{6\pi h}{16}) - 0.03 * \cos(\frac{8\pi h}{16}) \qquad (5.1)$$

As discussed earlier, the start time of the cycle has no impact on the value of the function for a specific flight departure time and is selected at a local minimum so that the function starts and ends at a local low point. For overnight bookings, the start time of the cycle was set at 5 a.m. for latent class 1 and 7 a.m. for latent class 2 based on half-hour increments.

The willingness to pay for a specific departure time is then obtained by dividing this fitted value by the estimate of the fare coefficient. Figure 5-3 below shows the estimated willingness to pay curves for overnight bookings. As expected, business-oriented (latent class 1) outbound travelers tend to prefer either a morning departure to conduct their business activities during the day at the destination or a late afternoon departure so that they can take advantage of most of the day at the origin city before heading to the airport.

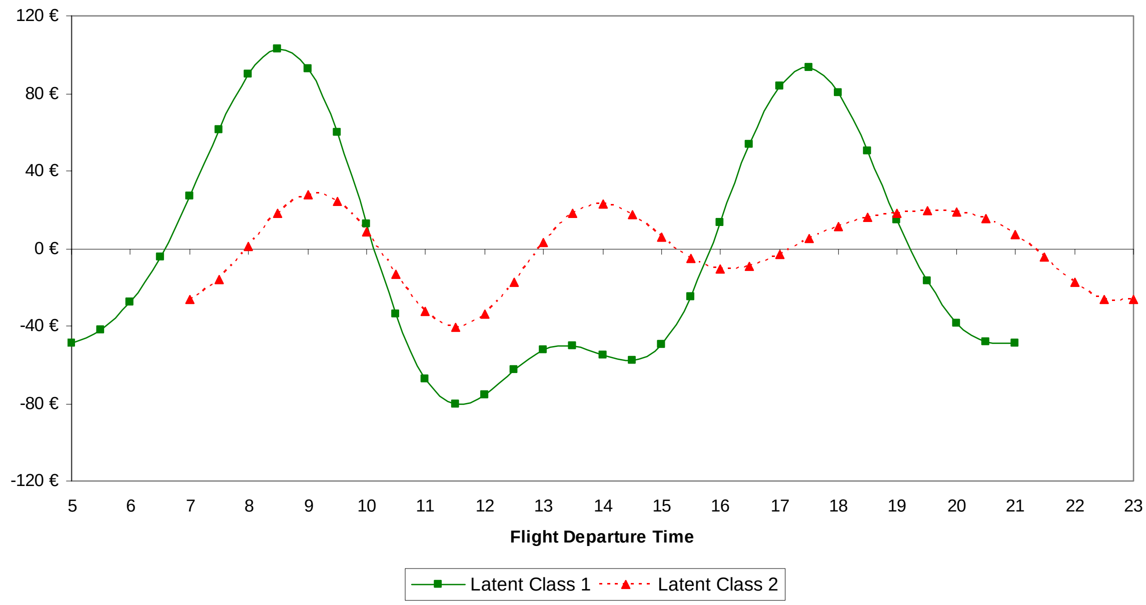


Figure 5-3: Willingness to Pay for a Flight Departure Time for Overnight Bookings

The time-of-day preferences of latent class 2 travelers have similarities and differences with those of latent class 1 travelers. Class 2 travelers also value morning flight departures, but with a peak point about 30 minutes later than class 1 travelers. They also have a preference for late afternoon flights although the peak period starts later than for class 1 travelers, is wider and extends into the evening. As they mostly travel for pleasure, many class 2 travelers may have to work for a full day before going to the airport and catch a flight in the evening while class 1 travelers may be able to leave the office earlier, catch a previous flight and arrive early enough for a late dinner in the destination city. In addition to peaks in the morning and in the afternoon, the willingness to pay curve of class 2 travelers has a third peak in the early afternoon. This probably corresponds to leisure-oriented travelers that take the afternoon off to benefit from a cheaper fare on a low-demand flight in the middle of the day and spend a full evening at the destination.

The other major difference between latent class 1 and 2 travelers is the magnitude of their respective willingness to pay curves. Price-conscious class 2 travelers are willing to pay far less for a specific departure time than business-oriented class 1 travelers. This is consistent with our expectations of the behavioral differences between time-sensitive business travelers and price-sensitive leisure travelers. However, the estimation results also indicate that class 2 travelers are not entirely focused on price and are willing to pay a premium for a specific flight departure time although only about one third that of class 1 travelers.

In addition to differences in the time-of-day preferences of overnight bookings across latent classes, differences in behavioral characteristics were also found within each latent class between day trip and overnight bookings. As day trip bookings primarily belong to latent class 1, Figure 5-4 below shows the willingness to pay curves for overnight and day trip bookings for class 1 travelers. There are two differences between the willingness to pay curves of day trip and overnight bookings: The peak of the curve is reached about 30 minutes earlier and the magnitude of the peak is higher for day trip bookings. This matches our expectations as outbound day trippers are likely to have a stronger preference for an early morning flight departure in order to maximize the time available during their short stay in the destination city.

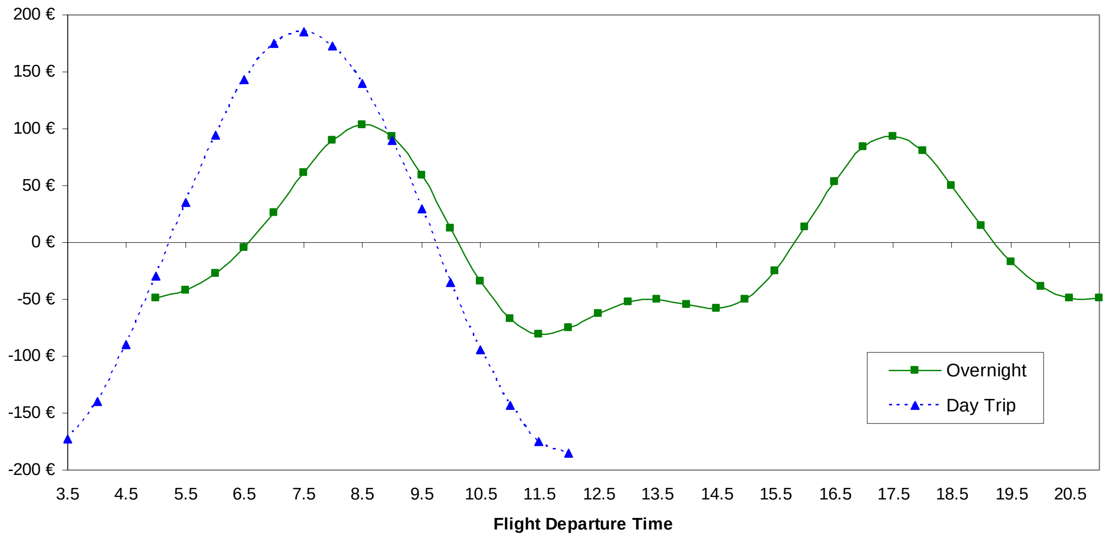


Figure 5-4: Willingness to Pay for a Flight Departure Time – Overnight and Day Trip Bookings (Latent Class 1)

The addition of a function of time specific to day trip bookings improved substantially the fit of the model as shown in Table 5-2 below.

Day Trip Booking Variables	Time-of-Day (TOD)	None
Log L	-3142.41	-3174.44
Number of Parameters	34	29
AIC	-3176.41	-3203.44
BIC	-3271.75	-3284.76

Table 5-2: Day Trip Bookings Explanatory Variables

The maximum log-likelihood of the model improved by more than 30 units although the number of estimated parameters increased by five including the estimated duration of the cycle for day trip bookings. The addition of a function of time to represent the specific time-of-day preferences of day trippers is then supported by model selection criteria such as the AIC and the more parsimonious BIC.

5.3. Benefits over Previous Models of Airline Passenger Choice

Let us now compare the estimation results of the latent class model of airline passenger choice to previous model specifications based on time-period dummies and a deterministic segmentation of the demand.

For models with time-period dummies, the day was divided into six time periods as the airline schedule is organized around six connecting banks with flights leaving to all destinations at about the same time within a bank. A different set of dummies is used for overnight and day trip bookings. In order to have a common base for both types of booking, the morning period (8 to 11 a.m.) is used as the base. Figure 5-5 below compares the willingness to pay curves of a continuous function and time-period dummies for overnight latent class 1 bookings.

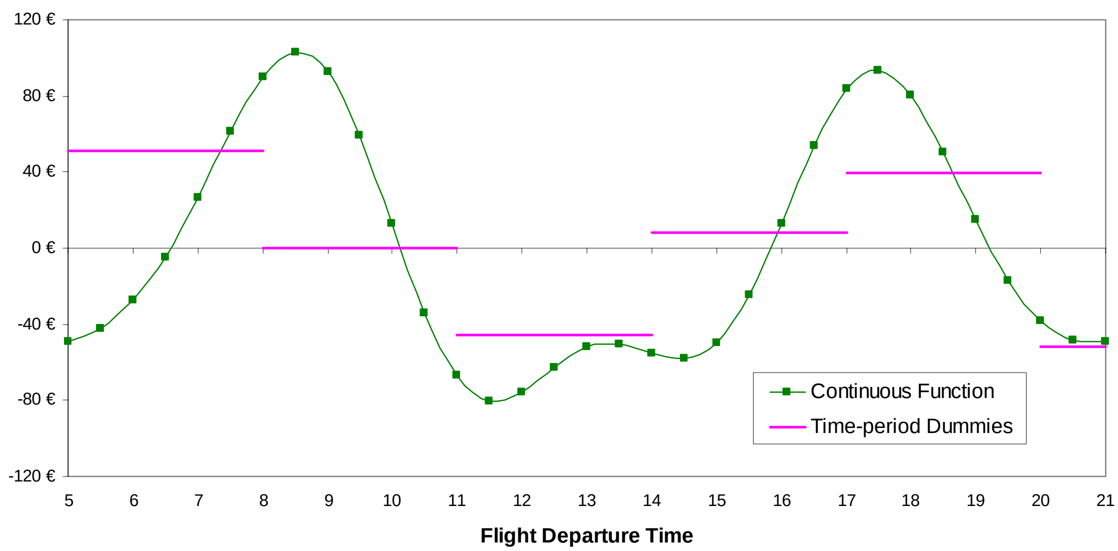


Figure 5-5: Willingness to Pay for a Flight Departure Time – Continuous Function of Time and Time-period Dummies (Latent Class 1 Overnight Bookings)

The resulting time-of-day preferences are similar for both models. However, a continuous function of time leads to a more precise measurement of willingness to pay values across flight departure times while time-period dummies appear to provide a form of average value for each period of the day.

The benefits of a continuous function of time are reflected in the large improvement in the fit of the model. As shown in Table 5-3 below, the maximum log-likelihood of the model improved by almost 60 points although ten additional parameters were used including the estimated duration of the cycle for both day trip and overnight bookings. This more than compensates for the increase in the number of parameters according to both the AIC and BIC criteria. The magnitude of the impact of a continuous function of time on the log likelihood of the model is related to how flight departures are scheduled in these airline markets. Given that the flight departure times are concentrated over a couple of connecting banks, the number of time-period dummies that can be specified is relatively limited. This leaves a large potential for improvement in the fit of the model when a continuous function of time is introduced. A continuous function of time can then provide a more precise estimate of the time-of-day preferences of travelers despite the relatively low level of variability in observed flight departure times.

	Latent Class Continuous Function	Latent Class Time-Period Dummies	Deterministic Continuous Function
Log L	-3142.41	-3200.99	-3172.65
Number of Parameters	34	24	30
AIC	-3176.41	-3224.99	-3202.65
BIC	-3271.75	-3292.29	-3286.78

Table 5-3: Benefits of the Latent Class Model of Airline Passenger Choice with a Continuous Function of Time

In order to assess the benefits of using latent classes, the estimation results of the latent class model of airline passenger choice are compared to a deterministic benchmark. In this benchmark model, the bookings are divided deterministically between two segments defined by one of the three factors used in the class membership model. Segmentation between bookings for travel within a week (MON to FRI) and non-week bookings was found to provide the best fit to the data.

Similar behavioral patterns are observed for the latent class choice model and the deterministic benchmark. However, the magnitude of the parameter estimates tends to be higher for the latent class choice model. For instance, the magnitude of the morning and afternoon peaks was found to be greater for the latent class approach although the peaks are observed at very similar times for both models as shown in Figure 5-6 below. Similarly, latent class 1 travelers are willing to pay more for the flexibility to change their travel plans without penalty than week travelers in the deterministic benchmark model.

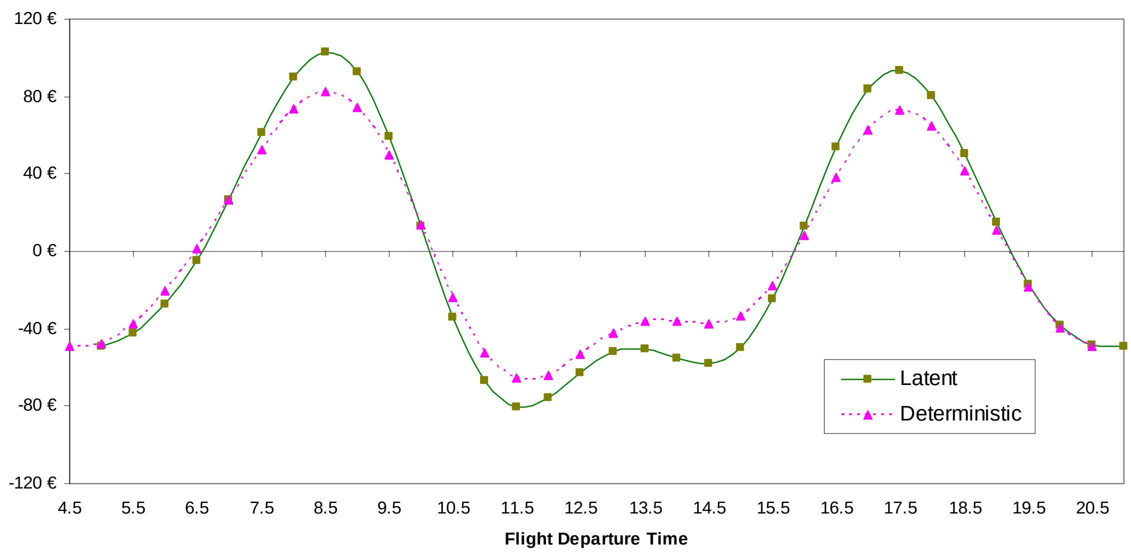


Figure 5-6: Time-of-Day Preferences in the Latent Class and the Deterministic Benchmark Models (Latent Class 1 Overnight Bookings)

Thus, the latent class choice model appears to provide a more distinct segmentation between time-focused business-style travelers and a mixed class of leisure-oriented and price-conscious business travelers. As mentioned earlier, about 35% of the bookings for products from the Week category are classified as belonging to price-sensitive latent class 2 while these bookings are considered as Week bookings in the deterministic benchmark model. As a result, the latent class choice model better represents the heterogeneity of behavior within the business segment of the demand by incorporating the difference in characteristics and the higher degree of price sensitivity from travelers preferring a less expensive albeit non-flexible product from the Week category. This more accurate segmentation of the demand is reflected into an improvement of almost 30 units of log-likelihood in the fit of the model that more than compensates for the four additional parameters used as explanatory variables of the class membership model according to both the AIC and BIC criteria.

6. Applications

Passenger choice models are a powerful tool to evaluate a range of airline planning decisions such as schedule planning, pricing and revenue management. Actually, the impact of any airline planning decision that involves either a change in the attributes of the alternatives or the choice set of the passengers can be analyzed using passenger choice models. However, the use of passenger choice models to support airline planning decisions has been fairly limited up to now. Since the previous studies of airline passenger choice using booking data have focused on the choice of an airline itinerary, the few airline applications developed so far have been related to schedule planning. In particular, Coldren (2005) has shown that the implementation by a U.S. major airline of a logit-based itinerary share model lead to a substantial improvement in the forecasting accuracy compared to previous models based on a quality-of-service index (QSI) methodology. Coldren further improved his model by exploring the complexity of substitution patterns across itineraries using models of the GEV family.

Although the model developed in this research does not allow for the type of substitution patterns provided by models from the GEV family, it improves on previous itinerary share models in several ways. Since it simultaneously models the choice of an airline itinerary and fare product, it captures the impact of fare product characteristics on substitutions patterns across flight departures. The mix of fare products is likely to vary across flights due to differences in passenger mix and the impact of revenue management controls. For instance, morning and late afternoon flights in short-haul markets appear particularly popular with business passengers while midday flights may attract more leisure-oriented demand. The revenue management system tends to reinforce the differences in passenger mix across flight departures by saving seats for late booking high-yield demand on peak flights and channeling low-yield demand to off-peak flights.

The model developed in this research also incorporates the impact of heterogeneity of behavior on substitution patterns across flight departures. Heterogeneity of behavior is expected to have a significant impact as the estimation results of the model show that core business passengers and day trippers are willing to pay a much higher premium to fly on their preferred departure time and far less willing to switch to other flights than the rest of the flying public. Combined with the differences in passenger mix across flights, the model developed in this research has then the potential to substantially improve the accuracy of itinerary share forecasts by capturing the difference in the time-of-day preferences across the two segments of the market.

In addition to existing applications to schedule planning, the changes brought by the rapid growth of low-cost airlines and amplified by web-based distribution of airline tickets have increased the need for new applications of passenger choice models to support a wider range of airline planning decisions such as pricing and revenue management. For instance, the relaxation of fare rules such as the Saturday night stay requirement due to growing low-cost competition has lead to the development of a new pricing strategy. Some airlines have started to offer several product packages differentiated by their characteristics such as the flexibility to change travel plans or advance seat selection rather than their fare rules. Air Canada has been a pioneer of this new pricing strategy and now offers four categories of branded products along with a menu of optional services also called a la carte pricing. Passenger choice models can provide valuable insight into the willingness to pay for the different product characteristics and optional features proposed to the travelers. For instance, the model developed in this research provides an estimate of how much passengers are willing to pay for the flexibility to change travel plans without penalty at different stages of the booking process for both latent classes of the model.

The relaxation of fare rules such as the Saturday night stay requirement not only stimulated the development of new pricing strategies, it also disrupted the revenue management models implemented by network airlines over the last two decades. Standard revenue management models such as the Expected Marginal Seat Revenue (EMSR) model proposed by Belobaba (1987) are based on the independent demand assumption. They assume that the airline is able to achieve through an appropriate set of fare rules a perfect segmentation of the market and that demand for higher-priced products does not depend on the availability of lower-priced albeit more restricted products. The relaxation of fare rules means that the independent demand assumption is increasingly violated in a growing number of markets, leading to a gradual spiral-down in revenues when EMSR-type models remain used (Boyd et al., 2001; Cooper et al., 2006).

As a result, most of the airline revenue management research in recent years has focused on developing new models that relax the independent demand assumption and incorporate sell-up behavior (Talluri and Van Ryzin, 2004; Fiig et al., 2005; Gallego et al., 2007). Sell-up occurs when a passenger able to satisfy the rules of a lower-priced product is also willing to purchase a higher-priced product if the lower-priced product is no longer available. However, all these models require as inputs an estimate of sell-up behavior. Passenger choice models provide an attractive approach to evaluate sell-up potential as choice probabilities for each travel alternative can be easily re-calculated when the passenger choice set is modified. Sell-up behavior can then be modeled through removing a lower-priced product from the passenger choice set and replacing it by a higher-priced product on the same itinerary.

Since it models the choice of a fare product and incorporates the impact of pricing and revenue management on the choice set of each booking, the parameter estimates of the model can be applied to estimate sell-up behavior. For instance, the sell-up rate from the cheapest fare product from the Week category (AWEEK21, 168 EUR) to the next higher-priced product (UWEEK14, 217 EUR) was calculated for an early morning departure that is highly valued by many business-oriented travelers. Figure 6-1 below shows the estimated sell-up rates for both a model specification based on a deterministic segmentation of the demand between week and non-week bookings and the two-class latent class choice model.

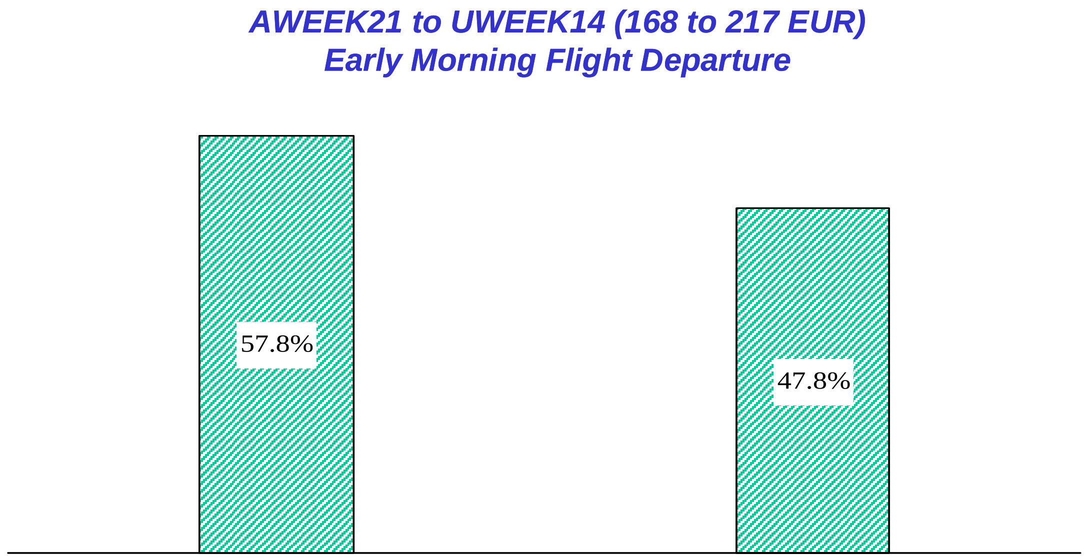


Figure 6-1: Estimated Sell-up Rates

The expected sell-up rate is lower for the latent class choice model than for the model based on a deterministic segmentation of the demand. While the deterministic model classifies the bookings for this non-flexible business-type product as primarily time-sensitive week bookings, the latent class choice model splits them between the two latent classes reflecting the higher degree of price sensitivity of this group of business travelers. As a result, the airline revenue management system will protect fewer seats for the higher-priced product if the sell-up potential is estimated according to the latent class choice model.

7. Summary and Conclusions

In this paper, we estimated a model of the choice of an airline itinerary and fare product that is based on actual airline booking data and is not subject to the risks of bias associated with the hypothetical nature of stated preference data. We developed an approach to overcome the two major limitations of booking data:

the lack of information about the availability of non-booked alternatives and the trip purpose that is typically used to segment demand in airline markets.

In order to reconstitute the choice set of each booking with a good level of accuracy, we developed a methodology to incorporate the impact of airline pricing and revenue management by combining booking with fare rules and seat availability data. While such a process proved effective to infer the choice set for a limited number of bookings and produced a unique dataset suitable for estimation purposes, any large-scale application of the model developed in this research would require implementing a new approach to data collection and management. As a result of this research, Amadeus has initiated a project to collect full choice set data and store it directly in all booking records for airlines hosted in its new inventory system called Altea Plan. These efforts will provide the basis for a potential full-scale implementation of this model.

In addition, we also exploited information available in booking records to replace trip purpose and segment the demand using multiple factors such as the characteristics of the trip and traveler in a latent class model of airline passenger choice. Compared to previous model specifications based on a deterministic segmentation of the demand, the latent class approach provides the flexibility to segment demand using several factors without dividing the bookings in a large number of small sub-segments that can be difficult to interpret. In addition to an improvement in the fit of the model, the latent class structure of the model leads to a more intuitive segmentation of the demand between a core of time-sensitive business travelers and a mixed class of price-conscious leisure and business travelers.

This model of airline passenger choice extends the scope of potential applications of passenger choice models to additional airline planning decisions, such as pricing and revenue management. As the challenges brought by increased competition and escalating energy and environmental-related costs continue to add pressure on this low-margin industry, airlines will increasingly need to exploit every opportunity to increase revenues through science-based pricing and a new generation of choice-based revenue management systems. The calibration of passenger choice models as illustrated in this research will then be crucial to support the development and optimize the effectiveness of these new revenue-enhancing strategies.

References

Adler T., Falzarano C.S., Spitz G. Modeling Service Trade-Offs in Air Itinerary Choices. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. *In 2nd International Symposium on Information Theory*, Petrov B.N. and Csaki F., eds., Akademiai Kaido, Budapest, 1973, pp 267-281.

Akaike H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control,* Vol. 19, No. 6, 1974, pp 716–723.

Belobaba P. *Air Travel Demand and Airline Seat Inventory Management.* Ph.D Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1987.

Ben-Akiva, M., Abou-Zeid, M. Methodological Issues in Modeling Time-of-Travel Preferences. Working Paper Presented at the 11th World Conference on Transport Research Society (WCTR), Berkeley, CA, June 2007.

Ben-Akiva, M., Lerman, S. *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, Cambridge, MA, 1985.

Boyd E.A., Kambour E., Koushik D., Tama J. *The Impact of Buy-Down on Sell-up, Unconstraining and Spiral-down.* Presented at the First Annual Meeting of the INFORMS Section on Revenue Management and Pricing, Columbia University, New York, 2001.

Coldren G.M. Modeling the Competitive Dynamic Among Air-Travel Itineraries with Generalized Extreme Value Models. Paper Presented at the 45th Annual Agifors Symposium, Sao Paulo, Brazil, 2005.

Cooper W., Hohem-de-Mello T., Kleywegt A. Models of the Spiral-Down Effect in Revenue Management. *Operations Research*, Vol. 54, No. 5, 2006, pp. 968-987.

Fiig T., Hopperstad C., Isler K., Cleaz-Savoyen R. *DAVN-MR: A Unified Theory of O&D Optimization in a Mixed Network with Restricted and Unrestricted Fare Products.* Presented at the Annual Meeting of the AGIFORS Revenue Management and Distribution Study Group, Cape Town, South Africa, June 2005.

Galindo-Garre F., Vermunt J.K., Bergsma W. P. Bayesian Posterior Estimation of Logit Parameters with Small Samples. *Sociological Methods and Research*, Vol. 33, No. 1, 2004, pp 88-117.

Gallego G., Li L., Ratliff R. *Revenue Management with Customer Choice Models.* Presented at the Annual Meeting of the AGIFORS Revenue Management and Distribution Study Group, Jeju, South Korea, May 2007.

Garrow L.A., Jones S.P., Parker R.A. How Much Airline Passengers Are Willing to Pay: An Analysis of Price Sensitivity in Online Distribution Channels, *Journal of Revenue and Pricing Management*, Vol. 5, No. 4, 2007, pp 271-290.

Grammig J., Hujer R., Scheidler M. Discrete Choice Modelling in Airline Network Management. *Journal of Applied Econometrics*, Vol. 20, No. 4, 2005, pp. 467-486.

Prossaloglou K., Koppelman F.S. The Choice of Carrier, Flight and Fare Class. *Journal of Air Transport Management*, Vol. 5, No. 4, 1999, pp. 193-201.

Schwarz G. Estimating the Dimension of a Model. *Annals of Statistics,* Vol. 6, No. 2, 1978, pp 461-464.

Smith B. *Helping Travelocity Transition to Travel Retailer.* Presentation to the MIT Operations Research Seminar, Cambridge, MA, September 2006.

Talluri K., Van Ryzin J. Revenue Management Under a General Discrete Choice Model of Consumer Behavior. *Management Science*, Vol. 50, No. 1, 2004, pp 15-33.

Theis G., Adler T., Clarke J., Ben-Akiva M. Risk Aversion to Short Connections in Airline Itinerary Choice. *Transportation Research Record*, Vol. 1951, 2006, pp. 28-36.

Vermunt J.K., Magdison J. *Technical Guide for Latent Gold Choice 4.0: Basic and Advanced.* Statistical Innovations Inc., Belmont, MA, 2005.

Vermunt J.K., Galindo Garre F. Avoiding Boundary Estimates in Latent Class Analysis by Bayesian Posterior Mode Estimation. *Behaviormetrika*, Vol. 33, No. 1, 2006, pp 43-59.

Acknowledgements

The author would like to acknowledge the financial support of the Sloan Foundation and the Airline Industry Consortium through the MIT Global Airline Industry Program. In addition, the author would like to acknowledge the support of Amadeus SAS for providing access to the data used for this research as well as for its financial support. Finally, I would like to acknowledge the guidance provided by the members of my doctoral committee, in particular my thesis supervisors Moshe Ben-Akiva and Peter Belobaba.