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

by

Emmanuel Carrier

Master of Science in Transportation, Massachusetts Institute of Technology, 2003
Master of Arts in Economics, Université Paris-I Panthéon-Sorbonne, 1997
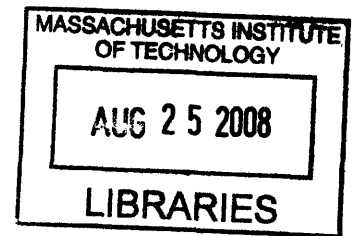Master of Science in Management, HEC School of Management, 1996

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Transportation

at the

Massachusetts Institute of Technology

JUNE 2008

Signature of Author...   ....
Department of Civil and Environmental Engineering
May 23, 2008

Certified by.............................../   ....
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by..
Peter P. Belobaba
Principal Research Scientist of Aeronautics and Astronautics
Thesis Supervisor

Accepted by ....
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

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

by

Emmanuel Carrier

Abstract

Over the last ten years, the rapid growth of low-cost airlines and the development of web-based distribution of airline tickets have transformed the competitive environment in the airline industry worldwide. The relaxation of fares rules by low-cost airlines has disrupted the pricing and revenue management models of large network airlines. A better understanding of passenger choice behavior is now required to support the development of new strategies to compete more effectively in the current marketplace.

In order to avoid the risk of bias associated with stated preference data, we focus in this research on how to develop a model of airline passenger choice based on booking data. Previous studies based on booking data have been limited to the sole choice of an airline itinerary and did not account for heterogeneity of behavior, a major characteristic of airline markets. This is due to the properties of booking data. For instance, only the chosen alternative is recorded in airline bookings and no information is available on other travel alternatives available at the time of the booking. Similarly, booking records contain no information on trip purpose that is traditionally used to segment airline markets. In this dissertation, we develop a modeling framework to overcome these limitations and extend booking-based passenger choice models to the joint choice of an airline itinerary and fare product.

Booking data was combined with fare rules and seat availability data to incorporate the impact of pricing and revenue management and reconstruct the choice set of each booking. Characteristics of the traveler and the trip were retrieved from the booking records and used to replace trip purpose. They were included as explanatory variables of a latent class choice model in which several

factors can be used simultaneously to segment the demand without necessarily dividing the bookings into many small sub-segments. In addition, a new formulation of a continuous function of time was proposed to model 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 account for the low attractiveness of some periods of the day such as nighttime.

Estimation results over a sample of 2000 bookings from three European short-haul markets show that the latent class structure of the model and a continuous function of time led to a significant improvement in the fit of the model compared to previous specifications based on a deterministic segmentation of the demand or time-period dummies. In addition, the latent class model provides a more intuitive segmentation of the market between a core of time-sensitive business travelers and a mixed class of price-conscious business and leisure travelers.

This research extends the scope of potential applications of passenger choice models to additional airline planning decisions such as pricing and revenue management. In particular, parameter estimates of the model were applied to forecast the sell-up behavior of airline passengers, a major input required by the newly proposed revenue management models designed to maximize revenues under less restricted fare structures.

Thesis Committee:

> Moshe E. Ben-Akiva, Edmund K. Turner Professor of Civil and Environmental Engineering, Thesis Supervisor

> Peter P. Belobaba, Principal Research Scientist of Aeronautics and Astronautics, Thesis Supervisor

> Cynthia Barnhart, Professor of Civil and Environmental Engineering and Engineering Systems and Associate Dean for Academic Affairs of the School of Engineering

> John-Paul B. Clarke, Associate Professor of Aerospace Engineering, Georgia Institute of Technology

Acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1   Introduction

The rapid growth of low-cost airlines over the last decade has changed the competitive landscape in many airline markets around the world. At the same time, the development of web-based distribution channels of airline tickets has made it easier for prospective travelers to compare the various travel options offered in the marketplace and has amplified the impact of low-cost competition. While the traveler has strongly benefited from new competition and the shift to web-based distribution channels, these changes have posed a considerable challenge to established network airlines. At the end of 2005, four out the six U.S. network airlines operated under bankruptcy protection.

Low-cost airlines not only generally offer lower fares, they also disrupt the pricing and revenue management strategies of network airlines by relaxing fare rules such as the Saturday night stay requirement of many discounted fare products. In this context, a better understanding of passenger choice behavior is needed to support the development of new pricing and revenue management strategies and compete more effectively in today's marketplace.

However, previous studies of airline passenger choice did not accurately represent the major characteristics of airline markets such as the impact of pricing and revenue management or the heterogeneity of behavior across different segments of the market. In this dissertation, we will then focus on how an airline can use its existing data sources to develop and estimate a model of airline passenger choice that better reflects the characteristics the choice environment in the airline industry.

1.1 Motivation and Research Objectives

Following the economic deregulation of the industry in 1978, the major network airlines pioneered a scientific approach to pricing based on product differentiation and the dynamic management of seat capacity referred to as revenue management. While airline pricing and revenue management are based on the heterogeneity of choice behavior across different categories of travelers, the use of passenger choice models to directly support airline planning applications has been relatively limited so far.

Previous studies of airline passenger choice have been based on two types of data, stated preference and booking data. In studies based on stated preference data, respondents are usually asked to choose between a limited set of two to three hypothetical alternatives. These hypothetical travel alternatives are designed to reproduce the typical impact of airline pricing and revenue management on the passenger choice set and supposed to be representative of the products offered in the marketplace. Nevertheless, the design of the experiment tends to overly simplify the passenger choice set and cannot reproduce the large number of travel alternatives viewed by prospective travelers in a real booking search. In addition, these studies suffer from the risk of discrepancy between the responses obtained in the context of a hypothetical scenario and actual passenger choice behavior. This is especially true for pricing applications when these studies are used to investigate the trade-off between price and other elements of airline service such as schedule or amenities.

While the impact of pricing and revenue management on the passenger choice set is modeled in a simplistic manner in studies based on stated preference data, it is entirely ignored in previous studies based on booking data. Previous studies based on the analysis of past booking data have been limited to the sole choice of

an airline itinerary and used for schedule planning applications. This is due to the properties of booking data: Only the booked alternative is 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 easily obtained 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 constantly updated 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 challenges brought by the rapid growth of low-cost airlines and amplified by the development of web-based distribution channels have also raised the interest for a better understanding of passenger choice behavior and new applications of passenger choice models to other areas of airline planning such as revenue management. By relaxing fare rules such as the Saturday night stay requirement, a new generation of low-cost airlines such as Jetblue in the U.S. or Ryanair in Europe has disrupted the foundations of the revenue-maximizing strategy developed by network airlines since the economic deregulation of the industry in 1978.

This strategy was based on two elements: product differentiation and revenue management. In order to attract more leisure passengers, the airlines started to offer lower discounted fares such as the Super Saver fares introduced by American Airlines in 1977. However, in order to prevent business passengers willing to purchase more expensive fares, discounted fares were systematically associated with a set of restrictions, also called fare rules, such as required round-trip travel, minimum stay (Saturday night stay), advance purchase requirements, change fees and non-refundability. In order to channel low-fare demand to off-peak flights and save seats for late-booking high-fare passengers on peak flights, product differentiation was supported by the dynamic management of airline seat inventory, also called revenue management. The expected marginal seat revenue (EMSR) model developed by Belobaba (1987) that has been widely used in the industry, is based on the assumption that the market is perfectly segmented and demand for the different fare products is independent. Thus, the availability of lower-fare products is assumed to have no impact on the demand for higher-fare products.

While it was recognized early on that the independent demand assumption was unlikely to fully hold in practice, fare rules such as the Saturday night stay requirement proved fairly effective to segment demand and prevent business passengers from purchasing lower-priced fare products. The relaxation of fare rules means that this assumption can no longer be considered valid in many airline markets and seriously undermines the revenue performance of EMSR-type models. Several alternative models have been proposed to maximize revenues under less restricted fare structures. However, they all require as inputs an estimate of sell-up behavior or how likely a passenger is to purchase a more expensive fare product if the revenue management system decides to no longer offer a lower-priced product. Discrete choice models provide a convenient approach to estimate sell-up potential as choice probabilities can be easily

recalculated when an alternative is added or removed from the passenger choice set. As a result, passenger choice models provide a means to capture a number of effects including sell-up when the choice set of the passenger is modified to reflect the potential decisions of the revenue management system.

The objective of this dissertation is then to build a model of airline passenger choice that overcomes the limitations of previous studies and extends the scope of potential applications to new areas of airline planning decisions such as pricing and revenue management. Before discussing further the objectives of this research, let us first describe in the next section the characteristics of the choice process in the airline industry.

1.2   The Traveler Choice Process

As shown in Figure 1-1 below, the traveler's choice process for an upcoming trip includes three major steps: Trip planning, booking search and the actual booking itself. In the first step, the prospective traveler defines the major characteristics of his future travel such as the purpose of the trip, the destination, the planned travel dates and the mode of transportation. These different elements are interdependent and can be flexible or not. For instance, if the purpose of the trip is to attend a convention, the venue and the dates of the trip are not set by the traveler but by the conference organizers. As a result, the traveler will have no flexibility over the destination and probably fairly little flexibility over the travel dates. However, if the traveler is planning a vacation sometime over the next couple months, he will have much more control over the selection of the destination and the potential travel dates. The traveler decisions may then be influenced by the results of the second step of the process, the booking search.

```
┌─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─┐        ┌─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─┐
│  ┌─────────────────────┐   │        │  ┌─────────────────────┐   │
│  │    TRIP PLANNING     │   │        │  │   BOOKING SEARCH     │   │
│  └─────────────────────┘   │        │  └─────────────────────┘   │
│                            │        │                            │
│  ┌─────────────────────┐   │        │  ┌─────────────────────┐   │
│  │ Why?  Trip Purpose  │   │        │  │ Distribution Channel│   │
│  ├─────────────────────┤   │        │  ├─────────────────────┤   │
│  │ Where?  Destination │   │        │  │  Seat Availability  │   │
│  ├─────────────────────┤   │  ⟺     │  ├─────────────────────┤   │
│  │  When?  Dates       │   │        │  │       Airline       │   │
│  ├─────────────────────┤   │        │  ├─────────────────────┤   │
│  │  How?  Mode         │   │        │  │    Fare Product     │   │
│  └─────────────────────┘   │        │  ├─────────────────────┤   │
│                            │        │  │  Itinerary/Schedule │   │
│        Unobserved          │        │  └─────────────────────┘   │
│                            │        │    Partially Unobserved    │
└─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─┘        └─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─┘
                                                      ⇓
                                           ┌─────────────────────┐
                                           │      BOOKINGS       │
                                           └─────────────────────┘
```
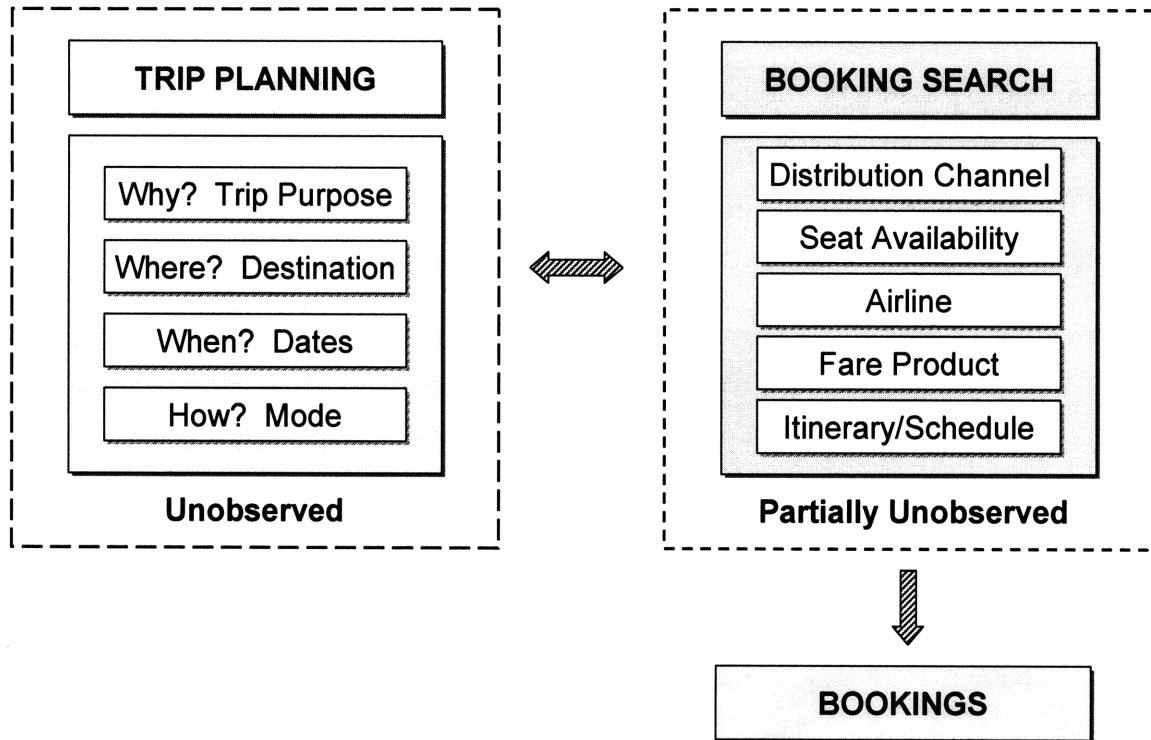
Figure 1-1:   The Traveler Choice Process

If air is considered as a potential mode of transportation, the trip planning phase of the process will trigger a booking search. The prospective traveler will then contact a travel agent or go online to a travel website and initiate a search for seat availability based on the characteristics of the trip set during the planning phase, such as destination and travel dates. Based on the results of the search, the traveler may revise some of his initial decisions. For instance, the traveler may adjust the dates or even the destination of a future trip in order to take advantage of a very attractive fare. After a number of iterations, the traveler will take one of the following decisions: not to travel at all, consider other modes of transportation, postpone his decision to a later time or make a booking.

At this stage of the process, the prospective traveler will then make a final decision about the major dimensions of the booking such as the choice of an airline, itinerary and fare product. For some travelers, the choice of an airline is

actually determined in part prior to the booking search as they are either required to travel on a specific carrier due to corporate travel policies or they restrict themselves to a preferred set of carriers based on airline loyalty programs or past experience. In addition to the carrier providing the service, the traveler will also choose the itinerary or flight schedule of his trip as well as a fare product that will determine the price paid and the characteristics of the product such as the range of amenities provided and the flexibility to change or cancel the booking at a later time.

While the resulting bookings are recorded in the airline reservations system, the different steps of the traveler choice process are mostly unobserved by the airline. For instance, the trip planning phase of the process is fully unobserved and the booking search process is also at least partially unobserved. While a booking search through a travel agent will typically go unrecorded, elements of an online booking search such as screen shots and click activity may be recorded by some travel websites. However, an airline will potentially have full access to this type of data for only a subset of the bookings made through its own website. In addition, the bookings made on the airline's website may be the results of a wider booking search initiated on another travel website that will also remain unobserved to the airline. Actually, as shown by Brunger (2008) in his qualitative study of online booking behavior, many travelers tend to check first one or several of the large online travel websites such as Expedia, Travelocity and Orbitz that market the inventory of many different carriers to gather information before logging to the website of the airline of their choosing in order to avoid the booking fees typically charged by these travel retailers. As a result, web analytics are not commonly used at this point in the airline industry and not in connection with a scientific analysis of passenger choice behavior.

1.3 Research Objectives

As mentioned earlier, we will focus in this research on how to develop a model of airline passenger choice that will provide the foundation for a wide range of potential applications. In order to avoid the potentially high risk of 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. This will provide the basis for the development of a booking-based model that will explore the trade-off between the major dimensions of airline passenger choice such as schedule convenience and price. In addition, since airline pricing and revenue management are largely based on taking advantage of the heterogeneity of behavior across different categories of airline travelers, elements of the booking records such as the characteristics of the trip and the profile of the traveler will be used to segment the demand and investigate the difference in choice behavior across different categories of bookings.

In this research, we will focus on how an airline can exploit its existing data sources 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 when relying on these data sources. For instance, given that no information is available in booking records about the different travel destinations considered by the traveler and data potentially collected during the booking search process is both partial and incomplete, substitution patterns across different origin-destination markets will be ignored. However, this assumption should have a

limited impact for most applications as airlines typically forecast demand based on historical bookings independently for each origin-destination market.

More importantly, since an airline cannot get full access to the bookings of its competitors, the choice of an airline will be taken as given and the model will concentrate on the other major dimensions of airline passenger choice such as the selection of an itinerary and fare product. This assumption may be more restrictive for some applications such as revenue management. For instance, sell-up behavior is likely to depend substantially on the attributes of competing travel alternatives offered in the marketplace. However, we will discuss how to incorporate the impact of competition when applying the parameter estimates of the model to forecast expected sell-up behavior.

1.4   Contributions

The contributions of this dissertation 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 provides the foundation for the development of a model of the choice of an airline itinerary and fare product based on booking data. It also 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.

This dissertation also provides two important contributions to the development of passenger choice models. First, we developed an alternative to the traditional 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 has several advantages over previous models based on a deterministic segmentation of the demand.

Latent classes allow the segmentation of the market using multiple factors without dividing the bookings into a large number of small sub-segments. These factors can be specified as explanatory variables of the class membership model and their parameter estimates provide insight on their respective weight to the segmentation of the market. In addition, 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. 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 than previous models based solely on trip purpose.

Second, we improved the measurement of the time-of-day preferences of airline travelers by using a continuous function of time instead 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 lack of demand for nighttime flights in these markets. This framework also provided a flexible approach to model the

time-of-day preferences of specific categories of travelers such as day trippers that travel exclusively in the morning on outbound flights due to the short duration of their trip.

Finally, this model of the choice of an airline itinerary and fare product extends the range of potential applications of passenger choice models to other areas of airline planning such as pricing and revenue management. In particular, the parameter estimates of the model were used to forecast the sell-up behavior of airline passengers, which is a key input to all newly proposed revenue management algorithms designed to maximize revenues under less restricted fare structures. The latent class choice model was found to provide lower estimates of sell-up potential than models based on a deterministic segmentation of the demand due to a more realistic split of business-type travelers between the time and price-sensitive segments of the market.

1.5 Thesis Outline

The remainder of the dissertation is organized as follows.

Chapter 2 provides an overview of the changes that have occurred in the airline industry over the last ten years, focusing primarily on the impact of growing low-cost competition and the development of web-based distribution channels of airline tickets on the choice environment in airline markets.

In Chapter 3, a literature review of airline passenger choice models is presented including a discussion of the benefits and limitations of different types of data such as stated preference and booking data.

Chapter 4 develops the modeling framework used in this research: We propose a methodology to reconstitute the choice set of each booking based on combining several data sources such as booking and seat availability data. We then develop a latent class model of the choice of an airline itinerary and fare product and propose a new formulation of a continuous function of time to model the time-of-day preferences of airline passengers.

In Chapter 5, we describe the data collected for this research including an exploratory analysis of the booking data as well as a detailed description of how the data was processed to generate the choice set of each booking.

Chapter 6 presents the estimation results of the model and focuses on the benefits of the latent class structure of the model compared to previous specifications based on a deterministic segmentation scheme.

In Chapter 7, we describe the potential applications of the model to a range of airline planning decisions, such as schedule planning, pricing and revenue management. We provide an example of how the parameter estimates of the model can be applied to forecast the sell-up behavior of airline passengers. We propose a framework for integrating passenger choice models and competitor availability data to obtain more accurate forecasts of sell-up potential.

Finally, Chapter 8 summarizes the findings of the research and discusses directions for future research.

Chapter 2   Context and Motivation of the Research

In recent years, the U.S. airline industry has experienced the most significant change since the economic deregulation of the industry in 1978. The rapid growth of low-cost airlines and the development of online distribution of airline tickets have put great pressure on the pricing and distribution strategies developed by the network airlines after deregulation. In this chapter, we first discuss the pricing strategies of low-cost airlines and their impact on the traditional revenue models used in the industry. We then highlight the development of online distribution of airline tickets and finally explore the impact of these changes on booking patterns and the choice behavior of airline travelers.

2.1   Airline Revenue Models

After 1978, the industry structured itself based on two elements: hub and spoke networks and price discrimination. On the supply side, the airlines used hub and spoke networks to consolidate traffic flows and serve with high frequency a large number of medium and small markets that could not sustain frequent non-stop service, a key driver of service quality, especially for business travelers. On the revenue side, since they were now able to set fares freely, the airlines took advantage of the behavioral differences between time-sensitive business travelers and price-sensitive leisure travelers to increase revenues by segmenting the demand and charging a different price to the business and leisure segments of the market. In order to achieve price discrimination and prevent business

passengers from purchasing tickets at discounted fares designed to lure leisure travelers, the airlines built a complex pricing system. Discounted fare products were systematically associated with a set of fare rules designed to make them unattractive to business travelers. In particular, airlines included a Saturday night stay requirement in the set of fare rules applied to most discounted fare products as the travel patterns of relatively few business travelers could satisfy that requirement. In addition, this new pricing strategy was supported by the dynamic management of aircraft seat inventory referred to as airline yield or revenue management. The objective of revenue management is to ensure that seats remain available, even at the last minute, for high-fare business passengers. This is especially important as typical booking patterns show that business passengers, willing to pay the most expensive fares, tend to book after most leisure passengers. Therefore, on high demand flights, revenue management systems limit the availability of lower-priced fare products and keep seats available for expected late-booking passengers willing to purchase the most expensive fares.

It is these two elements of the network airlines' strategy that the low-cost airlines have challenged. In order to take full advantage of the economies of density associated with hub operations, network airlines focused primarily on the development of their hubs and not on the introduction of non-stop service in point-to-point markets. As a result, some low-cost airlines such as JetBlue and Southwest have chosen to some extent a strategy based on providing point-to-point non-stop service. However, some other low-cost airlines also developed a hub-based network structure. For example, Air Tran established a hub in Atlanta where it competes with Delta. Even Southwest has developed over time a network of focus cities in which a large portion of its flight schedule is now concentrated and relies increasingly on connecting traffic for its growth.

More importantly, it is the revenue model i.e. the pricing and revenue management philosophy, policy and practices developed by network airlines that low-cost carriers have been challenging more aggressively in the recent period. During the nineties, the spread between the cheapest and the most expensive fare in a given market increased considerably. For instance, according to a study by the Transportation Research Board (Meyer et al., 1999), while the median fare in short-haul markets (less than 750 miles) decreased by more than 15%, the 95th percentile fare increased from $287 in 1992 to $316 in 1998 and was 3.3 times the median fare in 1998 compared to 2.5 in 1992. A similar trend had also been observed in medium and long-haul domestic markets. During the 1990s, network airlines had gradually increased their most expensive fares designed primarily for business travelers to levels that started to appear excessive. More and more business travelers became increasingly reluctant to purchase these very expensive unrestricted fares and, over the years, an increasing number of business travelers started to rely on ticketing strategies designed to circumvent fare rules such as the Saturday night stay requirement and make it easier to purchase cheaper restricted fare products. The growing dispersion of airfares gave low-cost airlines the opportunity to promote a different revenue model. For instance, David Neeleman, the founder of JetBlue repeatedly claimed that his objective was to bring "humanity" back to air travel.

Launched in early 2000, JetBlue has become the symbol of a new era in airline pricing practices. Although JetBlue, like all other low-cost airlines, is using differential pricing, there are two fundamental differences between its pricing strategy and that of the network airlines after deregulation. First, while JetBlue still offers a large range of fares, the dispersion of its fare structure is typically smaller in absolute terms. For instance, it initially did not charge more than $299 one way in any non-stop market including coast-to-coast markets in which network airlines used to charge up to $1200 one-way prior to its entry.

Apart from a compressed fare structure, the second major and maybe more fundamental difference in pricing practices between low-cost and network airlines are the strategies used to segment the market. In order to segment the market between business and leisure travelers, as mentioned earlier, network airlines have relied primarily on a set of fare rules designed to make discounted fares unattractive to business travelers. In addition to non-refundability and change fees, most low fares on network airlines typically required a Saturday night stay as the travel patterns of business travelers usually prevent them from satisfying that requirement. Although most low-cost airlines such as JetBlue offer primarily non-refundable tickets and charge a fee for changing flights, they do not impose a Saturday night stay requirement. Actually, many low-cost airlines sell tickets on a one-way basis removing de facto any minimum stay requirement.

Low-cost airlines base price discrimination not on restrictive fare rules but on the advance purchase of tickets as well as forecasts of future demand on a flight-by-flight basis. Many low-cost airlines still impose advance purchase requirements on their cheapest fare products. In addition, low-cost airlines also use revenue management to determine which fare product to offer on any particular flight at any particular time based on the current number of bookings and a forecast of future demand. For instance, JetBlue relies primarily on advance purchase of tickets and the level of demand for a given flight to achieve price discrimination.

This pricing strategy may be better accepted by the flying public as it translates to some extent into a single rule, easy to understand and that a large proportion of the traveling public may see as legitimate: The earlier you book, the cheaper the fare. In addition, the price discrimination practices of low-cost airlines are more acceptable to consumers as their most expensive fares are still in many cases relatively affordable and much cheaper than those offered by network

airlines. Thanks to the success of their operating and pricing strategy, low-cost airlines have been expanding rapidly in short-haul and medium-haul markets. Figure 2-1 shows that low-cost carriers gained 10 points of market share in the U.S. domestic market from 2000 to 2006.

U.S. Carrier Group Share of Domestic ASMs
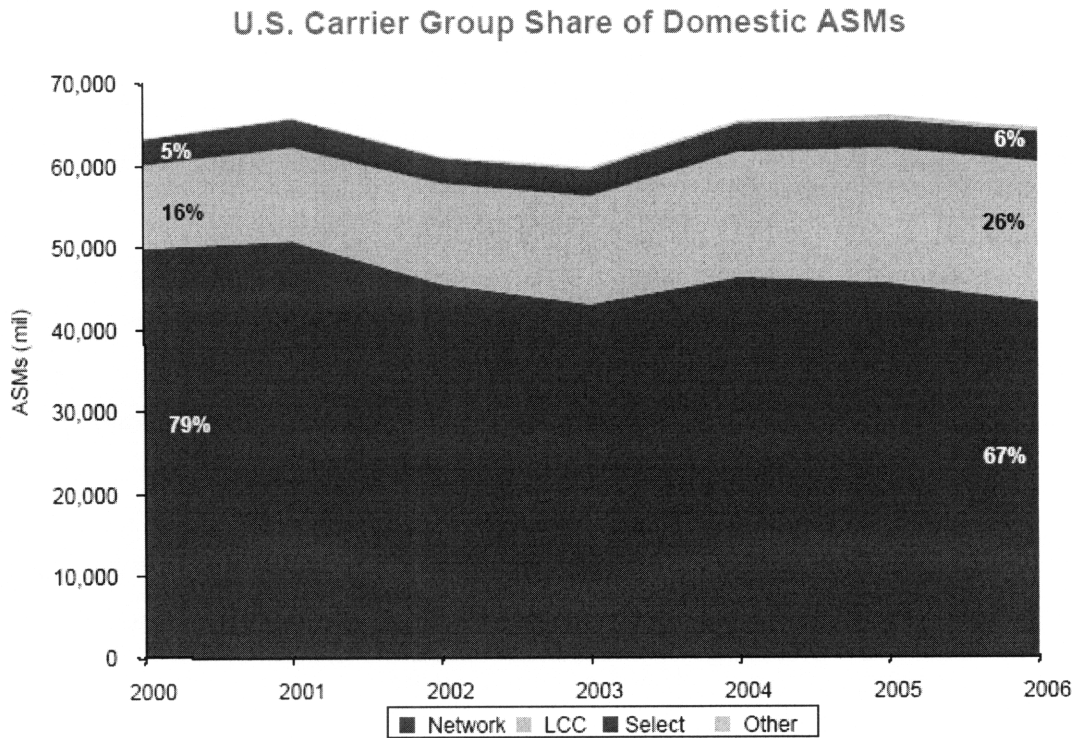
Figure 2-1: Low-Cost Carriers Market Share in the U.S. Domestic Market (Source: Swelbar, 2007)

In addition, over the last couple years, low-cost carriers have developed rapidly in most major U.S. urban centers, including cities outside of their traditional strongholds in Texas and the West. They have been particularly aggressive on the East Coast and in transcontinental markets and have gained a truly national presence covering most of the U.S. major domestic markets. Figure 2-2 below shows the change in low-cost airlines' market share in the 10 largest U.S. Consolidated Metropolitan Statistical Areas (CMSA) between September 1997 and March 2004.

27
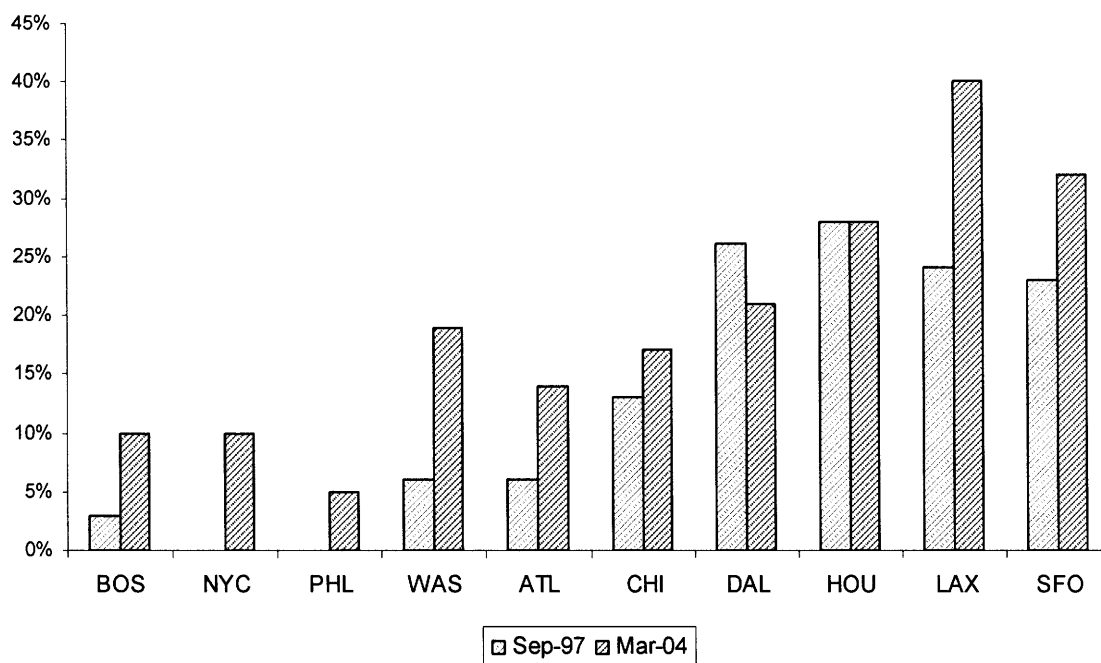
Figure 2-2:   Low-Cost Carriers Penetration in top 10 U.S. CMSA Markets
(Source: ECLAT consulting, Swelbar, 2004)

Faced with the rapid growth of low-cost airlines, especially newcomers such as JetBlue, and the growing popularity of their pricing philosophy, network airlines as well as established low-cost airlines have responded in a variety of ways, but all have aimed at mimicking JetBlue pricing practices on a more or less extended scale. For instance, in August of 2002, Southwest, the first and largest low-cost carrier in the U.S., announced a new system-wide fare cap of $299 one-way decreasing its former cap by $100 and matching JetBlue's cap policy.

As for network airlines, they also responded in a variety of ways but their pricing practices have been increasingly influenced by the "JetBlue" model as low-cost airlines have continued to expand and enter new markets. The first major shift in pricing policies by a network airline was initiated by a relatively small carrier, America West, the only major airline founded after the deregulation of the industry and still in existence as a stand-alone carrier at that

28

time, prior to its merger with US Airways in 2006. Taking advantage of its relatively lower cost structure compared to the rest of the industry and of its experience of head-to-head competition with low-cost airlines such as Southwest at its main hubs in Phoenix and Las Vegas, America West announced, in March 2002, a new system-wide low-cost style pricing policy decreasing maximum fares by 40 to 70 % and removing Saturday night stay requirements on most coach class fares. In February 2004, it extended this pricing philosophy to its first-class cabin with the introduction of two additional discounted non-refundable one-way first class fares. This initiative was followed in 2004 by a similar overhaul of pricing practices by another relatively small network carrier, Alaska Airlines

As for larger network airlines, their initial strategy was to match selectively low-cost airlines in competitive markets. In 2003, as low-cost airlines continued to increase their share of the U.S. domestic market, some network airlines decided to launch new low-cost subsidiaries to compete with low-cost airlines, especially in leisure markets. Despite the history of failure of most former low-cost subsidiaries of traditional network airlines such as Continental Lite, US Airways Metrojet, Shuttle by United and its own low-cost arm Delta Express, Delta launched in April 2003 a new low-cost subsidiary named Song. Song was designed specifically to compete with increasingly popular JetBlue, primarily on routes between the Northeast (Boston, Hartford, New York) and major resort destinations in Florida. Fares on Song were sold one-way, removing de facto minimum stay requirements, and capped at $299, mimicking JetBlue's fare structure. Song fleet was composed of dedicated 757 aircraft taken out of Delta's mainline fleet and refurbished with an all-coach class seating and entertainment systems similar to those installed on JetBlue's aircraft. In 2005, the Song fleet was expanded from 36 to 48 aircraft and Song took over Delta mainline service on transcontinental routes from New York JFK where it also competes with JetBlue.

After filing for bankruptcy protection, Delta retreated in 2006 from that strategy as Song service was discontinued and Song dedicated fleet was reintegrated into Delta's mainline fleet and reconfigured with a two-class layout. Some elements of Song service such as an enhanced entertainment system were retained and offered on Delta's medium-haul domestic flights. Following Delta's strategy, United launched in February of 2004 a new low-cost arm named Ted, replacing United mainline service to leisure destinations such as Las Vegas, Phoenix, Florida and Mexico from its hubs in Los Angeles, San Francisco, Denver, Chicago and Washington D.C. Like Song, Ted has a dedicated fleet of A320 aircraft taken from the United mainline fleet but reconfigured with all-coach seating and uses a low-cost carrier style pricing structure, selling tickets on a one-way basis. In June 2008, United finally decided to discontinue Ted service as well.

In early 2005, Delta announced a system-wide "simplification" of its fare structure in the U.S. domestic market, following an experiment at the airline Cincinnati hub in the fall of 2004. Delta's new fare structure named "Simplifares" still relied on selling round trip tickets but removed all Saturday night stay requirements replacing it by a far less restrictive 1-day minimum stay requirement. In addition, Delta decreased its standard change fee from $100 to $50 and established a cap of $499 one-way for coach class and $599 one-way for first-class tickets in domestic markets.

Delta's new pricing strategy, one of the most spectacular changes in pricing by a large full-service network carrier since the deregulation of the airline industry was harshly criticized by most of its competitors. However, other network airlines matched Delta's pricing overhaul either system-wide (American) or on a relatively large scale in many competitive markets (Continental, Northwest). As a result, there is no Saturday night stay requirement associated with discounted tickets in many U.S. domestic markets. As an illustration of the impact of these

changes, the following table shows the fare structure used by American in the Boston-Seattle market in the fall of 2001 and in the spring of 2004 after its main competitor, Alaska Airlines simplified its fare structure and American matched the change.

American Airlines, October 1, 2001				American and Alaska, May 1, 2004			
Roundtrip Fare	Advance Purchase	Minimum Stay	Change Fee?	Roundtrip Fare	Advance Purchase	Minimum Stay	Change Fee?
$458	21 days	Sat. Night	Yes	$374	21 days	1 day	Yes
$707	21 days	Sat. Night	Yes	$456	14 days	1 day	Yes
$760	21 days	Sat. Night	Yes	$559	14 days	1 day	Yes
$927	14 days	Sat. Night	Yes	$683	7 days	1 day	Yes
$1,001	14 days	Sat. Night	Yes	$827	3 days	none	No
$2,083	3 days	none	No	$929	None	none	No
$2,262	none	none	No				

Table 2-1: Fare Structure in the Boston-Seattle Market
(Source: Belobaba, 2005)

Prior to the "simplification" of the fare structure, a business traveler planning a trip wholly within a week from Boston to Seattle had to pay at least $2,083 (minus possibly a corporate discount of up to 20-30%) no matter how far in advance the booking was made. Under the simplified fare structure, a ticket for the same trip would cost between $374 and $929 depending on the date of the booking and fare class availability (before a corporate discount is applied although corporate discounts have been substantially decreased to usually less than 10% and only the most expensive fares are typically eligible). As a result, the "simplification" of the fare structure in domestic markets has made air travel during the week much more affordable and may have reduced the price of tickets substantially, especially for business travelers. An analysis by consultants at Sabre Airline Solutions for USA Today (De Lollis, 2006) has shown that the proportion of domestic tickets purchased that involve a Saturday night stay has

decreased from 51% in the first half of 2003 to 44% in the first half of 2006 as fewer business travelers extended their trip into the weekend to take advantage of lower restricted fares and some leisure travelers potentially switched to week travel.

More recently, most network airlines, including Delta, have gradually re-introduced Saturday night or three day minimum stay requirements for some of their discounted fares in select markets, primarily if they have a competitive advantage such as being the only airline offering non-stop service from a hub or serving a particular destination. However, due to both the large presence of low-cost competition and the competition among network airlines, deeply discounted fares with either a one-day or no minimum stay requirement are still offered by network airlines in many U.S. domestic markets. As reported in Business Travel News (Boehmer, 2007), an analysis of 4000 fares in the top 40 markets for the six U.S. network airlines by airfare consultancy Bob Harrell and Associates found that only 3% of inventory held Saturday-night stay restrictions in September 2007 compared to 7% six months after Delta introduced Simplifares and 16% six months prior to the launch of the fare simplification program.

Canada

In other parts of the world, such as Canada, Europe or Australia, low-cost airlines have also expanded aggressively since 2000 and the impact on pricing practices for short and medium-haul travel has been in many cases similar. In Canada, the entrance of Westjet has led to a complete restructuring of the fare structure in the Canadian domestic and the U.S. transborder market. Like its U.S. counterparts, Westjet introduced one-way pricing leading de facto to the removal of the Saturday night stay requirement.

Faced with this new competition in the domestic and U.S. transborder market, Air Canada developed an innovative approach to pricing that can be viewed as a mix between the low-cost and network airline pricing philosophies. In these markets, Air Canada now offers the choice between 4 categories of branded one-way fares designed to be differentiated not by fare rules but by product characteristics such as frequent flyer point accrual, re-booking flexibility and amenities such as advance seat selection, lounge access, upgrade eligibility or business/first class seating. For instance, the cheapest category named Tango offers advance seat selection for an extra $15 to $20 charge depending on the distance flown. In addition, Tango fares are eligible for only 25% mileage accrual and miles accrued do not count toward reaching elite status (non-status miles).

Within each category, Air Canada offers a wide range of fare levels and uses the same techniques as low-cost airlines i.e. advance purchase requirements and revenue management controls to determine which fare level to offer. As a result, the new pricing model developed and implemented by Air Canada can be viewed as a hybrid model. It is a combination between low-cost style pricing that relies on advance purchase requirements and inventory controls to segment the demand and fare product differentiation as in the network airline model except that product differentiation is based here on product features rather than on fare rules. The following table shows an example of the fare structure used by Air Canada in the Montreal-Edmonton market in summer of 2007. Air Canada operated two daily non-stop flights in this market and its main competitor, Westjet, was offering a single daily non-stop flight, except on Saturdays. Up to thirteen different fare levels in the Tango category were found on non-stop flights in this market. The difference in fare to upgrade from the Tango to the Tango Plus product was set at $45 in most of the cases. Similarly, a $240 difference in fare was observed between the Tango Plus and the flexible Latitude product for all fare levels, except the two most expensive ones.

Fare Levels	Tango	Tango Plus	Latitude	Executive
1	$219	$274	$514	$1,358
2	$229	$274	$514	$1,358
3	$249	$294	$534	$1,358
4	$324	$369	$609	$1,358
5	$334	$379	$619	$1,358
6	$354	$399	$639	$1,358
7	$362	$407	$647	$1,358
8	$394	$439	$679	$1,358
9	$402	$447	$687	$1,358
10	$434	$479	$719	$1,358
11	$442	$487	$727	$1,358
12	$584	$659	$884	$1,358
13	$592	$667	$884	$1,358

Table 2-2: Air Canada Fare Structure (Montreal-Edmonton Market, Summer 2007; Source: Air Canada Website)

More recently, Air Canada further refined its pricing strategy by offering a la carte pricing that let customers customize further the product. Depending on the product selected by the traveler, additional optional features are offered. For instance, customers purchasing a Latitude or Tango Plus product are offered the option to purchase an access pass to the airport lounge for an additional charge. On the other hand, Tango customers are offered the opportunity to get an additional discount if they elect not to accrue frequent flyer miles or if they accept not to make any itinerary changes or cancellations after the booking.

This innovative pricing strategy has proved popular with Air Canada travelers and seems to contribute to the airline's improved financial performance. As reported in the Wall Street Journal (Gutschi, 2007), speaking to an industry conference, Montie Brewer, CEO of Air Canada, declared that 48% of all passengers who flew in the fourth quarter of 2006 bought a higher price ticket, even if a cheaper fare product was available. In addition, 25% of the customers that paid the lowest fare purchased an additional service such as advance seat selection or a voucher for on-board meals and snacks.

The competitive structure of the Canadian domestic market has provided Air Canada with a favorable environment to introduce its new pricing strategy. Unlike the U.S. domestic market that is fragmented between six major network and several low-cost airlines, the Canadian domestic market is dominated by two players, Air Canada and Westjet. As a result, the pricing strategy pioneered by Air Canada has not yet become widespread in the U.S. domestic market.

However, several U.S. network airlines are taking steps to develop a multi-product or a la carte pricing strategy. For instance, during a meeting with investors (Compart, December 2006), Greg Taylor, United's senior vice-president for planning, reported that the airline is working on a plan to un-bundle its product offering into a base fare and a range of value-added optional features such as priority check-in and boarding, the option to change flights for a prepaid nominal fee or an access pass to the airline lounge. The airline already offers the option to purchase an upgrade to Economy Plus seating or a premium cabin at check-in. American now displays on his website five types of product categories that are differentiated by features such as refundability, flexibility to change travel plans and the opportunity to upgrade to the first class cabin.

Even low-cost airlines are moving toward more product differentiation. For instance, Southwest introduced in November 2007 a new product category designed to attract more business travelers. Named Business Select, it offers priority boarding, extra mileage accrual and free alcoholic beverages for a premium over the standard full fare. Similarly, Jetblue introduced in February 2008 a new category of refundable fares. The low-cost airline recognized that a single type of product does not satisfy the needs of all market segments, especially some business travelers that are prevented from purchasing non-refundable tickets by corporate travel policies.

Europe

In Europe, low-cost airlines started to develop in the mid-nineties after the liberalization of the intra-European market. Early entrants focused primarily on routes within the British Isles and linking Britain with the continent. The two largest European low-cost airlines, Ryanair and Easyjet emerged at that time: Easyjet was created from scratch by entrepreneur Stelios Haji-Ioannou and launched operations in 1995 while Ryanair transformed itself from a small money-losing Irish traditional carrier into a leading European low-cost airline linking Ireland with Britain but also establishing a large base at London-Stansted airport with many flights to continental Europe. Since the beginning of the century, new players have entered the market, especially in Germany, Scandinavia and Central Europe and Ryanair and Easyjet have expanded aggressively on the continent establishing bases in most major European markets.

Almost all of the 50 or so low-cost airlines that now operate in Europe follow a pricing strategy similar to that of their North American counterparts: European low-cost airlines typically sell one-way tickets removing de facto any minimum stay requirement. In addition, they also offer a wide range of fares associated with various advance purchase requirements and use revenue management to control the availability of the different fare levels.

Like U.S. network airlines, established European flag-carriers responded to the emerging low-cost challenge by a significant change in their pricing strategies. As in the U.S., a few European airlines chose to set up low-cost subsidiaries. For instance, KLM owns a carrier named Transavia that serves mostly leisure destinations in Southern Europe out of Amsterdam and Lufthansa is a major shareholder in German low-cost airline Germanwings. In addition, many

European flag-carriers decided to match some of the pricing practices of the low-cost airlines by removing restrictions such as the Saturday night stay requirement or decreasing their most expensive fares.

The most radical change came from carriers that had limited long-haul operations and were then more exposed to growing low-cost competition in short-haul markets. For instance, Aer Lingus, which derives more than 60% of its passenger revenues from European short-haul operations and is faced with intense competition from fellow Irish low-cost operator Ryanair, designed in 2001 a radical survival plan to transform itself from a loss-making traditional flag-carrier into a profitable airline modeled after its low-cost competitors. According to the airline annual reports, from 2001 to 2004, Aer Lingus revenues decreased by more than 200 millions Euros (-20%) because of a substantial decrease in average fares due to the shift in its pricing strategy. However, the Irish airline was able to reduce its operating costs by more than 350 millions Euros (-30%) and it achieved an 11.8% operating margin in 2004. Aer Lingus reduced its costs by increasing labor and aircraft productivity. The number of employees decreased by 30%, the fleet was reduced and homogenized with the shift to a single aircraft family on European operations, distribution costs were lowered by promoting aggressively online bookings (66% of the bookings at year-end in 2004), aircraft productivity was increased substantially by removing first-class seats and increasing aircraft utilization. In addition, capacity was redeployed from the U.K. routes to Continental European markets where the competition was less intense with the launch of more than 40 new routes between 2001 and 2006.

On the revenue side, Aer Lingus also revamped entirely its pricing and distribution strategy with the shift to one-way fares removing de facto all Saturday night stay requirements and by implementing a large decrease (up to

60%) in the most expensive fares charged to business travelers. Lower fares stimulated demand and led to an increase in load factors from 71% in 2001 to 82% in 2004 that partially compensated the decrease in average fares. In the fall of 2004, Aer Lingus was the first airline to extend this pricing strategy to its long-haul transatlantic operations. All transatlantic fares are now one-way and the airline reduced its business class fares by as much as 60% in order to make business class travel more affordable. This strategy has not been matched so far by its larger European and U.S. competitors in the transatlantic market.

Similar pricing and distribution strategies have been developed by other small or medium-size European carriers relying primarily on European traffic, for instance SAS or Finnair. Larger European airlines are less exposed to low-cost competition than their U.S. counterparts because European short-haul operations account for a much smaller share of their total operations thanks to their extended long-haul networks and strong cargo activity. As a result, large European airlines have responded to the low-cost challenge by increasing their market power through a wave of consolidation initiated by the merger between Air France and KLM and followed by the acquisition of Swiss by Lufthansa.

Large European airlines have also restructured their fare structures in European markets to compete more aggressively with low-cost airlines. For example, faced with increasing competition on European routes, in particular from Easyjet that established a base at the two main Paris airports of Orly and Roissy-Charles de Gaulle, Air France introduced in March 2004 in most domestic and European markets a new set of discounted round-trip fares that do not require a Saturday night stay. These new fare products are non flexible (non-changeable and non-refundable) and four fare levels are offered with respectively a 21, 14, 7 and 0-day advance purchase requirement. Air France still maintained another set of cheaper fares that require a Saturday night stay. However, this new set of fares

named "Gamme Semaine" has reduced the cost of short business trips substantially, especially for business travelers able to plan their trips relatively far in advance. However, the level of discount off the full unrestricted economy class fare varies widely based on the degree of competition in specific markets. For instance, in April 2005, the cheapest "Semaine" fare was 240 USD or 16% of a full economy class fare of USD 1478 in the Paris-Lisbon market and the similar fare was 483 USD (32% of a full fare of 1473 USD) in the then less competitive Paris-Vienna market.

2.2   Distribution of Airline Tickets

Along with the introduction of new pricing practices, low-cost airlines have been in many instances pioneers of major changes in the distribution of airline tickets. In this section, we will first discuss the development of direct and online distribution of airline tickets[1], then describe the evolving relationships between airlines and global distribution systems (GDS), and finally focus on the development of large online travel agents and its impact on the marketplace.

2.2.1   Direct and Online Distribution of Airline Tickets

Due to their strong focus on cost control, low-cost airlines have been at the forefront of the shift to direct distribution of airline tickets bypassing costly commercial or technological middlemen such as travel agents and GDS. Many European low-cost carriers have relied exclusively on direct distribution right from the start (for instance, Easyjet since its launch in 1995) and their inventory was not available through travel agents or listed in GDS. To market their

---

[1] Direct bookings are bookings made directly with the airline, either through an airport or city office, a call center or on the airline website. Online bookings are bookings made online, either on the airline website or through a third party online travel agent such as Expedia or Travelocity. As a result, a booking made on the airline website is classified both as a direct and an online booking.

products, they relied initially on their call centers. However, they took advantage of the technological advances of the late 1990s such as the development of the Internet and the introduction of electronic ticketing that eliminated the need for the distribution of physical paper tickets. Low-cost airlines moved quickly to online-based distribution and eliminated paper tickets system-wide reducing their significant distribution and processing costs. To accelerate the shift to online distribution, they often offered customers incentives such as discounts or extra frequent flyer credits for tickets booked on their website. These were replaced after some time by an extra charge for bookings made through call centers.

As shown in Table 2-3 below, online bookings accounted in 2006 for almost all bookings on the largest European low-cost carrier Ryanair and for the majority of the bookings on some North American low-cost airlines such as JetBlue or Southwest. However, since their inventory is not available in many instances through third party travel agents (traditional or online), low-cost airlines need to build a strong brand identity associated with low fares to drive prospective travelers to visit their website and usually incur relatively large marketing and advertising expenses, especially when they launch new destinations.

Compelled to reduce their own distribution costs to remain competitive and to take advantage of the same technological advances as their low-cost competitors, network airlines have responded in a variety of ways. Like their low-cost competitors, they rely increasingly on direct distribution of tickets through expanded call centers and the development of their own websites. To give an incentive to use their website, the most cost-effective of all distribution channels, some network carriers, particularly in North America, offered bonus miles for online bookings.

To make their website even more attractive, Northwest introduced at the end of 2004 a practice similar to that of some low-cost airlines i.e. a $5 booking fee for all tickets booked through its call centers. Northwest's new booking fee was quickly matched by all network airlines. In addition, many U.S. network airlines, such as Northwest, Continental and American introduced at the end of 2004 a lowest fare guarantee for online bookings, offering a refund and some form of incentive such as extra frequent flyer miles or a travel voucher if a traveler finds a lower published fare through another distribution channel. Despite all these incentives, network airlines remain in most instances far behind their low-cost competitors in terms of the proportion of direct and online bookings as many of their customers, especially business travelers, still rely heavily on traditional off-line travel agents.

In Europe, some small European carriers that do not rely heavily on long-haul traffic and are very exposed to low-cost competition on European short and medium-haul routes also developed aggressive strategies to increase online bookings. For instance, in 2006, Aer Lingus derived almost 75% of its total passenger revenues from tickets booked through its own website. As mentioned earlier, this shift in distribution strategy has led to a substantial decrease in distribution costs for the Irish carrier.

Carrier	2006	2004	2002
Southwest	70%	59%	50%
Jetblue	79%	75%	63%
Continental	24%	16%	8%
Ryanair	98%	97%	93%
Aer Lingus	73%	50%	28%

Table 2-3: Proportion of Sales on the Airline Website
(Source: Airline Websites and Annual Reports)

In addition to the development of their own website, network carriers have also gradually reduced and finally eliminated altogether base commissions paid in the U.S. to travel agents. A similar change has occurred in many European markets. For instance, effective April 1, 2005, Air France eliminated all commissions for tickets issued by travel agents in France, following similar changes that occurred earlier in other European countries. In order to cover their costs and remain profitable, off-line travel agents started to charge service fees. They rely increasingly on corporate customers that need a broad range of services such as account management or are required to book through travel agents to take advantage of negotiated corporate discounts.

2.2.2 New Airline-GDS Agreements

In order to further reduce costs and compete with their low-cost competitors, network airlines have started to look at how to decrease the fees paid to the GDS used by many travel agents, both online travel companies such as Travelocity or Expedia and off-line travel agents to access airline inventory and book travel on behalf of their customers. As most GDS were initially developed by a few major airlines, this industry was regulated in the U.S. in 1984 to prevent carriers that owned a GDS from using it as a competitive weapon by displaying more prominently their own inventory and barring other airlines from access to the marketplace. In addition to unbiased display of inventory, GDS were mandated to charge the same flat per-segment fee to any airline participating in the system. Since the deregulation of the industry starting in 2003, this rule has been lifted and airlines have started negotiating multi-year agreements with GDS companies, including new pricing conditions. In exchange for access to full content, the GDS generally agreed in these three-year agreements to reduce the per-segment fees. However, as network airlines still relied in 2003 on GDS for the overwhelming majority of their bookings and passenger revenues, they agreed to

limited discounts over previously regulated GDS fees and the GDS were able to lock-in favorable economic conditions for the three years to come.

In August 2004, along with the introduction of a $5 fee for ticketing through its call centers, Northwest announced the introduction of a $3.75 one-way "shared GDS fee" for all domestic tickets issued by travel agents through a GDS. Northwest argued that it paid on average $12.50 in fees on tickets issued through a GDS and that a significant portion of these fees were given back to travel agents by GDS companies as an incentive to use their systems. According to the airline, this new policy was necessary to compete in the U.S. domestic market with low-cost competitors. In response to Northwest initiative, Sabre, the largest GDS in North America, announced retaliatory measures that would make Northwest inventory less prominent in its displays, as now permitted after full deregulation of the industry. Faced with the risk of a significant drop in GDS-related bookings that include many of its most valuable bookings as off-line travel agents are a primary distribution channel for business travelers working for large corporations, Northwest had to renounce.

However, the high level of GDS fees remained a concern for most network airlines. This explains the strong interest expressed for lower-cost technologies under development by several companies dubbed GNE such as G2 Switchworks, Farelogix and ITA Software that planned to build alternative systems to existing GDS. In May 2005, all U.S. network airlines announced that their content would become available on G2 Switchworks when it was due to start operating late 2005. Combined with the growing popularity of their websites, this new distribution landscape provided the network airlines with some leverage for the negotiation of new multi-year agreements with the GDS companies in 2006. On the other hand, airlines remain highly dependent on GDS for access to key segments of the market such as business travelers or sales in distant overseas

markets where their brand identity is limited and the popularity of other distribution channels such as airline websites remains low.

Even several low-cost airlines, while still relying primarily on their own website, recognized the potential value provided by the GDS. For instance, Southwest Airlines announced in May 2006 a ten-year agreement with Galileo to get greater access to managed corporate travel. Similarly, after several years of absence, Jetblue made the strategic decision in August 2006 to participate again in three of the major GDS. In a conference call to discuss Jetblue's 2006 third quarter results (Compart, October 2006), David Neeleman indicated that 66% of GDS bookings were new customers and that the average fare of these bookings was about $35 higher than Jetblue's global average fare, net of GDS fees, providing the airline with improved access to the corporate market. Finally, Easyjet announced in November 2007 an agreement with two GDS companies (Amadeus and Galileo) in order to gain additional access to the European corporate travel market and to increase the proportion of business travelers on its flights, currently estimated around 20% of total passengers. Like for tickets booked through the phone, the airline will add a point-of-sale fee to GDS bookings to recover the additional cost and ensure that tickets are always cheaper on its website.

In this context, after highly publicized battles during the summer of 2006, airlines and major GDS companies finally agreed to new five-to-seven year agreements that include gradual reductions in GDS fees over the course of the agreement in exchange for continued access to full airline content. As reported in Travel Weekly (Schaal, 2006), GDS per-segment fees will initially decrease from almost $4 to slightly above $3 with a subsequent gradual decrease to the mid-2$ range by 2011, when most of these agreements expire. In order to reduce their operating costs and bring them in line with an expected gradual one-third decrease in segment fees, the GDS structured these agreements to reduce the

incentives paid to travel management companies by offering new so-called optional programs. Under the new airline-GDS agreement, travel agents would be charged by the airlines a $3.50 segment fee for all bookings made through the GDS unless they join these optional programs and agree either to the payment of a new per-segment fee of up to $0.80 or a decrease of a similar amount in incentives.

As the alternative systems developed by the GNE were not yet fully in operations or still had a very limited penetration among off-line travel agents when new agreements were negotiated with the airlines, the GDS were able to limit the decrease in per-segment fees and compensate it by a reduction in incentives offered to travel agents. This should give them the time necessary to adjust and reduce their own operating costs by shifting to lower-cost open-source technologies and compete aggressively when the current set of airline-GDS agreements expire in 2011-2013.

2.2.3 Development of Online Travel Agents and Websites

The other major change in the distribution of airline tickets is the development of large online travel agents that market a wide range of products such as airline tickets, hotel accommodations and car rentals sold separately or together as packages. These online travel retailers such as Travelocity or Expedia developed quickly at the end of the 1990s. They were later joined by competitors launched by the airlines themselves such as Opodo, owned by nine European flag-carriers and the European GDS Amadeus, and Orbitz, which was founded in 2001 by five U.S. network carriers and later sold to U.S. travel conglomerate Cendant in 2004.

Third party online websites have a large impact on the distribution of airline tickets and price transparency in airline markets. They have developed search

engines that allow prospective travelers to compare easily between fares offered by a wide range of carriers without going through an intermediary such as an off-line travel agent. In addition, some of these online retailers are using powerful search engines that can in some instances find lower fares by looking for combination of tickets on several carriers for a single trip. For instance, Orbitz introduced an advanced search engine developed by ITA software that takes advantage of the many one-way fares now offered in the market in response to low-cost airline pricing practices to construct itineraries involving several carriers if such a combination turns out to be cheaper than tickets on a single airline.

In addition, other web companies called scrapers have developed technologies that search quickly through both airline and online travel retailers, and let travelers compare fares across all these sites. These companies such as Sidestep and Kayak earn revenues through referral fees paid by the airlines or the online travel retailers for directing prospective travelers to their website. More recently, a new scraper travel website has gone a step further by providing travelers with guidelines on when to book their tickets. Launched in 2006, Farecast has developed a large historical database of airline fares and is using statistical techniques to predict whether a fare is likely to increase or drop in the near future.

2.3   The Impact on Booking Patterns and Airline Passenger Choice

The impact of the development of direct and online distribution on the choice behavior of airline travelers can be somewhat uncertain. The development of direct distribution of airline tickets may tend to reduce the universality of the passenger choice set since airline travelers could need to contact each carrier

directly to get access to fare and schedule information. For some airlines, schedule and fare information is available only through direct contact with the airline since, as mentioned earlier, to reduce cost and avoid GDS fees, some low-cost carriers do not participate in GDS. However, the quick development of large online travel websites has allowed many travelers to have a direct access to a wide range of airline fare and schedule information without going through a middleman such as an off-line travel agent that may be biased due to financial incentives from the airlines. Online travel websites such as Travelocity, Expedia or Orbitz make it much easier for consumers to compare across travel alternatives and choose the most attractive in terms of fare, schedule, airline and product characteristics. As a result, many including the airlines themselves argue that the development of online distribution has greatly increased price transparency in airline markets. In addition, the new generation of scraper websites further enhances price transparency by allowing prospective travelers to compare more easily across travel websites including the sites of low-cost carriers that do not appear on travel agent websites.

In addition, as low-cost airlines expand, their pricing strategy permeates a growing number of markets, impacts the choice set of an increasing number of travelers and influences over the longer-term the perception of the traveling public regarding fares and fare product characteristics. On one hand, due to lower fares offered by low-cost airlines, business travelers have been more and more reluctant to purchase the highly priced unrestricted business fares charged by network airlines. Since 2000, network airlines have experienced a significant drop in the number of business-type fare tickets purchased, which led to a rapid decline in yield. This forced them to overhaul their fare structures and finally adopt at least partially the pricing practices of their low-cost competitors. On the other hand, the removal of the Saturday night stay requirement, the decrease in walk-up fares through the compression of the fare structure and the

development of opaque web-based distribution channels like Priceline.com to sell distressed inventory may have influenced the behavior of many leisure passengers as well. Leisure trips during the week such as visits to friends and relatives as well as last-minute leisure travel have become more affordable. As a result, low-cost airline pricing practices have removed constraints that reduced the affordability of business trips and the flexibility of leisure trips and may have reduced the behavioral difference between the two segments of the market.

The impacts of these changes are reflected in many of the industry-wide data and have contributed, along with the aftermath of 9/11, to very large cumulative losses since 2001, especially for U.S. network airlines. As shown in Figure 2-2, while U.S. airline passenger revenues accounted for 1% of U.S. GDP in the 1990s, this proportion has decreased substantially since 2000 and remains well below that level despite a slight rebound over the last three years. From 2000 to 2002, U.S. airline passenger revenues dropped by almost $25 billion in current dollars. This shortfall more than persists today if one combines the remaining shortfall in revenues – at about 0.2% of GDP in 2006 i.e. $26 billion – with the increase in fuel costs that can be mitigated mostly only in the long-term through fleet renewal and advancements in aircraft technology and is usually passed onto consumers in other modes of transportation. Fuel costs at the industry level including all-cargo carriers rose from around $16 billion in 2000 to $38 billion in 2005.

This relative decrease in air travel spending tends to indicate that a structural shift occurred in the market. The shift toward lower airline passenger revenues is primarily driven by a decrease in yield that was partially mitigated by an increase in load factors and led to a lower decrease in revenue per available seat mile (RASM) as shown in Figure 2-4.
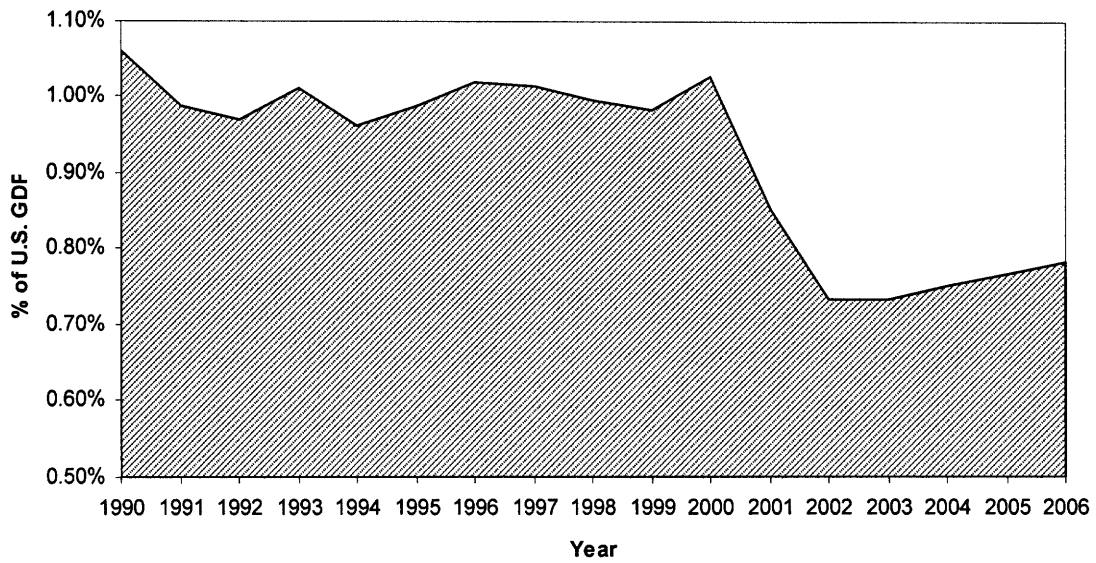
Figure 2-3:   U.S. Airline Passenger Revenues as a Proportion of U.S. GDP
                (Source: Air Transport Association, U.S. Department of Commerce)
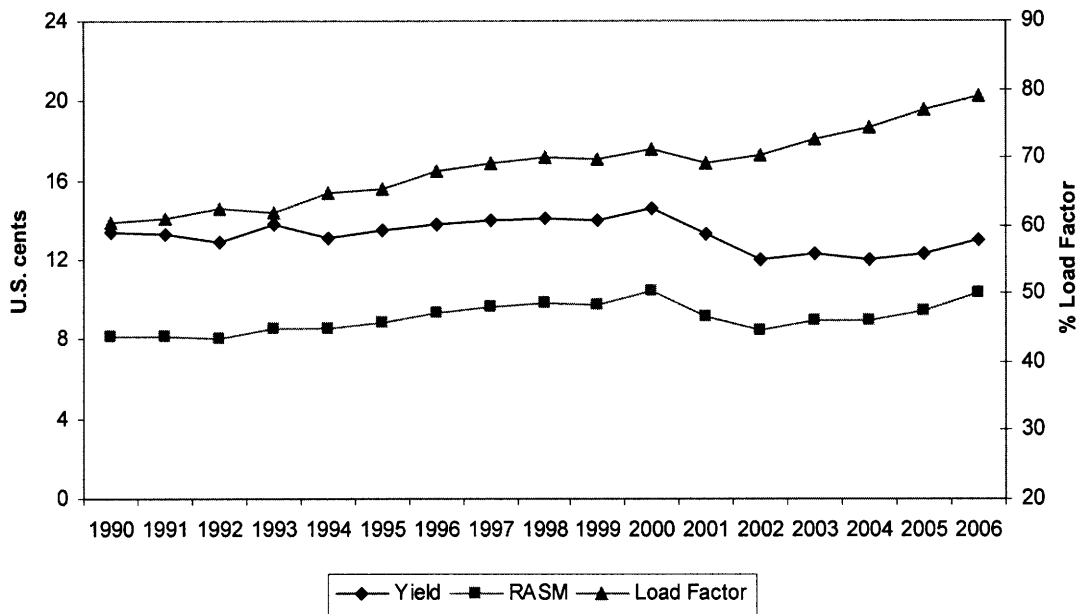
Figure 2-4:   Yield, RASM and Load Factor in the U.S. Domestic Market
                (Source: Air Transport Association)

49

This sharp decline in yield is largely due to increased price transparency in airline markets and the shift toward lower fares driven by growing low-cost competition and the pricing overhaul in the industry. For instance, according to a study by ECLAT Consulting (Swelbar, 2003), even before the overhaul of its fare structure following Delta's pricing initiative, the proportion of United's revenues from premium tickets (full fare economy and first class tickets) decreased from 41% in 1999 to only 20% in 2002. Similarly, a recent study of price competition in the U.S. airline industry (Pyrgiotis, 2008) reported that the fare premium of network airlines over their low-cost competitors in the top 856 U.S. domestic markets decreased sharply from $90.44 in 2000 to $54.67 in 2006 as shown in Figure 2-5. As mentioned, the decrease in yield was partially compensated by an increase in load factors that may be related to the impact of airline revenue management. If the fare structure is compressed and the difference between higher and lower fares is reduced, revenue management systems will tend to protect fewer seats for expected yet uncertain high-yield late-booking passengers and release additional low-yield seats leading to an increase in load factors.

Average Fare in Top 856 Markets
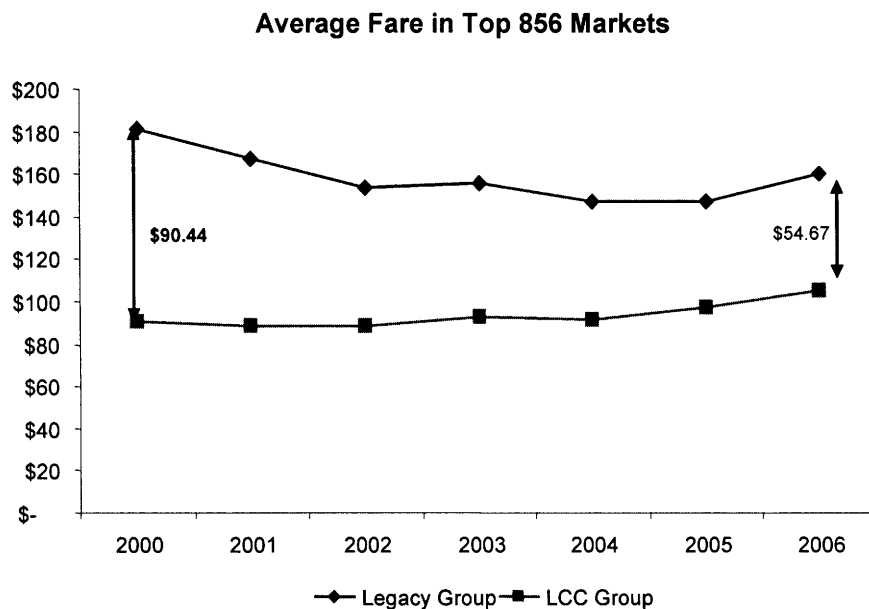
Figure 2-5:   Average Fare in the U.S. Domestic Market by Carrier Type
(Source: Pyrgiotis, 2008)

In order to gain additional insight on the impact of the development of online distribution and the growth of low-cost airlines, we studied the changes in booking patterns in a dataset provided by Northwest Airlines. This dataset includes all bookings (local traffic) on Northwest's non-stop flights in 22 U.S. domestic markets for the month of June 2000 and June 2004 including code-share passengers[2]. As indicated in Table 2-4 below, these markets were classified as either leisure or business-oriented and Northwest faced non-stop low-cost competition in a subset of these markets.

Market	Market Type	LCC NS Competition
AUS-DTW	Business	No
AUS-MSP	Business	No
BOS-DTW	Business	No
BOS-MSP	Business	No
DFW-DTW	Business	No
DFW-MSP	Business	Yes (Sun Country)
IAH-DTW	Business	No
IAH-MSP	Business	No
MCO-DTW	Leisure	Yes (Spirit)
MCO-MSP	Leisure	Yes (Air Tran, Sun Country)
MDW-DTW	Business	Yes (Southwest)
MDW-MSP	Business	Yes (ATA)
MSY-DTW	Leisure	No
MSY-MSP	Leisure	No
ORD-DTW	Business	No
ORD-MSP	Business	No
PHX-DTW	Leisure	Yes (Southwest, America West)
PHX-MSP	Leisure	Yes (Sun Country, America West)
STL-DTW	Business	Yes (Southwest)
STL-MSP	Business	No
TPA-DTW	Leisure	Yes (Spirit, USA 3000)
TPA-MSP	Leisure	No

Table 2-4:   List of Markets in the Northwest Dataset

---

[2] Code-share passengers are passengers with tickets sold by one airline called the marketing airline (here, for instance, Continental Airlines) valid for travel on flights wholly or partially operated by another airline called the operating carrier (in this case, Northwest Airlines)

Figure 2-6 below illustrates the shift observed toward direct and online distribution of airline tickets. The proportion of tickets booked through off-line travel agents decreased from about 60% to just over 40% of total bookings while direct bookings increased from 34% to 44% of total bookings and online bookings increased from just 11% to over 36% of total bookings from 2000 to 2004. There was especially a strong increase in online bookings though the airline website NWA.com from 7 to 22% of total bookings.
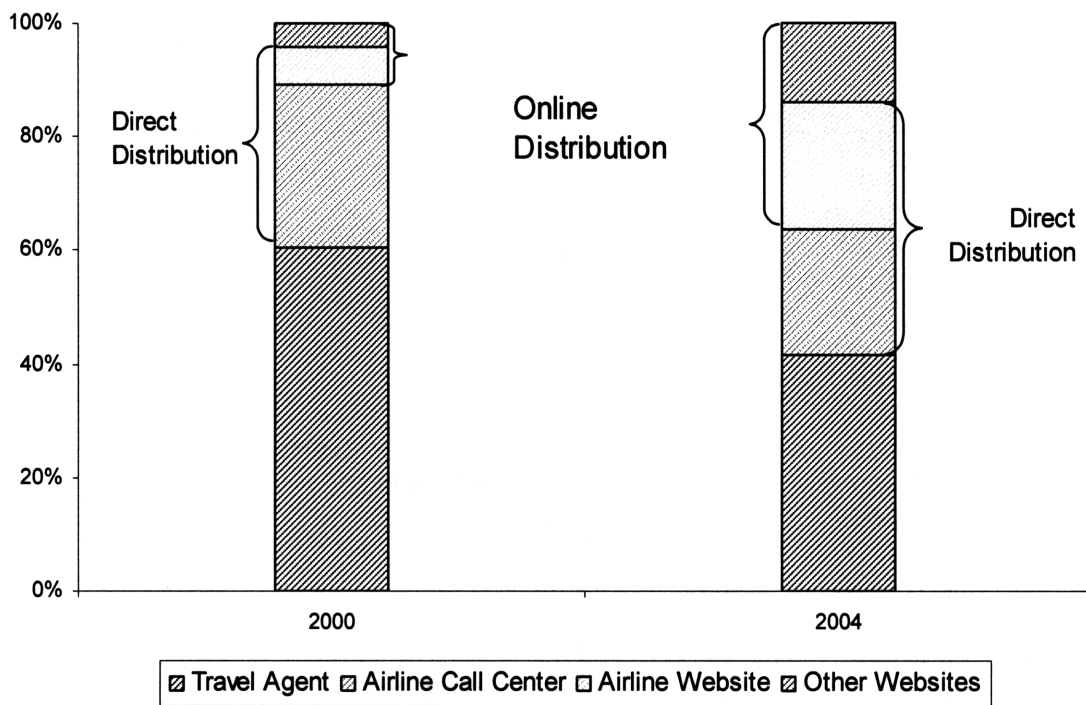
Figure 2-6:   Bookings by Distribution Channel

In addition, Figure 2-7 shows that both in 2000 and 2004, bookings made through off-line travel agents generated a much higher average fare than bookings made through other distribution channels. However, since the proportion of air travelers booking through travel agents fell sharply, the average fare remained almost stable increasing by only 2.7% from $161 to $165.30 in current dollars.
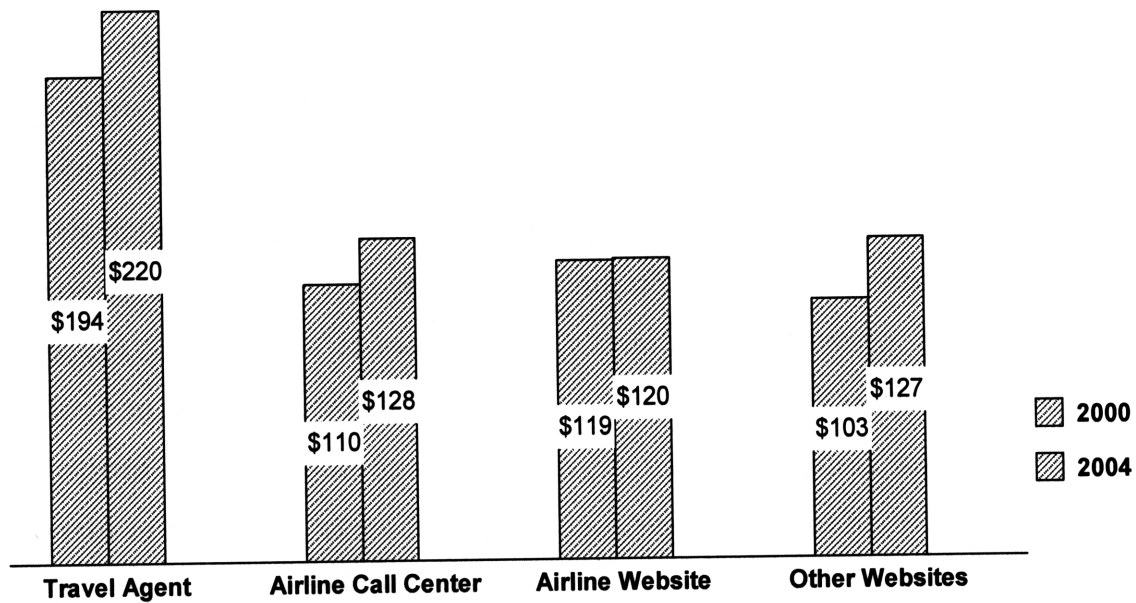
Figure 2-7:   Average Fare by Distribution Channel
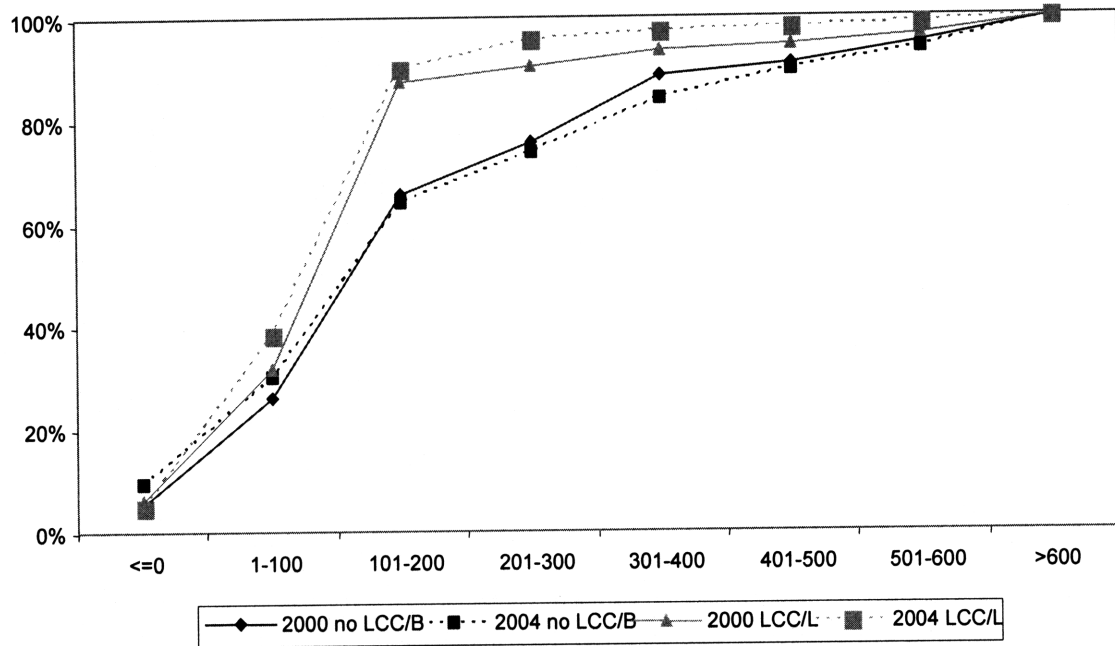
Figure 2-8:   Cumulative Distribution of Fares by Market Type and Competitive Environment

The analysis of the Northwest dataset also illustrates the influence of low-cost airlines on fare levels and the advance purchase airline tickets. As shown in Figure 2-8 above, the proportion of tickets with a one-way fare above $200 increased in business markets without low-cost competition but it decreased in leisure markets with low-cost competition. This illustrates the downward pressure on fares associated with the presence of a low-cost competitor and in particular how low-cost competition limits the ability of a network airline to sell relatively highly priced business fares. The presence of a low-cost competitor also has an impact on booking curves. As shown in Figure 2-9 below, in markets without low-cost competition, air travelers booked their tickets further in advance in 2004 than in 2000 while they booked closer to departure in markets with low-cost competition. Low-cost airlines usually offer more affordable fares closer to departure. To remain competitive and defend its market share, Northwest most likely matched its low-cost competitor fare structure leading travelers to delay the purchase of their tickets.
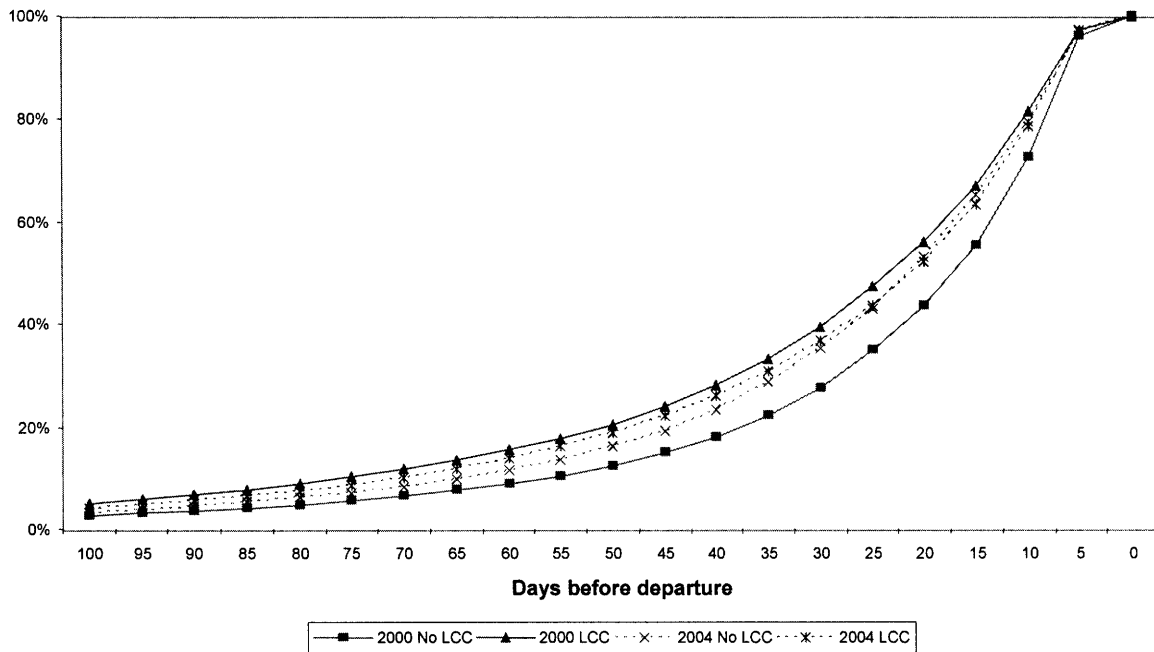
Figure 2-9:   Booking Curves by Competitive Environment

2.4   Conclusion

The rapid growth of low-cost airlines, the development of online distribution of airline tickets along with the economic downturn of the early years of the century and the events of 9/11 have led to a long period of turmoil and uncertainty for network carriers and the airline industry in general. In response to growing low-cost competition and increased price transparency in the marketplace, all network airlines in the U.S. and many European flag carriers initiated a major restructuring of their pricing practices in short and medium-haul markets. They have adopted to some extent the pricing practices of their low-cost competitors by removing fare rules such as the Saturday night stay requirement in many markets and reducing the spread between the most expensive and the cheapest discounted fares.

This major overhaul of airline pricing and distribution practices has an impact on the choice set of many air travelers and has led to observed changes in booking patterns. In addition, these elements have undermined the pricing and revenue management practices of network airlines and a better understanding of airline passenger choice is necessary to support the development of new pricing and revenue management strategies. For instance, new revenue management models designed to optimize revenues under less restricted fare structures require an estimate of sell-up behavior in order to determine when to stop selling a cheaper fare given expected sell-up to a higher fare. A better understanding of passenger choice behavior may provide additional insight on the price elasticity of demand and more reliable estimates of sell-up potential. Similarly, some airlines such as Air Canada have developed new innovative pricing strategies that combine fare product differentiation and fare customization with a low-cost style set of fare rules, expanding the set of options offered to airline travelers. Passenger choice models may also provide valuable insight to support the design of effective new

fare product differentiation and ancillary revenue strategies. The rest of the dissertation will then focus on how to develop a model of airline passenger choice that can be used to support a range of airline planning decisions, including new pricing and revenue management strategies designed to compete more effectively in the current marketplace.

Chapter 3   Literature Review

In this chapter, we review the existing literature on airline passenger choice. We first provide a brief overview of the literature on discrete choice models that have been used to study choice behavior in a variety of contexts. In particular, discrete choice models have been applied extensively to the transportation field to enhance the understanding of traveler choice behavior and improve the accuracy of travel demand forecasts. Then, we review more extensively previous studies of passenger choice behavior, focusing on the strengths and weaknesses of using different types of data. Finally, we identify some of the shortcomings of existing research and further discuss the objectives of this dissertation.

3.1 Discrete Choice Models

Discrete choice models based on random utility theory (RUM) are the tool most commonly used by researchers and practitioners to represent individual choice behavior. The concept of random utility theory originated in the work of Thurstone (1927) and was further interpreted and developed by Marschak (1960). The idea behind random utility theory is that while the utility of each choice alternative may be known by the decision-maker, it is not fully known by the researcher and therefore uncertainty must be taken into account. As a result, utility is modeled as consisting of two parts, a deterministic component observed by the analyst and a random component that remains unknown. Manski (1977) identified four main sources of uncertainty: Unobserved attributes of the

alternatives, unobserved individual attributes of the decision-maker, measurement errors and the use of proxy (instrumental) variables.

In addition to random utility theory, discrete choice models rely on several other assumptions. The decision-maker is assumed to choose among a set of finite, mutually exclusive and collectively exhaustive alternatives and to select the alternative with the highest utility. For each alternative, the deterministic part of the utility is specified as a function (generally a linear function) of its attributes, such as for instance travel time and travel cost, and the characteristics of the decision-maker (age, gender, income etc.). As mentioned above, utilities also include a random component and the output of a choice model is then the probability of an individual selecting each alternative. Econometric techniques, such as maximum likelihood estimation can be applied over a sample of observations to estimate the unknown coefficients of the deterministic part of the utility function.

Different assumptions on the distribution of the disturbances lead to different forms of choice models. Many of the early investigations of discrete choice models assumed a multivariate normal distribution for the disturbances leading to the so-called probit model. Although extremely flexible because it allows for an unrestricted covariance matrix of the disturbances, the probit model was difficult to estimate because the choice probabilities do not take a closed-form solution. As a result, the multinomial logit model developed by Marley, as cited by Luce and Suppes (1965), and McFadden (1974) is often preferred to probit due to its mathematical simplicity and hence, its high degree of tractability. The multinomial logit model assumes that disturbances follow an extreme value distribution and are independently and identically distributed (i.i.d.). Logit choice probabilities take a closed-form solution and can be easily calculated.

One of the most documented aspects of the logit model is its property known as independence from irrelevant alternatives (IIA) that originates in the i.i.d. specification of the disturbances. The IIA property can lead to an erroneous representation of consumer preferences when some alternatives are very similar and can be assumed to share some common unobserved characteristics (for example, the now famous red bus – blue bus textbook example).

There are several ways to overcome the IIA property and one of the most popular approaches has been to use models from the Generalized Extreme Value (GEV) family. The GEV family is a large class of models that offers a partial relaxation of the IIA property and is used to represent a variety of substitution patterns across alternatives. GEV models assume that the disturbances follow a joint extreme value distribution, which is a generalization of the univariate extreme value distribution used in the multinomial logit model, meaning that the logit model belongs to this larger class of models. GEV models have been popular with researchers as they retain a good level of tractability because the choice probabilities keep a closed-form solution. The nested logit model introduced by Ben-Akiva (1973) is the most widely used model from the GEV family due its simple functional form compared to other GEV models but more complex models such as cross-nested logit models are also increasingly popular.

In addition, the development of computational capabilities over the years has enabled the specification and estimation through simulation techniques of even more flexible models that were previously intractable. As stated by Train (2003), "simulation allows estimation of otherwise intractable models. (...) The researcher is therefore freed from previous constraints on model specification – constraints that reflected mathematical convenience rather than the economic

reality of the situation."[3] In particular, with the advent of simulation-based estimation, the logit kernel model, also called mixed logit has become increasingly popular among choice modelers. Logit kernel is a highly flexible model that combines the flexibility of probit-like forms and the mathematical simplicity of the logit model. Logit kernel disturbances are composed of two parts: The logit kernel (i.e. an i.i.d. extreme value disturbance) and another part, often probit-like that allows for flexibility. McFadden and Train (2000) have proved that any well-behaved RUM-consistent behavior can be represented as closely as desired with a logit-kernel specification. As a result, the logit kernel model can be used to represent any kind of substitution patterns. In particular, any GEV model can be approximated by a logit kernel specification using the appropriate choice of variables and mixing distribution. In addition to more flexible substitution patterns, logit kernel models also allow the specification of random coefficients to represent random taste variation and the estimation of panel data when unobserved elements may be correlated over a set of choice observations from the same decision-maker.

In addition to more flexible models such as logit kernel, advancements in computer technology have sparked the estimation of hybrid choice models that incorporate other components such as latent variables and latent classes into choice models and make it easier to combine several sources of data, for example, stated preference (SP) and revealed preference (RP) data. The hybrid choice model framework draws on ideas from many researchers including Ben-Akiva and Morikawa (1990) who developed methods to combine RP and SP data, McFadden (1986) who proposed ideas to include latent variables and psychometric data in choice models and Gopinath (1995) who developed the latent class choice model that incorporates unobserved discrete latent constructs (also called latent classes) to account for heterogeneity of choice behavior across

---

[3] Train, "Discrete Choice Models with Simulation", Cambridge University Press, 2003

latent segments of the population. Through these extensions, choice models can capture more realistically the complexity of many choice processes. For a more complete description of the hybrid choice model framework, the reader is referred to Walker (2001).

3.2   Passenger Choice Models

As mentioned in the previous section, 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. Two of these characteristics are actually not restrictive. An appropriate definition of the alternatives can nearly always ensure that the alternatives are mutually exclusive and that the choice set is exhaustive. For instance, by defining an extra alternative such as "none of the other alternatives", the researcher can effectively ensure that the choice set is exhaustive. However, the third condition, a finite number of alternatives, is restrictive since it is the defining property of discrete choice models and it separates the realm of applications between discrete choice and regression models.

As the choice set of air travelers is usually composed of a finite number of travel alternatives, including the option of not traveling at all, discrete choice models can be a valuable tool to represent the choice of individual air travelers among a set of travel options. As a result, discrete choice models have been used by researchers to represent the choice of airline passengers among a set of travel alternatives characterized by a range of attributes such as itinerary and fare.

Two main types of data have been used to study airline passenger choice, revealed preference and stated preference data. RP data are real market data that

reflect actual choice behavior in the marketplace whereas SP data are stated responses to pre-designed hypothetical scenarios and are collected through surveys. Both types of data have strengths and weaknesses. The major advantage of revealed preferences is that it represents actual choice behavior whereas stated preferences are hypothetical in nature and can be subject to several types of biases. However, stated preferences provide more flexibility due to the control over the survey design and the opportunity to build a set of hypothetical choice scenarios. For instance, stated preference data can be used to explore consumer choice behavior regarding non-existing alternatives or to introduce more variability compared to the actual existing alternatives. As a result, RP and SP data are highly complementary and estimation techniques have been developed to combine both types of data sources. As mentioned earlier, since the booking search process remains largely unobserved by the airlines, no literature was found on how to use web analytics and exploit data such as screen shots and click activity to analyze the choice behavior of airline passengers.

3.2.1 Stated Preference Data

Although discrete choice models applied to the airline industry have relied on either RP or SP data, the majority of studies of airline passenger choice have been based on the analysis of stated preferences collected through surveys of airline passengers. As mentioned earlier, very detailed data can be obtained through surveys such as socio-economic characteristics of the traveler (age, gender, income), travel history (previous trips, frequent flyer information), characteristics of the trip (trip purpose, preferred departure time), even indicators to represent attitudes toward elements such as the risk of misconnection (Theis, 2006).

Some of these elements, primarily trip purpose, have been used to segment air travel demand and investigate the differences in behavior across market

62

segments by estimating separate models for each segment. For instance, based on data collected through an on-line survey of 600 U.S. travelers, Adler, Falzarano and Spitz (2005) estimated separate models of the choice of an itinerary for business and non-business travelers. As expected by conventional industry wisdom, they show that travelers belonging to these two segments of the market have very different behavioral patterns. For instance, they estimated that while business passengers may be willing to pay $30 for a one-hour decrease in schedule delay[4], non-business passengers are only willing to pay less than $5 for a similar improvement.

In addition to the variety and wealth of data collected, surveys of airline passenger choice are typically designed to reproduce to some extent the characteristics of the choice environment in the airline industry, including the impact of airline pricing and revenue management on the passenger choice set. For instance, in their study of the choice of an airline, flight and fare class, Prossaloglou and Koppelman (1999) used different sets of fare products based on whether respondents were asked to select a travel alternative for a hypothetical winter vacation trip planned three weeks in advance or to attend a business meeting with three days advance notice. To reflect the traditional pricing practices used by network carriers at the time of the study, each group of respondents was asked to choose among three fare products. The products presented to business travelers did not require a Saturday night stay and carried either low or no change fees while two of the fare products proposed to leisure travelers did require staying overnight on Saturday and all of them carried increasingly stiff change fees.

---

[4] Schedule delay is defined as the time difference between a passenger preferred departure time (or arrival time) as stated by the passenger during the survey and the actual departure time (or arrival time) of the chosen alternative.

More recently, Garrow, Jones and Parker (2007) developed a study of airline passenger choice based on stated preference data collected through a survey of consumers recruited while they were shopping on a travel website. Survey respondents were presented with the choice of three hypothetical travel options tailored to the origin and destination of their current search and characterized by departure time, arrival time, number of connections, airline and aircraft type, comfort (legroom) and price. In order to reproduce the dispersion of fares found in airline markets, the survey design used a two-tier approach. They first draw for each respondent a multiplier of a pre-defined average fare in the market called the base fare (from 0.75 to 1.5) and then multiply it by a fare premium for each of the three itineraries included in the respondent's choice set (from 0.85 to 1.20). As a result, the highest fare could be almost three times the lowest one, representing the typical spread of fares found in airline markets.

Although most stated preference experiments have been designed to reproduce the characteristics of the passenger choice environment, they have done so in a very simplified manner. For practical reasons, these experiments only include the choice between a restricted set of two to three hypothetical travel alternatives. However, prospective travelers are likely to consider a large number of potential travel alternatives in a real booking search, especially given the increased search capabilities offered by online travel websites. As a result, none of these studies were able to fully incorporate the impact of airline pricing and revenue management and represent in a realistic manner the passenger choice set.

The other major weakness of stated preference data is the potential risk of bias associated with this type of data collection. As they are usually collected through surveys, stated preference data are subject to a risk of non-response bias, meaning that the people that agree to participate to the survey may not have the same distribution of attributes as the whole population. In addition, the

hypothetical nature of stated preference data may also lead to some level of response bias i.e. a risk of discrepancy between an individual's stated responses and his true choice behavior. This risk is increased if the design of the experiment involves the hypothetical disbursement of money, such as asking whether an air traveler prefers a cheaper itinerary on his second preferred carrier to a more expensive alternative on his most preferred carrier. The risk of response bias may be especially high in airline markets that are considered increasingly as commodity markets due to the rapid growth of low-cost airlines. As most studies of airline passenger choice are designed to quantify how much passengers are willing to pay for a variety of attributes such as non-stop flights, frequency, preferred airline or frequent flyer miles, these studies are exposed to a significant risk of response bias.

Findings from studies of airline passenger choice based on stated preference data have consistently indicated that both business and to a lesser extent leisure passengers are willing to pay fairly high premiums for various attributes of airline service such as path quality, airline and airport preferences. For instance, Adler, Falzarano and Spitz (2005) estimated that U.S. business passengers would be willing to pay an extra $70 for a one-hour decrease in flight time or $96 to travel on their most preferred airline relative to their least preferred airline while non-business travelers would be willing to pay respectively $31 and $38 for the same attributes. Such high values tend to run contrary to airline experience: Many airline professionals stress the paramount importance of price in the selection of a travel alternative, especially in the current competitive and distribution environment characterized by the strong presence of low-cost competition and the focus of most travel websites on search by price and not by schedule or carrier.

3.2.2 Revealed Preference Data

Although most studies of airline passenger choice are based on stated preference data, another stream of studies has been based on the collection and analysis of past booking data. For instance, Coldren et al. (2003) have used an extensive sample of bookings extracted in 2000 from a leading GDS to study the choice of itinerary in the top 500 U.S. domestic markets. Grammig, Hujer and Scheidler (2004) also relied on booking data provided by a GDS to study the choice of itinerary in ten German domestic markets. De Lapparent (2004) used booking data provided by a leading European airline to study the choice of an air route between the Paris and London metropolitan areas.

The major advantage of using booking data is that they reflect actual market preferences and, unlike survey data, are not subject to the risk of response bias. However, the major disadvantage of these studies is the limitations of the data retrieved from the booking records. First, the state of the airline inventory at the time of the booking is not recorded and only the chosen alternative is available in airline bookings. The passenger choice set must then be reconstructed using other sources of data such as the airline flight schedule or seat availability data. As a result, these studies have focused on a single dimension of airline passenger choice such as the choice of an itinerary. They usually assumed that seats were available on all itineraries and just used the airline flight schedule to infer relatively easily the passenger choice set. They focused on how to represent the complexity of substitution patterns across itineraries using models from the GEV family (Coldren and Koppelman, 2005) or probit models (Grammig, Hujer and Scheidler, 2004) and ignored the other key dimensions of airline passenger choice such as price. They either omitted fare from the passenger's utility function or relied on fare data collected at some level of aggregation such as at the market and/or carrier level. As a result, they do not provide insight on the fundamental

66

trade-offs such as between schedule convenience, fare and fare product characteristics.

In addition, since trip purpose is not available in booking records, booking-based studies of airline passenger choice did not account for heterogeneity of behavior, a major characteristics of airline markets. However, some elements available in airline bookings such as the characteristics of the trip or the profile of the traveler could provide useful information to segment the bookings and be used as an alternative to the traditional segmentation of air travel demand by trip purpose.

In their study of the choice of a flight and booking class, Algers and Beser (2001) recognized the risk of response bias associated with stated preference data and proposed to combine data collected through a 1994 survey of SAS passengers with booking data extracted from the airline reservation system for the same period. They implemented a sequential estimation method: They first estimated the parameters of the SP model and then applied them to calculate the fitted value of a general cost function for each alternative included in the RP dataset and estimate a scaling parameter to correct the scale of the utility function obtained from the SP data. While no details are provided on how the choice set was processed for each booking record included in the RP dataset, such an approach may offer an attractive and promising solution to the respective limitations of RP and SP data in the context of airline passenger choice. However, discrepancies may exist between the RP and SP datasets and they encountered difficulties when they combined the two datasets. In particular, a major explanatory variable of the choice of an itinerary included the SP model, the schedule delay, could not be found in booking records and was then missing in the RP dataset. They assumed its value to be zero for all passengers, meaning that the actual flight departure time was always supposed to be equal to the passenger's preferred departure time, reducing the benefits of using the RP

dataset to calibrate the SP model and potentially leading to substantial bias in the estimation results. In addition, as no information on the characteristics of the trip and the profile of the traveler was retrieved from the booking records, this study did not account for heterogeneity of behavior across different segments of airline travelers.

3.3   *Implications for this Research*

Discrete choice models have been successfully applied over the years to investigate the choice of airline passengers among travel alternatives. They have proved to be a valuable tool to gain insight into the determinants of airline passenger choice and have been used to support airline planning decisions such as schedule planning. In addition, previous studies also show that advanced discrete choice models provide a more accurate representation of passenger choice behavior. For instance, Coldren and Koppelman have shown that models from the GEV family better reflect the complexity of substitutions patterns across itineraries that share some common attributes such as departure time, airline and path quality. Adler, Falzarano and Spitz have demonstrated the benefits of using mixed logit models to capture heterogeneity of behavior with regard to several elements of the passenger's utility function including within specific segments of the market such as business and non-business passengers.

As mentioned earlier, while they provide the opportunity to collect very detailed information on the characteristics of the traveler and the trip, stated preference data of airline passengers cannot fully represent the complexity of the choice environment in the airline industry. They are also subject to a risk of response bias, especially since the objective of the research is in most cases to quantify how much passengers are willing to pay for elements of airline service. An alternative

to stated preference data is to analyze the choice behavior of airline passengers based on past booking data that reflect actual ticket purchase behavior and market preferences. However, previous studies of airline passenger choice based on booking data have been limited to a very partial description of the passenger choice environment. They have focused on a single dimension such as the choice of itinerary or route without taking into account the impact of other major attributes such as fare and fare product characteristics on the choice behavior of air travelers. In addition, booking-based studies of airline passenger choice did not incorporate the heterogeneity of behavior across different segments of travelers, a major characteristic of airline markets.

Since the limitations of stated preference data cannot be fully eliminated, the objective of this dissertation is to develop a multi-dimensional model of passenger choice behavior based on booking data that accounts for heterogeneity of behavior across bookings. A methodology will be developed to reconstruct the passenger choice set at the time of the booking, focusing, in particular, on how to incorporate the impact of airline pricing and revenue management. In addition, the model will exploit information available in booking records such as the characteristics of the trip and the traveler to segment the demand and investigate the heterogeneity of behavior across different segments of the market.

Chapter 4   Modeling Framework

In this chapter, we develop a model specification to reach the objectives of this research and investigate the preferences of different segments of airline travelers. We first define the major dimensions of airline passenger choice and the dependent variable of the model. We then discuss how the choice set of each passenger is inferred by combining booking data with several other types of data such as seat availability and fare rules. We propose an alternative approach to the traditional segmentation of airline demand by trip purpose based on the specification of a latent class choice model. We finally discuss how to represent the time-of-day preferences of airline passengers.

4.1   *The Choice of an Airline Itinerary and Fare Product*

The primary objective of this research is to develop a model of the major dimensions of airline passenger choice such as price and schedule. 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 airline providing the service. While Smith's research is limited to a subgroup of the market that is not fully representative of the global marketplace as passengers booking through online travel agents tend to be more likely to travel for non-business purposes, it still provides a strong indication of the most significant dimensions of airline passenger choice. It supports the conventional

industry wisdom on the major impact of price and schedule on the choice of airline passengers.

In this research, we will study air traveler choice behavior from the airline perspective. Unlike a travel agent such as Travelocity that markets the inventory of several carriers, an airline does not have access to the booking database 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 potentially biased subgroup of the market. We will then assume the choice of a particular airline as given and focus on how an airline can exploit its booking and inventory database to study the choice behavior of its own 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.

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 record based primarily 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.

4.2 The Passenger Choice Set

As mentioned in the previous chapter, one of the major limitations of booking data is that the choice set of the traveler is not recorded by airlines or global reservations systems in passenger bookings, also called passenger name records (PNR). As a result, the choice set must be inferred for each booking by combining booking data with other sources of data such as the airline flight schedule or seat availability data. If the dimension of the model is limited solely to the choice of an itinerary, it is relatively straightforward to reconstitute the passenger choice set based on the airline flight schedule. Previous studies of the choice of an airline itinerary based on booking data (Coldren, 2003; Grammig, 2005) made the implicit assumption that seats were always available on all itineraries and that the passenger choice set was then the same for all bookings in a specific directional origin-destination market.

Since the dependent variable of the model is here the choice of an itinerary and fare product, it becomes much more challenging to infer the passenger choice set at the time of the booking. Other airline planning decisions beyond the flight schedule such as pricing and revenue management have a major impact on the availability of the various fare products at the itinerary level. In order to incorporate the impact of airline pricing and revenue management, booking data was combined with flight schedule, fares rules and seat availability data to infer the choice set of each booking record.

Fare rules are often used to differentiate the different fare products and, as mentioned earlier, include both product features such as the flexibility to change travel plans and access conditions such as minimum stay or advance purchase requirements. Fare rules are typically set by the airline pricing department based on a range of criteria such as the company pricing strategy and the competition

in the marketplace. As a result, fare rules tend to be relatively stable and are likely to remain unchanged over the course of a booking period spanning several months.

However, airlines do not rely solely on differential pricing and fare rules to maximize revenues, as high-demand flights tend to be popular for both low-fare and high-fare passengers. In addition, leisure-oriented passengers purchasing primarily lower-fare tickets tend to book earlier than higher-fare business-oriented passengers, a trend that is often reinforced by some fare rules such as the advance purchase requirements of many heavily discounted airfares. In order to maximize revenues, most airlines have associated differential pricing with the dynamic management of airline seat inventory known as revenue management.

Each fare product gets assigned to a specific booking class and booking classes are ranked by their average expected revenue. Based on the level of current bookings and forecasts for future demand at the booking class level, the airline restricts the availability of lower-priced booking classes on high-demand flights to protect seats for expected late-booking higher-fare passengers. Unlike fare rules, these inventory controls (also called seat availability) are updated constantly based on both the booking activity and regular adjustments made by the airline revenue management system and staff to reflect changes in current and expected future demand. As a result, to obtain a reasonably accurate view of the airline inventory and availability of the different fare products at the time of each booking, seat availability data was collected daily over the booking period for all the itineraries considered in the booking data.

As shown on Figure 4-1 below booking, fares rules and seat availability data are combined to reconstitute for each booking record the passenger choice set at the time of the booking.
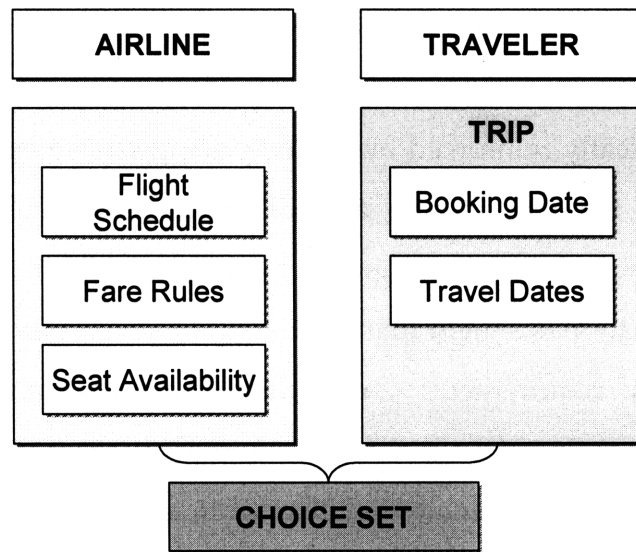
Figure 4-1:   Passenger Choice Set

For instance, a fare product is excluded for all itineraries in the passenger choice set if its advance purchase requirement is longer than the time difference between the dates of outbound travel and the date of the booking. In addition, a fare product on a specific itinerary is eliminated from the choice set if a search in the seat availability data indicates that, on the date of the booking, the airline was not accepting bookings on this itinerary in the fare class associated with this fare product. A detailed description of the fare products and the choice set generation process is provided in Chapter 5.

4.3   Heterogeneity of Behavior

There are two major approaches to model heterogeneity of behavior across observations: random coefficients and discrete segmentation. The basic idea of models with random coefficients is that each observation of the sample has its own preferences that differ from the average preferences by an unknown and, hence random amount. However, random coefficients may not be the most

appropriate approach to model heterogeneity of behavior in airline markets. The differences of behavior may not be randomly distributed across observations but are likely to be driven by specific elements that can be identified such as trip purpose. This is typically reinforced by airline pricing strategies that are usually designed to divide the market into a small number of fairly homogeneous segments. For instance, the most inexpensive fare products are often associated with rules designed to make them unattractive to business travelers such as the Saturday night stay requirement. Consequently, an alternative approach in which bookings are assigned to a discrete number of segments may be more appropriate to capture heterogeneity of behavior in airline markets than random coefficients.

Unlike models with random coefficients that capture random taste variation across observations, models based on a discrete segmentation group observations into meaningful segments that have similar needs, capabilities and preferences. The advantage of discrete segmentation models is that heterogeneity of behavior can be related to a set of specific causal variables. This approach has usually been applied to study heterogeneity of choice behavior in airline markets as demand for airline travel is typically segmented by trip purpose: Business travelers are assumed to select a travel option that best fits their schedule with less emphasis on cost while leisure travelers shop for air travel primarily based on price. In addition, business travelers usually have less flexibility and control over their schedule as they travel to attend business meetings with clients or business partners and are, in many instances, not able to plan their travel needs a long time in advance.

As mentioned earlier, most previous studies of airline passenger choice based on stated preference experiments have collected data on trip purpose and used that criterion to assign observations into two groups on a deterministic basis. For

instance, Prossaloglou and Koppelman (1999) and Adler et al. (2005) segmented the market between business and leisure travelers while Garrow, Parker and Jones (2007) divided their sample between reimbursed business travelers on one hand and self-pay business and leisure travelers on the other hand. Separate models were then specified for each of the segments (Prossaloglou and Koppelman, Adler et al.) or interaction variables were used for specific elements of the passenger utility function such as price and schedule-related explanatory variables (Garrow et al.).

4.3.1 Deterministic Segmentation

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 booking data, other elements of the booking record may be correlated with trip purpose and provide valuable information to segment airline bookings. There are two types of data included in booking records that 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 and travel-related characteristics such as frequent flyer membership and status. Gender is the only socio-economic characteristics of the traveler that is easily available in booking records. Gender is in many instances explicitly requested and recorded or can be inferred with a good level of accuracy from the passenger's first name. Frequent flyer information is also recorded in the PNR, if the traveler has provided his frequent flyer number at the time of the booking. As members of airline loyalty programs are required to provide their frequent flyer number to accrue credits, it can be assumed that most active frequent flyer members will actually do so while

77

making their travel arrangements. The frequent flyer number may also be used to determine the status of the traveler in the airline loyalty program. Most airline loyalty programs usually divide members between three or four tier levels based on their recent activity and reward them with increasing benefits such as extra mileage accrual or additional services such as priority check-in or complimentary lounge access. In many instances, the frequent flyer status is also explicitly included in the booking to enable customer service representatives to better serve the needs of the most loyal travelers at each step of their trip.

In addition to the traveler's profile, characteristics specific to the trip can be exploited to segment air travel demand. In particular, for roundtrip tickets, 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 wholly within a week is expected to be strongly correlated with business travel as business travelers tend to return home before the weekend, especially in short-haul markets. In addition, the airline fare structure tends to reinforce this trend as many leisure travelers are strongly discouraged to travel within a week in order to access cheaper fare products that require a stay over the weekend. The distribution channel of the ticket may also provide a fairly strong indicator of trip purpose. The distribution channel of the ticket, while not directly recorded in the PNR, can be inferred from the identification number of the booking agency (also called office identification number). 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 travel management services such as billing, enforcement of company travel policies or access to discounted corporate fares. As a result, although trip purpose is not recorded in airline bookings, these elements may provide an alternative to the traditional segmentation of air travel demand by trip purpose. Figure 4-2 below summarizes

the data that was extracted from booking records and used to capture heterogeneity of behavior in airline markets.

AIRLINE	TRAVELER	
Flight Schedule	**TRIP**	**PROFILE**
Fare Rules	Booking Date	Gender
Seat Availability	Travel Dates	Frequent Flyer Membership
	Distribution Channel	Frequent Flyer Status
CHOICE SET	**LATENT CLASSES**	
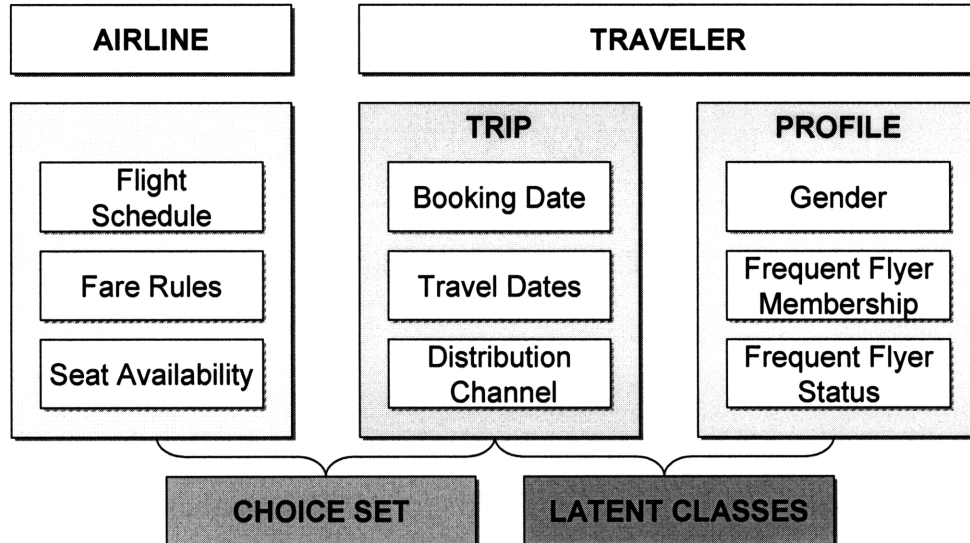
Figure 4-2: Heterogeneity of Behavior

While all these factors may contribute to segment demand, in practice, only a subset of them can be used under the conventional deterministic approach found in the literature as the number of segments would otherwise become too large. For instance, if three criteria are used such as week travel, distribution channel of the ticket and frequent flyer membership, the bookings need to be divided into eight segments. It may not be possible to identify differences in choice behavior for each of these segments due to the limited size of the sample or because sufficient variation in choice behavior may not exist across such small sub-segments. As a result, we will focus primarily on estimating univariate two-segment models. Estimation results will be compared to determine which criterion is the most effective to segment airline demand using a combination of goodness-of-fit measures and interpretation of the estimation results. In addition, a number of multivariate segmentation schemes will also be explored and compared to the latent class choice model described in the next section.

79

4.3.2 Probabilistic Approach: The Latent Class Choice Model

As mentioned in the previous section, the use of multiple factors to segment airline bookings under a deterministic approach automatically leads to an increase in the number of segments that may become too small to identify differences of choice behavior across bookings. As a result, a probabilistic approach based on latent classes provides the opportunity to use all the information available in a booking record without necessarily increasing the number of classes.

Latent classes are unobserved segments. Since we cannot directly identify to which class a particular booking record belongs, a probabilistic assignment process is used also called the class membership model. A range of observed factors that may affect class membership can be specified as explanatory variables of the class membership model. Since trip purpose is unobserved in airline bookings, latent classes appear to be an attractive and flexible approach to model heterogeneity of behavior for studies based on booking data. Several criteria extracted from booking records such as characteristics of the trip and the traveler's profile can be specified as explanatory variables of the class membership model while assigning the bookings to as few as two latent classes. Even if data on trip purpose were available, latent classes may still provide a valuable approach over a deterministic segmentation by trip purpose by supplementing it with other elements such as distribution channel of the ticket or elements of the traveler's profile.

4.3.3 Heterogeneity of Behavior within a Segment of Airline Demand

In addition to the difference in choice behavior between different segments of the market, 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 relate to 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 outbound itinerary is restricted to morning flight departures so that the traveler may have sufficient time to conduct his activities at the destination and travel back at the end of the day to the origin city and all other itineraries should be excluded from the passenger choice set. Even within this framework, the time-of-day preferences of day trippers may still differ from the rest of the travelers. Day trippers are expected to have a stronger preference for an early morning departure in order to have as much time as possible during their short stay in the destination city. However, differences in behavior may not be significant between day trip and overnight bookings regarding the rest of the utility function such as the disutility associated with the lack of flexibility to change travel plans of some fare products.

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 all roundtrip tickets. For models based on a deterministic segmentation of airline demand, interaction variables will be added to the part of the utility function related to the choice of an itinerary to determine whether the time-of-day preferences of day trippers differ from other bookings. Additional interaction variables may also be used to test whether day trippers have specific behavioral characteristics associated with other parts of the utility function.

For latent class choice models, there are two potential options to explore the heterogeneity of behavior associated with day trippers. The first option is to introduce a day trip dummy into the class membership model. This may be appropriate if the day trip dummy is expected to supplement other variables and

better capture heterogeneity of behavior across the different classes. However, no parameter estimates of the specific preferences of day trippers will be obtained. In order to capture the specific preferences of day trippers, interaction variables are introduced into the class-specific choice models. This second option seems especially attractive if the heterogeneity of behavior is related to a specific part of the utility function such as the time-of-day preferences. Parameter estimates of the specific preferences of day trippers can then be obtained for each latent class. Finally a combination of the two options can be used to test whether specific preferences of day trippers can be identified within each latent class while a day trip dummy may improve the class membership model and capture additional heterogeneity of behavior across latent classes.

In addition to heterogeneity of behavior that can be related to some specific variables such as a day trip dummy, there may remain additional heterogeneity of behavior for which a particular causal effect cannot be easily identified. Random coefficients will then be used within the class-specific choice models to test whether additional random taste variation can be identified.

4.4   Model Specification

As mentioned earlier, as the passenger choice set includes a finite number of travel alternatives, discrete choice models will be used to study the preferences of airline passengers. The choice model is combined with a class membership model to segment the market and assign bookings to latent classes. Finally, random coefficients will be added to test whether remaining random taste variation is observed within latent classes. The proposed model specification integrates these three components into a latent class model of airline passenger choice as shown in Figure 4-3 below.
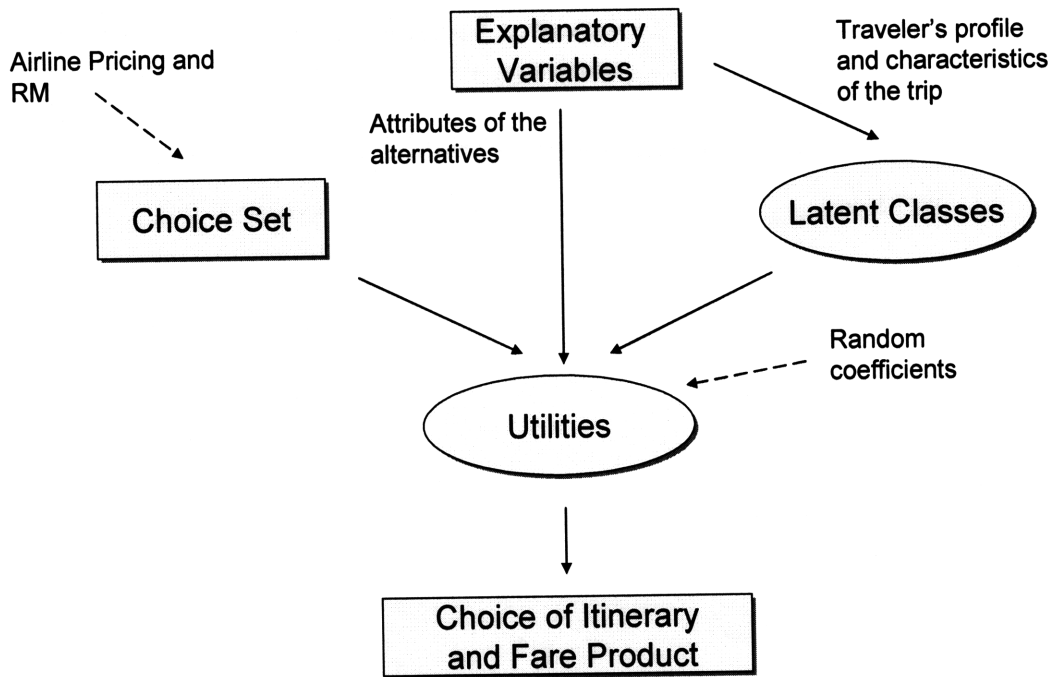
Figure 4-3:   The Latent Class Model of Airline Passenger Choice

As discussed earlier, each travel alternative in the choice set is defined as a combination of an itinerary and fare product. 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. A trip on the same flight but at a higher refundable fare of 400 EUR will be considered as another travel alternative.

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

$$P(i/X_M, X_C) = \sum_{s=1}^{S} \underbrace{P(s/X_M)}_{\text{Class Membership}} \underbrace{P(i/X_C, s)}_{\text{Choice}} \quad \forall i \in C \tag{4.1}$$

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:

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

$$U_{ib} = X_{Cib}\beta_C + \varepsilon_{ib} \tag{4.2}$$

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:

$$P\left(i / X_{Cib}, s, C_b\right) = \frac{e^{X_{Cib,s}\beta_{C,s}}}{\sum_{j=1}^{J_b} e^{X_{Cjb,s}\beta_{C,s}}} \qquad \forall \ s \in S \tag{4.3}$$

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

If random coefficients are added, $\beta_{C,s}$ becomes a random vector with variance $\Sigma_{\beta_{C,s}}$ and distribution $f(\beta_{C,s})$. The choice probabilities of this logit kernel model with random coefficients become:

$$P\left(i \, / \, X_{Cib}, s, C_b\right) = \int \frac{e^{X_{Cib,s}\beta_{C,s}}}{\sum_{j=1}^{J_b} e^{X_{Cjb,s}\beta_{C,s}}} f(\beta_{C,s})d\beta_{C,s} \qquad \forall \;\; s \in S \qquad (4.4)$$

The parameters that need to be estimated in this model include the mean class-specific parameters $\beta_{C,s}$ and the parameters of the variance-covariance matrix $\Sigma_{\beta_{C,s}}$. If the coefficients are assumed to be independent, only the standard deviation of each coefficient is estimated.

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:

$$P\left(s \, / \, X_{Mb}\right) = \frac{e^{X_{Mb,s}\beta_M}}{\sum_{t=1}^{S} e^{X_{Mb,t}\beta_M}} \qquad (4.5)$$

Where β_M are the unknown parameters of the class membership model

These three components are integrated together to form a latent class choice model with random coefficients:

$$P\left(i \, / \, X_{Mb}, X_{Cib}, C_b\right) = \sum_{s=1}^{S} \int \frac{e^{X_{Mb,s}\beta_M}}{\sum_{t=1}^{S} e^{X_{Mb,t}\beta_M}} \frac{e^{X_{Cib,s}\beta_{C,s}}}{\sum_{j=1}^{J_b} e^{X_{Cjb,s}\beta_{C,s}}} f(\beta_{C,s})d\beta_{C,s} \qquad (4.6)$$

85

In addition to parameter estimates of the class membership model, a set of parameter estimates of the choice model is then obtained for each latent class. The Latent Gold Choice software by Statistical Innovations that is designed for the estimation of latent class choice models was used to estimate the parameters of the model. For models without random coefficients, parameter estimates were obtained through maximum likelihood estimation techniques using a combination of the expectation-maximization (EM) and Newton-Raphson algorithms. The estimation process starts with a user-defined number (250) of EM iterations. The software then switches to a Newton-Raphson algorithm. The software exploits the advantages of both algorithms, i.e. the stability of EM at the beginning of the estimation process when it is far away from the solution with the speed of the Newton-Raphson when it is close to the optimum.

The specification of random coefficients in Latent Gold Choice is fairly restrictive: Random coefficients are always assumed to be normally and independently distributed. Since the multi-dimensional integral does not take a closed-form, it is approximated by means of Gauss-Hermite numerical integration. Due to the computational burden associated with this numerical procedure, a maximum of three random coefficients can be specified. For a complete description of the algorithms used by Latent Gold Choice, the reader is referred to the software's technical guide (Vermunt and Magdison, 2005).

4.5   Time-of-Day Preferences of Airline Travelers

While we have discussed so far the general structure of the model, we focus in this section on a specific part of the utility function, more specifically on how to model the time-of-day preferences of airline travelers. We first present the conventional approach based on time-period dummies. We describe an

alternative approach in which dummy variables are replaced by a continuous function of time over a 24-hour period. We then propose to estimate the duration of the cycle to better fit the time-of-day preferences of specific types of markets such as short-haul markets. We finally discuss how this new approach can be used to model the time-of-day preferences of specific types of bookings such as day trips.

4.5.1 Conventional Approach: Time-period Dummies

Another limitation of using booking data is that a passenger preferred departure or arrival time, i.e. the departure time or arrival time that would best fit his schedule requirements for the trip is not available in booking records. While such data has often be collected in stated preference data, only actual flight departure and arrival times on the booked itinerary are recorded in airline bookings while departure and arrival times for other itineraries can be obtained from the airline flight schedule. As a result, the time-of-day preferences of airline passengers are modeled in this research based on the observed flight schedule set by the airline and not on the underlying passenger ideal departure and arrival times.

The conventional approach to represent passenger preferences for a specific departure or arrival time (also called here time-of-day preferences) is to divide the day into a finite number of time periods and specify in the utility function a dummy variable for each period. For instance, Coldren et al. (2003) 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. As many network airlines tend to group flight departures at hub locations into connecting banks with flights arriving from all destinations at similar times at the beginning of the bank and leaving at similar times at the end of the bank after passengers connect between flights, an alternative could be to group flight

87

departure times by connecting banks. However, since time is continuous, a continuous function of time may provide a more precise estimate of the time-of-day preferences and offer an attractive alternative to discrete time-period dummies.

4.5.2 Continuous Function of Time

In their study of the choice of time of day in activity and tour based models, Abou Zeid et al. (2007) recognize that, since time is a continuous variable, the effect of any time-related variable included in the utility function should also be continuous. They propose to replace time-period dummies by a continuous function of time. Since 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. They propose to take advantage of the properties of the trigonometric operators and use a function of the following form:

$$
U(h) = \beta_1 \sin(\frac{2\pi h}{24}) + \beta_2 \sin(\frac{4\pi h}{24}) + \beta_3 \sin(\frac{6\pi h}{24})
$$

$$
+ \gamma_1 \cos(\frac{2\pi h}{24}) + \gamma_2 \cos(\frac{4\pi h}{24}) + \gamma_3 \cos(\frac{6\pi h}{24}) + ... \tag{4.7}
$$

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. As suggested by Ben-Akiva and Abou-Zeid (2007), 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 unattractive periods of the day where demand for air travel is extremely low. 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. While only the flight departure time is recorded and no data on the passenger ideal departure time is available in airline bookings, previous studies of airline passenger choice based on stated preference data have shown that extremely few passengers want to depart during the night in this type of market.

For instance, in their recent study of the choice of an airline itinerary based on a survey of visitors of an online travel website, Garrow et al. (2007) collected data on passenger ideal departure time for outbound travel in U.S. domestic markets. They classified the markets into three categories, East-West, West-East and North-South/South-North to account for the impact of time differences on time-of-day preferences. Figure 4-4 below shows the ideal departure time for passengers traveling in North-South/South-North markets where there is no time difference between the origin and destination cities and the average passenger length of haul is expected to be fairly short.
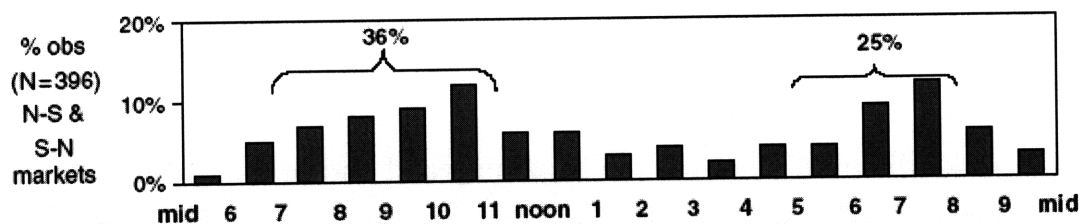
Figure 4-4: Passenger Ideal Departure Time in U.S. North-South Markets
(Source: Garrow, Parker and Jones, 2007)

As shown in the figure above, very few passengers stated that their ideal departure time was between midnight and 6 a.m. In addition, the travelers

sampled for this survey were using a travel website designed to search for the lowest fare in the marketplace. They may not necessarily be representative of the distribution of time-of-day preferences of the whole population traveling in these markets. As a result, airline travelers as a whole may be even more reluctant to travel at night than this specific segment of bargain-hunters.

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 (4.7) becomes:

$$U(h) = \beta_1 \sin(\frac{2\pi(h-s)}{d}) + \beta_2 \sin(\frac{4\pi(h-s)}{d}) + \beta_3 \sin(\frac{6\pi(h-s)}{d})$$

$$+ \gamma_1 \cos(\frac{2\pi(h-s)}{d}) + \gamma_2 \cos(\frac{4\pi(h-s)}{d}) + \gamma_3 \cos(\frac{6\pi(h-s)}{d}) + ... \qquad (4.8)$$

Where $\quad 1 - e \le d \le 24$

$\qquad 0 \le s \le e$

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 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 (4.8), the passenger utility function is not linear in the parameters any more. As Latent Gold Choice does not support a

90

utility function of this type, 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 any 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. When the start time of the cycle is modified, observed flight departure times are all shifted by the same amount of time relative to the start of the cycle. Their position relative to each other in the daily cycle remains then unchanged. 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 while the other parameters of the model were estimated using Latent Gold Choice.

As mentioned earlier, interaction variables will be used to capture the specific time-of-day preferences of day trip bookings. For a model specification based on time-period dummies, a second set of dummies specific to day trippers is introduced. However, the number of time-periods will be reduced to cover only the morning part of the day as outbound day trippers are observed to travel exclusively on morning flight departures. If a continuous function is used as an alternative to time-period dummies, a function specific to each type of booking is included in the utility function. The duration and start time of the cycle are then estimated for each category of bookings. For instance, the duration of the cycle is expected to be much shorter for day trip bookings. As a result, estimating the duration of the cycle provides a flexible approach to model the time-of-day preferences of specific categories of travelers such as day trippers.

4.6 Summary

In this chapter, we focus on how an airline can exploit its existing data to analyze the choice behavior of its own passengers. As a result, we take the choice of a particular airline as given and develop a model of the two major remaining dimensions of airline passenger choice, the choice of an itinerary and fare product. We discuss how booking data can be combined with other data sources such as seat availability data to incorporate the impact of airline decisions such as pricing and revenue management and reconstruct the choice set at the time of the booking. In addition, we develop an alternative to the traditional deterministic segmentation of airline demand by trip purpose and propose to use latent classes to segment the market based on a range of factors extracted from the booking records such as the profile of the traveler and the characteristics of the trip. We also propose a generalized formulation of a continuous function of time, in which the duration of the daily cycle is estimated rather than set to a full 24 hours to represent the characteristics of time-of-day demand in specific types of markets and categories of bookings.

In the next chapter, we analyze the data collected for this research and describe in more detail how it was processed to reconstruct the choice set of each booking.

Chapter 5   Data Collection and Choice Set Generation

As mentioned in the previous chapter, several sources of data are needed to estimate the choice of an airline itinerary and fare product. In this chapter, we will first describe how data was collected and the different types of data obtained for this research. We will then focus on an exploratory analysis of the booking data with an emphasis on the elements of the booking records that may provide an alternative to trip purpose to segment the demand in airline markets. Finally, we will describe in more detail how the different types of data are combined to reconstruct the passenger choice set for each booking in the dataset.

5.1   Data Collection Process

As shown in Figure 5-1 below, four types of data were collected to support this research on passenger choice models. Three of them were used to incorporate the impact of airline decisions such as scheduling, pricing and revenue management on the passenger choice set. Passenger decisions were obtained from the booking data and include the date of the booking, the dates of travel and flight itinerary and the fare product selected. These decisions were combined with the airline decisions to reconstruct the passenger choice set at the time of the booking for each record included in the dataset.
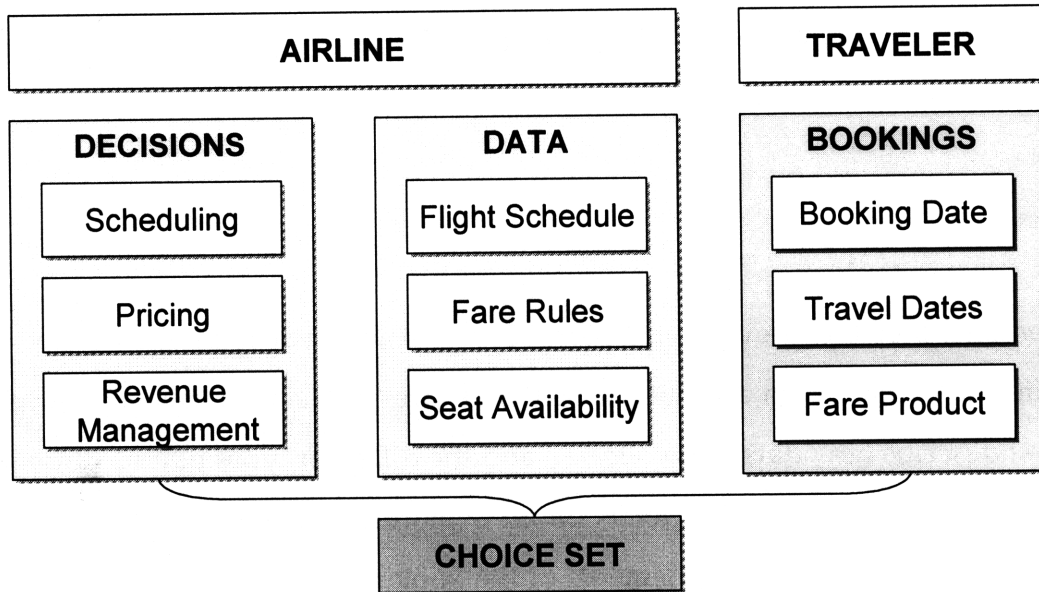
Figure 5-1:   Data Sources

5.1.1 Booking Data

Booking data for this research was obtained through a partnership with Amadeus, the leading European global distribution systems (GDS) used by many travel agents worldwide to book airline tickets on behalf of their customers. In addition to the distribution of airline tickets through its network of affiliated travel agents, Amadeus offers airlines a range of additional services depending on their level of participation. In particular, Amadeus offers airlines the opportunity to outsource all their ticketing activities in its Amadeus System User (ASU) program also called Altea Sell. By becoming an Amadeus System User, participating airlines still make their own pricing and inventory control decisions but pass this information regularly to Amadeus who manages all communications with the various distribution channels on the airline's behalf, completes all ticketing transactions and sends the final booking records back to the airline. This means that, for ASU airlines, Amadeus has a record of every booking including tickets booked by travel agents affiliated with a competing GDS or made through direct distribution channels such as the airline's website.

In order to have a complete set of bookings, 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 German business destinations: Dusseldorf (DUS), Frankfurt (FRA) and Stuttgart (STR). The distance and flight time is similar on these three routes: 257 miles to DUS, 291 miles to FRA and 312 miles to STR. Rail travel from Paris to these cities was, at the time of data collection, fairly unattractive with either no direct service (DUS) or a 6-hour minimum travel time (FRA, STR).

The booking data was processed through several steps to obtain a dataset that fulfills the needs of this research. First, only local demand in these three origin-destination markets was considered and all passengers traveling on flights between Paris and these cities connecting from or to other destinations were eliminated. In addition, the final dataset includes only travelers originating in Paris as the fare structure depends on the direction of travel meaning that the fare products offered for travel originating in Germany are different than for travelers leaving starting their trip in France. Data was collected 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, 574 in the PAR-DUS, 909 in the PAR-FRA and 532 in the PAR-STR market.

While all tickets were sold on a roundtrip basis, only the outbound leg of the journey was considered in order to reduce the number of alternatives to a tractable size and get a sufficient number of bookings for each alternative in the universal choice set. Although this may seem as a strong assumption, its impact is largely mitigated as different fare products can be combined on the inbound and outbound legs of the journey. In addition, even if a fare product requires a roundtrip purchase, the itinerary and fare information are presented on the

airline's website sequentially for each leg of the journey which means that the prospective traveler selects an itinerary and fare product on a leg by leg basis. A different fare product was actually observed on the outbound and inbound legs of the trip for 5% of the bookings in the dataset. For all roundtrip tickets, the equivalent one-way fare for the outbound leg was calculated as half of the roundtrip price of the fare product selected for that part of the trip.

5.1.2 Flight Schedule

Only non-stop itineraries were considered as the structure of the airline network is such that very few connecting itineraries are available in the markets considered. In addition, in such short-haul markets, travel time on connecting itineraries is extremely unattractive compared to non-stop flights. Like many hub network carriers, the airline structured its flight schedule around six connecting banks in order to offer short connection times at its major hub. As a result, similar departure times are observed in all three markets under consideration. Flight departure times were inferred from the booking records and cross-checked with the flight schedule provided in the Official Airline Guide (OAG). Table 5-1 below provides a list of daily non-stop flights with their departure time from Paris. Since local demand is fairly large in the Paris-Frankfurt market and the airline relies exclusively on narrow-body aircraft to serve short-haul European routes, two flights were scheduled by the airline during the early morning bank to serve peak demand leaving from Paris in the morning and three flights during the early afternoon bank to have sufficient capacity for passengers returning from Frankfurt in the late afternoon peak period (5 to 7 p.m.).

Schedule Bank	DUS	FRA	STR
Early Morning	7:20 AM	7:00 AM 7:35 AM	7:25 AM
Late Morning	10:20 AM	9:45 AM	10:30 AM
Midday	1:05 PM	1:05 PM	12:45 PM
Early Afternoon	3:35 PM	3:25 PM 4:00 PM 4:50 PM	3:30 PM
Late Afternoon	6:30 PM	6:25 PM	7:05 PM
Evening	8:35 PM	8:10 PM	

Table 5-1:   Flight Schedule (Source: OAG)

5.1.3   Fare Rules

As already discussed, airlines use differential pricing to segment the demand and take advantage of the differences in behavior between business and leisure travelers to increase load factors and revenues. They offer a range of fares (also called fare structure) and discounted fare levels are typically associated with a set of restrictions or fare rules designed to make them unattractive to business-type travelers.

The list of fares offered in these markets and their fare rules were obtained from Sabre, another GDS company and accessed through the Travelocity website. The airline was using a similar fare structure in many European short-haul markets from Paris, including the three markets considered in this research. This fare structure is a mix of the traditional pricing strategy of 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 supplement the unrestricted fare with several price points associated with different levels of required advance purchase.

The fare structure includes the following four categories of products:

- 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 during the 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 fare 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 5-2 below provides a list of all fare products with their respective fare rules and equivalent one-way fare in the directional 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 5-2:   Fare Structure

5.1.4 Seat Availability Data

As discussed in the previous chapter, the airlines have combined differential pricing with the dynamic management of seat capacity to protect seats for late-booking high-fare demand. As a result, airlines have built over the years sophisticated revenue management systems designed to manage in real-time the supply of seats available to each fare product on each flight or itinerary based on current booking activity and forecasts for future demand. The state of the airline inventory, also called fare class or seat availability is constantly updated by the airline reservations system based on current booking activity and updates made by the airline revenue management system and staff and communicated to all distribution channels, including the GDS such as Amadeus.

In order to incorporate the impact of revenue management, 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 in the three markets under consideration departing between July 1 and July 7. Table 5-3 below shows an example of seat availability data for flight 1306 from Paris to Dusseldorf departing at 7:20 a.m. on July 4, 2005 at various stages of the booking process.

Availability Date	Advance Purchase	Seat Availability
4 APR 2005	90	Y9S9B9K9R9M9H9Q9T9V9L9X9U9W9E9A9N9I9
5 MAY 2005	60	Y9S9B9K9R9M9H9Q9T9V9L9X9U9W9E0A0N0I0
6 JUNE 2005	30	Y9S9B9K9R9M9H9Q9T9V9L9X9U9W9E0A0N0I0
13 JUNE 2005	21	Y9S9B9K9R9M9H9Q9T9V9L9X9U0W0E0A0N0I0
20 JUNE 2005	14	Y9S9B9K9R9M9H9Q9T9V9L9X9U0W0E0A0N0I0
27 JUNE 2005	7	Y9S9B9K9R0M9H9Q0T9V6L6X6U0W0E0A0N0I0

Table 5-3:   Example of Seat Availability Data

Each fare product gets assigned to a booking class based on the first letter of the product code and the seat availability data shows the maximum number of seats available for each class. For instance, 30 days before departure, the NAP30 and EAP21 products were unavailable and the WAP14 was the cheapest available product from the Saturday Night Stay category. Similarly, for passengers traveling within a week, the AWEEK21 product was unavailable and UWEEK14 was the cheapest available product at that point of the booking process.

5.2 Exploratory Analysis of the Booking Data

In this section, we describe in more detail the booking data collected for this research. We will first analyze the distribution of the booking data along the two major dimensions of airline passenger choice modeled in this research, itinerary and fare product. We will then analyze elements of the booking records that will be used to segment bookings such as the characteristics of the trip and the profile of the traveler both across the whole dataset and by fare product category.

A corporate discount was used in 1049 bookings or 52% of the bookings of the dataset. Since no information is available in booking records on eligibility for a corporate discount, all eligible travelers were assumed to select a corporate discount fare product. This means that for corporate discount bookings, all other fare products are excluded from the passenger choice set and the dependent variable is reduced to the sole choice of an itinerary. While the implications of this assumption will be further discussed in Section 5.3, corporate discount bookings are excluded from the analysis of the distribution of bookings by fare and fare product category but are included in the rest of the exploratory analysis relative to the distribution of bookings by flight departure time, characteristics of the trip and traveler profile.

5.2.1 Analysis of the Booking Data by Fare and Schedule

The distribution of bookings by fare category is similar in all three markets and is largely oriented toward business travel during the week. Even after excluding corporate contract bookings, Flex and Week fare products account for over 60% of the bookings in these markets as shown in Figure 5-2 below.
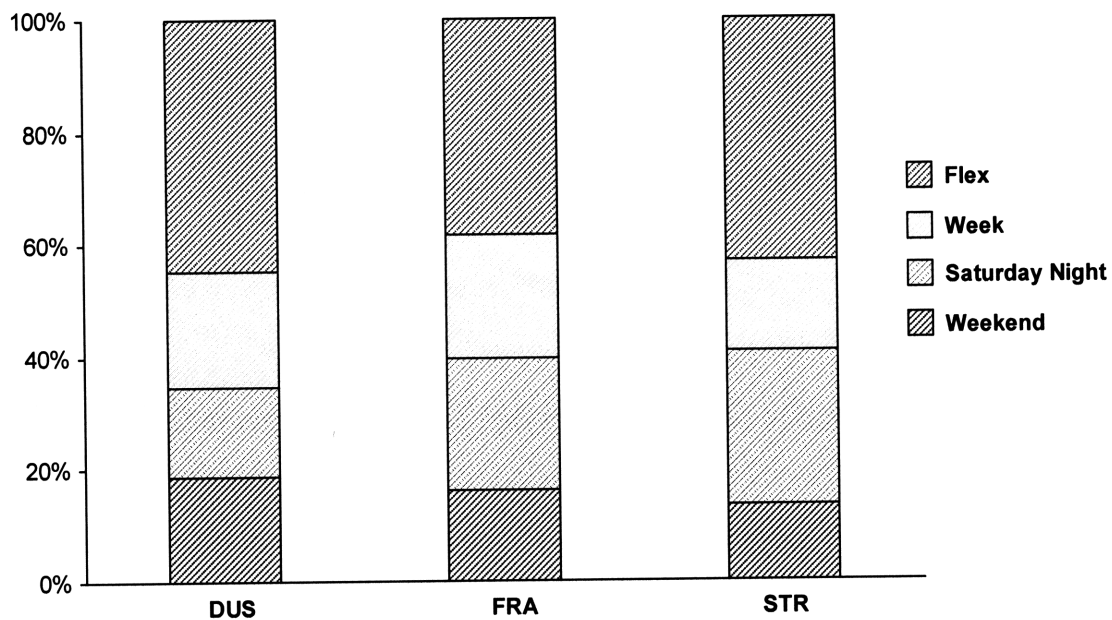
Figure 5-2:   Distribution of Bookings by Fare Product Category

Given the relatively large number of Week and Flex bookings in these business-oriented markets, the average fare including taxes and charges over the entire dataset is 243.19€ one-way with a standard deviation of 109.69€. As shown on Figure 5-3 below, about 60% of the bookings had a ticket price of 200 Euros and above in all three markets considered.
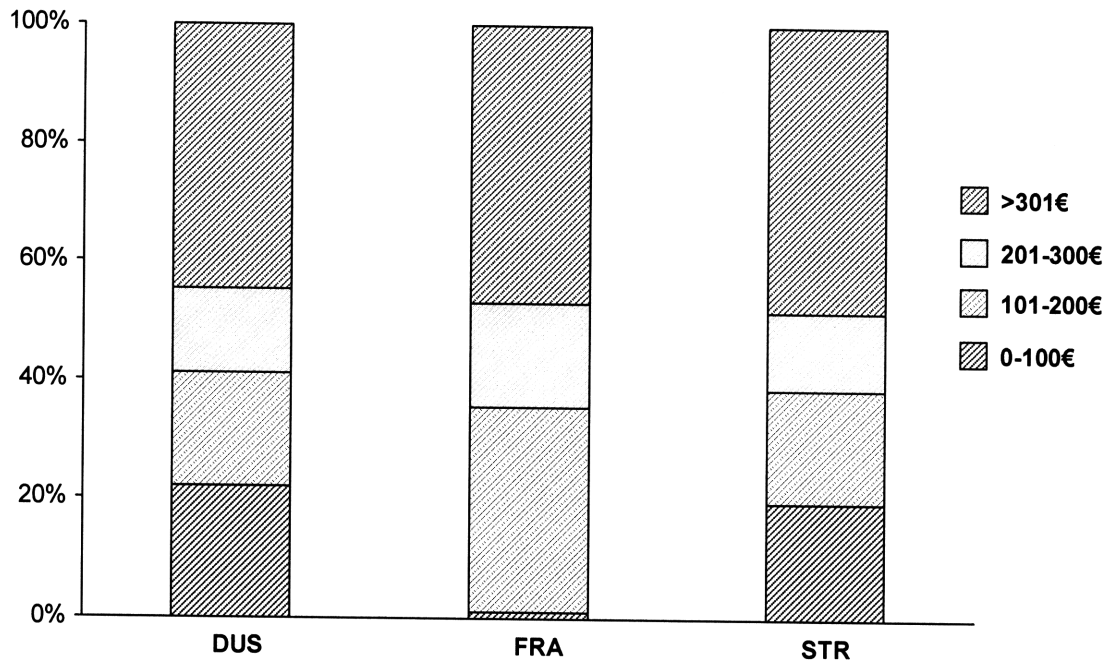
Figure 5-3:   Distribution of Bookings by Ticket Price

Similarly, the distribution of bookings by flight departure time reflects the strong orientation of these markets toward business travel. Early morning departures tend to be the most popular departure time for outbound travelers in these markets. While early morning flights require an early arrival at the airport in most cases before 6:30 a.m., they allow these early bird travelers to reach the destination airport by 8:30 a.m. and to conduct business activities at their final destination by the middle of the morning. Afternoon and late afternoon flights are the second most popular departure time, while fewer travelers selected a midday or evening flight departure.
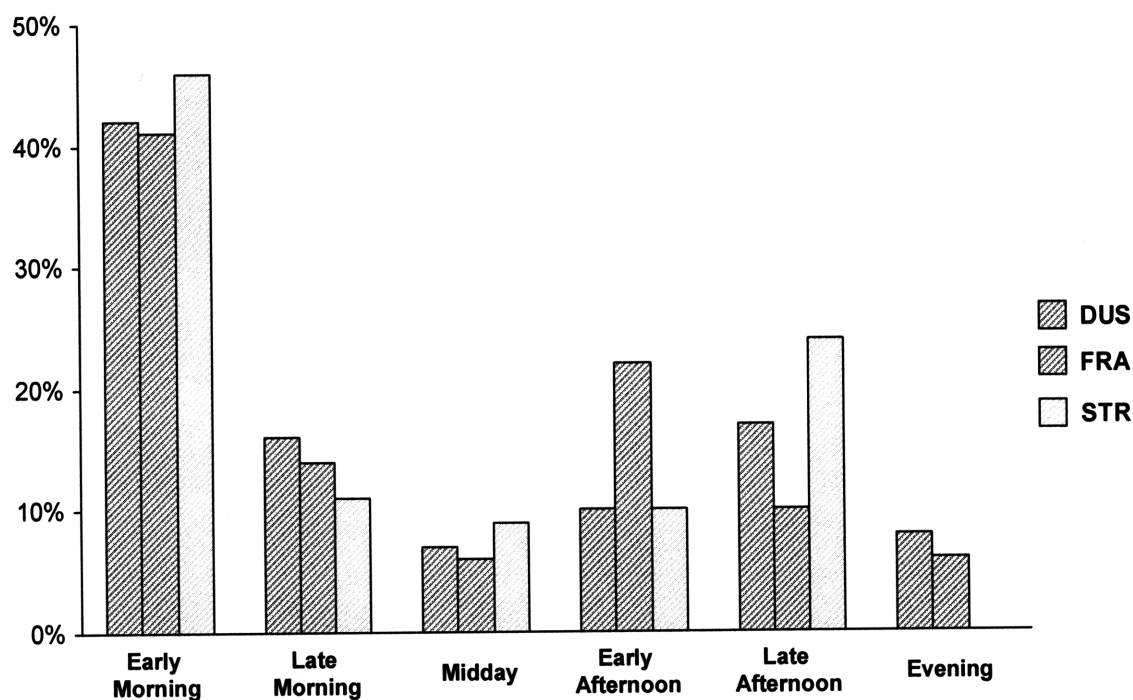
Figure 5-4: Distribution of Bookings by Flight Departure Time

5.2.2 Profile of the Traveler and Characteristics of the Trip

The profile of the typical travelers in these markets also tends to reflect the orientation toward business travel. As mentioned earlier, two elements of the traveler's profile were retrieved from the booking records: Gender and the membership and status in the airline loyalty program. As shown in Figure 5-5, a large proportion of the travelers in all three markets are of the male gender. In addition, many travelers in the dataset belong to the airline loyalty program. A large proportion of travelers in these markets, about 40% of all travelers, have reached elite status in the airline loyalty program. This means that many travelers in these markets fly on a frequent basis with the airline, an element that is usually related to business-type travel.
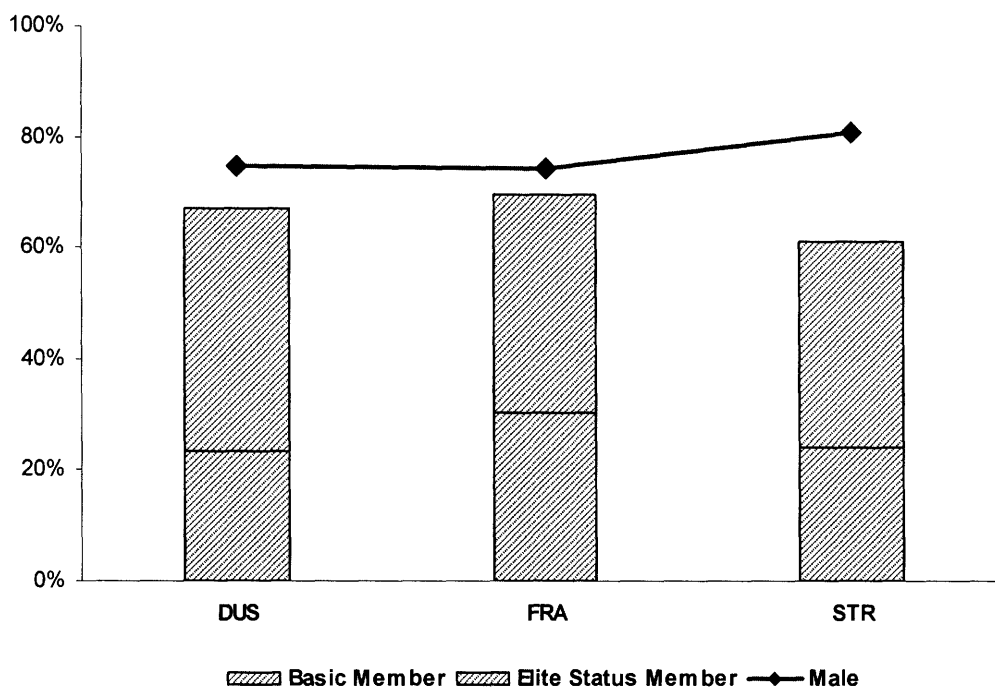
Figure 5-5:   Gender and Frequent Flyer Information

As shown on Figure 5-6 below, over 80% of the bookings in these markets were made through a traditional offline travel agent. Traditional European network airlines, unlike their low-cost competitors, were still in mid-2005 at the early stage of the development of online and direct distribution channels. This is also an indication of the business orientation of the market as many business travelers are required to book through a travel agent, especially if they want to take advantage of a corporate discount. Similarly, as business travelers tend to be less able to anticipate their travel plans and book tickets a long time in advance, the analysis of the booking curves shows that a large proportion of the travelers booked their tickets in the last two weeks before departure (Figure 5-7).
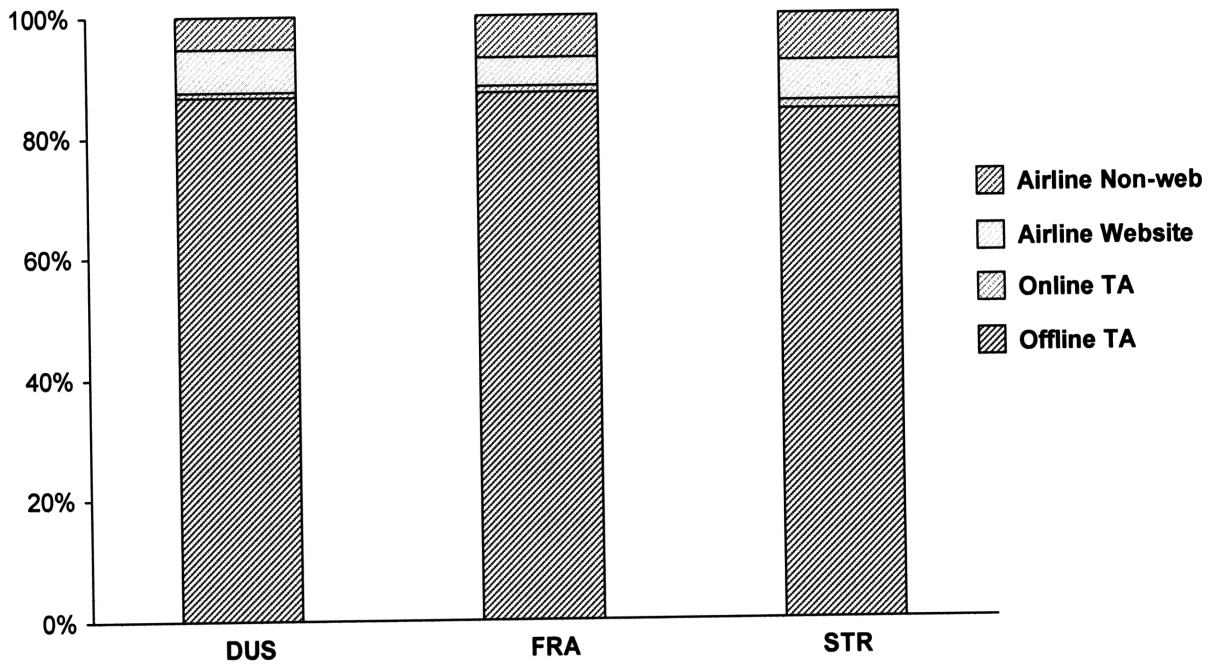
Figure 5-6:   Distribution of Bookings by Distribution Channel
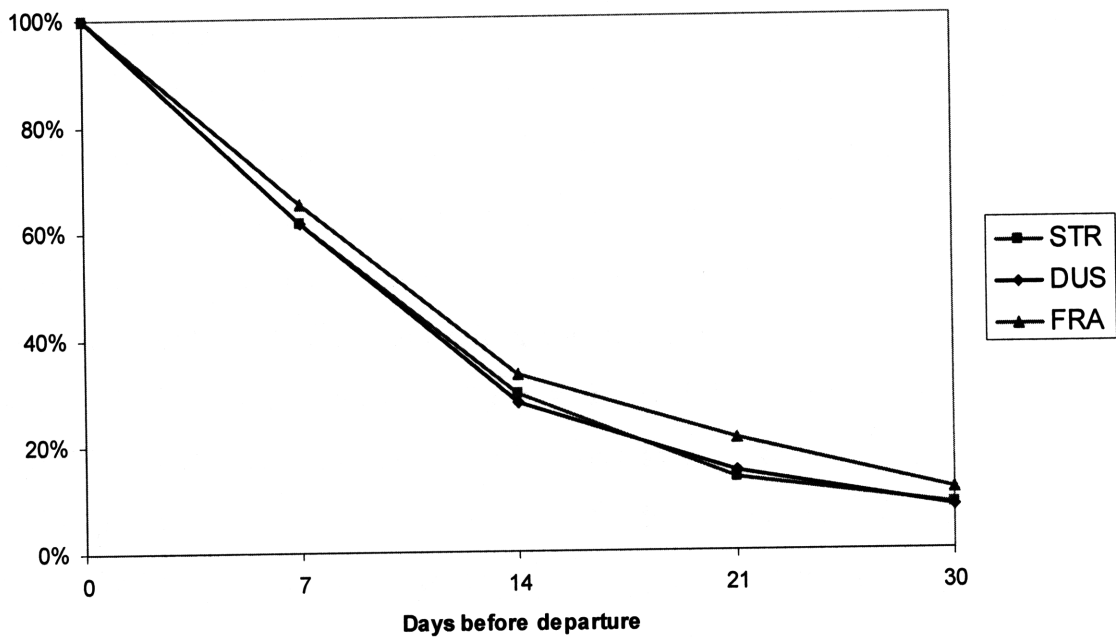
Figure 5-7:   Distribution of Bookings by Advance Purchase (Booking Curves)

5.2.3 Analysis of the Booking Data by Fare Product Category

While the analysis of the booking data at the market level reveals the strong orientation of these markets toward business travel, we will focus in this section on the characteristics of the bookings by fare product category. As expected, the Saturday Night Stay and Week categories have the highest average advance purchase as travelers are required to book early in order to take advantage of a larger discount (Figure 5-8). While the average advance purchase of the Weekend fare product that requires only 1-day advance purchase is lower than for the Saturday Night Stay category, it is largely higher than for the Flex category. This suggests that, while some of the week and non-week travelers anticipate their travel plans to take advantage of lower fares, week travelers also tend on average to book their tickets later than non-week travelers.

In addition to advance purchase behavior that is partially driven by the airline fare structure, fare product categories also differ with regard to characteristics of the trip such as the distribution channel of the ticket or elements of the traveler's profile such as frequent flyer membership. As expected, the proportion of loyalty program members is higher for the fare product categories designed primarily for business-type week travelers. Also as expected, the proportion of bookings through a traditional offline travel agent is the highest for the Flex category. However, the proportion of offline travel agents bookings is lower for the Week than for the Saturday Night Stay category. This suggests that the Week category attracts at least partially some technology-savvy price-sensitive week travelers whose needs may not have been previously satisfied with flexible albeit more expensive fare products.
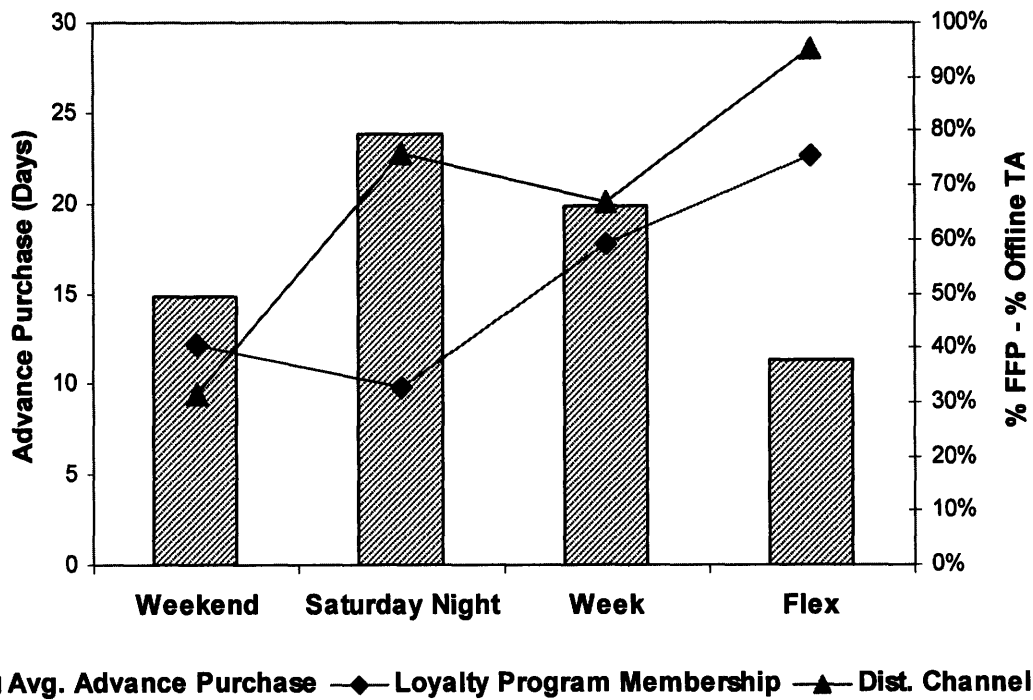
Figure 5-8:   Characteristics of the Bookings by Fare Product Category

While the characteristics of the bookings vary widely across fare product categories, the large proportion of corporate discount and non-discounted Flex bookings shows the strong orientation of these three European short-haul markets toward business travel during the week. In the next section, we will describe in more detail how some of the characteristics of the bookings analyzed in this section, such as the advance purchase of the ticket, are combined with airline decisions such as pricing and revenue management to reconstruct the passenger choice set for each booking.

5.3 Choice Set Generation Process

As discussed earlier, while the universal choice set in a market includes all possible combinations of an itinerary (a non-stop flight departure) and a fare product, the actual choice set varies for each booking based on the interaction between airline decisions such as pricing and revenue management and the characteristics of the bookings, primarily the dates of travel and the date of the booking (see Figure 5-1). In this section, we describe in more detail how the choice set is generated for each booking in the dataset.

As shown in Figure 5-9 below, the passenger choice set was generated through a five-step process.

Booking Loop

Fare Rules

Seat Availability

Dominance Rules

Corporate Contract

Day Trip

Next Booking
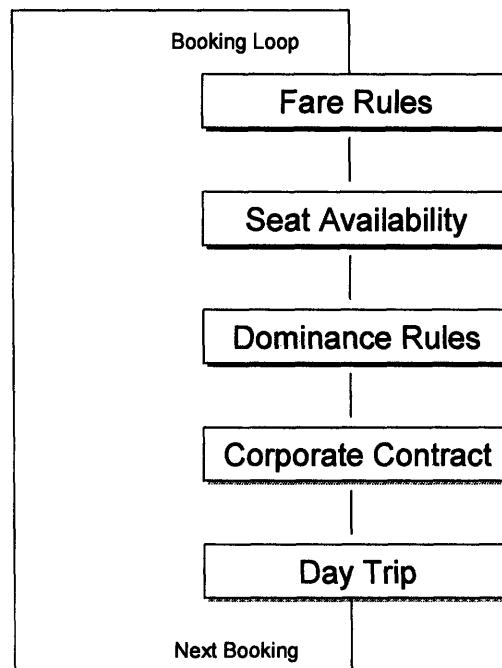
Figure 5-9:   The Five-Step Choice Set Generation Process

Step 1: Apply the Fare Rules

As mentioned earlier, the airline is using fare rules to differentiate fare products, segment demand and increase revenues. Two elements of the booking record were considered to apply the fare rules set by the airline pricing department: dates of travel and date of the booking.

The dates of outbound and inbound travel are used to determine if the trip satisfies the minimum stay and maximum stay restrictions of some fare products,. For instance, if the dates of travel show that the trip did not include a Saturday night stay, all fare products from the Saturday Night Stay category were eliminated from the choice set. Similarly, the Weekend product was eliminated unless the date of outbound travel was a Friday or Saturday with a return on the following Sunday or Monday. Finally, all Saturday Night Stay products were eliminated if the duration of stay exceeded one month.

In addition, the dates of travel were combined with the creation date of the booking to determine whether the advance purchase requirements of some fare products were satisfied. Fare products that required an advance purchase longer than the time difference between the creation date of the booking and the date of outbound travel were also eliminated from the choice set.

Step 2: Incorporate the Seat Availability Controls

For each booking, the seat availability data was searched to determine on the date the booking the availability of each fare product on all itineraries in the choice set: Products with no available inventory were eliminated. No booking in the dataset was observed to occur more than three months prior to travel, when the collection of seat availability data for future departures started.

Step 3: Observed Dominance Rules

In addition to airline pricing and revenue management decisions, a set of three dominance rules was observed from the booking data.

First, while several fare products of a same category may potentially be available on the same itinerary, the cheapest fare product was always chosen as it is cheaper and has similar characteristics to other products in the category. As a result, only the cheapest available fare product within a category was included in the choice set and all other products of the same category were eliminated.

Second, when the dates of travel of the booking and seat availability controls are such that the Weekend fare product is included in the choice set on at least one itinerary, it was always selected. The attractiveness of the Weekend product is due to both its low fare and, unlike heavily discounted fare products from the Saturday Night Stay category, the possibility to book it even at the last minute subject to seat availability. As a result, for many travelers, the alternative fare product from the Saturday Night Stay category is likely to be much more expensive explaining the systematic preference observed for the Weekend fare product. As a result, if a Weekend fare product is included in the choice set on any itinerary, all other fare products were eliminated and the model reduces to the sole choice of an itinerary.

The third observed dominance rule relates to bookings with dates of travel such that products from the Saturday Night Stay category may be included in the choice set. For such bookings, products from the Week and Flex categories are eliminated for all itineraries in which a Saturday Night Stay product is available. Week fare products are always less attractive as they are less flexible and more expensive than the product from the Saturday Night Stay category with the same

level of advance purchase requirement. In addition, the large difference in price explains why the Saturday Night Stay fare product is always preferred to the Flex fare when both are offered on the same itinerary. However, the Flex product was selected in five observations where the Saturday Night Stay product was not available on the chosen itinerary but was included in the choice set on other itineraries.

Step 4: the Corporate Contract Assumption

As mentioned earlier, since no data is available in booking records on the eligibility for a corporate discount, it is assumed that all eligible travelers have booked a corporate discount Flex product. As a result, if a corporate discount fare product is selected, all other fare products are eliminated from the choice set and the model is reduced to the sole choice of an itinerary. While this may be viewed as a strong assumption, especially since corporate discount bookings account for 52% of the bookings in these markets, its impact is somewhat mitigated as corporate discount products are highly attractive for passengers traveling wholly within a week. As discussed below, corporate contract fares provide full flexibility to change or cancel travel plans for either a low premium over the discounted Week fare products or even a cheaper fare in many instances.

For 97.4% of corporate discount bookings, fare products from the Weekend and Saturday Night Stay categories are excluded from the choice set as the dates of travel do not satisfy the fare rules of these products. For all these bookings, only a potentially somewhat cheaper albeit totally non-flexible fare product from the Week category would potentially be included in the choice set along with the corporate discount alternative. Moreover, the corporate discount product is the dominant alternative if a BFIRME fare product was purchased less than 14 days

in advance as it is cheaper and offer more flexibility than the UWEEK7 and RWEEK products from the Week category. Similarly, a SFIRME fare product is dominant if it was purchased less than 7 days in advance as it is cheaper than the RWEEK alternative. The corporate discount fare product was then dominant in 49.2% of all corporate discount bookings, mitigating the impact of this assumption.

Step 5: Day Trip

Unlike previous steps in which select fare products were removed from the choice set, in this step, specific itineraries are eliminated to account for the characteristics of day trip bookings. Day trip bookings occur when the outbound and inbound legs of the trip are booked for travel on the same day. In that case, it is observed that either an early or a late morning outbound itinerary is always selected so that the traveler may have sufficient time to conduct his activities at the destination and travel back at the end of the day to the origin city. As a result, for day trip bookings, all other itineraries are eliminated from the choice set.

To illustrate the different steps of the choice set generation process, the following two bookings were extracted from a similar dataset. Both trips were booked on June 22, 2005, for departure on Thursday, July 7, 2005. Let us first apply the fare rules based on the dates of travel and date of the booking. Since the trip was booked 15 days in advance, all fare products that require more than 15 days of advance purchase from the Saturday Night Stay (NAP 30 and EAP 21) and Week (AWEEK21) categories are eliminated from the choice set for both bookings. However, while she returned on Saturday, July 9, 2007, he stayed over Saturday night at the destination city and returned on Monday, July 11, 2007. As a result, all Saturday Night Stay fare products are eliminated from her choice set while fare products from this category remain included in his choice set.

Let us then incorporate the seat availability controls set by the airline. The seat availability data on June 22, 2005 for travel on July 7, 2005 is displayed in the lower middle part of Figure 5-10 for the first three flights of the day. Revenue management controls restricted the availability of fare products from the Saturday Night Stay and Week categories on the two later flights as only Flex fare products remained available. As a result, all remaining Saturday Night Stay and Week products were eliminated from the choice set for these two flights while they remained available on the first flight of the day.

Let us then apply the three dominance rules described above. First, fare products from the Saturday Night Stay (QAP7 and MSX0) and Week (UWEEK7 and RWEEK) categories were eliminated from the choice set on flight 1 as a cheaper product of the same category was available. Since the dates of travel of the bookings did not satisfy the requirements of the Weekend fare product, this dominance rule did not apply. Finally, products from the Week and Flex categories were eliminated on Flight 1 from the choice set of the traveler returning on Monday, July 11, 2007. However, it should be noted that Flex fare products were not eliminated from the choice set on Flights 2 and 3 as no product from the Saturday Night Stay category was available on these flights.

Finally, since neither of these bookings involved a corporate discount fare product or a day trip, the last two steps of the choice set generation process did not apply to this specific example.

Given their respective choice set, she selected the UWEEK14 product from the Week category on Flight 1 departing at 6:40 a.m. for a one-way fare of 178.58€ while he chose the WAP14 fare product on the same flight and paid a one-way fare of 116.08€.

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

Outbound Departure Date: Thursday, July 7, 2005

Inbound Departure Date: Saturday, July 9, 2005
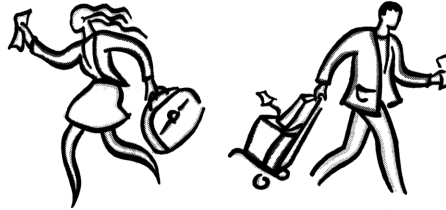
Booking Date: June 22, 2005 (AP=15)

Outbound Departure Date: Thursday, July 7, 2005

Inbound Departure Date: Monday, July 11, 2005

	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 5-10:   An Example of the Choice Set Generation Process

5.4   Summary

In this chapter, we have described how several sources of data are combined to incorporate the impact of key airline decisions such as pricing and revenue management and reconstruct the passenger choice set at the time of the booking. In addition, the analysis of the booking data has shown the strong orientation of these European short-haul markets toward business travel and the interactions between the airline fare structure and the characteristics of the trip and the profile of the travelers.

In the next chapter, we will use this dataset to estimate a latent class model of the choice of an airline itinerary and fare product and we will discuss how the trends found in the exploratory analysis of the booking data are reflected in the estimation results of the model and the choice preferences of the travelers.

Chapter 6 Estimation Results

In the previous two chapters, we have developed a latent class model of the choice of an airline itinerary and fare product and discussed how different sources of data were collected and processed to incorporate the impact of airline pricing and revenue management on the passenger choice set. We present here the estimation results of the model based on booking data collected in three business-oriented European short-haul markets. We first introduce in more detail the explanatory variables of the model and the methodology used for hypothesis testing and model selection. We then analyze the estimation results of a two-class latent class model of airline passenger choice with a continuous function of time. We discuss why such a model specification is preferred over a deterministic segmentation of the bookings between week and non-week travelers as well as the benefits of using a continuous function of time over time-period dummy variables. We finally discuss potential extensions of the model such as the addition of random coefficients.

6.1. Model Variables

As mentioned in Chapter 4, the dependent variable of the model is the combination of an itinerary and a fare product. As the airline fare structure in these markets includes 13 fare products and only non-stop flights are considered, the number of alternatives in the universal choice set varies from 117 in the PAR-FRA (9 daily flights) to 78 in the PAR-DUS (6 daily flights) and 65 in the PAR-STR markets.

6.1.1 Class-Specific Choice Model

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 in this study and flight 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 an adjustable duration of the cycle is used to model the time-of-day preferences of the passengers as described in Equation (4.8). As suggested by Ben-Akiva and Abou-Zeid (2007), the number of parameters included in the trigonometric expansion is determined empirically based on whether the resulting profile of the function matches our expectations of the typical time-of-day preferences of outbound airline travelers in short-haul markets. A trial and error method with half-hour increments is used to determine the duration of the cycle that maximizes the log-likelihood of the model. As mentioned in Chapter 4, a different 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, expressed on a one-way basis including taxes and surcharges. In addition to the fare, other attributes include the fare rules set by the airline pricing department. As discussed earlier, there are two types of fare rules. Some fare rules are conditions that have to be satisfied to access specific fare products such as minimum and maximum stay or advance purchase requirements. These fare rules are not included as explanatory variables in the utility function as their impact is incorporated in the choice set generation process. The other type of fare rules includes features of the products such as the

116

flexibility to modify the booked itinerary and/or cancel the trip. As indicated in Table 5-2, fare products from the Saturday Night Stay and Week categories are associated with restrictions on modifying itineraries or canceling the trip, while the Flex fare product allows for unlimited changes to the passenger itinerary and is fully refundable. The flexibility to change travel plans is then used by the airline to differentiate fare products and segment the demand, in particular for week travelers as fare products from the Week category are highly restricted with no changes or cancellations permitted in most instances.

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 passenger utility function. Since the Weekend and corporate discount products are excluded from the analysis of the choice of a fare product, a total of ten different fare products are considered and a maximum of nine dummy variables can be included in the utility function with the fully unrestricted Flex fare used as a base. The coefficients of these dummy variables are expected to be negative reflecting the disutility associated with these rules relative to the fully unrestricted Flex fare.

Since some of the fare products offered by the airline 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 fare rule is then obtained. These estimates provide a metric to evaluate the effectiveness of the airline pricing strategy. The rules regarding the flexibility in travel plans were combined with the advance purchase requirements of the fare products. This is to reflect that the uncertainty in travel plans and hence the need to change an itinerary or cancel the trip tends to increase with how long in advance a fare product has to be booked and the traveler needs to anticipate its travel plans.

Grouping the fare products into categories is equivalent to constraining the value of product-specific dummy variables to be equal across all the products in a category. A likelihood ratio test can 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 (used as a base).

Table 6-1 below provides a list of the fare products included in each of the four categories used in the model.

Explanatory Variables	Saturday Night	Week	Flex
NON-FLEX + 21AP	NAP30 EAP21	AWEEK21	
NON-FLEX + 7-14AP	WAP14 QAP7	UWEEK14 UWEEK7	
NON-FLEX + 0AP	MSX0	RWEEK	
FLEXIBLE			S

Table 6-1:   Fare Product Classification

Thus, the utility function of class-specific choice models contains three set of explanatory variables:

- The attributes of the fare products described above (FARE, and NON-FLEX & 21AP, NON-FLEX & 7-14AP, NON-FLEX & 0AP).
- A continuous function of time for the overnight bookings including the variables from the trigonometric formulation (SIN2PI-OV, COS2PI-OV,...) and the duration of the overnight cycle (DUR-OV).
- A similar function for the day trip bookings (DUR-DT & SIN2PI-DT, COS2PI-DT,...).

6.1.2 Class Membership Model

In addition to the attributes of the alternatives, other elements of the booking records were extracted to segment demand and capture heterogeneity of behavior across air travelers. As mentioned earlier, demand for air travel is usually segmented by trip purpose. However, since trip purpose is not observed in booking records, characteristics of the trip and the traveler are used as substitutes to segment demand. They are 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.
- Week travel (MON to FRI): If the outbound travel started on or after Monday and inbound travel occurred on or before Friday of the same week, this dummy variable will be equal to 1.
- Distribution channel of the ticket (OFFLINE TA): A dummy variable is added to indicate if the ticket was booked through an offline travel agent.

6.2   Hypothesis Testing and Model Selection

The development of a model specification is an iterative process that relies on rigorous statistical methods as well as a priori assumptions and judgmental assessment by the model-builder. In this section, we describe the statistical tools that were used to support the development of the latent class model of airline passenger choice proposed in this research and compare it to model specifications similar to those found in the literature.

Our analysis was based on two sets of tools, tests of hypotheses and measures of goodness of fit. In addition to the asymptotical t-test used to assess individually the significance of the parameters, the likelihood ratio test was used to compare across model specifications when the hypotheses are nested. The hypotheses are said to be nested if a model also called the restricted model is a special case of a more general model called the unrestricted model under the assumption that the restrictions can be expressed as linear constraints on a subset of the parameters of the unrestricted model. The likelihood ratio test compares the values of the log likelihood functions for the restricted and unrestricted model. The statistic for the likelihood ratio test follows asymptotically a chi-squared distribution with r degrees of freedom and is equal to:

$$-2 \left(\text{Log } L \, (R) - \text{Log } L \, (U) \right) \tag{6.1}$$

Where   Log L (R) is the log likelihood of the restricted model

Log L (U) is the log likelihood of the unrestricted model

r is the number of independent restrictions imposed on the parameters of the model

If the value of the test statistic is lower than the critical value obtained from the chi-squared distribution with r degrees of freedom for a specific level of confidence, the assumption that the restrictions are true cannot be rejected.

As mentioned in the previous section, a model specification in which fare products are grouped by fare rules is a special case (restricted) of a model with one dummy variable for each fare product if we assume that the value of the dummies is the same across all products grouped into a category. For instance, grouping the three non-flexible fare products requiring at least 21 days of advance purchase (NAP 30, EAP 21 & AWEEK 21) into a single category can be expressed as follows:

$$\beta_{NAP30} = \beta_{EAP21} = \beta_{AWEEK21} \qquad (6.2)$$

In order to test this set of two independent restrictions, two estimation runs are needed: One for the unrestricted model with a dummy variable for each of the three fare products and one for the restricted model with a common dummy for all three products. If the test statistic shown in equation (6.1) is less than 5.99, we cannot reject the null hypothesis that the coefficients of these three fare products are equal at a 0.05 level of significance. The likelihood ratio test would then support grouping these three fare products into a single category and estimate jointly the disutility associated with their characteristics such as the lack of flexibility to change travel plans without penalty.

Similarly, a likelihood ratio test can also be applied to test for heterogeneity of behavior across observations. If observations are grouped into several market segments on a deterministic basis, the model with all the observations pooled together can be considered as a special case of the model with deterministic segments under the assumption that each parameter of the model is of equal value across the different segments. As a result, a likelihood ratio test can be

121

used to determine whether or not a model with a deterministic segmentation of the observations is needed to capture heterogeneity of behavior. For a more complete description of hypothesis testing, the reader is referred to Ben-Akiva and Lerman (1985), Chapter 7.

When the hypotheses are non-nested meaning that one of the two models is not a special case of the other, likelihood ratio tests cannot be used and we relied on goodness of fit measures to compare competing models. Everything else being equal, a model with a higher maximum value of the log likelihood function is considered to be better. However, since the log likelihood will always increase or at least stay the same when new variables are added to the utility function, the maximum value of the likelihood function cannot be used to compare models with different number of parameters. Thus, goodness of fit measures need to be adjusted for the number of parameters used in the model. Akaike (1973, 1974) proposed the Akaike Information Criterion (AIC) that penalizes the maximum log-likelihood of the model by the number of estimated parameters:

$$AIC = Log\ L\ (\beta^*) - k \qquad (6.3)$$

Where   Log L (β^*) is the maximum log likelihood of the model

k is the number of estimated parameters of the model

The Akaike criterion is similar to the adjusted value of the likelihood ratio index also commonly called Rho-bar-squared and designated in this dissertation as Rho-bar-squared AIC:

$$Rho\text{-}bar\text{-}squared\ AIC = 1 - \frac{Log\ L(\beta^*) - k}{Log\ L(0)} \qquad (6.4)$$

Where   Log L (0) is the log likelihood of the model with no parameters

122

However, Gourerioux and Monfort (1995) show that the foundations of the AIC are not fully satisfactory and they consider the AIC to be inconsistent in the following sense. They take the example of two nested normal linear models M_1 and M_2 and show that the probability of preferring M_1 over M_2 using the AIC does not converge to 1 when M_1 is the true model.

Various modifications of the AIC have been proposed. Using Bayesian arguments, Schwarz (1978) proposed the Bayesian Information Criterion (BIC) that penalizes the log likelihood of the model as a function of both the number of estimated parameters and the number n of observations in the dataset.

$$\text{BIC} = Log\ L(\beta^*) - \frac{k}{2}\log(n) \qquad (6.5)$$

Similarly, the BIC can be expressed as a likelihood ratio index and will be called Rho-bar-squared BIC:

$$\text{Rho-bar-squared BIC} = 1 - \frac{L(\beta^*) - \frac{k}{2}\log(n)}{L(0)} \qquad (6.6)$$

The penalty associated with the number of parameters is higher in the BIC than in the AIC as soon as the number of observations is greater than 8. Unlike the AIC, the BIC is consistent.

These goodness of fit measures were used to provide guidance on the number of latent classes as models with the same specification other than the number of classes cannot be considered as nested. As mentioned in Walker and Li (2006), the BIC is often favored to determine the number of latent classes. Given that 2015 bookings are included in the sample, the BIC will penalize each parameter of the model by about 3.8 units of log likelihood compared to just one unit for the

AIC. As a result, the BIC will tend to favor more parsimonious model specifications and suggest that a lower number of classes should be used. However, these statistics should be used in conjunction with an examination of the estimation results to determine the number of latent classes that provides the most satisfying behavioral interpretation, for instance, in terms of how the classes can be interpreted and compared to a priori assumptions on the segmentation of the market.

6.3 Estimation Results

The estimation results of a two-class latent class model of airline passenger choice are presented in Table 6-2 below. We first examine the estimation results of the class membership model. We then focus on the estimation results obtained for the class-specific choice models starting with the parameters relative to the choice of a fare product. We finally use the parameter estimates of the continuous function of time to calculate the fitted value of the function and discuss the time-of-day preferences of outbound travelers in these short-haul markets.

6.3.1 Class Membership Model

In addition to the intercept, three dummy variables are used as explanatory variables of the class membership model. They indicate respectively whether the booking was made through an offline travel agent, the traveler is a member of the airline's loyalty program and the trip occurred within a week meaning that the trip started on or after Monday and ended on or before Friday of the same week.

	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	0.0968
Rho-bar-squared BIC	0.0697

Table 6-2:   Two-Class Latent Class Model of Airline Passenger Choice

125

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 that a booking belongs to each latent class is calculated as the logit probability associated with the parameter estimates of the class membership model.

For one of the two latent classes also called here Latent Class 1, the parameter estimates are positive for all three explanatory variables of the class membership model while the parameter estimate of the intercept is negative. As the sum of the coefficient estimates for these three parameters largely exceeds the absolute value of the intercept, bookings that exhibit all three characteristics used in the class membership model are more likely to belong to Class 1 than to Class 2. On the opposite, a booking that does not share any of these three characteristics is more likely to belong to Class 2. Figure 6-1 below shows the probability of belonging to Class 1 for the eight different covariate patterns.

	NONE	FFP	MON to FRI	Offline TA	FFP + MON to FRI	FFP + Offline TA	Offline TA + MON to FRI	ALL
FFP	0	1	0	0	1	1	0	1
MON to FRI	0	0	1	0	1	0	1	1
Offline TA	0	0	0	1	0	1	1	1

Bar values: NONE 0.0%, FFP 0.1%, MON to FRI 7.5%, Offline TA 12.8%, FFP + MON to FRI 37.5%, FFP + Offline TA 52.1%, Offline TA + MON to FRI 99.2%, ALL 99.9%
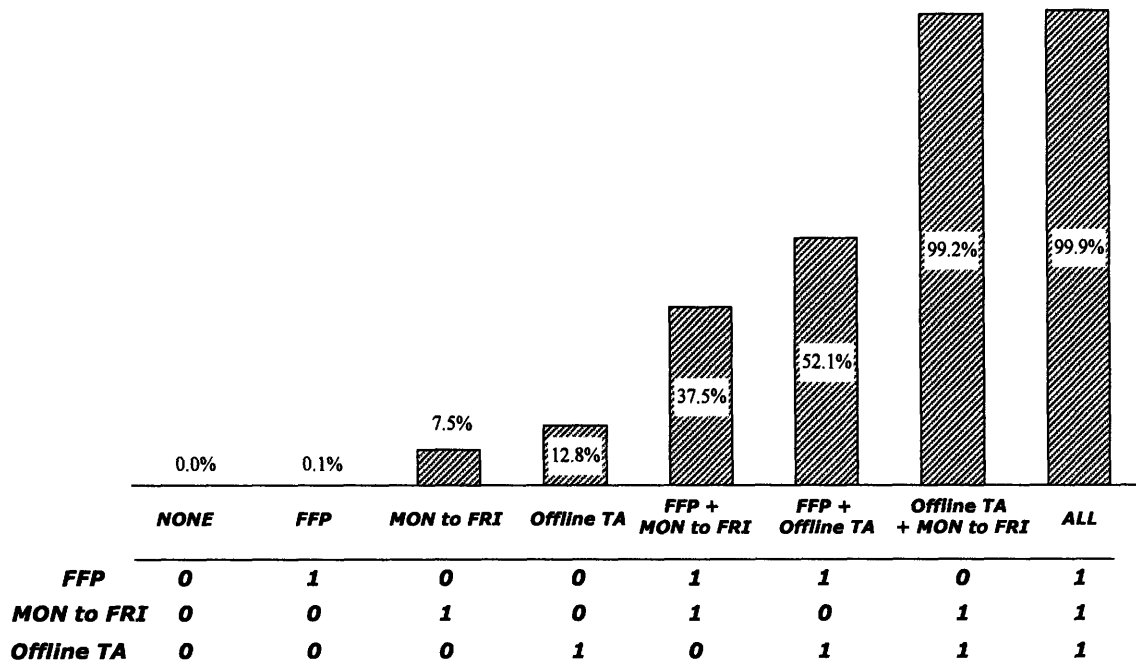
Figure 6-1:   Latent Class 1 Membership Probabilities

Covariate patterns can be grouped into three categories. On one hand, bookings made both for travel within a week and through a traditional offline travel agent (ALL and Offline TA + MON to FRI covariate patterns) are estimated to almost always belong to Class 1. On the other hand, bookings from the two covariate patterns that do not involve any of these two variables (NONE and FFP) are almost always classified as belonging to Class 2. Finally, bookings that belong to the other four covariate patterns are estimated to be split to various degrees between the two latent classes of the market.

Class 1 can then 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 Class 2 appears to be primarily oriented toward leisure travelers. The quasi-deterministic probabilities found here 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 fare products.

Class 1 groups slightly over 75% of the observations as a large number of bookings belong to the two covariate patterns that combine an offline travel agent distribution channel and week travel. This is consistent with a priori expectations as these markets are strongly oriented toward business travel as discussed in the exploratory analysis of the booking data in Chapter 5. As shown in Figure 6-2 below, these two covariate patterns represents almost 70% of all bookings included in the dataset. However, if corporate contract bookings for which the model reduces to the sole choice of an itinerary are excluded from the analysis, the distribution of bookings by covariate patterns is less concentrated and the proportion of bookings belonging to these two covariate patterns falls to

slightly below 50%. Furthermore, if we analyze 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 business fare from the Week category. About 35% of the bookings of a product from the Week category are then estimated to belong to Class 2, compared to only 10% for bookings made by travelers that purchased the Flex fare product.
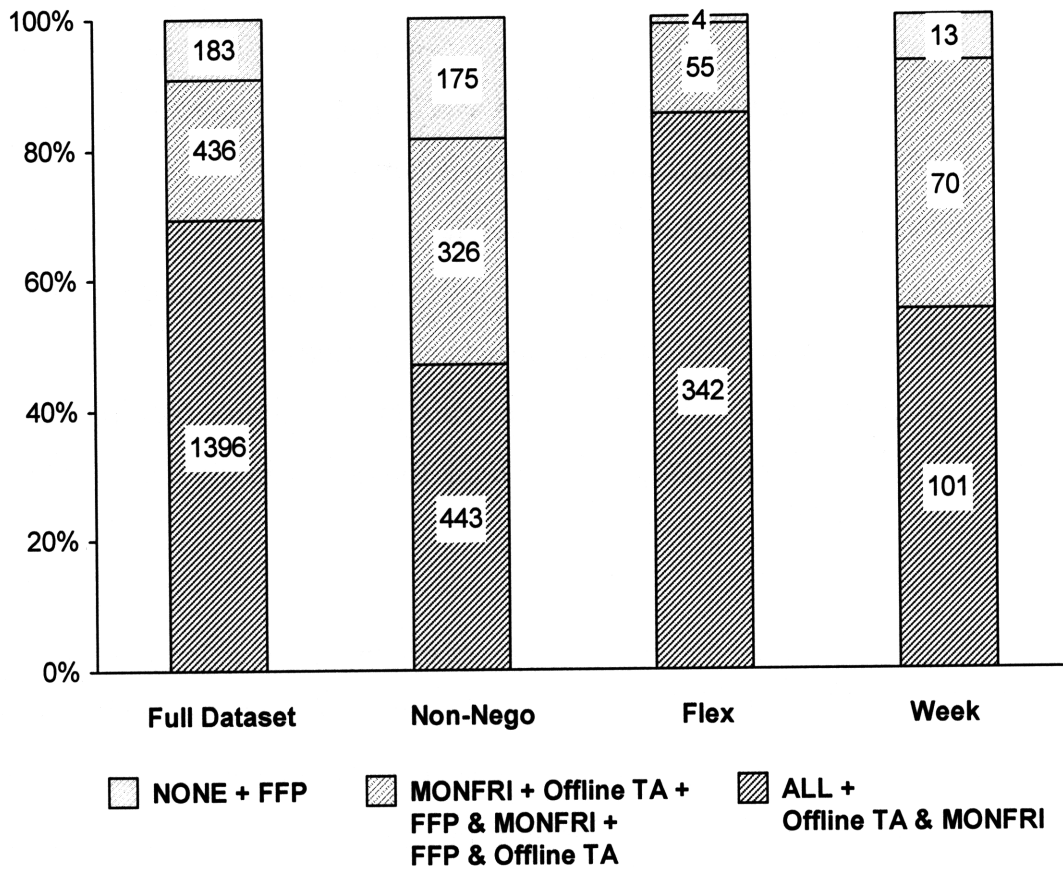
Figure 6-2:   Distribution of the Bookings by Covariate Patterns

The quasi-deterministic class membership probabilities observed for half of the covariate patterns 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 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 is also positive but has a much lower estimated value. This is line with our expectations as business travelers tend to travel mostly within a week and use much more frequently the services of a traditional travel agent to book their trip 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 benefit more from the rewards of an airline frequent flyer program. However, this effect is not expected to be that strong as membership in airline loyalty programs is complementary and some leisure travelers may also be able to accumulate enough to obtain a reward such as a complimentary award ticket. Finally, as mentioned earlier, the quasi-corner solution obtained here for half of the covariate patterns reflect our a priori expectations on the strong segmentation of airline markets.

In addition to the three explanatory variables included in the class membership model, three additional parameters were tested but excluded based on a likelihood ratio test as shown in Table 6-3 below. Similarly, these parameters would also be dismissed using goodness of fit measures such as the Rho-bar-squared AIC or the more restrictive Rho-bar-squared BIC.

Class Membership Model	3 Variables	6 Variables
Max Log-likelihood	-3142.41	-3140.90
Number of parameters	34	37
Rho-bar-squared AIC	0.0968	0.0964
Rho-bar-squared BIC	0.0697	0.0669
Likelihood Ratio Test = $3.02 < \chi^2_{(0.95,3)} = 7.82$		

Table 6-3:   Number of Explanatory Variables of the Class Membership Model

These additional explanatory variables include a male dummy variable which means that gender does not appear to significantly influence class membership, even in these markets that are strongly oriented toward business travel. This may reflect the growing presence of women in the corporate world. The use of two dummies to differentiate between standard and elite members of the airline's loyalty program also did not add sufficient explanatory power over a single variable grouping together all participants to the airline loyalty program. Finally, a day trip dummy was also dismissed. As discussed below, the behavioral characteristics of day trippers were captured using a specific set of explanatory variables in the class-specific choice models. However, the addition of a day trip dummy was not found to capture additional heterogeneity of behavior across latent classes.

6.3.2   Fare and Fare Product Characteristics

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 of the market. Three dummy variables were included in the utility function to represent the lack of flexibility to change travel plans of some fare products in interaction with their required level of advance purchase. As expected, they are also all 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 advance purchase requirements. 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.

All but one of these parameters are significant at the 90% confidence level. The parameter estimate for fare products that do not require any advance purchase has a very slightly negative value and is not significantly different from zero for the leisure-oriented Latent Class 2. This is not unexpected as most leisure-style travelers are likely to have fairly little uncertainty over their travel plans when making a booking so close to the departure date. Consequently, they are expected to place little value on the flexibility to change travel plans.

In order to compare how the disutility value of the lack of flexibility to change travel plans varies across latent classes, the ratio of these coefficients to the fare coefficient is calculated and reported in Figure 6-3 below. It represents the estimated monetary value of the lack of flexibility to change travel plans for each of the three levels of required advance purchase. It can also be interpreted as how much airline travelers are estimated to be willing to pay for the opportunity to change their travel plans without penalty depending on how early they were required to purchase their ticket. The values obtained are largely higher for Class 1 than for 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 price-sensitive 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 or modify their inbound or outbound itinerary.

LATENT CLASS 1   **LATENT CLASS 2**

196.84 €

178.62 €

106.74 €

135.78 €

55.79 €

0.45 €

21 DAYS   7 & 14 DAYS   0 DAYS   21 DAYS   7 & 14 DAYS   0 DAYS
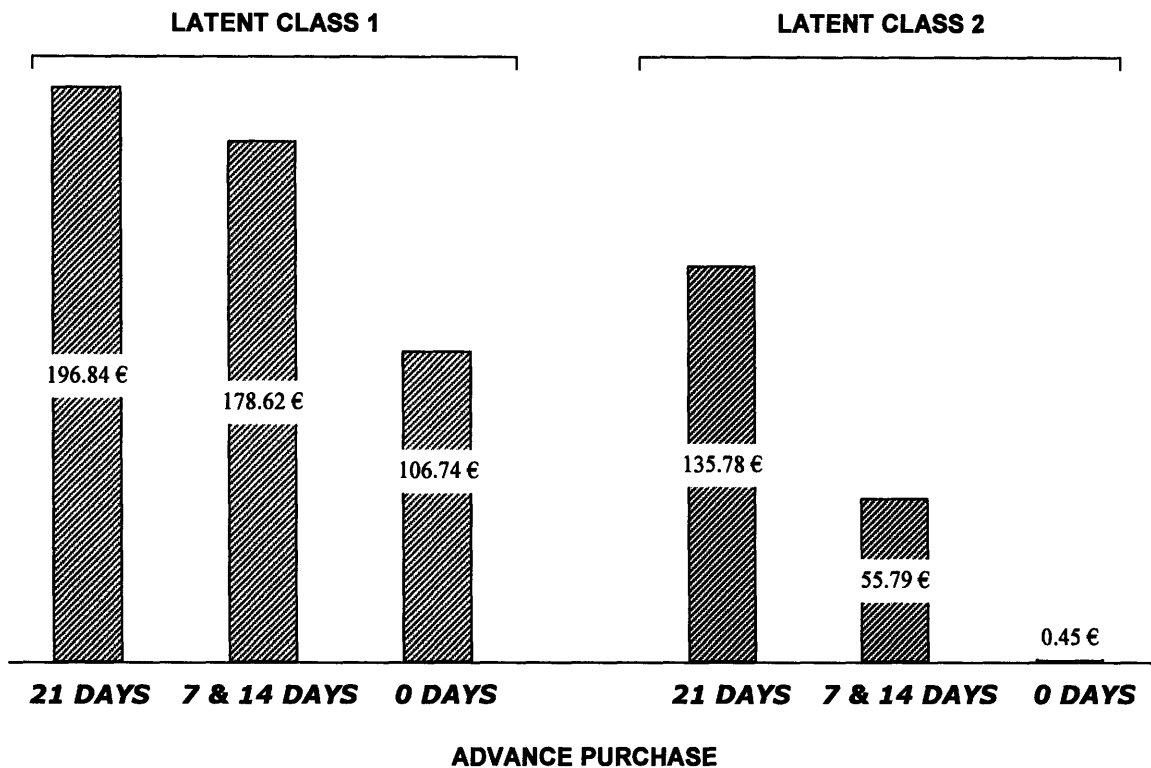
ADVANCE PURCHASE

Figure 6-3   Willingness to Pay for the Flexibility to Change Travel Plans without Penalty

The estimated value of this disutility varies between almost nothing for Class 2 travelers for fare products that require no advance purchase to about 200 euros for Class 1 travelers that purchased a product with at least a 21-day advance purchase requirement. Based on these estimates, the average value of this disutility was calculated using sample enumeration techniques over all bookings from the Saturday Night Stay and Week categories. The average estimated value of the lack of flexibility to change travel plans was found to be around 108 euros.

This estimate is compared with the fare difference charged by Air Canada in markets of similar length of haul between the fully flexible Latitude and the Tango Plus fare product that is non-refundable and carries a $40 CAD change fee. The fare difference between the Latitude and the Tango Plus fare product is advertised to start as low as $100 CAD on the airline's website. Based on a sequence of booking queries on the Air Canada website (www.aircanada.com), the fare difference between the Latitude and the Tango Plus product was found to vary between $100 CAD and $180 CAD (about 68 to 122 euros[5]) in select Canadian domestic markets of less than 500 miles depending mostly on the level of competition in the marketplace. For instance, the difference in fare was found to be only $100 CAD in the heavily contested Montreal-Toronto market (313 miles) but is set to $160 CAD in the Montreal-Mont Joli market (330 miles) that is served solely by Air Canada regional affiliate, Air Canada Jazz. The fare difference between the Latitude and Tango Plus product was not found to vary with the level of advance planning of the trip in these short-haul markets.

These values seem slightly lower compared to the estimated value of the flexibility to change travel plans obtained in the model. On one hand, the Latitude fare product is a bundle and while the flexibility to change travel plans is its major characteristics, it also includes a few extra features such as a complimentary snack and some limited extra mileage accrual although these elements can probably be valued around $10 to $15 CAD. On the other hand, the non-flexible Tango Plus product carries a $40 CAD change fee (about 27 euros) that is much lower than the 60 euros change fee or the full loss of the ticket value that apply to non-flexible fare products in the markets considered in this research.

---

[5] At an exchange rate of 1.4667 CAD to one Euro obtained on November 24, 2007 on the Yahoo website, http://fr.finance.yahoo.com/convertisseur.

No information is publicly available on the methodology used by Air Canada to set the prices of its different fare products in general and the premium of the Latitude over the Tango fare product in particular. The approach used by Air Canada is likely to be extremely different from the passenger choice models developed in this research. However, given the large difference in change fees, the estimated value of the willingness to pay for the flexibility to change travel plans without penalty provided by the model seems to be fairly consistent with the range of values currently observed on the Air Canada website for markets of similar length of haul.

As mentioned earlier, several alternatives were considered to model the disutility associated with the lack of flexibility to change travel plans. In order to test whether the different fare products can be grouped based on similar fare rules, a likelihood ratio test is used as described in Section 6-2. Table 6-4 below shows the value of the log likelihood for the unrestricted model in which a dummy variable is used for each fare product and for the restricted model in which the fare products are consolidated into the four categories as shown in Table 6-1.

Fare Product Categories	10 Categories	4 Categories
Max Log-likelihood	-3134.68	-3142.41
Number of parameters	46	34
Rho-bar-squared AIC	0.0956	0.0968
Rho-bar-squared BIC	0.0589	0.0697
Likelihood Ratio Test = $15.45 < \chi^2_{(0.95,12)} = 21.03$		

Table 6-4:   Number of Fare Product Categories

Given that these restrictions are used for each latent class of the model, the restricted model has a total of 12 fewer parameters than the unrestricted model. Based on a likely ratio test, we cannot reject the hypothesis that the coefficients are of equal value within the four categories used in the restricted model. The

same conclusion is reached from model selection criteria such as the Rho-bar-squared AIC and Rho-bar-squared BIC.

Other scenarios were tested such as classifying the products into six categories by separating totally non-flexible products that can be neither changed, nor cancelled and semi-flexible products that require paying a fee to modify or cancel the booking. However, this scheme was also rejected based on a likelihood ratio test in favor of the scenario in which the fare products are grouped into four categories with no distinction between these two types of fare rules. This suggests that airline travelers do not place a significantly different value on the disutility associated with totally non-flexible and semi-flexible products. This matches our expectations as most semi-flexible products belong to the Saturday Night Stay category and are likely to be purchased by price-sensitive leisure travelers. These travelers may have a fairly low level of uncertainty about their travel plans and more importantly may be very reluctant toward any additional charge such as a change fee.

Furthermore, the consolidation of the fare products into a smaller set of categories by reducing the number of levels of required advance purchase was also rejected based on likelihood ratio tests.

6.3.3 Time-of-Day Preferences

We now consider the parameters of the continuous function of time used to represent the choice of an itinerary. We first use the parameter estimates to calculate the fitted value of the continuous function of time and discuss the time-of-day preferences of the travelers based on the resulting profile of the function. We then compare this approach to previous models based on the specification of time-period dummies.

6.3.3.1 Continuous Function of Time

As discussed in Section 6.1, the flight departure time is used as the sole attribute of the alternatives to model the choice of an airline itinerary and a continuous function of time is specified to represent the time-of-day preferences of airline travelers. A separate function is used for overnight and day trip bookings as day trippers are expected to have a stronger preference for early morning departures in order to have as much time as possible to conduct their activities during their short stay at their destination.

For each type of booking, the number of parameters included in the trigonometric formulation was determined by trial and error based on both an empirical analysis of the resulting profile of the utility function and the statistical significance of the parameters. A trigonometric formulation with eight parameters was preferred for overnight bookings based on the profile of the function. For day trip bookings, only two parameters could be identified as the number of observations is reduced to a relatively few morning flight departures. A trial and error method with half-hour increments was also 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 estimated values of the parameters are reported in Table 6-2 for each latent class of the model.

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

$$U(h) = 0.33 * \sin(\frac{2\pi h}{16}) + 0.55 * \sin(\frac{4\pi h}{16}) + 0.09 * \sin(\frac{6\pi h}{16}) + 0.24 * \sin(\frac{8\pi h}{16})$$

$$- 0.15 * \cos(\frac{2\pi h}{16}) + 0.68 * \cos(\frac{4\pi h}{16}) - 0.33 * \cos(\frac{6\pi h}{16}) - 0.03 * \cos(\frac{8\pi h}{16}) \qquad (6.7)$$

The willingness to pay for a specific departure time is then obtained by dividing this fitted value by the estimate of the fare coefficient.

As discussed in Chapter 4, 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 Class 1 and 7 a.m. for Class 2 based on half-hour increments. Figure 6-4 below shows the estimated willingness to pay curves for overnight bookings.
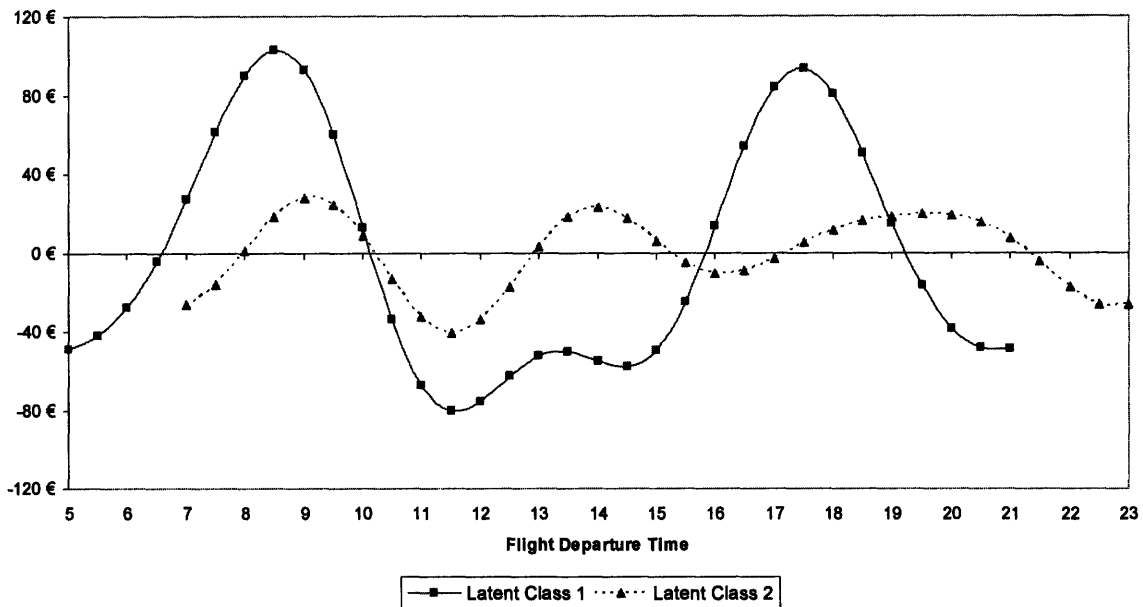
Figure 6-4   Willingness to Pay for a Flight Departure Time for Overnight Bookings

As expected, business-oriented (Class 1) outbound travelers tend to prefer either a morning departure to be able 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.

The time-of-day preferences of Class 2 travelers have some similarities and differences with those of 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 Class 1 and 2 travelers is the magnitude of their respective willingness to pay curves. As shown in Figure 6-3, 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 core business 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 class between day trip and overnight bookings. As day trip bookings primarily belong to Class 1, Figure 6-5 below shows the willingness to pay curves for overnight and day trip bookings for Class 1 travelers.
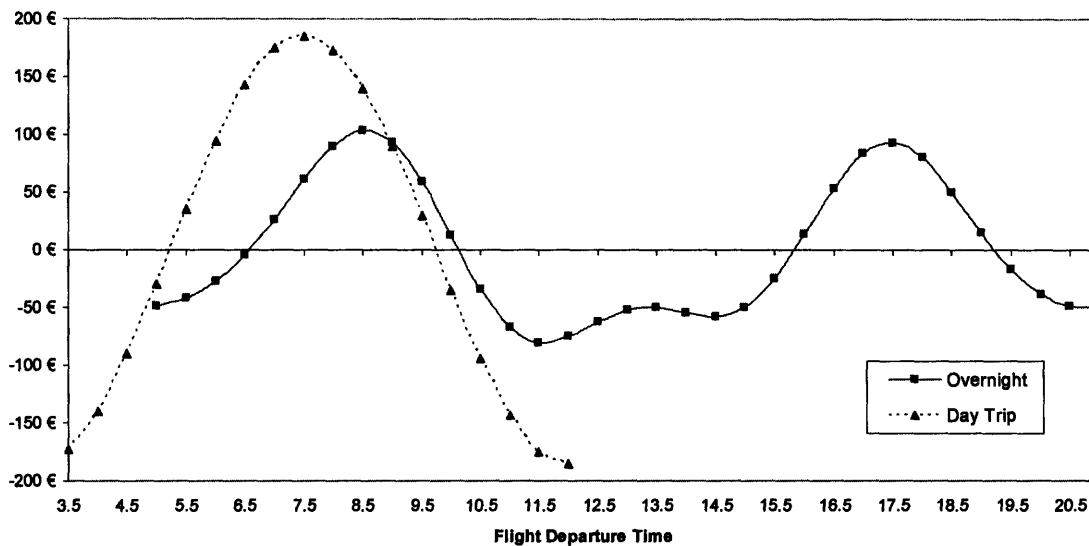
Figure 6-5:   Willingness to Pay for a Flight Departure Time – Overnight and Day Trip Bookings (Class 1)

For day trip bookings, the duration of the cycle is reduced to 9 hours as day trippers are observed to select only morning flight departures. 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.

The addition of a function of time specific to day trip bookings improved substantially the fit of the model as shown in Table 6-5 below. The maximum log-likelihood of the model improved by more than 30 units although the number of parameters increased by five including the duration of the cycle for day trip bookings. The use 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 Rho-bar-squared AIC and the more restrictive Rho-bar-squared BIC.

Day Trip Booking Variables	All	Time-of-Day (TOD)	None
Max Log-likelihood	-3136.68	-3142.41	-3174.44
Number of parameters	42	34	29
Rho-bar-squared AIC	0.0961	0.0968	0.0891
Rho-bar-squared BIC	0.0627	0.0697	0.0660
Likelihood Ratio Test	11.46 < $\chi2(0.95,8) = 15.51$		N/A

Table 6-5:   Day Trip Bookings Explanatory Variables

However, no statistically significant difference was found between day trip and overnight bookings regarding other parameters of the class-specific choice model based on a likelihood ratio test or model selection criteria. As a result, parameters specific to day trippers were not added to the rest of the utility function.

6.3.3.2 Time-Period Dummies

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. Since no flight departures are observed in the nighttime, it is not possible to identify the parameters for time-periods for that part of the day. As a result, nighttime (11 p.m. to 5 a.m.) is excluded. We make the implicit assumption that the utility value of nighttime flight departures tends to minus infinity, as in models with a continuous function of time for flight departures outside of the estimated cycle.

	LATENT CLASS 1 74.4%			LATENT CLASS 2 25.6%		
	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
Class Membership						
INTERCEPT	-4.60	3.66	-1.3	5.51	7.87	0.7
FFP MEMBER	1.00	0.50	2.0	-4.20	8.12	-0.5
MON to FRI	3.34	3.67	0.9	-1.25	1.24	-1.0
OFFLINE TA	3.64	3.59	1.0	-4.00	8.29	-0.5
Choice Model						
FARE	-0.0131	0.0052	-2.5	-0.0272	0.0074	-3.7
NON-FLEX & 21AP	-2.58	1.06	-2.4	-3.73	1.20	-3.1
NON-FLEX & 7-14AP	-2.28	0.68	-3.3	-1.62	0.79	-2.0
NON-FLEX & 0AP	-1.36	0.26	-5.3	-0.03	0.47	-0.1
OVERNIGHT BOOKINGS						
EARLY MORNING 5-8a	0.67	0.11	5.9	-0.50	0.26	-1.9
MORNING 8a-11a	0.00	0.00	0.0	0.00	0.00	0.0
MIDDAY 11a-2p	-0.60	0.18	-3.4	-0.01	0.18	-0.1
AFTERNOON 2-5p	0.11	0.12	0.9	-0.17	0.18	-1.0
LATE AFTERNOON 5-8p	0.51	0.12	4.2	0.45	0.16	2.7
EVENING 8-11p	-0.68	0.20	-3.4	0.32	0.19	1.7
DAY TRIP BOOKINGS						
EARLY MORNING 5-8a	1.90	0.14	13.6	0.33	0.55	0.6
MORNING 8a-11a	0.00	0.00	0.00	0.00	0.00	0.00

Summary Statistics

Log L (0)	-3516.81
Log L	-3200.99
Number of parameters	24
Rho-bar-squared AIC	0.0830
Rho-bar-squared BIC	0.0638

Table 6-6:   Two-Class Latent Class Model with Time-Period Dummies

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. For overnight bookings, five dummies are included in the utility function while a single dummy variable is used for day trip bookings as day trippers were never observed to select afternoon flights. Estimation results of the model with time-period dummies are shown in Table 6-6 above.

First, the use of time-period dummies has little impact on the estimated value of the rest of the parameters. Parameter estimates of the fare and fare product characteristics are very similar to those obtained in the model with a continuous function of time. For models with time-period dummies, the willingness to pay for a flight departure can be calculated directly by dividing for each period the parameter estimate of the dummy by the fare coefficient. As the morning period is used as a base, the willingness to pay values should be interpreted relative to a morning flight departure.
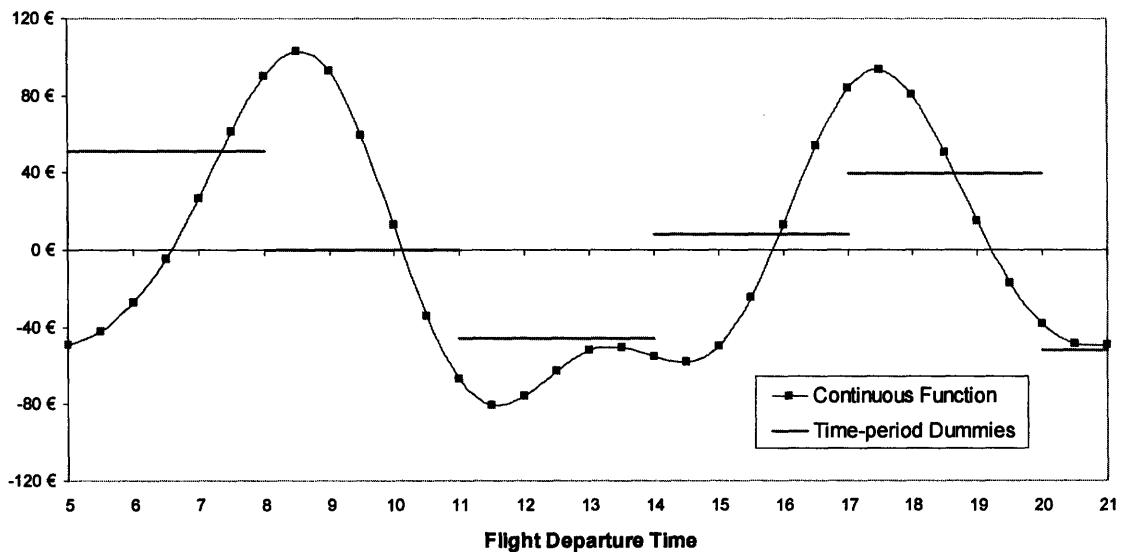
Figure 6-6: Willingness to Pay for a Flight Departure Time – Continuous Function of Time and Time-period Dummies (Class 1 Overnight Bookings)

Figure 6-6 above compares the willingness to pay curves of both model specifications for overnight Class 1 bookings. The resulting time-of-day preferences are similar for models with a continuous function of time and time-period dummies. For instance, Class 1 travelers are found to prefer early morning and late afternoon flights in both model specifications. 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 of the passenger willingness to pay for a specific departure time for each period of the day.

This more precise measurement of the time-of-day preferences provided by a continuous function of time is reflected in the large improvement in the fit of the model over a specification with time-period dummies. The maximum value of the log-likelihood function is improved by almost 60 points although ten additional parameters are used including the estimated duration of the cycle for both day trip and overnight bookings. This large improvement in the fit of the model more than compensates for the increase in the number of parameters according to both the Rho-bar-squared AIC and Rho-bar-squared BIC.

The magnitude of the impact of a continuous function of time over the log likelihood of the model may be related to how flight departures are scheduled in these airline markets. Given that the observed departure times are concentrated over a couple of connecting banks, the number of time-period dummies that can be identified is relatively limited. This leaves a large potential for improvement in the fit of the model when a continuous function of time is introduced. It has the capability to provide a more precise estimate of the time-of-day preferences of airline travelers despite the relatively low level of variability in observed flight departure times. As flight departures are scheduled around connecting banks in

many airline markets, this approach is particularly useful to investigate the time-of-day preferences of airline travelers.

However, a continuous function of time may reproduce too closely the patterns observed in the dataset and have limited prediction capabilities. This may be especially true as the duration of the cycle is determined to maximize the log-likelihood of the model. In order to verify whether a continuous function of time provides a genuine improvement in the fit of the model, we tested whether this approach is subject to a risk of overfitting.

Overfitting occurs when a model has so many parameters that it fits the data perfectly and yet it performs very poorly to predict future data drawn from the same distribution. In order to detect overfitting, a validation study is used. The dataset is divided into two parts, a smaller test set that comprises generally 25 to 30% of the observations and a larger training set. The model is estimated over the training set and the prediction capabilities of the model are assessed on the test set using goodness of fit criteria such as the maximum log likelihood of the model. For a more complete description of the validation methods available to detect and prevent overfitting, the reader is referred to Moore (2001).

In order to test for overfitting, the dataset was divided into two subsets based on the second letter of the passenger's first name. If the letter was either an A, B, C or a D, the booking was assigned to the test set. Otherwise, the booking was included in the training set. The test set included 522 observations, about 26% of the total number of bookings in the dataset. The model was estimated on the training set and the coefficient estimates were applied to calculate the choice probabilities for each booking and the log likelihood of the test set. Table 6-7 below shows the values of the log-likelihood obtained over the training and test sets for models with time-period dummies and a continuous function of time.

144

	Time-period Dummies	Continuous Function
Training Set Log L	-2389.27	-2342.24
Test Set Log L	-816.92	-808.71
Training + Test Set Log L	-3206.19	-3150.95
Number of parameters	24	34
Rho-bar-squared AIC	0.0815	0.0944
Rho-bar-squared BIC	0.0624	0.0673

Table 6-7:   Log-Likelihood of the Model Calculated over the Test Set

As expected, the use of a continuous function of time leads to an improvement of the log-likelihood of the model over the training set. More importantly, it also leads to an increase in the log-likelihood calculated over the test set compared to a model with time-period dummies. Thus, the use of a continuous function of time to represent the time-of-day preferences of airline travelers does not appear to be subject to a risk of overfitting.

As a result, a model specification with a continuous function of time is preferred over earlier models with time-period dummies. It has the ability to provide a more precise estimate of the time-of-day preferences of airline travelers despite the low variability of flight departure times observed in many airline markets. In addition, it provides a more scientific approach to determine the duration and start time of the cycle. Unlike models with time-period dummies in which the length and breakpoints of each period are selected arbitrarily, the duration of the cycle is estimated by maximizing the log-likelihood of the model and the start time of the cycle is selected at a local minimum of the function.

Estimating the duration of the cycle appears as the major driver of the successful specification of a continuous function of time in short-haul markets. In addition to a substantial decrease in the fit of the model, a continuous function of time with a 24-hour cycle did not lead to a profile of the time-of-day preferences that matched our expectations, especially for day trip bookings. Since the schedule constraints of day trippers are such that flight departure times are concentrated

on a short period of the day in the morning, the specification of a function of time with a 24-hour cycle leads to unrealistic high values of the willingness to pay for morning flight departures while unrealistic low values are obtained for the rest of the day. As a result, the use of a continuous function of time with an estimated duration of the cycle provides a flexible approach to model the time-of-day preferences of specific groups of travelers or specific types of airline markets.

6.4   Segmentation of Airline Demand

As mentioned earlier, demand in airline markets is traditionally segmented by trip purpose. Since trip purpose is not recorded in airline bookings, it is replaced by several factors available in airline bookings that are presumably more or less strongly correlated with trip purpose. In order to include several factors without dividing the dataset into a growing number of small sub-segments, bookings are assumed to be distributed over a limited number of latent classes. A probabilistic class membership model is used to estimate how each booking is likely to belong to the different latent classes based on its observed characteristics. In this section, we will first focus on how we determined the number of latent classes used in the model. We will then discuss the benefits of the latent class model of airline passenger choice by comparing its estimation results to previous model specifications based on a deterministic segmentation of airline demand. We will finally conclude by examining whether additional heterogeneity of behavior can be captured by using random coefficients within the class-specific choice models.

6.4.1   Determining the Number of Classes

As mentioned earlier, the number of classes is determined based both on model selection criteria and on whether the parameter estimates of the model define

latent classes that can either be easily interpreted or tend to match a priori assumptions on the segmentation of the demand. Table 6-8 below shows the maximum value of the log-likelihood function for models with a number of latent classes varying from one to three. All selection criteria indicate that a model with several classes is preferred over a model with a single class. In addition, the Rho-bar-squared BIC that is often used in latent class choice models to determine the number of classes suggests that the two-class model is superior while the Rho-bar-squared AIC favors a model with three classes.

Number of Classes	1 Class	Two-Class	Three-Class
Max Log-likelihood	-3236.33	-3142.41	-3080.61
Number of Parameters	16	34	52
Rho-bar-squared AIC	0.0752	0.0968	0.1092
Rho-bar-squared BIC	0.0624	0.0697	0.0678

Table 6-8:   Goodness of Fit Measures by Number of Latent Classes

The estimation results of the class membership model for the three-class model shows that Class 1 gets split into two sub-classes (also called here Class 1A and 1B) while the size and parameter estimates of Class 2 remains almost unchanged compared to the two-class model. As shown in Table 6-9 below, all the parameters of the class membership model have the same sign for both sub-classes. No major difference in the distribution of bookings by fare product category is observed between the two sub-classes as bookings for the fully flexible and non-flexible products from the Week category are split among the two sub-classes in proportion similar to their respective size. In addition, around one-third of bookings of products from the Week category remain classified as belonging to Class 2 as in the two-class model meaning that price-conscious business travelers are not recaptured into one of the other two sub-segments.

SEGMENT SIZE	Latent Class 1A 57.3%			3-class Model Latent Class 1B 20.8%			Latent Class 2 21.9%		
	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
INTERCEPT	-2.40	2.91	-0.8	-3.78	5.57	-0.7	6.18	3.24	1.9
OFFLINE TA	1.47	2.89	0.5	3.63	5.55	0.7	-5.10	3.22	-1.6
FFP MEMBER	0.61	0.25	2.4	0.58	0.23	2.5	-1.19	0.42	-2.9
MONFRI	2.71	0.88	3.1	0.72	0.87	0.8	-3.43	1.71	-2.0
FLEX BOOKINGS	68.9%			21.4%			9.7%		
WEEK BOOKINGS	50.2%			15.8%			34.0%		

Table 6-9:   Estimation Results of the Class Membership Model for the Three-Class Model

Thus, the split of business-style travelers into two sub-classes cannot be easily interpreted and does not match a priori assumptions such as the expected behavioral differences between business travelers that book fully flexible products and price-conscious business travelers that prefer a discounted non-flexible product. A model specification with two latent classes is then preferred.

6.4.2 Deterministic and Latent Segmentation of the Demand

In order to assess the benefits of using latent classes, the estimation results of the two-class model are compared to a deterministic benchmark. In this benchmark model, the bookings are divided deterministically across two segments defined by one of the three variables of 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.

Estimation results of the deterministic model are presented in Table 6-10 below. Similar behavioral patterns are observed for the latent class model and the deterministic benchmark. For instance, the disutility associated with the lack of flexibility to change travel plans increases with advance purchase requirements in both models. In addition, week travelers are willing to pay more for the flexibility to modify their travel plans without penalty than non-week travelers.

148

	MON to FRI 74.8%			OTHER 25.2%		
	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
FARE	-0.0145	0.0045	-3.2	-0.0177	0.0063	-2.8
NON-FLEX & 21AP	-2.48	0.89	-2.8	-2.94	1.09	-2.7
NON-FLEX & 7-14AP	-1.98	0.56	-3.5	-1.46	0.72	-2.0
NON-FLEX & 0AP	-1.19	0.21	-5.6	-0.19	0.47	-0.4
DUR-OV = 16 HOURS						
SIN2PI-OV	0.26	0.09	2.8	0.26	0.13	2.0
SIN4PI-OV	0.49	0.11	4.4	0.32	0.18	1.7
SIN6PI-OV	0.08	0.09	0.9	-0.46	0.13	-3.5
SIN8PI-OV	0.17	0.08	2.2	0.39	0.11	3.7
COS2PI-OV	-0.14	0.05	-2.8	0.10	0.07	1.4
COS4PI-OV	0.67	0.09	7.1	0.17	0.12	1.4
COS6PI-OV	-0.33	0.07	-4.4	-0.35	0.11	-3.1
COS8PI-OV	-0.08	0.13	-0.6	-0.01	0.20	-0.1
DUR-DT = 9 HOURS						
SIN2PI-DT	-1.44	0.10	-13.7	-0.29	1.34	-0.2
COS2PI-DT	1.86	0.33	5.6	-0.97	8.71	-0.1

Choice Model

Summary Statistics

Log L (0)	-3516.81
Log L	-3172.75
Number of parameters	30
Rho-bar-squared AIC	0.0893
Rho-bar-squared BIC	0.0654

Table 6-10:   Deterministic Segmentation of Airline Bookings by Week and Non-Week Travel

The profile of the time-of-day preferences is also similar for both model specifications as shown in Figure 6-7 below. However, the magnitude of the parameter estimates tends to be greater for the latent class model. For instance, the magnitude of the morning and afternoon peaks is greater for the latent class model although the peaks are observed at similar times for both models. Similarly, Class 1 travelers are willing to pay more for the possibility to change their travel plans without penalty than week travelers in the deterministic model. A reverse trend is observed for the other segment of the market: Class 2 travelers are not willing to pay as much as a non-week travelers for the flexibility to change their travel plans at no additional cost.
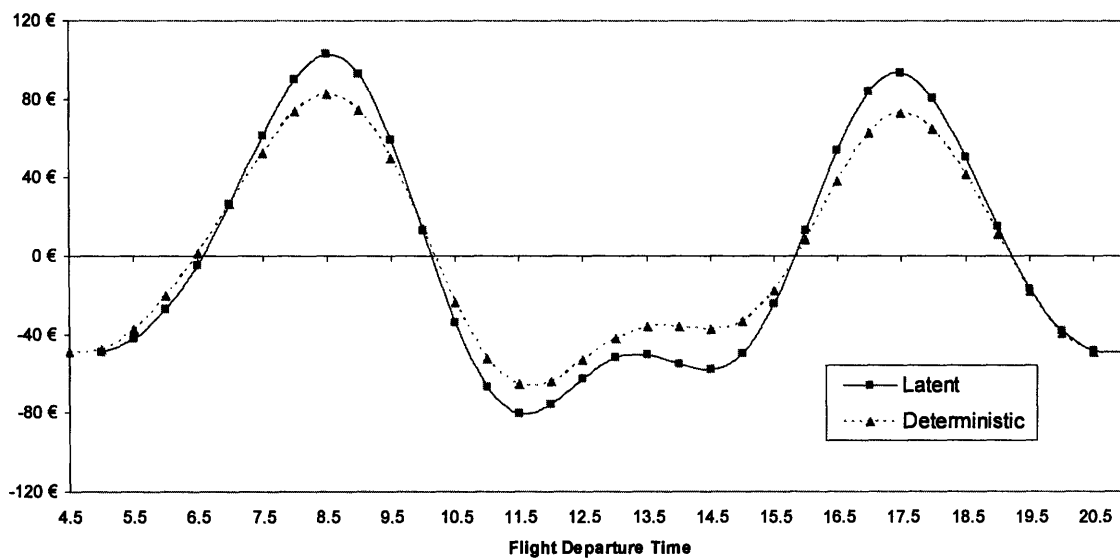
Figure 6-7:   Time-of-Day Preferences in the Latent Class and the Deterministic Benchmark Models

Thus, the latent class choice model appears to provide a more polarized segmentation between time-focused business travelers and a class of leisure-oriented and price-conscious business travelers. This is due to the impact of the two additional explanatory variables of the class membership model that have the same sign as the MON to FRI dummy and contribute to further segment the

150

bookings between the two latent classes of the market. This results into a large improvement in the fit of the model of almost 30 units of log-likelihood.

The deterministic model may be viewed as a special case of the latent class model in which the value of the parameter estimate of the MON to FRI dummy tends to infinity while the other parameters of the class membership model are equal to zero. However, standard statistical tests such as the likelihood ratio test are only valid to test whether the value of a parameter is equal to a finite value of interest. As a result, model selection criteria were used to compare the fit of the latent class choice model and the deterministic benchmark.

Although the latent class model includes four additional parameters used as explanatory variables of the class membership model, this increase in the number of parameters is more than compensated by the improvement in the maximum log-likelihood of the model according to both the Rho-bar-squared AIC and BIC. Similarly, the latent class model was also found to be preferred to models with a higher number of deterministic segments according to both of these criteria.

While model selection criteria were used to compare the latent class choice model to a deterministic benchmark, a likelihood ratio test can be used to determine whether the observations should be divided into deterministic market segments. As mentioned earlier, a model in which all the observations are pooled together is a special case of a model with two deterministic segments if the estimated value of each parameter is equal across segments. The statistic of the likelihood ratio test is calculated based on the value of the maximum log-likelihood for the model specifications with and without market segmentation as shown in Table 6-11 below. Based on the test statistic, we can reject the hypothesis that the value of the parameters is equal across segments and the model should account for heterogeneity of behavior across bookings.

151

Segmentation	None	Week Travel
Max Log-likelihood	-3236.33	-3172.65
Number of parameters	16	30
Rho-bar-squared AIC	0.0752	0.0893
Rho-bar-squared BIC	0.0624	0.0654
Likelihood Ratio Test = 127.36 > $\chi^2_{(0.95,14)}$ = 23.69		

Table 6-11:   Deterministic Segmentation of Airline Bookings

Based on model selection criteria, the latent class model of airline passenger choice is preferred to a deterministic benchmark in which bookings are segmented between week and non-week travelers. In addition, the latent class model was also found to be preferred to model specifications with multivariate deterministic segmentation schemes as the number of estimated parameters is increasing quickly when the bookings are divided in a large number of increasingly small market segments. As a result, latent classes provide a parsimonious way to segment the bookings into many underlying latent sub-segments using a wide range of factors available in booking records without increasing rapidly the number of estimated parameters. It also leads to a more distinct segmentation between a core of time-focused business travelers and a mixed class of price-conscious business and leisure travelers.

6.4.3   Random Coefficients

So far, two types of heterogeneity of behavior have been captured by the model. Heterogeneity of behavior across different segments of the market was captured through the latent class structure of the model. In addition, characteristics of the trip were included in the utility function of class-specific choice models to capture systematic taste variation within each latent class of the model. However, there may be additional heterogeneity of behavior remaining within each latent class. Random coefficients will be used to test whether random taste variation can be found in class-specific choice models.

As mentioned earlier, a maximum of three independent normally distributed random coefficients can be specified in Latent Gold Choice. In practice, only an even lower number of random coefficients could be estimated due to computational limitations. To limit computational time, models with a single random coefficient were successively estimated. For random coefficients, estimates are obtained for both the mean and standard deviation of the parameters. Table 6-12 below shows the estimated value of the standard deviation for the three coefficients used to represent the characteristics of the fare products for both latent classes of the model.

| | Latent Class 1 | | | Latent Class 2 | | |
	Est.	Std. Er.	t-stat	Est.	Std. Er.	t-stat
SIGMA_NON-FLEX & 21AP	2.82	2.34	1.2	0.35	1.79	0.2
SIGMA_NON-FLEX & 7-14AP	3.53	3.11	1.1	0.00	1.01	0.0
SIGMA_NON-FLEX & 0AP	0.00	2.03	0.0	1.19	0.91	1.3

Table 6-12:   Latent Class Choice Model with Random Coefficients

According to their t-statistic, none of the standard deviations of these random coefficients is significantly different from zero. Regarding the rest of the utility function relative to the time-of-day preferences of airline travelers, the value of the standard deviation of the coefficients of the continuous function of time cannot be interpreted individually. However, no significant random taste variation was found for the coefficients of a model with time-period dummies.

As a result, the heterogeneity of behavior in these markets appears to be captured primarily by the class membership model and interaction variables in the class-specific utility function. This is due to the characteristics of the fare structure used by the airline in these markets that largely leads to a bimodal distribution of the bookings based on whether or not the trip qualifies for products from the Saturday Night Stay or Weekend categories. However, in

153

airline markets with less restricted fare structures, random coefficients may prove useful to supplement the latent class structure of the model and capture a broader level of heterogeneity of behavior that cannot be easily related to observed characteristics of the bookings.

6.5   Summary

In this chapter, we have shown the benefits of a latent class model of airline passenger choice compared to previous deterministic segmentation schemes found in the literature. A range of factors extracted from the booking records could be used as explanatory variables of the class membership model and contributed to segment the demand without dividing the bookings into a large number of increasingly small sub-segments. It also led to a greater differentiation between a core of time-focused business travelers and a mixed class of price-conscious business and leisure travelers.

We have also shown the benefits of using a continuous function of time to model the time-of-day preferences of airline travelers over previous models based on time-period dummies. We proposed a generalized formulation of a continuous function of time in which the duration of the cycle is estimated providing a scientific approach to model the specific characteristics of time-of-day preferences in short-haul airline markets and sub-segments of the demand such as day trip bookings. Compared to time-period dummies, a continuous function of time provides a more precise measurement of the time-of-day preferences of the travelers. This is especially beneficial in many airline markets due to the low level of variability observed in flight departure times as the flight schedule is structured around a few connecting banks.

In the next chapter, we will discuss how the latent class model of airline passenger choice developed in this research can be used to support a range of airline planning decisions such as schedule planning, pricing and revenue management. In particular, we will focus on the impact of the latent class structure of the model on the predicted behavior of airline passengers and ultimately the planning decisions suggested by the model.

Chapter 7   Applications

In previous chapters, we have developed and estimated a model of the choice of an airline itinerary and fare product that incorporates the impact of pricing and revenue management on the passenger choice set and captures heterogeneity of behavior across several segments of the market. In this chapter, we discuss how this model can be used to support a range of airline planning decisions such as schedule planning, pricing and revenue management. In particular, we will show how the parameter estimates of the model can be applied to forecast the proportion of airline passengers willing to sell-up to a more expensive fare product. Sell-up estimates are required by new revenue management methods designed to maximize revenues in less restricted fare structures now common in many airline markets. We will discuss whether the specification of a latent class model of airline passenger choice leads to different estimates of sell-up behavior compared to a deterministic benchmark model.

7.1   Introduction

As mentioned earlier, discrete choice models may be used to support a wide range of short to medium term airline planning decisions. Actually, the impact of any planning decision that involves either a change in the attributes of the alternatives or the passenger choice set can be analyzed using discrete choice models. Changes in the attributes of the alternatives include, for instance, an increase or decrease in the price of a fare product or a change in flight departure time. Pricing decisions such as a new set of fare rules or a new revenue

management algorithm will impact the availability of the various fare products and the passenger choice set.

The use of discrete choice models to support airline planning decisions has been very limited up to now. Since 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. However, the relaxation of fare rules such as the Saturday night stay requirement due to the growing competition of low-cost airlines and the development of more complex product offerings based on a menu of optional services have brought the need for a better understanding of passenger choice behavior. Discrete choice models can provide valuable insight into the willingness to pay of airline passengers for different elements of airline service and the price elasticity of the demand for various segments of the market. The development of a model of the choice of an airline itinerary and fare product that incorporates the impact of fares rules and seat availability on the passenger choice set extends the scope of potential applications of discrete choice models in the airline industry beyond schedule planning to pricing and revenue management.

The rest of this chapter is organized as follows. Section 7.2 and 7.3 discuss how discrete choice models can be applied to support airline planning decisions in schedule planning and pricing. In Section 7.4, we show how the parameter estimates of the model can be used to forecast the sell-up behavior of airline passengers that is a key input to the new revenue management methods under development to optimize revenues under a less restricted fare structure. Finally, Section 7.5 concludes with a brief discussion of future research directions that could provide benefits for some of the applications described in this chapter.

7.2   Schedule Planning

Schedule planning has been the first area of application of consumer choice models in the airline industry. Schedule planning decisions include setting the flight schedule, increasing or decreasing frequency or even introducing new non-stop service in a market. Demand forecasts at the itinerary level are used by the airline planning department to evaluate these decisions and adjust the airline schedule. Itinerary-level demand forecasts are usually obtained in two steps. First, a demand forecast at the market level is produced based on historical data and other elements such as economic growth. Second, this total market demand is allocated to the itinerary level using an itinerary share model.

Many itinerary share models use a demand allocation methodology called quality of service index (QSI) (Coldren, 2005b). In QSI models, the share of an itinerary is proportional to an overall index summarizing the "quality" or attractiveness of each itinerary. Quality is defined as a weighted function of various service attributes such as elapsed time, number of stops or connections or airline preferences. The weights of the different variables of the model are obtained exogenously through statistical techniques or based on the analyst knowledge or intuition. The share of an itinerary is then equal to the quotient of the itinerary's QSI over the sum of the QSI for all itineraries in the market.

QSI models have two major drawbacks. First, the preferences weights are usually obtained exogenously and independently from each other. As a result, these models ignore the interactions that are likely to exist among itinerary characteristics such as between elapsed time and number of connections. Second, the preference weights are defined at the aggregate market level or even at a higher degree of aggregation and do not capture the specific characteristics and behavior of individual passengers.

In order to model preferences at the individual booking level, the Boeing Commercial Airplane Group (1993) developed a new itinerary share model called the Decision Window Model. In this model, each passenger gets assigned a decision window that represents his range of preferred travel times for. Two parameters define a decision window: its width and position. The width of a decision window depends both on the shortest elapsed time in the market and the characteristics of the traveler. For instance, decision windows are on average shorter for business travelers than for leisure travelers as passengers traveling on business tend to be more time-sensitive. The position of the decision window is defined such as to reproduce the typical time of the day distribution of demand in every market. All itineraries that fall within the boundaries of the window are equally schedule-attractive to the passenger and are preferred to other itineraries. Travelers choose among itineraries included in their time window based on a rating that includes elements such as number of stops and preferences for a particular airline.

Although Boeing's approach models the choice of an itinerary at the booking level, its decision rule is based on a rating system that has the same drawbacks as QSI models. As mentioned by Coldren (2005b), discrete choice models provide an attractive alternative to QSI models as they use statistical techniques to simultaneously estimate the parameters of the explanatory variables and the share of each itinerary based on disaggregated booking-level data. His research shows that the implementation of a logit-based itinerary share model by a major U.S. carrier led to a substantial improvement in the carrier's forecasting accuracy compared to previous QSI-based models. Coldren further explored the complexity of substitution patterns across itineraries using models of the GEV family. However, his model did not incorporate heterogeneity of behavior across bookings as no information was collected on the characteristics of the trip and the profile of traveler.

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 three 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 products is likely to vary across flight departures due to differences in passenger mix and the impact of revenue management. For instance, morning and late afternoon flights in short-haul markets appear particularly popular with business passengers while midday flights may attract more leisure demand. The revenue management system tends to reinforce the variations 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.

In addition, the model developed in this research 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 than the rest of the flying public. Combined with variations in passenger mix, our model has the potential to substantially improve the accuracy of itinerary share forecasts by capturing the different time-of-day preferences of the two segments of the market.

Finally, this model provides forecasts of expected demand at the itinerary-product level, a lower level of aggregation than previous models based solely on the choice of an itinerary. As a result, a revenue forecast can be obtained for each itinerary based on a weighted average of revenue by itinerary-product forecasts rather than by multiplying the total itinerary demand by the average fare in the market. Although the accuracy of itinerary-product forecasts is expected to be

lower than forecasts at the itinerary level especially if the size of the sample is fairly limited, this more disaggregate approach may improve the accuracy of revenue forecasts. While a model of the choice of an itinerary and fare product is expected to provide an enhanced tool to support schedule planning decisions, it also extends the range of potential applications to additional airline planning decisions such as pricing as discussed in the next section.

7.3 Pricing

Prior to the economic deregulation of the airline industry starting in the U.S. in 1978 and spreading later throughout many parts of the world, airfares were regulated by the government and airlines competed primarily on service such as flight schedule and in-flight amenities. In each market, a single standard fare was set by the regulatory authority and some limited discounts were allowed for specific categories of travelers such as children or senior citizens. Freed from government control over pricing, the airline industry quickly adopted a differential pricing strategy. The airlines took advantage of the variations in the sensitivity to time, price, comfort and service across travelers to offer several fare products, carry additional passengers and increase revenues.

In order to prevent passengers willing to purchase the most expensive fare product from buying lower-priced products, discount fares were systematically associated with a set of restrictions such as non-refundability, advance purchase and minimum stay requirements designed to make them unattractive to business travelers. A few studies (Boeing, 1988) have shown that these "fences" were relatively successful at discouraging business travelers from purchasing fare products designed for the price-sensitive leisure segment of the demand. Although differential pricing is rooted in economic theory as described by

Belobaba (1998) and has been widely adopted by airlines and other industries worldwide, fares are in most cases set in response to the competitive environment and not based on a scientific analysis of the price-sensitivity of the demand.

This relatively simple competition-based approach to airline pricing is due to the fact that differential pricing was associated with the dynamic management of airline seat inventory, also known as revenue management, which restricts the availability of lower-priced fare products. As a result, the lowest available price offered in a market not only depends on the list of published fare products established by the airline pricing department but also on the number of seats available for sale for each product set by the airline revenue management system and staff.

Although a well designed fare structure would likely improve the effectiveness of the airline revenue management system, the impact of how the fares are set is relatively limited when a large number of price points are used. For instance the potential incremental revenue of setting a mid-end fare to $102 instead of $100 is low compared to the benefits of deciding whether the airline should continue selling this fare or close it down in favor of a more expensive fare product with a price of $130 in a particular market. In addition, the revenue management system takes into account the expected revenue of each fare product when setting seat availability controls. Finally, increasing the price even by a small amount may be detrimental to the airline's market share if the competitors in the marketplace fail to match the increase. This is especially true for low-fare products as that part of the demand is extremely sensitive to price and the availability of low-fare products generally indicates that demand for some particular flights is likely to be fairly low relative to capacity.

This explains why much of the scientific attention has been focused on how to improve airline revenue management models rather than on airline pricing itself. In addition, Boyd (2007) argues that revenue management science can remain less complicated than it would otherwise be thanks to the division between pricing and revenue management. As a result, revenue management scientists can focus on whether lower-priced or higher-priced products should be made available without getting concerned about setting the actual prices. The division between competition-based pricing and science-based revenue management has shown to provide a good approximation to the very complex problem of pricing airline seats.

Even when a major overhaul of the airline fare structure is planned, airlines seem to rely mostly on trial and error rather than on a scientific approach. For instance, Delta Airlines first introduced its new fare structure dubbed "Simplifares" in markets out of its Cincinnati hub in 2004 before rolling it out in all U.S. domestic markets in January 2005. As mentioned in Chapter 2, in addition to fewer products and a compressed price differential between the most and least expensive product, Simplifares replaced the Saturday night stay requirement, one of the most effective fences used for more than two decades, by a much less restrictive one-day minimum stay. As expected, the other major airlines responded by matching Delta's new fare structure either system-wide or in competitive markets. Faced with a large decrease in average fares that was only partially offset by an increase in passengers, Delta gradually removed the various features of Simplifares, once again increasing the number of fare products used in domestic markets and reintroducing a three-day minimum stay in many markets where the presence of low-cost competition is relatively minor.

While airline pricing has remained largely driven by competition, there is a growing need for a more scientific approach to airline pricing due to the

development of a new product strategy sometimes referred to as un-bundling or a la carte pricing. As mentioned in Chapter 2, airlines such as Air Canada have started to market product packages that are differentiated not by a set of restrictions designed to make them unattractive to some categories of travelers but by a set of product features designed to match the needs of the various segments of the market. In addition, this multi-product strategy is completed with a menu of options that can be added and subtracted to further tailor each product to the needs of the traveler.

This strategy is getting more and more attention in the industry both among network airlines and low-cost carriers. For instance, Southwest Airlines (Reed, 2007) recently introduced a new product dubbed "Business Select" that offers priority boarding, extra frequent flier credits and a complimentary cocktail on-board for a $10 to $30 premium over the standard full fare. The airline expects Business Select to generate a $100 million in extra revenues and is part of a wider strategy designed to bolster revenues and offset the impact of high labor and increasing fuel costs. The airline has set the ambitious goal of achieving a $1 billion increase in incremental revenues by 2009 through a series of initiatives designed to generate ancillary revenues or attract more business travelers paying higher fares, a very significant figure for a company with revenues slightly under $10 billion in 2007.

A more scientific approach to the pricing of these new product features may help the airlines achieve their ambitious goals of large incremental revenues. Unlike airline seats, most of these product features are not subject to capacity constraints and their pricing cannot be supported by the use of standard revenue management techniques. They are also, at least for now, less subject to the competitive pressures of the marketplace as they are not part of the standard price displayed by the major travel websites such as Travelocity or Expedia.

Actually, most of these product features are only available on the airline's website and proposed to a more captive segment of the market relative to travelers using other distribution channels.

Discrete choice models can be used to measure the value different categories of travelers place on these features and can provide valuable insight and guidance on how to price this new component of the airline product offering. For instance, the estimation results of our model provide an estimate of the passenger willingness to pay for the flexibility to change travel plans. Furthermore, our estimation results show that the value of flexibility tends to increase with the level of advance planning of the trip. These results provide guidelines on how to price product packages that offer this type of feature such as Air Canada's flexible Latitude product. For instance, this suggests that Air Canada may be able to increase the sales and revenues of its Latitude product by replacing its current flat fee structure by a scheme in which the price to upgrade to the Latitude package depends at least in part on how far in advance the ticket is booked.

However, the range of applications in airline pricing may be relatively limited for models based solely on booking data. As discussed in Chapter 4, discounted fare products carry two types of fare rules, product features such as the flexibility to modify or cancel a ticket and conditions such as minimum stay or advance purchase requirements. While explanatory variables can be included in the utility function to represent the features of a product, the disutility value of conditions cannot be captured by the model as they are embedded into the passenger choice set. In addition, discrete choice models based on booking data cannot be used to explore the value of new product features considered by the airline. As a result, the flexibility provided by stated preference data may be useful for many airline pricing applications. Stated preference experiments can

be designed to explore the value of fare rules that restricts the access to discounted fare products and can include non-existing product features.

Stated preference data should nevertheless be combined with booking data to reduce the risk of bias associated with the hypothetical nature of the experiment as discussed in Chapter 3. This is especially important as pricing applications involve exploring the trade-off between the characteristics of products and the hypothetical disbursement of money. As a result, the full potential of discrete choice models for applications to airline pricing decisions may only be realized by combining booking and stated preference data. However, we will discuss in the next section how discrete choice models based on booking data can provide a valuable tool to support the development of new revenue management techniques designed to maximize revenues under the new pricing strategies now commonly used in the industry.

7.4 Revenue Management

7.4.1 Revenue Management for Less Restricted Fare Structures

As mentioned earlier, airline pricing decisions are divided into two steps. The airline pricing department manages the list of fare products offered in each market and sets fare levels and fare rules based mostly on market conditions. Then, the airline revenue management system takes these as given and sets limits on the availability of lower-priced products to protect seats for expected future demand of higher-priced products. The basic revenue management decision is then whether to accept at a given point in time a booking request for a specific fare product or reject it in anticipation of future, yet uncertain higher-fare demand. Revenue management systems require both a forecast of future demand

and a resource allocation mechanism to decide whether or not to accept booking requests given the remaining seat capacity at the time of the request.

Littlewood (1972) proposed the first scientific approach to the seat allocation problem for a two-class model based on the concept of expected marginal revenue. Along with other assumptions, Littlewood's decision rule assumes that the airline is able to achieve perfect segmentation of the market and that demand for the higher-priced product does not depend on the availability of the lower-priced product. In his Ph.D dissertation, Belobaba (1987) extended Littlewood's rule to the n-class framework in a model known as expected marginal seat revenue (EMSR) that also relies on the independent demand assumption. Most airline revenue management systems developed in the late eighties and the nineties were based on EMSR-type models.

While the use of well-designed fare rules can provide a good level of segmentation of the market, it was recognized early on that the independent demand assumption does not hold fully in practice. A passenger that is able to satisfy the rules of a lower-priced albeit more restricted fare product may also be willing to purchase a higher-priced product if the lower-priced product is no longer available, an effect called here sell-up. In his dissertation, Belobaba proposed a modified version of the EMSR model that accounts for sell-up across adjacent classes by adjusting downward the fare of the lower-priced class based on the probability of sell-up to the higher class also called sell-up rate. Belobaba and Weatherford (1996) generalized this framework by accounting for sell-up to all higher classes. However, in these models, demand forecasts for higher-priced products do not depend directly on the availability of a lower-priced product. Brumelle and McGill (1990) proposed a decision rule for a two-class model that explicitly accounts for the impact of the availability of the lower-priced product on the demand for the higher-fare product.

These models were not widely adopted by the airline industry for two reasons. First, as discussed earlier, fare rules such as the Saturday night stay and advance purchase requirements proved fairly effective at preventing passengers with a high willingness to pay from purchasing deeply discounted tickets. Second, little guidance was provided on how to estimate the sell-up rates required by these models. Thus, much of the attention of the airline revenue management community switched on how to maximize revenues at the network level.

However, as mentioned in Chapter 2, the expansion of low-cost airlines has led to the relaxation of fare rules in many markets, especially the Saturday night stay requirement. This has greatly reduced the performance of EMSR-type models. When the independent demand assumption is increasingly violated, fewer bookings for the higher-priced product are observed which, in turn, leads to lower forecasts for future high-price demand. This results in lower protection levels and even fewer bookings for high-end products. This pattern leads over time to a progressive decline in the sales of higher-priced products and a substantial decrease in airline revenues called the "spiral-down effect". Boyd et al. (2001) first demonstrated the impact of the spiral-down effect using simulation techniques while Cooper et al. (2006) provided a formalized approach of the problem.

The development of a la carte pricing and a multi-product strategy initiated by Air Canada and used by an increasing number of airlines does not reduce the need for a revenue management model that does not rely on the independent demand assumption. This new pricing strategy provides the industry an alternative to segment the demand without imposing a Saturday or minimum stay requirement to discounted products. The extra features used to differentiate across the various product packages are typically priced as an add-on to the most basic product. However, a range of price points is used within each product

category and revenue management is used to set what can be viewed as the price of using seat capacity in a specific market. Since the different price levels within a product category are not differentiated by anything but price and possibly some advance purchase requirements, the success of this strategy remains highly dependent on the ability of the revenue management system to maximize revenues under a less restricted fare structure in which the independent demand assumption does not hold.

As a result, most of the airline revenue management research in recent years has focused on how to maximize revenues given a less restricted fare structure. Talluri and Van Ryzin (2004) proposed a model to maximize revenues for the single flight-leg problem under a general model of consumer choice and Liu and Van Ryzin (2008) extended this approach at the network level. Fiig et al. (2005) proposed an adjustment to the displacement adjusted virtual nesting (DAVN) model widely used in the industry to set revenue management controls at the network level to account for the different levels of product differentiation across markets. Gallego, Li and Ratliff (2007) proposed an extension of the decision rule of Brumelle et al. to the n-class problem. Like previous models proposed in the early nineties, all these models require an estimate of the sell-up rates across fare products.

7.4.2 Estimating Sell-up Behavior

Belobaba et al. (2006) proposed a methodology to estimate sell-up behavior directly from observed bookings. They calculate at different points of the booking process the sell-up rate from the lowest-priced to a particular product p as the quotient of observed bookings for product p and higher-priced products over the sum of all bookings. They use a series of ordinary least squares regressions to estimate an average value of the sell-up rate at various stages of

170

the booking process based on an exponential demand assumption. However, these models have proved fairly difficult to calibrate.

Discrete choice models provide a flexible alternative means to evaluate the sell-up behavior. Observed bookings are not used to estimate sell-up potential directly but to calibrate a model of airline passenger choice. The choice probabilities for each travel alternative can easily be re-calculated based on the parameter estimates of the model when the passenger choice set is modified. Discrete choice models can then be used to capture a number of effects such as sell-up behavior. For instance, sell-up behavior can 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.

In order to estimate sell-up behavior, a model of the choice of a fare product is then needed. In order to provide more accurate predictions of sell-up potential, the model should also incorporate the impact of pricing and revenue management on the choice set of each booking. The model developed in this research is then well-suited to estimate the sell-up behavior of. The parameter estimates of the model were applied to forecast the sell-up rate for a pair of fare products from the Week category. The sell-up rate from the AWEEK21 (168 EUR) to the UWEEK14 (217 EUR) product was calculated for an early morning departure that is highly valued by many business-oriented travelers.

The following five-step process was used to calculate the expected sell-up rate:

- Step 1: Select all bookings for which a AWEEK21 product was chosen on an early morning flight departure.

171

- Step 2: For each booking, calculate the choice probabilities and the total share of the lower-priced product (AWEEK21). First we calculate the value of the utility function for each of the alternatives in the passenger choice set. For instance, for an overnight booking by a week traveler, the utility value of the AWEEK21 product on the 7:35 a.m. flight departure in the Paris to Frankfurt market is given by Equation (7.1) below:

$$
\begin{aligned}
U(AWEEK21, 7:35AM) &= -0.0145*161.04 - 2.48 + 0.26*\sin(\frac{2\pi(7.583)}{16}) + 0.49*\sin(\frac{4\pi(7.583)}{16}) \\
&+ 0.08*\sin(\frac{6\pi(7.583)}{16}) + 0.17*\sin(\frac{8\pi(7.583)}{16}) - 0.14*\cos(\frac{2\pi(7.583)}{16}) + 0.67*\cos(\frac{4\pi(7.583)}{16}) \\
&- 0.33*\cos(\frac{6\pi(7.583)}{16}) - 0.08*\cos(\frac{8\pi(7.583)}{16})
\end{aligned} \tag{7.1}
$$

The share of the lower-priced product $S_{AWEEK21}$ is then equal to the sum of choice probabilities for the AWEEK21 product over all bookings b selected in the first step of the process:

$$
S_{AWEEK21} = \sum_b \frac{e^{U(AWEEK21, 7:35AM)}}{\sum_{j=1}^{J_b} e^{U(j)}} \tag{7.2}
$$

Where J_b is the number of alternatives in the choice set of booking b

- Step 3: Adjust the choice set by replacing the AWEEK21 product by the higher-priced UWEEK14 product.

- Step 4: For each booking, calculate the choice probabilities given the adjusted choice set. Equation (7.1) becomes:

172

$$U(UWEEK14,7:35AM) = -0.0145*21304 - 2.48 + 0.26*\sin\left(\frac{2\pi(7.583)}{16}\right) + 0.49*\sin\left(\frac{4\pi(7.583)}{16}\right)$$

$$+0.08*\sin\left(\frac{6\pi(7.583)}{16}\right) + 0.17*\sin\left(\frac{8\pi(7.583)}{16}\right) - 0.14*\cos\left(\frac{2\pi(7.583)}{16}\right) + 0.67*\cos\left(\frac{4\pi(7.583)}{16}\right) \quad\quad (7.3)$$

$$-0.33*\cos\left(\frac{6\pi(7.583)}{16}\right) - 0.08*\cos\left(\frac{8\pi(7.583)}{16}\right)$$

The share of the higher-priced product (UWEEK14) is then equal to:

$$S_{UWEEK14} = \sum_b \frac{e^{U(UWEEK14,7:35AM)}}{\sum_{j=1}^{J_b} e^{U(j)}} \quad\quad (7.4)$$

- Step 5: The sell-up rate is equal to the share of the higher-priced product ($S_{AWEEK21}$) divided over the share of the lower-priced product ($S_{UWEEK14}$).

Figure 7.1 below shows the estimated sell-up rate obtained for a model based on a deterministic segmentation of the demand between week and non-week bookings and the two-class latent class choice model developed in this research. The expected sell-up rate is lower for the latent class model than for the deterministic benchmark model. While the bookings for this non-flexible business-type product are classified as primarily time-sensitive week bookings by the deterministic benchmark model, they get split by the latent class model between the two latent classes reflecting the higher degree of price sensitivity of these business travelers. As a result, the airline revenue management system will protect fewer seats for the higher-priced product if sell-up behavior is estimated according to the latent class model.

173

AWEEK21 to UWEEK14 (168 to 217 EUR)
Early Morning Flight Departure

57.8%   47.8%

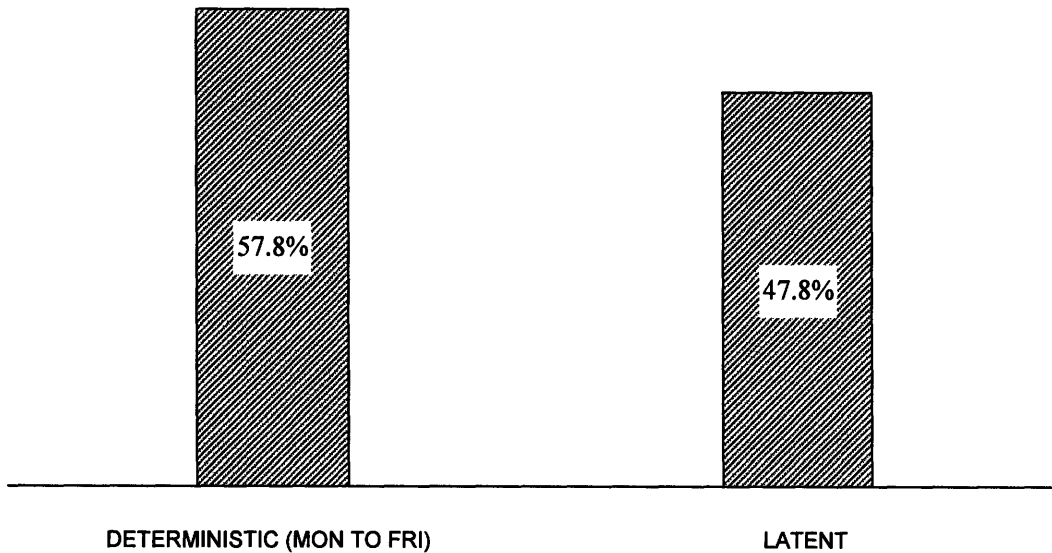DETERMINISTIC (MON TO FRI)   LATENT

Figure 7-1:   Estimated Sell-up Rates

However, these estimates are highly dependent on some assumptions of the model. For instance, some travelers may decide not to travel at all if the lower-priced product is no longer available. This effect is ignored as the model is based on observed booking data and no information is available on prospective travelers that did not find a travel alternative that satisfied their needs and budget and chose not to travel. More importantly, the impact of competition is not taken into account as the model takes the choice of a particular airline as given and focuses on the choice of an itinerary and fare product. Competition is expected to have a large impact on sell-up behavior. Passengers will be less likely to purchase a higher-priced product and may spill to the competition if a cheaper alternative is available on another airline, in particular at a similar departure time. As a result, in order to provide a more realistic estimate of sell-up rates, the impact of competition should be incorporated.

Gallego, Li and Ratliff (2007) proposed to create a pseudo-alternative and calibrate its utility value to represent the competition in an aggregate manner based on the respective market share of the airline and its competitors. However, the development of screen scraping techniques provides the airlines with the means to obtain a fairly accurate picture of the attributes of travel alternatives offered by the competition. The use of competitor availability data has sparked some debate in the revenue management community, with some fearing that this could trigger a spiral-down in fares if every airline starts matching its competitors on a real-time basis. While the use of competitor availability data to directly match the competition does not incorporate the complexities of the network effects specific to each airline and could potentially have a negative impact on airline revenues, competitor availability data may be a very valuable resource to capture the impact of competition on sell-up behavior.

Since an airline cannot fully access the bookings of its competitors, the calibration of a passenger choice model is likely to remain based on its own booking records. However, competitor availability data could be used to calculate for each booking the choice probabilities for the travel alternatives offered both by the airline and its major competitors. Sell-up rates could then be obtained by replacing in the choice set a lower-priced product by a higher-priced alternative while maintaining competitor availability and prices constant. As a result, this approach could provide a more accurate estimate of sell-up potential that captures the impact of competition at the disaggregate booking level. By using competitor availability data to estimate sell-up behavior, the airlines may prevent over-protecting seats for higher-priced demand that spills to the competition and fails to materialize while at the same time avoiding the risks of a large downward pressure on fares associated with a large-scale real-time matching strategy of competitor availability and prices.

7.4.3 An Integrated Approach to Choice-based Revenue Management

Figure 7-2 below presents an integrated framework of a choice-based revenue management system. In this framework, the airline's booking and seat availability data are used to calibrate a choice model of itinerary and fare product. Real-time competitor seat availability data is then used to extend the passenger choice set to the travel alternatives offered by the competition and parameter estimates of the model are applied to forecast sell-up behavior. Estimated sell-up rates are then used to feed a choice-based revenue management optimizer designed to maximize revenues under a less restricted fare structure in which demand for higher-priced products is assumed to depend in part or entirely on the availability of lower-priced products.
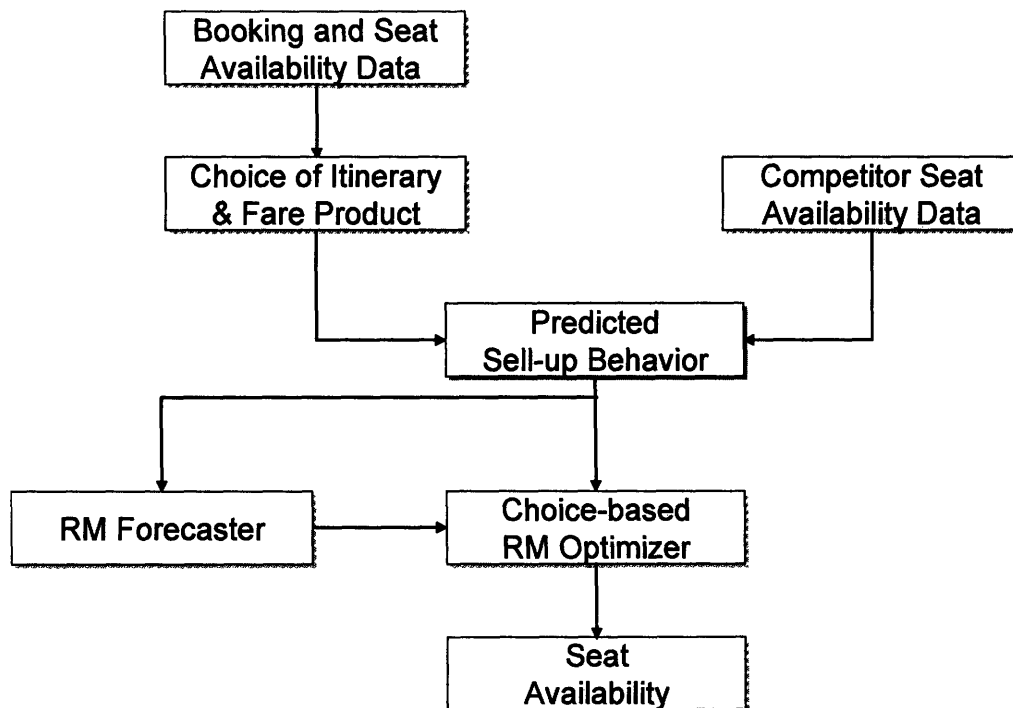

Figure 7-2:   Structure of a Choice-Based Revenue Management System

While the performance of this integrated choice-based approach to revenue management remains unknown, simulation studies (Belobaba, 2007) based on

176

the Passenger Origin-Destination Simulator (PODS) have shown that the revenue losses associated with the relaxation of fare rules can be partially recovered by combining an estimate of sell-up behavior with a choice-based revenue management optimizer. This integrated approach may have the potential to further reduce the revenue losses associated with a less restricted fare structure by improving the accuracy of the sell-up estimates. Simulation studies are then needed to assess the revenue performance of this framework.

7.5 Summary

As mentioned earlier, discrete choice models can be applied to analyze the impact of any airline planning decision that involves a change in either the attributes of the alternatives, the passenger choice set or both. By modeling the choice of an airline itinerary and fare product, the model developed in this research has the potential to both improve existing applications in schedule planning and extend the scope of applications to other areas of airline planning such as pricing and revenue management.

The development of a multi-product strategy along with a menu of optional product features designed to segment demand in lieu of rescinded fare rules such as the Saturday night stay requirement has raised the need for a more scientific approach to airline pricing. Discrete choice models can provide a valuable tool to support the pricing of these product packages and their various options. However, the range of applications in pricing could be expanded by combining booking and stated preference data to test the market response to product features embedded in the passenger choice set or new non-existing products considered by the airline.

In addition, the calibration of a choice model of airline itinerary and fare product that incorporates the impact of pricing and revenue management on the passenger choice set has the potential to improve the estimation of sell-up behavior if combined with real-time competitor availability data. More accurate estimates of sell-up behavior are needed to improve the performance of revenue management algorithms designed to maximize revenues for less restricted fare structures and contribute to recover the losses associated with the relaxation of fare rules such as the Saturday night stay requirement.

In the next chapter, we will discuss in more detail further research directions to improve the accuracy of the model and some of the steps needed to implement the choice-based revenue management framework described in this chapter.

Chapter 8   Conclusion

In the preceding chapters, we developed and estimated a model of the choice of an airline itinerary and fare product that incorporates the impact of airline pricing and revenue management on the passenger choice set and captures the choice behavior of different segments of airline travelers. We then discussed how this model can provide a valuable tool to support new pricing and revenue management strategies under development to offset the negative revenue impact of relaxed fare rules and increased price transparency in the marketplace. This chapter concludes this dissertation by summarizing the important findings of this research and their relevance to the current revenue challenges faced by both network and low-cost airlines. We then discuss future research directions to improve the accuracy of the model and provide a better representation of the choice environment in the airline industry.

8.1   Research Findings and Contributions

Previous studies of airline passenger choice based on booking data have been restricted to the sole choice of an airline itinerary and have ignored the fundamental trade-off between itinerary, fare and fare product characteristics when air travelers select a travel alternative for a future trip. This is due to the properties of booking data, in particular the lack of information on the availability of other fare products at the time of the booking. In this research, booking data was combined with other data sources to model the joint choice of an airline itinerary and fare product and explore the trade-off between price and

schedule in airline markets. Since only the booked itinerary and fare product are currently recorded in airline bookings, booking data was combined with fare rules and seat availability data to incorporate the impact of pricing and revenue management on the passenger choice set and represent more accurately the choice environment in the industry. For each booking, the choice set was inferred by applying the fare rules such as minimum stay and advance purchase requirements and checking seat availability on the date of the booking. The choice set generation process also incorporated the characteristics of specific categories of bookings. For instance, only morning flights were included in the choice set of outbound day trippers. A sample of slightly over 2000 bookings in three short-haul European markets was developed and used for this research.

While such a process proved effective to infer the choice set for a limited number of bookings and produced a unique dataset to test the model proposed in this research, any large-scale implementation would require developing a new approach to data collection and management. Ideally, information about the choice set should be collected and attached to the passenger name record created at the time of the booking. Online travel retailers such as Travelocity already collect that type of data by recording snapshots of the screens viewed by the site's visitors (Smith, 2006). However, to get a more complete view of the marketplace, choice set information needs to be collected for all distribution channels including bookings made through offline travel agents. In part as a result of this research, Amadeus has initiated a project to collect choice set data and store it directly in all booking records for airlines hosted in its new inventory system called Altea Plan. Competing GDS are also making the technology investments to collect choice set information for both web-based and non-web bookings. These efforts should provide the basis for a potential large-scale implementation of the model developed in this research.

Previous studies of airline passenger choice based on booking data also ignored the heterogeneity of behavior across different categories of travelers, a major characteristic of airline markets. Airline pricing strategies are typically designed to segment the market between time-sensitive business and price-sensitive leisure travelers. However, information on trip purpose is not available in airline bookings. In order to replace trip purpose, several elements of the booking records that tend to be correlated with trip purpose were used such as the characteristics of the trip and the profile of the traveler. A latent class choice model was specified to capture heterogeneity of behavior in the dataset. Compared to deterministic segmentation schemes, latent classes provide the flexibility to segment the demand based on several factors without dividing the bookings in a large number of small sub-segments. In addition, these various factors are included as explanatory variables of the class membership model and their relative contribution to the segmentation of the demand can be evaluated based on the estimated value of their parameters.

Estimation results of the latent class model of airline passenger choice showed that characteristics of the trip such as travel wholly within a week or booking through a traditional travel agent were the two most significant drivers of heterogeneity of behavior followed to a lesser extent by the membership in the airline loyalty program. Other elements such as gender were not found to contribute significantly to the segmentation of the demand in these markets. Two classes of travelers were identified, the former being more time and product-sensitive while the choice behavior of the latter was more influenced by price.

This is in line with the traditional segmentation of airline markets between time and service-sensitive business travelers and price-sensitive leisure travelers. However, compared to previous models based on a deterministic segmentation of the demand by trip purpose, the latent class choice model was found to split

the bookings between a core of time-sensitive business travelers and a mixed class of leisure and price-conscious business travelers willing to give up the flexibility to change travel plans in exchange for a discount off the fully flexible albeit more expensive fare product. These non-flexible discounted fare products that do not require an overnight stay during the weekend were introduced by the airline in response to growing low-cost competition in European short-haul markets. By capturing the differences in choice behavior between sub-segments of business-type travelers, the estimation results of the model support the development of such a multi-product pricing strategy to defend market share in a more competitive marketplace while limiting the impact on revenues.

Latent classes also provide an attractive and flexible framework to capture heterogeneity of behavior in airline markets as it mimics typical airline pricing strategies based on product differentiation. The latent class choice model was found to improve the fit of the model over a deterministic benchmark according to model selection criteria such as the AIC and BIC. The benefits of the multi-criteria approach to segmentation provided by latent classes should become even greater as airlines evolve from traditional pricing strategies based on fare rules to more sophisticated multi-product and a la carte pricing strategies designed to better serve the specific needs of the different segments of the market.

In addition, the model developed in this research also improved the measurement of the time-of-day preferences of airline travelers. A generalization of a trigonometric continuous function of time was proposed to incorporate the specific characteristics of time-of-day preferences in short-haul markets. In this new formulation, the duration of the cycle is estimated rather than set to a full 24 hours. This approach takes into account the lack of demand for nighttime flights in short-haul markets and results in a profile of time-of-day preferences that is consistent with previous studies. In addition, a continuous function of time was

found to provide a more precise measurement of the time-of-day preferences and a substantial improvement in the fit of the model compared to previous specifications based on time-period dummies. This new formulation also provides a flexible approach to model the time-of-day preferences of specific categories of travelers such as day trippers for which demand for outbound travel is concentrated in the morning.

The model of the choice of an airline itinerary and fare product developed in this research provides the foundation for a range of airline planning applications. It has the potential to improve the prediction capabilities of existing itinerary-based choice models used for schedule planning applications by incorporating the trade-off between fare and schedule and the impact of heterogeneity of behavior. More importantly, it expands the scope of potential applications to additional airline planning decisions such as pricing and revenue management that are essential to meet the revenue challenges faced by the industry.

It provides a scientific approach to the development of a multi-product and a la carte pricing strategy that is increasingly popular among both network and low-cost carriers. For instance, the estimation results of the model suggest that the value of the flexibility to change travel plans tends to increase with the level of advance planning. This type of insight could be used by the airline pricing department to set the premium of a product package offering that kind of feature. Estimation results of the model were also applied to evaluate the sell-up potential from a lower-priced to a higher-priced product, a major input to the new revenue management models being designed to maximize revenues in a less restricted fare structure. Compared to previous model specifications based on a deterministic segmentation of the demand, the latent class structure of the model was found to provide lower estimates of sell-up behavior reflecting the higher level of price-sensitivity of some business-type passengers.

8.2 Future Research Directions

This research has focused on the development of a model of the choice of an airline itinerary and fare product that incorporates the impact of pricing and revenue management on the passenger choice set. A latent class model was estimated based on a sample of actual bookings to illustrate how an airline can exploit its existing data sources to better understand the choice behavior of different categories of travelers. While this dissertation can be viewed as a proof-of-concept of the potential benefits of a booking-based multi-dimensional model of airline passenger choice, the findings of this research should be further validated by applying the model to a larger set of more heterogeneous bookings.

As described in Chapter 5, the data collected for this research focused on a particular set of three business-oriented short-haul markets that are largely dominated by traditional network airlines. In these markets, bookings were found to be divided between two fairly homogeneous groups. In addition, data was collected over a short period of time resulting in a relatively limited sample size. By applying the same methodology to a larger dataset collected over a longer period of time, additional effects could be captured such as the impact of seasonality or day-of-the week.

More importantly, the validity of the approach should be tested over a wider range of markets that differ with regard to some major characteristics such as length of haul, business/leisure mix, fare structure or penetration of low-cost competition. Bookings are expected to be more heterogeneous in markets with a more balanced distribution of the demand between business and leisure travelers or a higher degree of low-cost competition. A larger number of latent classes could then be potentially identified, especially if a larger sample is used to

estimate the model. Additional heterogeneity of behavior could also potentially be captured within each latent class through random coefficients.

In addition to testing the validity of the model under various market conditions, extensions to the model may also provide the means to relax some of the assumptions made in this research and improve the accuracy of the model. Regarding the choice of an itinerary, some assumptions can be fairly restrictive and extensions of the model could provide substantial benefits, especially for schedule planning applications. For instance, only non-stop flights were included in the passenger choice set and connecting itineraries were not considered.

The impact of this assumption is minor in the short-haul markets selected for this research as connecting itineraries are very unattractive in these markets. However, as the model is applied to a more diverse set of markets, including markets with different lengths of haul, connecting itineraries should be considered. The number of connections and elapsed travel time are expected to have a significant impact on passenger choice behavior in many medium and long-haul markets, especially on the trade-off between price and itinerary. Some passengers may be willing to pay a premium for traveling on a non-stop flight rather than on a connecting itinerary, especially time-sensitive business-style travelers. As a result, connecting itineraries should be included in the choice set and explanatory variables should be added to estimate the impact of the number of connections and the total travel time for each segment of the market.

In addition to connecting itineraries, the model should also be extended to allow for more flexible substitution patterns across itineraries compared to the logit specification used in this research. For instance, adjacent flight departures are expected to be closer substitutes than flights departing in the morning and in the evening, especially for time-sensitive business travelers. As described in Ben-

Akiva and Lerman (1985), the IIA property of the multinomial logit model only applies within each market segment and not to the bookings as a whole. As a result, the use of a latent class model to account for heterogeneity of behavior mitigates to some extent the impact of the IIA property. As additional heterogeneity of behavior could be captured through the use of a larger and more diverse dataset and a higher number of latent classes could potentially be identified, the impact of the IIA property would be even further reduced and a logit-based model could still yield relatively realistic forecasts of passenger choice behavior.

However, the impact of the IIA property within each latent class of the model may still be significant, especially for the time-sensitive segment of the market. In order to overcome the limitations of the IIA property, a nesting structure in which itineraries are grouped based on their departure and/or arrival times could be introduced to better represent the complexity of the substitution effects across itineraries. Similarly, itineraries could also be grouped by path quality to reflect the potential higher degree of substitution among non-stop versus connecting itineraries. A nested structure could be introduced through a choice kernel from the generalized extreme value family (GEV). Alternatively, this could also be modeled using a mixed logit specification. As mentioned in Chapter 3, McFadden and Train (2000) have shown that any choice model consistent with random utility theory can be approximated by a mixed logit model by using an appropriate set of variables and mixing distribution. Such an approach could be especially attractive if a mixing distribution is already used to test for random taste variation within each latent class.

The use of additional data sources such as survey, stated preference and competitor availability data has also the potential to enrich and further improve the accuracy of the model. Survey data provide the opportunity to collect

additional information unavailable in booking records and mitigate some of the limitations associated with the sole use of booking data. For instance, data could be collected on characteristics of the trip such as trip purpose or characteristics of the traveler such as age and income that would supplement the information already contained in booking records and contribute to a more precise segmentation of the demand. Similarly, survey participants could be asked about their preferred departure or arrival time supplementing information on the actual flight departure and arrival times found in airline bookings. Preferred departure time information could be used to model the impact of schedule delay on the time-of-day preferences of airline travelers and would improve the accuracy of the model, especially for schedule planning applications.

In addition to more information about the profile of the travelers, surveys of airline passengers provide the opportunity to conduct stated preference experiments that could be very valuable, especially for pricing applications. As mentioned earlier, the impact of some fare rules that restrict the access to specific fare products such as minimum stay requirements cannot be captured easily in models based solely on booking data as they are embedded in the passenger choice set. Stated preference experiments provide a means to capture these effects directly by designing scenarios in which passengers are asked to choose between alternatives that carry or not the fare rule of interest. Stated preference experiments can also be used to capture the impact of new non-existing product features. This could provide a valuable tool to support the introduction of a la carte pricing strategies that are becoming increasingly popular in the industry. However, stated preference data should be combined with booking data to mitigate the risk of response bias. This is especially true for pricing applications for which stated preference experiments are typically used to explore the passenger willingness to pay for extra product features.

Data on the state of the competition could also be very useful, in particular for revenue management applications. As discussed earlier, the potential for sell-up to a higher-priced product is expected to depend on the attributes of travel alternatives (price, schedule, path quality) offered by major competitors operating in the market. Competitor availability data is typically publicly available either through a GDS for participating carriers or through the airline website for non-participating carriers and can be captured automatically through scrapping techniques. Ideally, the attributes of the travel alternatives available on competitors should be recorded in each booking as part of the choice set information along with data on the different alternatives offered by the airline.

Travel alternatives on competitors cannot be included in the choice set for estimation purposes as booking records are proprietary. As a result, an airline cannot get full access to the bookings of its competitors although partial information can be obtained through the Market Information Data Tapes (MIDT) that can be purchased from GDS companies for tickets booked through their systems. However, travel alternatives on competitors should be included in the choice set at a disaggregate booking level to predict sell-up behavior in a competitive environment as described in the framework proposed in Chapter 7. Simulation studies should be used to determine the relative revenue performance of the choice-based approach developed in this research compared to other methodologies used to estimate sell-up behavior for the various optimization algorithms proposed to maximize revenues with less restricted fare structures.

Finally, the dynamics of the interaction between passenger choice behavior and airline planning decisions could be a very interesting and promising area of future research. In this dissertation, we have used past booking data to calibrate a passenger choice model and support a range of airline planning decisions such

as schedule planning, pricing and revenue management. However, these airline decisions will in turn modify the passenger choice set and ultimately affect the choice behavior of future travelers. As a result, simulation studies could be a very valuable tool to identify these dynamic effects and determine whether the interaction between airline passenger choice and planning decisions converges to an equilibrium.

Enriched with stated preference and competitor availability data, the latent class model of airline passenger choice developed in this research has the potential to provide an effective choice-based decision support tool for many airline planning applications including pricing and revenue management decisions. Some of these applications have been envisioned for a long time. For instance, the use of passenger choice models to capture effects such as sell-up behavior and support revenue management decisions was already proposed in early works on revenue management such as Belobaba (1987). 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 a science-based new pricing strategy 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 performance of these new revenue-enhancing strategies.

References

Abou Zeid M., Rossi T.F., Gardner B. Modeling Time of Day Choice in the Context of Tour and Activity Based Models. Presented at 11th World Conference on Transport Research, Berkeley, CA, 2007.

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.

Algers S., Beser M. Modelling Choice of Flight and Booking Class – a Study Using Stated Preference and Revealed Preference Data. *International Journal of Services Technology and Management,* Vol.2, Nos. 1/2, 2001, pp 28-45.

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

Belobaba P. *Airline Differential Pricing for Effective Yield Management.* In Handbook of Airline Marketing, Butler G.F. and Peel J., eds., Aviation Week Group (Mc Graw-Hill), 1998, pp 349-361.

Belobaba P. What's Wrong with U.S. Airlines – Recent Airline Industry Challenges. Annual Presentation to MIT IAP Seminar, Cambridge, MA, January 2005.

Belobaba P., Hopperstad C., Zerbib G. Methods for Estimating Sell-up. Presented at the Annual Meeting of the AGIFORS Revenue Management and Distribution Study Group, Cancun, Mexico, May 2006.

Belobaba P. Comparisons of RM Methods for Reversing Spiral Down. Presented at the Annual Meeting of the AGIFORS Revenue Management and Distribution Study Group, Jeju, South Korea, May 2007.

Belobaba P., Weatherford L. Comparing Decisions Rules that Incorporates Customer Diversion in Perishable Asset Revenue Management Situations. *Decision Sciences*, Vol.27, No. 2, 1996, pp 343-363.

Ben-Akiva M. *Structure of Passenger Travel Demand Models*. Ph.D Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1973.

Ben-Akiva, M., Abou-Zeid, M. Methodological Issues in Modeling Time-of-Travel Preferences. 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.

Ben-Akiva M., Morikawa T. Estimation of Travel Demand Models from Multiple Data Sources. *Transportation and Traffic Theory*, 1990, pp 461-476.

Boeing Commercial Airplane Group. 1988 Domestic Fare Survey. Seattle, WA, 1988.

Boeing Commercial Airplane Group. Decision Window, Path Preference Methodology, Time Mode. Seattle, WA, 1993.

Boehmer J. Simplifares Dead but Legacy Remains. *Business Travel News*, 24 September 2007.

Boyd E.A. *The Future of Pricing, How Airline Ticket Pricing Has Inspired a Revolution*. Palgrave Macmillan, New York, 2007.

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.

Brumelle S., McGill J., Oum T., Sawaki K., Tretheway M. Allocation of Airline Seats between Stochastically Dependent Demands. *Transportation Science*, Vol. 24, No. 3, 1990, pp. 183-192.

Brunger W. *The Impact of the Internet on Airline Fares*. Ph.D Dissertation, Case Western University, Cleveland, OH, 2008.

Coldren G.M., Koppelman F.S., Kasturirangan K., Mukherjee A. Modeling Aggregate Air-Travel Itinerary Shares: Logit Model Development at a Major U.S. Airline. *Journal of Air Transport Management*, Vol. 9, No. 6, 2003, pp. 361-369.

Coldren G.M., Koppelman F.S. Modeling the Competition among Air-Travel Itinerary Shares: GEV Model Development. *Transportation Research, Part A*, Vol. 39, No. 4, 2005, pp. 345-365.

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.

Compart A. GDS Agreements Paying Dividends for Jetblue. *Travel Weekly*, 25 October 2006.

Compart A. United Propose to Offer Lower Base Fares with Bevy of Add-Ons. *Travel Weekly*, 15 December 2006.

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.

De Lapparent M. *De la Valeur du Temps à la Valeur du Risque de Perte de Temps de Transport : le Cas des Déplacements Professionnels*. Ph.D Dissertation, University Paris-I Panthéon-Sorbonne, Paris, 2004.

De Lollis B. Fare Changes Gives Fliers Weekends Back. *USA Today*, 22 August 2006.

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.

Gopinath D.A. *Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand.* Ph.D Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1995.

Gourerioux C. and Monfort A. *Statistics and Econometric Models, Volume Two.* Cambridge University Press, Cambridge, UK, 1995.

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.

Gutschi, M. Air Canada Flyers Buy the Extras --- Nearly Half Accept Higher Fares to Get Meal or Choice Seat. *The Wall Street Journal*, 29 March 2007.

Liu Q., Van Ryzin G. On the Choice-Based Linear Programming Model for Network Revenue Management. *Manufacturing and Service Operations Management*, Vol. 10, No. 2, 2008, pp. 288-310.

Luce R.D., Suppes P. *Preference, Utility and Subjective Probability.* Handbook of Mathematical Psychology, Vol. 3. R.D. Luce, R.R. Bush, E. Galanter, ed. Wiley, New York, 1965.

Manski C. The Structure of Random Utility Models. *Theory and Decisions*, Vol. 8, 1977, pp 229-254.

Marschak J. *Binary Choice Constraints on Random Utility Indicators.* Stanford Symposium on Mathematical Methods in the Social Sciences. K. Arrow, ed. Stanford University Press, Stanford, CA, 1960.

McFadden D. *Conditional Logit Analysis of Qualitative Choice Behavior. Frontiers in Econometrics.* P. Zarembska, ed. Academic Press, New York, 1974, pp105-142.

McFadden D. The Choice Theory Approach to Marketing Research. *Marketing Science*, Vol. 5, No. 4, 1986, pp 995-1026.

McFadden D., Train K., Mixed MNL Models for Discrete Responses. *Journal of Applied Econometrics*, Vol. 15, No. 5, 2000, pp 447-470.

Meyer J. et al. *Entry and Competition in the U.S. Airline Industry: Issues and Opportunities*. Special Report 255, Transportation Research Board, National Research Council, July 1999.

Moore A., Cross-validation for Detecting and Preventing Overfitting. Tutorial Materials, Pittsburgh, October 2001. Website: www.cs.cwu.edu/~awm/tutorials, Accessed May 31, 2007.

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.

Pyrgiotis N., *Price Competition in Top U.S. Domestic Markets: Revenues and Yield Premium*. Master Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2008.

Reed D. Southwest Hopes Changes Add up to Ch-cha-ching. *USA Today*, 28 December 2007.

Schaal, D. Air Contract Details: Fees Dipping, Content Missing. *Travel Weekly*, 23 October 2006.

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.

Swelbar W. No Ordinary Time: The Airline Industry in 2003. Growth of Low Fare Carriers. Presentation to the 2nd Annual MIT Airline Industry Conference, Washington D.C., April 2003.

Swelbar W. Low Cost Carriers in the U.S.: Shifting Strategies Required for the next Growth Phase. Presentation to the 3rd Annual MIT Airline Industry Conference, Washington D.C., April 2004.

Swelbar W. Evolving Network Structures: Data and Analysis behind the Changes. Presentation to the Airline Management Class, Massachusetts Institute of Technology, Cambridge, MA, March 2007.

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.

Thurstone L. A Law of Comparative Judgement. *Psychological Review*, Vol. 34, 1927, pp 273-286.

Train K. *Discrete Choice Models with Simulation*. Cambridge University Press, New York, 2003.

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.

Walker J.L. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structure and Latent Variables*. Ph.D Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.