Consumer choice behaviour and strategies of air transportation service providers

Identifying opportunities in aviation for a transportation research institute and providing a case-study on itinerary choice based on revealed-preference data.

Michael A.B. van Eggermond

November 2007

Master thesis 'Transport, Infrastructure and Logistics'

TU Delft
Delft University of Technology

IVT Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Consumer choice behavior and strategies of air transportation service providers

Michael A.B. van Eggermond
Institute for Transport Planning and Systems
ETH Zürich

Civil Engineering – Transport & Planning
Technology, Policy and Management
TU Delft

Phone - NL: +31 (0) 6 24 52 8383
Phone - CH: +41 (0) 79 595 71 84

michael@eggermond.com
November 2007

Committee

Chair

Professor Piet H.L. Bovy	*Delft University of Technology*

Supervisors

Professor Kay W. Axhausen	*Swiss Federal Institute of Technology*
Dr. ir. Alexander R.C. de Haan	*Delft University of Technology*
Dr. ir. Rob van Nes	*Delft University of Technology*
Dipl. Wirt. Ing. Nadine Schüssler	*Swiss Federal Institute of Technology*

Submitted to the Faculty of Civil Engineering and Geosciences and the Faculty of Technology, Policy and Management in partial fulfillment of the requirements for the degree of Master of Science in Transportation, Infrastructure and Logistics at Delft University of Technology.

Keywords
Aviation, Discrete choice modeling, Airlines, Departure time modeling, Pricing, Choice Set Formation

Cover picture
Interior Swiss Airlines. Taken from the Vitra Design Museum/Swiss Airlines

Summary

Abstract

The Institute for Transport Planning and Systems (IVT) is a research institute affiliated to the ETH Zurich and is specialized in, amongst others, transportation demand forecasting. The IVT expressed its interest to apply its knowledge in the field of aviation, following a general trend in the scientific community, thereby extending and expanding its knowledge of traveler choice behavior in general and air traveler behavior specifically. From an actor and network analysis, it is derived that numerous actors influence air traveler decision-making across a range of choice stages. It is chosen to highlight the itinerary choice stage in a case study, as only two actors are directly involved, namely airlines and travel portals. Furthermore, it is possible to address prevailing issues in the discrete choice community and the aviation industry.

The focus of the case study is on understanding the relative valuation of non-monetary and monetary characteristics of an itinerary based on revealed preference data and investigating the role of different choice sets. A discrete choice modeling approach is followed.

The research presented in the case study makes use of a number of datasets: a dataset that contains tickets bookings through computer reservation systems (CRS) for November 2006, a dataset with fares observed in the period September 2006 – November 2006 for flights departing in November 2006 on 70 origin-destination pairs and the Official Airline guide (OAG). They have been combined to form a comprehensive dataset for the analysis of itinerary choice.

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

In general it is concluded that the IVT can contribute significantly to current issues in the aviation industry. The fact that actual behavior in the case study is analyzed can prove both important for convincing both practitioners and researchers of the findings.

Summary by chapter

Chapter 1 provides an overview of developments in aviation. In recent years air traveler behavior research has seen a strong increase in activity, stimulated by the aviation industry who perceive an increased need for an understanding of traveler behavior on a micro-level. Such an understanding can aid airlines with decisions involving network and schedule planning, pricing strategies and improvement of revenue management controls. Furthermore, an increased understanding of traveler needs can aid online travel agents with their competitive advantage over carrier websites by offering a complete, tailor-made overview of available itineraries. Finally, an increased understanding of traveler behavior can aid airports by providing insight in access mode and airport choice. The increasing competition between flag and low-cost carriers, simplified fare structures and the widespread availability of fare information through on-line distribution channels has led to an increased need in the understanding of the role of fare and fare product (i.e. flexibility, overnight stay) in decision-making.

Chapter 2 then continues with an analysis of the aviation system, by means of an actor analysis. Amongst the other actors included in the analysis were: airlines, airports, travel portals, high speed train operators and policy-makers. It is concluded that numerous actors influence air traveler decision-making across a range of choice stages. In a competitive and complex environment, it may be difficult for a research institute to obtain complete and objective data, a requirement for scientific work. Based on several analyses, it is chosen to highlight a single choice stage, namely itinerary choice.

Chapter 3 presents the discrete choice modeling framework used and discusses several concepts regarding choice sets and choice set formation. Several dimensions have been proposed that influence the composition of the choice set, namely the booking time dimension, the information acquisition (travel portal/ carrier) and frequency dimension, the preferred arrival time dimension and the fare dimension. Furthermore, a distinction is made between objective and subjective choice sets from a researcher's perspective and a traveler's perspective.

Chapter 4 forms the start of the case study. The focus of the case study is on understanding the relative valuation of non-monetary and monetary characteristics of an itinerary based on revealed preference data and investigating the role of different choice sets.

The case study makes use of a number of datasets: a dataset that contains tickets bookings through computer reservation systems (CRS) for November 2006, a dataset with fares observed in the period September 2006 – November 2006 for flights departing in November 2006 on 70 origin-destination pairs and the Official Airline guide (OAG). They have been combined to form a comprehensive dataset for the analysis of itinerary choice.

Chapter 5 presents an analysis of the available datasets. In the analysis, it is assumed that travelers have knowledge of all itineraries listed on Expedia on the day of booking and departing on the preferred departure date. This leads to an observed objective choice set containing a maximum of 50 itineraries, representing the available supply in the aviation network. However, if constraints would be added to the choice set generation algorithm, such as a constraint incorporating a time window of an hour, 90% of the choice sets would contain 20 alternatives or less.

Travelers returning the same day depart on weekdays, traveler returning the next day depart from Monday to Thursday. Travelers returning after 6 days do not show a clear preference for departure day. 5% of the travelers book their ticket up to 36 days in advance, 50% of the tickets are booked 8 days before departure and 85% of the tickets are booked up to 3 days in advance.

Travelers do tend to chose an itinerary with a lower fare, however do not chose an itinerary with the lowest fare available. This indicates that other, non-monetary, attributes of the itinerary play a role as well.

Travelers returning the same or next day prefer departing in the morning and returning in the evening, travelers returning after 6 days do not show a clear preference for departure day. Carrier preferences are less expressed. A preference for mainline jets and regional aircraft is observed.

Chapter 6 discusses the results of several MNL-model estimations. Fare yields a relative large contribution to utility in an average chosen itinerary. Over 50% of the utility is contributed by fare. In a typical non-chosen itinerary, a transfer is responsible for the largest part of utility. Approximately 80% of the disutility is contributed by a transfer. Travel time is important as well, but will not vary much on any given route.

Chapter 7 evaluates the effect of overlap between itineraries. This effect is evaluated by adding a measure to the utility function that captures the overlap in multi-dimensional way (i.e. fare, journey time and differences in arrival time). Estimated parameters for the independence measure are highly significant and the model performance increases. Thus, it can be shown that the independence measure captures the similarity between alternatives at least to a certain extent without increasing estimation time.
For all models the sign of the parameter indicates that passengers perceive similar alternatives as negative. This is however dependent on the specification of the utility function: in a utility function specified per departure period of day (i.e. morning, afternoon, evening), passengers perceive similar alternatives as positive.

Chapter 8 discusses the findings and implications of the case study and as such forms the conclusion of the case study. Travelers are willing to pay a premium for itineraries departing in the early morning and returning in the early evening. Yet, revenue management systems remain important, as does the estimation of the willingness-to-pay of a passenger. A simplified one-way fare structure, however, may be the best direction for the future, with a differentiation of fares per weekday, departure time and booking period. Airlines not focusing on business passengers but on leisure traffic can avoid airports during congested moments: passengers staying longer at their destination have a low preference for departure time. The results presented advocate strongly against hub-and-spoke systems, at least on intra European flights.

Chapter 9 presents a synthesis, conclusions and recommendations. Based on the synthesis, it is concluded that it is possible for the IVT to contribute substantially to prevailing issues in the scientific community and the aviation industry. It is recommended to airports, airlines and travel portals to offer a complete overview of journey costs and door-to-door travel time. Such an overview can lower transaction costs for a passengers, may prevent window-shopping and aid a traveler in his or her decision. The fact that actual behavior is analyzed can prove both important for convincing both practitioners and researchers of the findings.

Chapter 10 discusses the possibilities of air demand assignment and the development of a similarity measure in more detail.

On page 142 and 142 relevant definitions are presented. The *Appendices* provide background information to the different chapters. Where necessary, they are referred to in the main text.

- This page left blank intentionally -

Preface

The research presented in this thesis is the results of several months of work, months during which I came to enjoy research and which have been a very educational period, both personal and professional. Despite only being authored by one, this thesis could not have come to be without the help of an entire series of people, which I would like to thank in the ensuing.

Professor Axhausen, I would like to thank you for giving me the opportunity to conduct my research at the Institut für Verkehrsplanung und Transportsysteme (IVT), your guidance, your everlasting enthusiasm and getting me to appreciate the expressions 'Spannend' and 'Machen Sie Mal'. Nadine, thanks for the warm welcome at the ETH and the IVT, your help and support over the last months, and for having all the obscure papers readily available. I really enjoyed my time at the IVT, much thanks to you. From Delft University of Technology I would like to thank Professor Bovy for all his support, helping ordering my thoughts about itinerary choice, the enjoyable time in Lausanne and reminding me that we are actually modeling behavioral processes of real-life people, and that it's not all about the data. Rob, during the last two years you have been a great and patient mentor and reader. Thanks for your humor and your writing tips. I guess you'll recognize a lot or your suggestions in this thesis and I admire your ability for maintaining the overview. Last but certainly not least, Alexander thanks for giving me pleasure in policy analysis again, really, and keeping me up with me, despite a hectic and messy start.

From Swiss International Airlines I would like to thank Karl Isler and Henrik Imhof for providing the dataset containing the ticket booking through CRS systems and the webbots to observe the fares. Their contribution was central to the success of this study. In addition, I would like to thank Laurie Garrow, Stephane Hess and John Rose for their help with various issues.

I would like to express my gratitude to everybody at the Gruppe Verkehrsplanung, your hospitality and friendship and your warm welcome. I admire you all, learned from you and had a lot of fun. Thank you for the coffees, lunches, drinks, movies, barbecues and chocolate cakes. Especially, I would like thank the guys in F33.1, Claude and Alex, for helping me out with all my questions. Of course you need a reference category…. Next time I'll read the Biogeme manual and the standard works on discrete choice modeling before dropping by. Also, I would like to thank Philipp with his help with interpreting results, his profound knowledge of silly facts of Dutch colonial and Austrian history and his hospitality during my last week in Zurich. Ernst Happel is the man and we rule(d) the world! Adrian, thanks for your support on the IT-side, keeping Leibnitz online, correcting my German pronunciation, and the unforgettable discussions about Swiss politics. Ruth, next time I expect to be picked up from the airport just as every other student. No, thanks for all your help! Michael and Andi, thanks for the numerous enjoyable dinners at holiday park Hönggerberg. I really appreciated your company.

There is an entire list of people who I need to thank for their endless hospitality and amazing flexibility during my various stays in the Netherlands and Switzerland. In no particular order, I would like to thank Tom & Rachel, Marten, Paul, Reinout, Roel P. and Luis in the Netherlands and in Switzerland Marcus (Wahn, Wahn, überall Wahn!), Alex & Lily, Philipp and Petra of the Rotary House. In addition, I would like to thank Bas and Marten for their periodic calls and friendship, heck, the boss is paying, who cares?, Roel G. for his medical advice when I fell off my bike again, Jost of the Unispital for getting me back on my bike again, San Yin for being my best friend, Ian and Reinout for the great weekends in Switzerland, Jacco of SundayAfternoon for the artwork, the entire J.C. Sinister, Arnoud, Hans, Tim, Tom and various Piefkes of the Oude Langendijk in Delft and my fellow roommates of the Rotary House and Schaffhauserstrasse in Zurich for giving me a very, very nice, interesting, educational and most important fun student time.

Finally, Mom, Dad and Anja, thanks for all your support in all possible and imaginable ways from the very beginning and especially over the last years, which is impossible to express in these two sentences. Without you, I would never have brought my studies and this exercise to a successful completion.

November 2007
Michael van Eggermond

- This page left blank intentionally -

Contents

- This page left blank intentionally -

List of Figures

List Of Tables

Part I Demarcation and Conceptualization

- This page left blank intentionally -

Chapter 1 Introduction

1.1 Background of the Research

The United States Airline Deregulation Act of 1978 marked the beginning of competition in the aviation industry. Before 1978, prices for all airlines serving the same routes were about the same and transparent. For most travelers, the outcomes of deregulation are advantageous. Deregulation has led to lower fares, safer journeys, improvement in airline productivity and to some extent a higher quality of service (Kahn, Hanlon 1999).

In the European Union, liberalization came in three packages of measures, agreed in 1987, 1990 and 1992 (Dodgson 1994). The opening of access to routes, full unrestricted cabotage rights, applies since January 1997. The implications for the European market are expected to be same as the ones observed in the United States (Good*, et al.* 1993).

To illustrate the sheer size of the aviation market: in 2005 the total number of passengers transported by air in the EU25 rose by 8.5%, compared to 2004, to more than 700 million. Passenger numbers rose by 8.8% in 2004 and by 4.9% in 2003. Of these passengers, 23% were carried on national flights, 42% on intra-EU25 flights and 35% on extra-EU25 flights (Eurostat 2007).

In both the United States and Europe deregulation has led to more competition between traditional airlines. Furthermore, traditional airlines were faced with a new type of competition, namely low cost carriers. In the U.S. the best example of a low-cost carrier is Southwest Airlines, which operates since 1967. Ryanair (1985) and EasyJet (1995) were the first low-cost carriers to enter the European market. Besides the increased competition between different airlines, carriers in Europe are faced with increasing competition from high speed trains, as trans-national networks are on the verge of completion and offer a competitive service between a number of city pairs (Wittwer 2007). In the airline industry, given the current competition levels, there exists an increased need for improving the understanding of travelers' choices on a micro-level (van Ryzin 2005, Westermann 2006).

In the model of competing airlines, competition leads to concern about costs both within airlines and service providers to airlines, including airports and services at airports, as these also offer the possibility to realize cost reductions. Airlines competing on costs may choose to cut on inline flight services, seat reservations, airport lounges and airport costs (Barrett 2000). Airlines competing on quality may choose airports with higher quality access & egress and high slot availability. Carriers only view passengers as their customer group, whereas airports both regard airlines and passengers as their key customers (Graham 2001). The unregulated market may lead to an increasing volatility in airport traffic demand, and future airport traffic volumes and traffic composition may be increasingly difficult to predict. Instead of the supply driven approach, as is the case with traditional airport planning, airports should focus more on flexible, demand driven, planning (de Neufville & Odoni 2003, Burghouwt 2007).

From the above it can be concluded that interest exists in understanding consumer behavior on a microscopic level and for flexible demand forecasting in the aviation industry. Demand forecasting is not something new; it is essential in the analysis of transportation systems and is concerned with the behavior of consumers of transportation services and facilities (Ben-Akiva & Lerman 1985, Train 2003) and has its roots in microeconomic theory. Reliable demand forecasts can aid policy makers, and (non)profit organizations with their medium and long-term planning of infrastructure, space allocation and transportation services. A major breakthrough in the field of demand forecasting was the development of disaggregate travel demand models, also known as discrete choice models which analyze the behavior of a decision-maker (i.e. individual, household, organization) on a microscopic level (McFadden 1974, McFadden 1978). Traditionally, studies carried out by researchers and practitioners have focused on transportation related areas, e.g. individual & public transport and location choice (Ben-Akiva & Lerman 1985, Fox*, et al.* 2003, Chorus 2007). Disaggregate models are also being applied in other fields, such as marketing (Erdem & Swait 2004).

Compared to the number of studies that consider urban and regional situations, the number of research projects and applications carried out on long-distance travel, trips more than 100 kilometers from home, has been limited using the discrete choice methods. Only 1% of all trips fall in this category; however, these trips are of economic importance (holiday and business industry) and represent 20-25% of the total passenger kilometers made. Car and plane are the two dominant modes, with a market share each of approximately 30% (Hubert & Potier 2003). For these reasons, this area may be of great interest for research institutes which focus on transportation planning: besides being an attractive, challenging and relatively uncultivated field of research, the aviation system offers possibilities for collaboration with both governmental institutions and the airline industry.

From the aforementioned, it may have become clear that a need exists to extend and explore the advantages of disaggregate transportation demand forecasting in the aviation sector. As mentioned, various studies have already been conducted in this field. More specifically, combined airport and airline modeling has been addressed in the San Francisco Bay area (Hess & Polak 2005) and more recently in the Greater London area (Hess & Polak 2006b). These studies confirm the forecast capability of discrete choice models. Research towards itinerary choice modeling has been carried out in the United States (Proussaloglou & Koppelman 1999, Coldren*, et al.* 2003, Coldren & Koppelman 2005). These studies try to capture the complexity of behavioral processes in the aviation sector, as it involves choices along a multitude of dimensions (booking time, departure time, access mode choice, carrier choice), by implementing advanced model structures.

In addition to adopting innovative model techniques and contributing to the understanding of decision-making and science, it is also necessary to address the applicability of disaggregate demand models in the aviation industry. As already discussed, the power of disaggregate demand models lies in the fact that they model and aim to predict the behavior of the decision-maker. By putting the results of transportation demand studies in a broader context: by showing the concerned actor(s) what the possibilities and opportunities of disaggregate forecasting are and especially focusing on aspects that the concerned actor(s) can influence directly or indirectly, the results of a transportation demand study are much more valuable to relevant actors. This especially holds in an environment where the applied technique is relatively new, such as the aviation industry.

1.2 Problem owner: a transportation research institute

As problem owner, a transportation research institute, the Institute for Transport Planning and Systems (*Institut für Verkehrsplanung und Transportsysteme, IVT)* is chosen. The institute belongs to the civil engineering department of the ETH Zurich, which is one of sixteen departments of the ETH Zurich. The activities of the institute concern education, research & consulting and could be said to coincide with the mission of ETH Zurich in general:

"The ETH Zurich imparts to its students the highest state of knowledge and practical skills… The ETH Zurich is not content with mere participation in solving already known problems […]. In doing so, it depends on the spirit of discovery, innovative force, and flexibility in its members." (Executive Board of the ETH Zurich 1996).

Although the ETH Zurich financially supports the research carried out at the departments, it also encourages industrial collaboration:

"The ETH Zurich encourages partnerships and interdisciplinary co-operation among members of its community, with other educational and research institutions, with industry, and with the public administration, and it believes in keeping the public informed regarding these activities […]. It endeavors to gain additional financial support, beyond the allotted public funds, from industry and private sources" (Executive Board of the ETH Zurich 1996).

If the aforementioned would be translated to the objectives specific to the transport and spatial planning group the main objective could be said to be *continuity* of the group. Goals that can help realize this goal are *innovative & relevant research, interesting education opportunities* and *industry collaboration.*

Education given by the group includes introduction to transportation planning, evaluation of transport planning and an introduction to choice modeling. Recent research and consulting projects of the group include the evaluation of the effect of road pricing (mobility pricing), large-scale multi-agent simulation of travel behavior, social networks and travel behavior and the simulation of land use and transportation.
Modeling choice behavior is an essential ingredient to all of these projects. These can be choices regarding mode, departure time, destination and route. More recently, discrete choice modeling has been applied in a field new to the IVT (Frick & Meister 2006), using data provided by Swiss International Airlines, following a general trend in the discrete choice modeling community towards the application of choice modeling in the aviation sector (Proussaloglou & Koppelman 1999, Coldren, *et al.* 2003, Coldren & Koppelman 2005, Garrow, *et al.* 2007). Other studies of the IVT in the field of aviation have paid attention to dynamic pricing (Kisselef 2006) and demand distribution in Europe (Erath 2004, Hackney 2005).

As Professor K.W. Axhausen, head of the transport and spatial planning group, puts it: *itinerary modeling is an interesting, exciting and challenging application of discrete choice theory. It offers the opportunity to analyze the role of similar alternatives in decision-making. Furthermore it's an* *interesting topic for students as either a bachelor or master project.* He also puts itinerary choice modeling in a broader context and sees itinerary models as a step towards a European travel demand model.

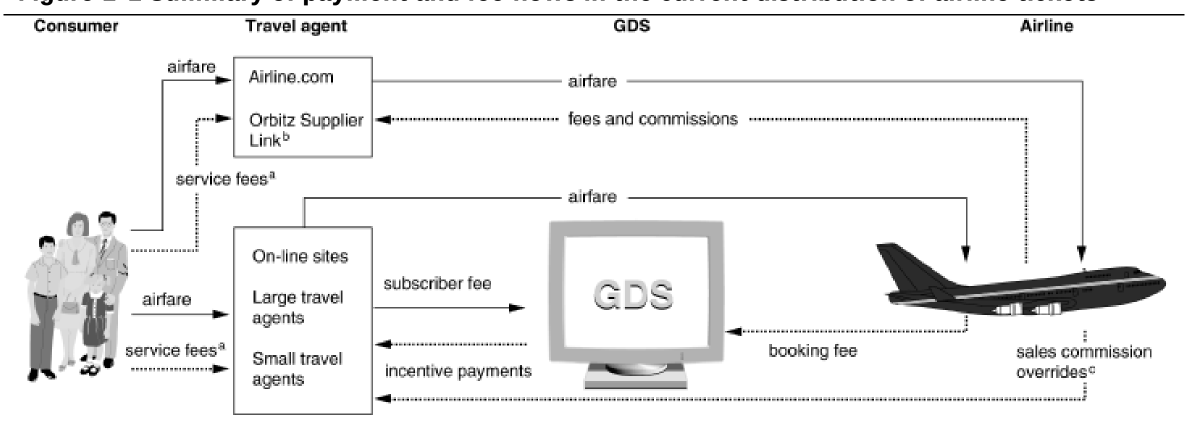
The problem perceived by the problem owner follows from preceding discussion. The IVT possesses thorough knowledge of choice modeling in a transportation context. To further develop and expand this knowledge in general and in aviation specific, resources of industry are needed, preferably data and/or funds. However, actors and stakeholders are reluctant to adopt relatively new approaches to understanding traveler choice behavior and share their resources for competitive reasons and therefore less willing to provide these resources.

The starting point of this research is considered to be that the aviation system provides a challenging research field and that by putting the knowledge of a transportation research institute with regard to disaggregate forecasting in a broader context and showing actors possibilities and opportunities relevant to them, more support for research in the aviation sector in Europe can be found.

1.3 Actors, Stakeholders and the Aviation System

From the discussion in the previous section, it may have become clear that in addition to the IVT, various actors and stakeholders play a role in the aviation system. Enserink (2002) provides a set of guidelines to recognize relevant actors and stakeholders. These guidelines regard to which extent actors are effectively involved in the system at stake, what actors influence the outcomes of the system, what actors have the necessary resources to influence the outcomes, which actors are actually willing to help and which actors will be influenced by the problem.
For instance, by following the payment and fee flows in the distribution of airline tickets (Figure 1-1), a series of relevant actors can be recognized. The abbreviation GDS stands for global distribution systems, systems that assist travel agents and smaller airlines with ticket bookings and seating availability.

Figure 1-1 Summary of payment and fee flows in the current distribution of airline tickets



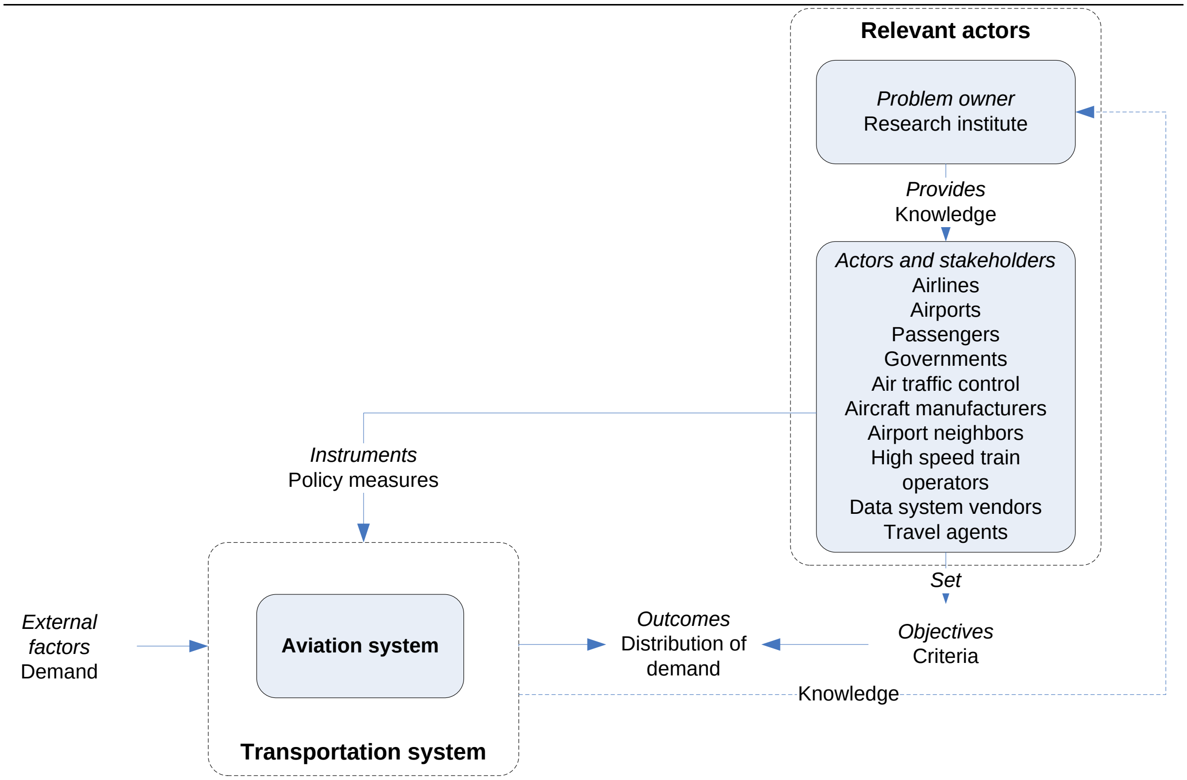
Source: United States General Accounting Office (2003)

A starting point for a discussion of the aviation system can be a system diagram. Such a diagram provides an overview of the problem owner, other actors and stakeholders, the system to be analyzed, policy measures, external influences and outcomes at a quick glance. In addition, a system diagram also helps to define system boundaries. Bots (2002) provides an extensive discussion of system diagrams. A highly aggregated system diagram is presented in Figure 1-2.

The list of actors and stakeholders presented in Figure 1-2 is constructed by using the aforementioned guidelines and contains the following actors: airlines, airports, passengers, governments, air traffic control, aircraft manufacturers, airport neighbors, high speed train operators and (online) travel agents.

Most noticeable is that the problem owner does not directly influence the system. The problem owner can provide solutions to actors that can influence the system. Knowledge of the system is required in order to be able to provide these solutions. Further more, two systems are visualized. This is done to indicate that the aviation system is part of a larger transportation system.

Figure 1-2 System diagram



In this research, the main focus lies on the aviation system, on the solutions a transportation research institute can provide to relevant actors and the required knowledge of the aviation system. The aviation system is also considered to be a system boundary. Where necessary, the link to the transportation system in general will be discussed.

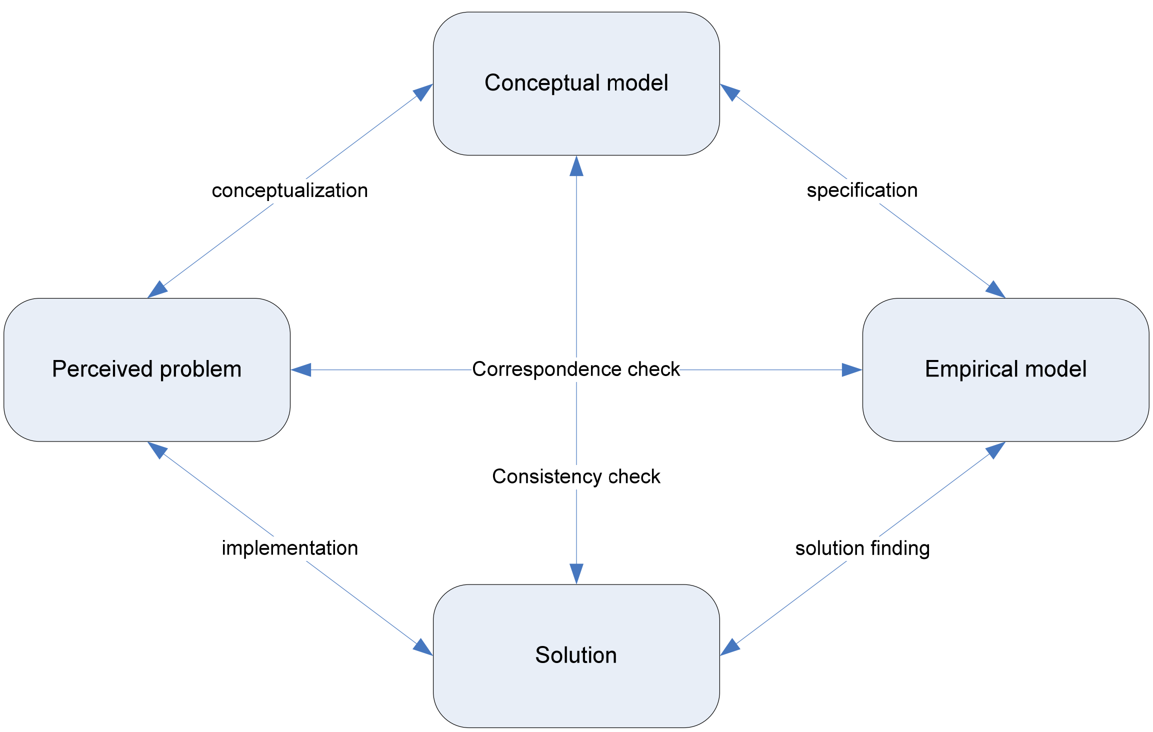
In the first paragraph, special attention was paid to airlines, airports and passengers. In addition, travel portals are considered important as these provide an interface to the traveler. This focus will be kept throughout the research and is considered to be one of the system boundaries. An outlook to where the research could be extended for the other actors will be given in the final chapter.

1.4 Problem solving process

In Figure 1-3, a systems view of problem solving is shown. The four elements which form the different stages of the problem solving process are highlighted in the blocks; the arrows emphasize the different activities.

The four stages of problem solving are the perceived problem, the conceptual model, in which the variables, that specify the nature of the problem in broad terms, are defined, the empirical model in which the conceptual model is further specified, and the solution. The activities are conceptualization, specification, solution finding, implementation, correspondence check and consistency check.
Perhaps the most important fact, and yet subtle fact presented in the figure is that, while all model cycles have a beginning, not every beginning is the same (Jacobs 2005).

Figure 1-3 A systems view of problem solving



Source: Mitroff *et al.* (1974)

As could be read in the first paragraph and in the research objectives, a problem is perceived, on the one hand by actors in the aviation sector. On the other hand, a problem can be recognized with regard to the application of discrete choice models in the aviation system. Furthermore, several conceptual models are available, such as models regarding decision-making (e.g. man as a utility maximizing agent, discrete choice models, models of decision-making). However, the perceived problem needs to be translated towards a conceptual model and the conceptual model needs to be specified as an empirical model.

1.5 Research Objectives

From the previous sections, several arenas can be recognized in which the IVT operates. On the one hand, the IVT is dedicated to research and the subsequent contribution to science. On the other hand, the continuation of research requires contribution of third parties, either in data, funds or knowledge in general. The research objectives take into account both the scientific environment and industry. Furthermore, the internal perspective of the IVT is relevant. From these three arena's the objectives which the research should meet are derived:

1. *From an internal perspective:*
    a. The research should form a continuation of available knowledge within the IVT;
    b. The research should deliver a proof of concept for current and potential industry partners of the IVT;
2. *From a scientific perspective:*
    a. The research should contribute to current research in the aviation system;
    b. The research should incorporate active and prevailing issues in the scientific community;
3. *From an industry perspective:*
    a. The research should address current issues in aviation;
    b. The research should take into account objectives and instruments of relevant actors.

For a research institute specialized in transportation planning and research, the first steps in a new application field should go hand in hand with a clear view of *what is possible, what is demanded for by whom and what is required for further research.* Such an approach may lead to more intensive industry collaboration and may make the transition from academic research towards aviation planning practice shorter.

1.6 Report Outline

The remainder of this thesis is split into three parts. These parts are not in the same order as the research objectives, however do reflect them.

Part I of the thesis consists of three chapters, which serve as a demarcation and conceptualization of relevant concepts for this research. This section forms the conclusion of the first chapter of Part 1 and provided the background of this research, the choice of problem owner and an overview of relevant actors and stakeholders
Chapter 2 will get into detail on the aviation system, discussing the objectives and instruments of the different actors, thereby providing an extensive literature review reflecting current issues in the aviation industry that have common ground with the activities of the IVT. The chapter will conclude with an assessment of state-of-the-art research in the light of traveler decision-making and industry issues. These two combined lead to a demarcation of the research and a set of specific research questions, reflecting current issues in aviation. *Chapter 3* presents a theoretical background on

decision-making and choice set formation and concludes with choice set terminology relevant for itinerary choice modeling. Chapter 2 and Chapter 3 lead to a set of requirements for data necessary to answer the research question, taken into account the discrete choice framework.

Part II of the thesis will present a demonstration, or case study, thereby focusing on itinerary choice modeling, taking into account the role of similarities among itineraries.
Chapter 4 presents datasets meeting the requirements for the data necessary to the case study and provides a brief description of the contents and steps necessary to make the suitable for the research questions. *Chapter 5* provides an analysis of the descriptive statistics of the generated choice sets and gives insight into preferences and preference structures in chosen and non-chosen itineraries. *Chapter 6* presents a MNL-model of itinerary choice, which shows the relative valuation of non-monetary attributes or service characteristics of an itinerary and fare. *Chapter 7* addresses the problem of correlation between alternatives, provides an overview of possible ways to account for this correlation and gives a demonstration of a measure that explicitly takes into account the multi-dimensional nature of an itinerary choice. *Chapter 8* forms the conclusion of Part II of the thesis and provides a discussion of the findings of the case study and the implications for airlines, airports, travel portals and the IVT.

Part III forms the conclusion of the thesis. As such, conclusions and recommendations will be presented in *Chapter 9.* These conclusion and recommendations are presented as a synthesis between Part I and Part II. *Chapter 10* discusses tangible recommendation in the form of roadmaps. These roadmaps highlight the steps necessary to realize the recommendations.
On page 142 and 142 a graphical representation of relevant definitions and definitions can be found.
Page 154 and further contain the appendices. The appendices form a supplement to the different chapters and contain background information on the various topics discussed in this thesis. Where necessary, they are referred to in the main text.

Chapter 2 Actor analysis

2.1 Introduction

In the previous chapter a system diagram of the aviation system was presented in Figure 1-2. This system diagram contained, amongst others, an overview of actors and stakeholders. These were airlines, airports, passengers, governments, air traffic control, aircraft manufacturers, airport neighbors, competing modes and travel agents.

In the ensuing sections these actors and stakeholders will be discussed, starting with travelers. Per actor, the following outline is followed: first, the objectives of the actor will be discussed, and then a more in-depth overview of the actor will be given with a focus on forecasting and the actor's instruments. This approach makes it possible to present a structured literature overview. Finally, an overview of the actor network will be presented and will help to identify relations, interaction and issues between the different actors. The goal is to identify possible areas of collaboration between actors and the IVT and what the expectations and needs are in the case of collaboration.

Chapter 3 will focus on modeling choice behavior in an aviation setting. Together with the concepts presented in this chapter and the available datasets, a case study will be formulated in Part II of the thesis.

2.2 Concepts of Traveler Decision-Making

2.2.1 Objectives

It is assumed that the primary objective of an air traveler is to maximize the utility of his or her total trip from origin to destination. The origin of the trip can for instance be home or work, the destination is the location where an activity takes place, such as a meeting, or leisure. The motivation of a traveler to undertake the trip stems from the fact that the perceived utility at the activity end of the trip (e.g. leisure, business) minus the disutility of the trip, resulting in the net-utility, is higher than the net-utility of not taking part in the activity or the net-utility of other possible destinations and corresponding activities.

The trip utility can be described with a generalized cost function. The generalized cost function is a function of, amongst others, travel costs, travel time, waiting time, number of transfers, comfort, reliability and transaction costs (i.e. the costs for the transaction, including search time). With regard to public transport in general and air transport in particular, this function can be extended to include access/egress of the airport, handling at the airport , number of destinations and frequency of the different air services. Each traveler will make a different trade-off between the different components of the generalized cost function. It is common to make a classification of different types of travelers.

In reaction to the stringent behavioral assumptions behind the presented perspective of an individual being a utility-maximizing agent, a class of theories describing human decision-making has emerged in economics and psychology, grouped under the name behavioral economics. In this group of theories, it is assumed that individuals use simple heuristics or make mistakes when making choices, instead of making complex trade-offs and applying optimization procedures. A well-known stream of theories within this group was introduced by Simon (1955), who asserts that bounds to rationality exist and that individuals apply simple heuristics, in order to reach a satisfactory decision at low decision making costs. A more extensive discussion of this stream and other streams of theories in behavioral economics can for instance be found in Chorus (2007).

A brief overview of common traveler classification will be given in the next section. Following the discussion of possible traveler segmentation, different choice stages will be discussed in section 2.2.3.

2.2.2 Traveler segmentation

Each traveler will have a different trade-off between the different components of the generalized cost function. In this section, several types of traveler classifications will be discussed.

First, a distinction can be made with regard to the *payment method* for the trip. For instance, Garrow *et al.* (2007) make a distinction between leisure/self-paid business travelers and reimbursed business travelers. Garrow shows that the latter category revealed a lower sensitivity to price.
A second criterion is the *trip purpose*, which is closely related to the aforementioned. Here the categories can be short-notice business trips, long-notice business trips (conferences, seminars) and leisure trips (holiday), where leisure can include trips that have a social aspect, such as a family visit.
Both these categorizations can be extended with an extra dimension, *the frequency of the traveling*. This extra dimension has a twofold implication for air transportation service providers. On the one hand, travelers flying frequently may be more interested in loyalty schemes of airlines, such as frequent flyer programs. This is also called the repurchase intent of customers and is researched in marketing departments. On the other hand, frequent travelers have a higher probability of knowing the ins and outs of the system, and thus are well aware of the availability of itineraries matching their preferences.
By adding the monetary value (the share of profits a traveler group accounts for) of a passenger Boland *et al.* (2002) obtain 16 possible passenger segments, varying from corporate masses (4 trips per year) to global stars (12 trips per year). They argue that such an extensive segmentation is necessary for marketing reasons.
Finally, it is possible to make a distinction regarding the *decision-maker*. It is possible that the traveler does not make the booking choice himself, but that a third person is responsible for the decision-maker. In a leisure situation this could be the 'most' experienced traveler in for instance a family, whereas business travelers may make use of a corporate travel agency.

2.2.3 Elements of decision-making

Before or during a trip, journey or sequence of activities, individuals make several choices with regard to their journey. Amongst these choices are destination choice, mode choice, departure time choice and route choice. These considerations can be made sequential or simultaneous. As individuals seek to maximize their perceived utility, the result of the deliberation is a trip that is assumed to have the highest net utility.

In this section, a conceptual framework provided by van Zuylen (2005) is used to illustrate these different choices and their relation to time.

In Figure 2-1, the freedom of choice with respect to mode choice, route choice and departure time is depicted. The upper schemata shows the case of an airline ticket, the lower figure shows the case of the trip to the airport.

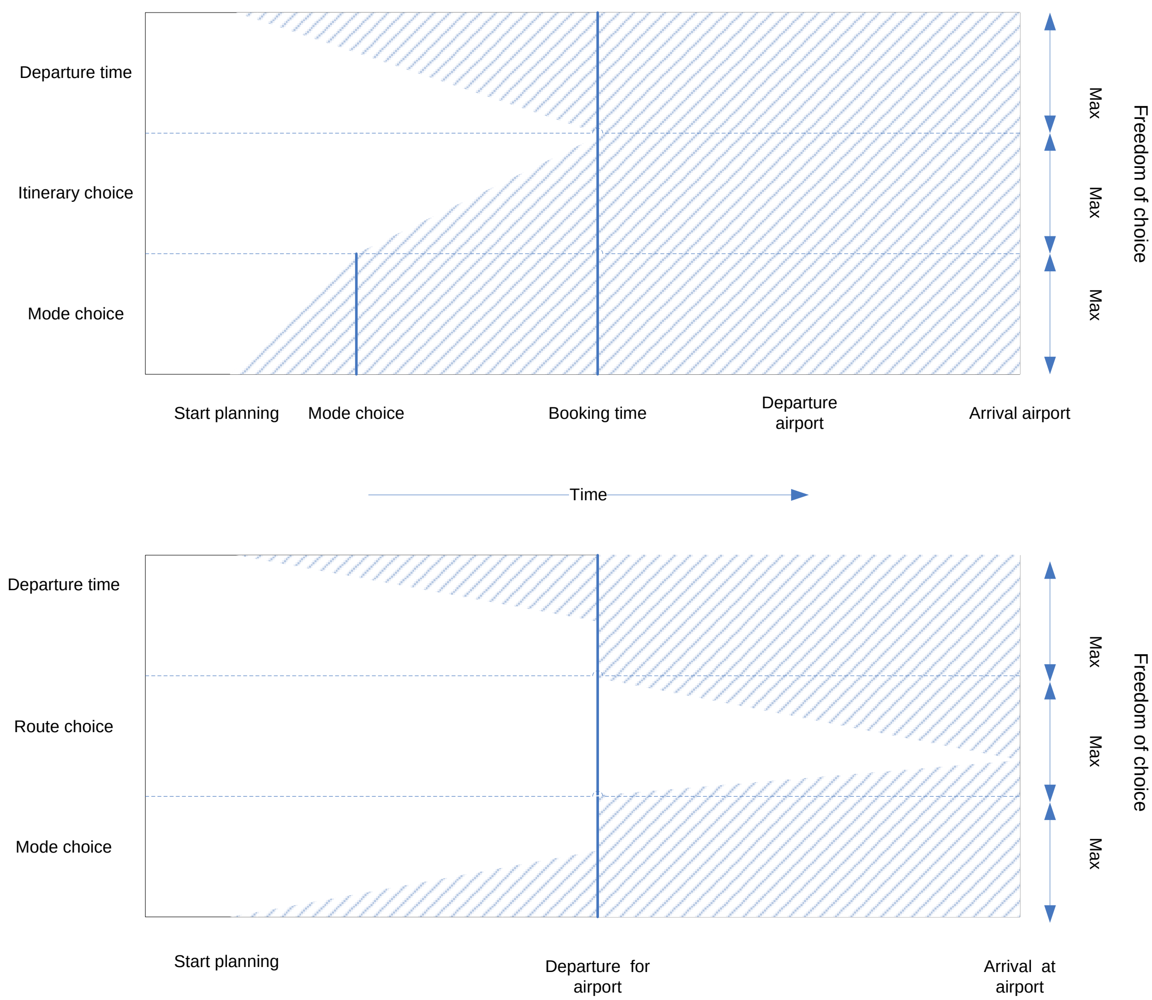
At a certain moment in time, an individual starts planning his journey. At the beginning of the planning period, all known modes, itineraries and departure times will be considered. As time passes by, several modes will no longer be under consideration any more and not all departure times will be considered. This leads to a smaller solution space. Once the mode is chosen, itinerary choice and departure time choice still have to be made. In this case, the individual chooses to travel by plane. Over time, certain alternatives will not be offered because they are not available anymore and the individual chooses for a combination of itinerary and departure time. An individual is only aware of itineraries after obtaining information. For the moment, it is assumed that a traveler obtains information through an information portal (e.g. internet site, travel agent). The decision cannot be changed anymore until the moment of departure. The only option left are either cancelling their flight or rebooking. This will depend on the restitution offered by the airline and the ticket chosen.

The lower figure depicts an individual undertaking a journey by public and/or private transport and depicts the access to the airport. As compared to the top figure, two main differences can be observed. First, the booking time is absent. Second, route choice options still exist until arrival at the airport.

Again, an individual will start planning his trip at a certain moment. At the moment of planning, the choice exists between an 8:00 public transport trip and an 8:15 individual transport trip. At 8:10, the former alternative will not be available alternative anymore, thus the mode choice is constrained. At 8:15, the individual departs for the airport. Several route alternatives remain, e.g. the high way or a rural road. Coming closer to the airport, the number of alternative routes reduces until only one alternative remains.

In the aforementioned, two series of decisions were described. The first series considered the decision for an airline ticket, the second for the route to the airport. However, as already stated in the first paragraph of this section, these choices can be made simultaneous or sequential. Therefore, these are likely to be correlated with each other.

Figure 2-1 Freedom of choice as a function of time



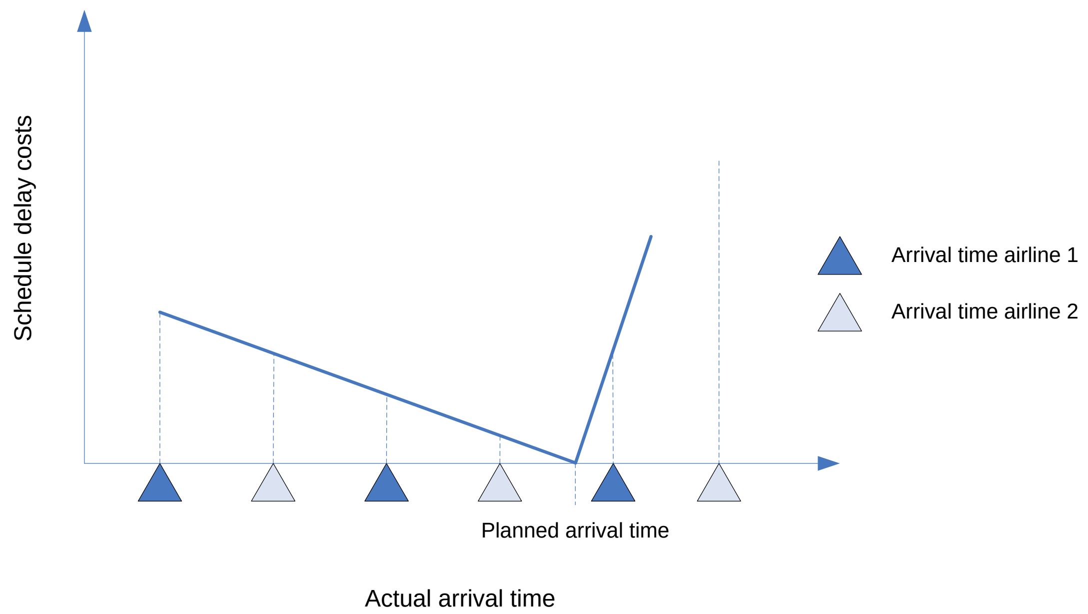
Source: Adjusted from van Zuylen (2005)

In the aforementioned, some relevant concepts and questions for this research emerge:

- All alternatives and choices known to the individual up to the moment of booking need to be known. This therefore includes choice of mode.
- Choices regarding access and egress and airport should also be known.
- The process underlying the formation of the different choice sets should be investigated further.

A constraint added to the considerations made by an individual regarding destination, mode choice, departure time and route choice, is the arrival time. An individual will normally not prefer to arrive late. Reasons can vary from having an appointment, another connection in a sequence of trips, etc. Arriving too late or too early will add to the generalized costs of the trip. Frequently, this is referred to as schedule delay cost; the delay a traveler experience because of a time table. This concept stems from public and private transport and is also visualized in Figure 2-2. It should be noted that the time tables of airlines are different from those in public transport.

Figure 2-2 Schedule delay costs as a function of arrival time



In Figure 2-2, two airlines are considered that both serve a city-pair frequently. The airline offering the lowest generalized costs will be preferred by the decision-maker. In this case, airline 2 will be preferred. This concept can further be extended to flights arriving a day earlier or later. It can also be seen in the figure that an airline offering a high frequency, has a higher probability of fulfilling customer needs. De Neufville and Odoni (2003, p. 150) highlight a number of empirical studies that show that when more than one airline competes on a route, the airline with the greater frequency, will attract more passengers than its share of the market, everything else being equal (e.g. service, fare).

2.2.4 Factors influencing travelers' choice

In the previous sections traveler segments and elements of decision-making were discussed. Different types of travelers exist which make different considerations. In Table 2-1 an overview is presented of the goals of travelers and instruments they have at their disposal in order to reach their goals. Two choice levels can be recognized: itinerary and airport. The amount of effort a traveler will put in the search of a transportation service (transaction costs) will depend on the type of product, as does the valuation of travel time, comfort, costs and deviation from preferred arrival time.

Table 2-1 Goals & instruments of air travelers

	Sub-goal		Instrument
G_TR_1	Minimize transaction costs	I_TR_1	Select information portal with highest probability of relevant information.
G_TR_2	Minimize travel time	I_TR_2	Select itinerary with short travel time
		I_TR_3	Select airport with low access and egress time
G_TR_3	Maximize comfort	I_TR_4	Select carrier with high quality standards
		I_TR_5	Select airport with high quality standards
G_TR_4	Minimize costs	I_TR_6	Select itinerary with lowest fare
		I_TR_7	Select airport with lowest access and egress costs
G_TR_5	Minimize schedule delay	I_TR_8	Select itinerary close to preferred arrival time

2.3 Airlines

2.3.1 Objectives

Objectives of airlines can perhaps best be highlighted by stating their mission. For instance, Lufthansa, the German flag carrier, states:

"The Group's overriding objective is to achieve long-term value creation with profitable growth. To that end, its efforts are directed towards positioning Lufthansa as the leading network carrier in Europe." (Lufthansa 2007)

"We operate in the market under the core Lufthansa brand and other brands. All those brands manifest our commitment to providing customers with a service noted for safety, reliability, punctuality, technical competence, quality, flexibility and innovation.
We are committed to creating sustainable value for our investors. The norms are set by the capital market. With aim at a performance level that stands as a benchmark for the European airline industry." (Lufthansa 2007)

KLM, Royal Dutch Airlines and member of the Air France/KLM group states a similar mission:

"KLM's strategic goal is profitable and sustainable growth. Together with Air France, it will achieve this through the further development of its three core activities, passenger transport, cargo transport and aircraft maintenance, in the most attractive markets, through cooperation within SkyTeam and through further reductions in unit costs."

"By striving to attain excellence as an airline and by participating in the world's most successful airline alliance, KLM intends to generate value for its customers, employees and shareholders." (KLM 2007)

Both carriers state profitable growth as their main objective. KLM clearly distinguishes sub-goals how to achieve her goal of profitable and sustainable growth, namely: development of its core activities, cooperation and cost reduction. In addition, both name a third actor, namely investors for whom they generate value. If a closer look is taken to the factors that influence share-price, occupancy rate of the aircraft (number of passengers per kilometer) is considered by analysts and consultants, as is fuel usage.

Development of core activities and cost are influenced by long- and short-term decisions an airline makes; these will be discussed in the ensuing section. Different airlines will chose a different positioning. Possible segmentations of airlines are discussed in 2.3.3. In section 2.3.4 an overview is given of different forms of cooperation between airlines.

2.3.2 Airline planning decisions

The airlines business is characterized by high fixed costs; unit costs of an airline are therefore strongly influenced by the strategic to tactical planning decisions airlines make. Ordered by long term to short term decision (or strategic to tactical), the following decision stages can be recognized (Belobaba 2006):

- *Fleet planning*, which regards the number and type of aircraft to acquire or retire. Criteria for aircraft evaluation include technical performance and characteristics, economics of operation and revenue generation, marketing and environmental issues and political and international trade concerns;
- *Route evaluation*, which regards what network structure to operate and which city-pairs to serve. Considerations include forecasts of potential demand and revenues, airline's market share of total demand and network implications for costs and revenues;
- *Schedule development*, which regards frequency planning, timetable development, fleet assignment and aircraft rotation planning. In this stage the demand per itinerary is necessary and the response of demand to a decrease or increase in service level per time period (Lohatepanont & Barnhart 2004);
- *Pricing*, which considers the products, fares and restriction for each origin-destination market. Current challenges lie within the field of price elasticity estimation and willingness-to-pay;
- *Revenue management*, how many bookings should be accepted, by type of fare to maximize the revenue of each flight and over the network. This can than also be seen as inventory control for airlines. It is estimated that revenue management systems increase revenues by 4-6% (Talluri & van Ryzin 2005). A brief overview on revenue management is included in Appendix A.1.

It should be noted, that this these different planning stages form an iterative process, where some planning stages are more closely related to each other than others (e.g. fleet planning and route evaluation or pricing and revenue management). Also it is imaginable, that some airlines will first evaluate routes and then adjust their fleet planning subsequently. It is thought that this is the case with low cost airlines.

Nevertheless, in all planning stages, forecasting demand plays an important role. The level of detail varies from aggregate (i.e. development of air demand, in general, origin-destination market) to disaggregate (i.e. origin-destination pair, leg). This is also shown in Appendix A.2. The following discussion covers literature regarding the aforementioned planning levels and air traveler decision-making. It should be noted that, that in addition to the literature stemming from scientific journals and conference proceedings, it is mostly authored or co-authored by people affiliated to major carriers, travel portals, and aircraft manufacturers. It is therefore thought to be representative of current issues in practice and science.

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

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

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

Carrier (2006) argues that previous studies have not included fare and schedule convenience on a detailed level, which ultimately influences passenger choice and sees as a potential application area of discrete choice models pricing policy and revenue management. Such a level of detail might, however, be unnecessary for strategic and tactical planning, as also argued by Grammig *et al.* (2005). Boeing for instance, offers a high and low resolution discrete choice model (Parker 2007), and apply them to different purposes and planning levels. Carrier analyzes the joint choice of an itinerary and a fare product based on past booking data in several origin destination markers. However, he does not reveal any details where the estimated models actually may be applied. Talluri and van Ryzin (2004) step into more detail and apply a simple discrete choice model to revenue management and compare it to a current revenue management method. The incorporation of a discrete choice model in revenue management algorithms lead to an increase in revenue. Important to note is that they consider an individual making a choice for a fare product on an itinerary and not a choice between itineraries, as earlier studies (Coldren*, et al.* 2003, Coldren & Koppelman 2005, Garrow*, et al.* 2007, Parker 2007) do. Talluri and van Ryzin argue that the choice for itinerary is followed by the choice for a fare-product.

A considerable part of revenue management literature covers standby and overbooking forecasting. The role of standby and overbooking forecasting is discussed in Appendix A.1. Discrete choice modeling is applied here by Garrow and Koppelman (2004a, 2004b). These studies offer a more detailed description of standby and no-show behavior, as they use disaggregated data and offer an analysis of rescheduling behavior. Ratliff and Vinod (2005) give a further overview of the incorporation of choice models in revenue management systems and see advanced forecasting methods that automatically categorize flight requests and selling channel to automatically select the most appropriate demand function as essential for obtaining maximum revenue performance. Furthermore, they discuss the necessity of adopting a restricted fare product (RFP) structure in revenue management systems in order to maintain competiveness against low-cost carriers. Westermann

(2006) also discusses the deficits of current revenue management systems and proposes a pricing-model based on the willingness-to-pay of passengers and calls this real-time dynamic pricing. Dynamic is defined as the decision of which price to display to the consumer is dynamically influenced by the availability of seats, the expected competing demand and willingness to pay, the prices of competitors, alternatives for the consumer and other relevant and observable criteria. Westermann further stresses that airlines should adjust their pricing towards the market they operate in and the network they serve.

The preceding discussion covered different airline planning levels. Accurate passenger forecasts are a requirement for all these planning levels. However, the level of detail of the different forecasts and thus the required inputs differ. Most notably, a revenue management practitioner may be interested in only modeling the choice of fare product or the choice of itinerary amongst itineraries of the same carrier, whereas for pricing the elasticity and willingness-to-pay are necessary. The latter aspects have not been addressed up to now. Schedule developers need demand levels per period of day and aircraft usage, and the effects of adding or removing flights to/from the schedule. On a strategic level, aggregate demand forecasts are more beneficial and less detailed information is necessary.

2.3.3 Airline segmentation

As already stated in the first chapter, low cost carriers pursue a strategy different from traditional carriers. This is an example of classifying an airline by her service level. In addition, several other classifications exist. Joppien (2003) offers several other criteria for segmenting airlines:

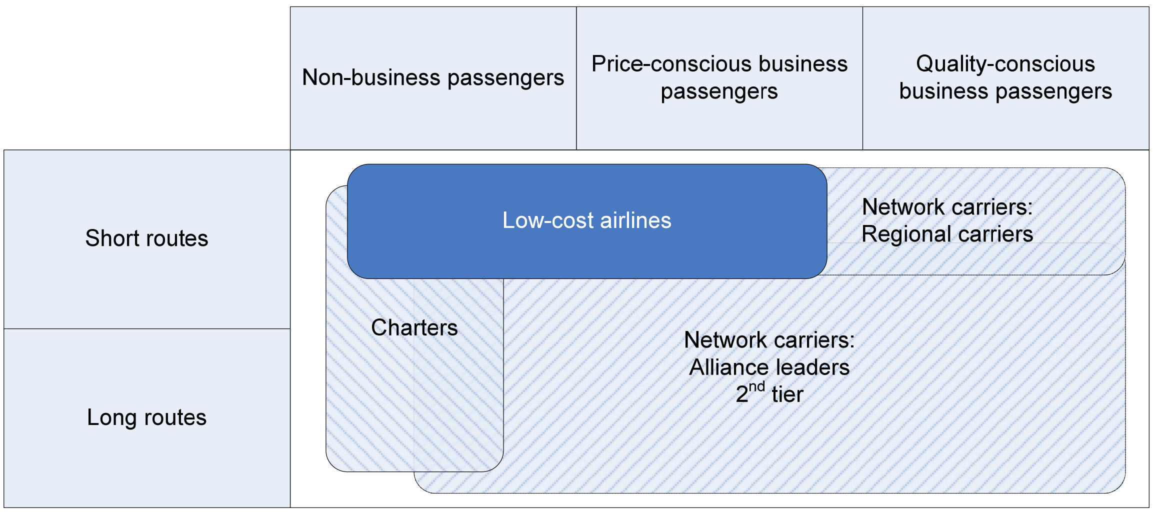
- *Size*: trunk airlines, local service carriers and commuter airlines;
- *Rights:* intercontinental, continental and national carrier;
- *Legal position:* designated carrier, non-designated carrier, point-to-point carrier and charter;
- *Ownership:* state, private or public;
- *Business focus:* passenger airline, cargo carrier, mix of passenger and cargo airline, with or without cargo fleet and an aircraft leasing company;
- *Network structure:* multi-hub airline (Lufthansa, Air France/KLM), single-hub airline, main route point-to-point airline, peripheral point-to-point airline;
- *Organization structure:* pure airline (Ryanair), multiple business segments (Lufthansa), virtual airline (airline that has other airlines operating under her name, British Airways);
- *Brand presence:* own brand airline (Lufthansa), Wet-lease airline, Franchise airline (Lufthansa Cityline)
- *Service level:* full-service airline, low frills airline, no-frills airlines;
- *Alliance membership level:* Alliance leader (Lufthansa), Alliance member (Austrian airline), Alliance free-rider (Lauda airline, an airline owned by Austrian airline)

From the classification it can be seen that some of them are based on cooperation between airlines: this will be addressed in the next section 2.3.4.

It is possible to combine some of the aforementioned classifications of airlines. A combination of network structure with service level with the earlier traveler classification (section 2.2.2) is presented in Figure 2-3.

The network structure is assumed to have two levels: long routes and short routes, service level is varied from low-cost airline to network carrier and a distinction is made between three types of passengers: non-business passengers, price-conscious business passengers and quality-conscious business passengers.
An airline might be most interested in the areas where overlap exists: this indicates competition between different types of carriers and therefore potential customer loss or gain. The entrance of low-cost airlines will further push segmentation and will lead to a sharper focus on customer segments, as there is a clear overlap between segments at the moment. Furthermore, the competition for price conscious passengers on short routes will be more severe, as most overlap between competitors exists in on short routes.

Figure 2-3 Segmentation of airlines by route and passenger type



Source: Mercer Management Consulting (2002)

Therefore, a closer look is given to the elements of low-cost carrier, which is presented in Table 2-2. Table 2-2 shows three examples of low-cost airlines. The low cost concept rests on three pillars: no frills, low operating costs and positioning.
From this overview the differences with traditional carriers become obvious. Especially Ryanair positions itself as an extreme low-cost carrier, as it offers no frills and serves only secondary airports. EasyJet offers no frills, but serves major airports. As these charge higher taxes and have higher turn-around times, the distinction with traditional airlines becomes less clear.

Table 2-2 Elements low cost airlines

Elements	Ryanair	easyJet	Virgin Expres
Simple product	- Genuine no frills offerings	- Genuine no frills offerings	- Hybrid design (charter, low cost)
Low operating costs	- Secondary airports - Homogenous fleet - Minimum cost base	- Services major airports, hence higher turnaround times and fees	- Services major airports, hence higher turnaround times and fees
Positioning	- Straight forward, aggressive low cost positioning	- Low cost position, except for major airports	- Unclear position (code share with SN Brussels airlines)

Source: Mercer Management Consulting (2002)

2.3.4 Cooperation between airlines

Cooperation between airlines mostly takes place in the form of alliances. Two types of alliances can be recognized (Park 1997): complementary alliances and parallel alliances. The main distinguishing features are that complementary alliances have non-overlapping routes, whereas parallel alliances' routes overlap (Morrish & Hamilton 2002).

Apart from the increase in network, which is often emphasized in advertisements, alliances use a number of joint services (Burton & Hanlon 1994):

- *Franchising*: this is the practice of one airline permitting another to use its name, uniforms and brand image. A common arrangement is the one between major carrier and regional carrier, where a major airline sells these privileges to regional airlines. The major carrier undertakes sales management and marketing. In return, the regional carrier pays a fee and acts as a feeder for the major carrier;
- *Blocked-spaced agreements*: under this agreement one airlines allocates to another a number of seats on some of its flights;
- *Code sharing*: this is an agreement between two airlines, under which an airline operating a service allows another airline to offer that service to the traveling public under its own flight designator. The practice is mostly used to show connecting flights as being on one airline (Hannegan & Mulvey 1995).

By customers, the latter can be perceived as an advantage and a disadvantage. Advantages for passengers may include a higher service quality and more frequent flyer points. Disadvantages for customers also exist. For instance, customers may not be able to identify the precise product they buy.

Another issue regarding code sharing practices regards fare setting. With parallel alliances, competition may decrease on certain routes, which may lead to higher fares. Complementary alliances may lead to lower fares, as efficiencies can be realized. Park (1997) conducted an econometric analysis with panel data from the years 1990-1994 and found that fare levels decreased (complementary alliances) or increased (parallel alliances) dependent on the type of alliance agreement.

Airlines perceive alliances as a strategic advantage. More specifically, four advantages can be identified: access to new markets, traffic feed into established gateways, defense of current markets, and costs and economies of scale through resource pooling. However, no evidence has been found that it is profitable for an airline to join an alliance (Hanlon 1999, Morrish & Hamilton 2002).

2.3.5 Instruments

In Table 2-3 an overview is presented of goals and sub-goals of airlines and summarizes the previous sections. All the presented goals in the left column, except *reduce costs*, regard the development of passenger transport. The instruments presented in the right column can help realize these sub-goals. Partly, these regard airline positioning, such as the reduction of airport costs and reduction of flight costs. Other the instruments stem from airline cooperation. In addition, they have their origin in scheduling and revenue management and thus the different planning decisions.

Not all airlines will be willing to reduce airport costs by serving cheaper airports, changing timings or cutting in-flight service levels. This will be dependent on the one hand of the positioning of an airline and on the other hand the sensitivity of a traveler with respect to timings, airports and in-flight levels.

Instruments such as the optimization of aircraft deployment, fleet assignment and alliance networks have a strong technical nature. However, customers' preferences for aircraft type and network layout should be taken into account.

Table 2-3 Sub-Goals & instruments of an airline

	Sub-Goals		Instruments
G_AL_1	Reduce transaction costs	I_AL_1	Increase commission costs to third parties
G_AL_2	Reduce airport costs	I_AL_2	Reduce airport costs by serving cheaper airports
		I_AL_3	Reduce airport costs by flying on off-peak hours
		I_AL_4	Reduce airport costs by cutting on airport service levels
G_AL_3	Reduce flight costs	I_AL_5	Reduce in-flight service level
		I_AL_6	Optimal fleet assignment
G_AL_4	Increase revenues from tickets	I_AL_7	Enhance revenue management systems
G_AL_5	Increase passenger volumes	I_AL_8	Marketing
		I_AL_9	Increase market presence
		I_AL_10	Increase routes
		I_AL_11	Enlarge perceived network by code-share and franchise agreements
		I_AL_12	Take-over competitors
G_AL_6	Increase utilization rate	I_AL_13	Shorten turn-around times
G_AL_7	Increase occupancy rate	I_AL_14	Optimize aircraft deployment
		I_AL_15	Optimize alliance networks

2.4 Airports

2.4.1 Objectives

The mission statements of Amsterdam Schiphol and Zurich Airport will be highlighted to clarify possible objectives airports. Schiphol has the following mission:

"The mission of Schiphol Group is to create sustainable value for its stakeholders by developing AirportCities and by positioning Amsterdam Airport Schiphol as the leading AirportCity.
Schiphol Group develops airports based on the vision that an airport should provide a perfect stopover in the travel process where the visitor is offered a unique experience. We call this the AirportCity concept." (Schiphol 2007)

The operator of Zurich Airport, Unique, has the following mission:

"We regard our main responsibility as the continued expansion of Zurich Airport in response to demand while maintaining the highest quality standards. Zurich Airport has established itself as the national and international air traffic hub for Switzerland and as an important commercial centre ." (Unique 2007a)

In the annual report of Unqiue (Unique 2007b) a clear focus is put on its business orientation: *[...]* "*focuses on increasing its corporate added value and sustainable development".* Throughout the annual report, several areas are mentioned where the increase can be realized, i.e. aviation-related activities, development of non-aviation business and real-estate management. Sustainable development primarily concerns growth within allowed noise and emission limits. Finally, it is interesting to notice that Unique sees herself in competition with other European airports.

The strong focus of Schiphol on the AirportCity-concept is remarkable. However, several authors (Graham 2001, de Neufville & Odoni 2003) note that a substantial part of airport's revenues stem from non-aviation business activities (concessions, car-parking revenues, advertising, management fees). For Schiphol, these activities result in 35% of its total revenue and 30% of its profits (Schiphol 2007).
Most important remain the revenues from aeronautical activities such as landing fees, terminal area air navigation fee, aircraft parking, airport noise charges, passenger & cargo service charges and security charges. Some of these fees are variable per period of day and day of week. The fact that these charges are variable partially stems from congestion at some airports; demand management is the term referred when airports try to reduce congestion. This will be discussed in section 2.4.3.
De Neufville and Odoni mention that airports are criticized at the fees they charge their users, especially hub airports. In 2006 Schiphol built a new terminal for low cost carriers, which has spartan facilities and is available during hours when air traffic is low. Unique, on the other hand, is not willing to differentiate between airlines and offers all airlines the same service level.

Competition between airports can exist on a regional level and, with the increased number of hub-and-spoke operations of airlines, long-distance competition is also not uncommon. In both cases, the

airport is, as Schiphol calls it, a stopover in the travel process. In a recently published report regarding the future of Schiphol (Schiphol*, et al.* 2007), Schiphol addresses the competition against other European hubs and plans to build so-called front ports, where travelers can check-in and travel to the airport by public transport. The objective is both to relieve the Randstad area from congestion and increase the accessibility of the airport.
Airports are positioned between airlines and travelers. This will be highlighted in section 2.4.4, Moreover, airports are also positioned between several levels of government. For instance, Schiphol Airport is owned by the national government (75.8%), the municipality of Rotterdam (2.4%) and the municipality of Amsterdam (21.8%) (Schiphol 2007), which leads to a complex range of interests and goals. A causal diagram presented in Appendix A.7 illustrates this.
The main goal formulated by airports sustainable growth. Sub-goals to realize this growth include 'maintain and improve competitive position', 'increase of income' and 'serve of public interest'. 'Increase of income' can be realized by increasing the income from aviation and non-aviation activities. The first sub-goal can be divided into competitive position as perceived by airlines and/or by travelers, which will be discussed in 2.4.4. In section 2.4.2, a closer look will be given to multi-airport regions. This gives a chance to address both the competition between airports and the role of access to the airport. It should be kept in mind that due to the complex environment of an airport, the relative attention paid to each topic can strongly vary over time.

2.4.2 Multi-airport regions

The San Francisco Bay Area, the greater London area, New York and Washington D.C. are well known examples of multi-airport regions, with several airports in each others vicinity. De Neufville & Odoni (2003) mention the existence of 30 multi-airport regions with over 80 airports. The definition of multi-airport region should not be narrowed down to one single city. For instance, an engineer at Philips Lighting, living in Maastricht and working in Eindhoven, both located in the south of the Netherlands, and a frequent flyer (> 3 times per month) considers Bonn, Cologne, Dusseldorf, Brussels and Eindhoven as possible departure airports, all of which lie within approximately 130 kilometers of his home. Maastricht/Aachen airport is also considered, but usually does not offer the requested flight. This simple single example indicates that the definition of a multi-airport region should not be too restrictive.
If such competition exists, airport planners need to know what passengers consider when choosing a flight (assuming the choice for flying has been made). For example, an investment in accessibility may or may not lead to increase in market share. Another example is the presence of a severely congested airport and other less congested airports. Authorities may not be willing to invest in more capacity and are interested in redistributing passengers over different airports. The latter is for instance the case in the greater London area, where the London Heathrow is congested and Stansted still has capacity remaining. It is thus in the interest of the authorities how the attractiveness of Stansted may be improved.
Several studies have been carried out concerning airport choice behavior. Bondzio (1996) conducted a study regarding airport choice in Southern-Germany and showed that travel time to the airport played an important role and that access time was more important for business passengers than for leisure travelers. Pels *et al.* (2001, 2003) analyzed the combined choice of airport and airline in the San Francisco area and the combined choice of access-mode and airport. In the first study they found that

airline choice is nested within airport choice, i.e. the competition between airlines departing from the same airport is more severe between airlines departing from different airports. In their second study they analyze the joint choice of access-mode and airport, showing high sensitivity to access time, especially for business travelers. Business passengers also consider frequency of the flight to be important. Leisure travelers consider access cost and itinerary fare more important. A case study of the London area is presented by Hess and Polak (2006b). Their study reveals that business travelers are very reluctant to accept increases in access journey times; outlying airports depend heavily on good-access connections and/or low air fares. The results of these studies show that strong differences exist between preferences of leisure and business travelers. Tron *et al.* (2007) conducted a stated-preference survey on airport choice in the Southern Ontario market. The Greater Toronto Airports Authority (GTAA) has been examining the possibility for a new airport, located 55 kilometers north-east of downtown Toronto. Most of the results of their study confirm the results of the previously discussed studies: business travelers prefer the current, more centrally located airport. In addition, business travelers tend to perceive increases in in-vehicle time to the airport as negative. The results show that the shorter the flight, the less tolerance exists towards in-vehicle time. In general, variation in the return schedule has less impact on utility than variations in the outbound schedule. This, however, does not hold for business travelers.

In addition to considering the choice of passengers, it is also important to take into account airlines: de Neufville and Odoni (2003, p. 144) discuss that airlines are mostly unwilling to switch their services to secondary airports for several reasons. Airlines are interested in a high frequency of service, and consider competition and network compatibility as important. A secondary airport that focuses on a certain segment, such as business travelers or customers of low cost airlines, may have a better chance to succeed.

2.4.3 Demand management

Demand management refers to any set of administrative or economic measures and regulations aimed at constraining the demand for access to a busy airfield and/or its modifying temporal characteristics (de Neufville & Odoni 2003) Three approaches are available: purely administrative, purely economic and hybrid approaches, which combine the previous two. The fundament of the administrative approach is a slot: an interval of time reserved for the arrival or departure of a flight. Airlines do not necessarily have to use assigned slots. Economic approaches utilize congestion pricing (Brueckner 2002, Pels & Verhoef 2004), which internalizes external costs. A hybrid approach would consist of the assignment of slots and congestion pricing, where the landing fees would be published prior to the slot assignment. These slots could then be auctioned. Without getting into detail, a potential application of the itinerary choice models of Coldren *et al.* (Coldren*, et al.* 2003, Coldren & Koppelman 2005) can be seen here. Coldren showed that the choice probabilities of different itineraries, holding all other attributes constant, differed by time-of-day. For airlines such models can help in determining their willingness-to-pay for a certain slot and airports with the differentiation of their fees. However, airports are restrained by regulation with regard to slot concessions and fees (Appendix A.4); if a slot was occupied for 80% by an airline of the days, an airport cannot deny the slot rights for the following year.

2.4.4 Factors affecting the choice of airport

From the previous sections, it may have become clear that airports are positioned between passengers and airlines. Therefore, an overview is given in Table 2-4 of factors influencing the choice for airports by travelers and airlines. This overview is taken from Graham (2001). The terms in italic are factors an airport can more or less directly influence. Factors as destinations of flight, flight availability, frequency of service and flight abilities & timings depend on the willingness of airlines to offer these services. However, an airport is in the position to accommodate and allocate these services in the case of scarce capacity.

Table 2-4 Factors affecting the choice of airport

Passengers	Airlines
Destinations of flight	*Slot availability*
Image of airport	Network compatibility
Flight fares	*Airport fees and availability of discounts*
Frequency of service	*Other airport costs*
Flight availability and timings	*Competition*
Services by type of passengers	*Marketing support*
Airline alliance policy and frequent flyer programs	*Range and quality of facilities*
Image and reliability of the airline	*Ease of transfer and connections*
Range and quality of shops, catering and other commercial facilities	*Maintenance facilities*
Surface access and cost and ease of access to airport/car parking costs	Environmental restrictions

Source: Graham (2001, p. 184)

2.4.5 Goals and Instruments Airports

Table 2-6 presents an overview of sub-goals and instruments of airports, thereby summarizing the previous sections. It can be argued that overlap exists between the instruments of 'maintain competitive position' and 'increase passenger volumes'. It is chosen to formulate these separately, as airports can compete on quality and/or quantity (Porter 1985).

Table 2-5 Sub-goals and instruments airports

	Sub-goals		Instruments
G_AP_1	Maintain and improve competitive position	I_AP_1	Improve access by public transport
		I_AP_2	Improve access by private transport
		I_AP_3	Increase number of airlines operating on airport
G_AP_2	Increase passenger volumes	I_AP_4	Increase number of destinations
		I_AP_5	Increase frequency of flights
		I_AP_6	Decrease airport costs
		I_AP_7	Improve ease of transfer and connections
		I_AP_8	Improve airport image
G_AP_3	Increase income from non-aviation activities	I_AP_9	Improve range and quality of shops, catering and other commercial facilities

2.5 Travel portals

2.5.1 Objectives

Due to the increased possibilities of information technology and the increased usage of internet-based services by, amongst other, travelers, online travel portals have seen a tremendous growth in popularity over the last decade. For instance, Expedia Inc. showed a growth of 10% in 2006 (Expedia 2007). The possible objectives of travel portals will be discussed by two examples, Expedia and Orbitz.

Expedia sees its mission as:
"Expedia, Inc.'s mission is to get the world going by building the world's largest and most intelligent travel marketplace…. Expedia, Inc. plays a leading role in facilitating travel, whether for business or for pleasure, and is committed to providing travelers with the very best resources to serve their travel needs." (Expedia 2007)

Orbitz (Orbitz 2007), a second leading travel portal, does not clearly state a mission but mentions *"offering leisure and business travelers a wide selection of low airfares, as well as deals on lodging, car rentals, cruises, vacation packages and other travel. The site was created to address consumers' need for an unbiased, comprehensive display of fares and rates in a single location.*" Orbitz is owned by Travelport, which *"delivers great content and cost savings to travelers, travel professionals and travel suppliers every day.*"

As can be seen, both companies put the traveler as the centre of attention, but also see their suppliers (hotels, airlines, car rental agencies) as important. Another aspect worth noticing is that both offer a wide range of services, addressing the needs of different user groups, such as corporate travelers, luxury travelers and travel suppliers. Markets tapped into by their search systems include hotel booking systems, car rentals and airline booking systems. These are sectors classically using some form of revenue management and use global distribution systems (GDS). Finally, both see their itinerary search systems as marketing and value adding feature.

Albeit not explicitly mentioned, profitable and sustainable growth can be seen as the main goal of travel portals. Sub-goals include maintaining and extending travel supplier relations, increasing revenue from bookings and cutting costs. Increase of its revenue from airline bookings can be realized by offering itineraries in which the traveler is interested. As discussed earlier, a traveler is interested in minimizing his generalized costs. For a travel portal, this boils down to offering a product against the price a potential customer is willing to pay.

Expedia sees this market as increasingly competitive: suppliers list their products more and more on their own websites, offering lower prices and reduced commission costs. Furthermore, downward pressure exists on commissions from suppliers. Finally, they see risks regarding information technology, such as system interruptions, proprietary rights, legal issues regarding privacy and changing search engine algorithms. This is confirmed by practitioners (Appendix A.4): travelers use travel portals for window shopping, but purchase their preferred product directly from the supplier.

Travel portals however can collect click and stream information and use this information to analyze customer choice behavior (Ratliff & Vinod 2005, Appendix A.4).

2.5.2 Instruments

Table 2-6 presents goals and instruments of travel portals. Three goals are highlighted in the left column: increase revenue from bookings, maintain & extend supplier basis and minimize costs. In the right column instruments are listed. The instruments strongly reflect the information a travel portal presents to its visitors, such as itineraries and packages. The instrument *maintain & extend suppliers in which customers are interested* partly reflects the increased demand for low cost carriers, which are not always covered by travel portals.

Table 2-6 Goals & instruments travel portals

	Goals		Instruments
G_TP_1	Increase revenue from bookings	I_TP_1	List itineraries that will be considered by the traveler; Offer better itineraries from the competitors;
		I_TP_2	Provide information that the traveler perceives as necessary for decision-making;
		I_TP_3	Increase revenue per transaction by offering package deals.
G_TP_2	Maintain and extend travel supplier base	I_TP_4	Maintain & extend suppliers in which customers are interested;
		I_TP_5	Provide value to suppliers;
		I_TP_6	Offer arrangements on favorable terms.
G_TP_3	Minimize costs	I_TP_7	Decrease costs of data processing and call centers;
		I_TP_8	Decrease costs per transaction;
		I_TP_9	Decrease inventory costs.

2.6 High Speed Train Operators, GDS-vendors, Policy-Makers and Farecasters

2.6.1 High speed train operators

In the context of air transport and airlines several competitors can be recognized. For instance private transport, on-demand aviation and video-conferencing offer a substitute for the product offered by both traditional carriers and low-cost carriers. Most notably, however, are high speed train lines. These offer a competitive service in terms of travel time between certain city pairs and offer leg-room, the ability to walk about, dining facilities and do not have extensive security checks and check-in times.
Currently, Thalys, Eurostar and SNCF have a revenue management system in place for their tickets, which involves several fare classes, as is the case with airlines. Policy-makers in some countries (Germany, Switzerland) have objected to this because railway companies, as publicly owned companies, should treat all customers equally.

Recently, the high speed rail services of Austria, Belgium, Germany, France, the Netherlands and Switzerland formed a new international marketing alliance, named Railteam, including existing international rail services such as Thalys and Eurostar. The aim is to offer passengers coordinated timetables and prices for their entire journey by the end of 2008 (Railteam 2007, The Economist 2007).

2.6.2 Global distribution system vendors

Global distribution systems (GDS) were developed by airlines in the 1960's and 1970's in order to shorten handling times, lower transaction costs and cope with the growing amounts of flight data. GDS are also known as customer reservation systems (CRS). Airlines made their GDS publicly available, while in the pre 1978 regulated market competition was not a serious issue. Over the last decades, consolidation has taken place in the GDS market and four major players remain: Sabre, Galileo, Amadeus and Worldspan. Most major airlines are affiliated to each of them.

GDS vendors sell their information technology related services to airlines, railway companies, travel agencies, airports and other travel related companies. Some GDS vendors offer a front-end, a travel portal, to their systems or affiliated to one through holding companies. The position of GDS in the booking chain is shown in Figure 1-1; GDS are positioned between airlines and travel agents. It can be seen that customers can either book at an airline directly, or through GDS systems.

2.6.3 Aircraft manufacturers

Several aircraft manufactures produce aircraft suited for passenger transport, most notably Boeing and Airbus. Other producers are Bombardier, BAE systems and Aerospatiale. In addition, aircraft are in use of manufacturers that do not exist anymore such as Fokker and McDonnell Douglas.

As these are publicly listed companies, sustainable and profitable growth can be said to be their main objective. Revenues stem from different markets, such as defense systems, satellites and commercial aircraft. In the latter case, airlines and aircraft leasing companies could be said to their main customers. As such, aircraft manufacturers want to show their primary customers the added value of a new aircraft. The added value can be expressed in terms of operation costs, but also marketing value. For this reason, Boeing for instance has a unit concerned with demand forecasting (Parker 2007). This unit can also help to give direction to new aircraft design, as for example was the case with the hub-and-spoke versus point-to-point point of view of Airbus respectively Boeing.

2.6.4 Policy-makers

In the aviation system, numerous levels of policy makers are involved, amongst which the municipalities, regional authorities, national governments and supra-national organizations, such as the European Union. In general, the main objective of all these policy-making instances could be said to be 'overall welfare'. The interpretation of this objective is likely to vary per level. For instance, the European Union may be interested in the accessibility of a region, as new infrastructure can change accessibility, affecting the attractiveness and potential development of a region. In addition, good international connections in international networks are important for the distribution of welfare.

A local government is more concerned about noise and emission levels or local employment. The role of policy-makers will remain fairly limited in this thesis.

2.6.5 Farecasters

Farecasters are relatively new to the aviation market. They try to capture the price dynamics of airline tickets and predict fares to customers. Due to increase in calculation power, cheap storage and ticket prices being published on the Web, the prediction of prices is possible (Bray 2007, Heiss 2007, Tedeschi 2007).

2.7 Actor network

In the previous sections, various actors involved in the aviation system were discussed. This gave an extensive overview of their objectives, instruments and other relevant issues. In this section, the relationship between the different actors will be discussed. An actor-instrument diagram is presented in Figure 2-4, including the previous discussed actors (spheres) and their instruments (connectors). It is a very simplified view of the actual world: airlines will interact with each other, with multiple airports and multiple customers. The same holds for customers: customers interact with multiple airlines. In addition, customers will not have information about the entire network. From the diagram can be seen that the customer, or traveler, interacts with four actors directly. For the time being, these interaction points will be called interfaces. The four interfaces are:

1. The interface customer-travel portal;
2. The interface customer-airline;
3. The interface customer-high speed train operator and;
4. The interface customer-airport.

Based on the information offered through these interfaces which is gathered and considered by a customer, a travel decision is made. The interfaces can either provide customers with information based on their needs, for instance when they make a travel query. It can also be that travelers obtain information through newspaper and television advertisements and either make travelers aware of an alternative (i.e. new routes, new services) or influence the perception of the customer of an alternative (i.e. airline or airport image). In Appendix A.5 an overview of some customer - airlines and customer - travel portal interfaces is given.

From the discussion in the previous sections, it can be distilled that it is beneficial for two actors to offer a comprehensive transport advice (i.e. door-to-door or region-to-region, see also Appendix A.6) to the traveler through these interfaces: travel portals and airports situated in multi-airport regions. In addition, airlines serving all airports in a multi-airport region may offer passengers comprehensive travel advice. This especially holds if the airports are managed by the same owner, as is the case in the Greater London area and in the Netherlands; the Schiphol Group operates the airports Amsterdam Schiphol, Rotterdam, Eindhoven and Maastricht. Therefore, it is remarkable to see that neither the websites of the airports of the Greater London Area or Schiphol offer the option to select all airports as departure point. The websites of airlines and travel portals do offer this information for the Greater London Area, but not for the Netherlands as a whole. Both airports and travel portals have nothing to lose by offering more comprehensive product to a customer. The margins for travel portals will remain the same for each ticket sold, whereas airports can either increase ticket sales or redirect traffic to secondary airports.

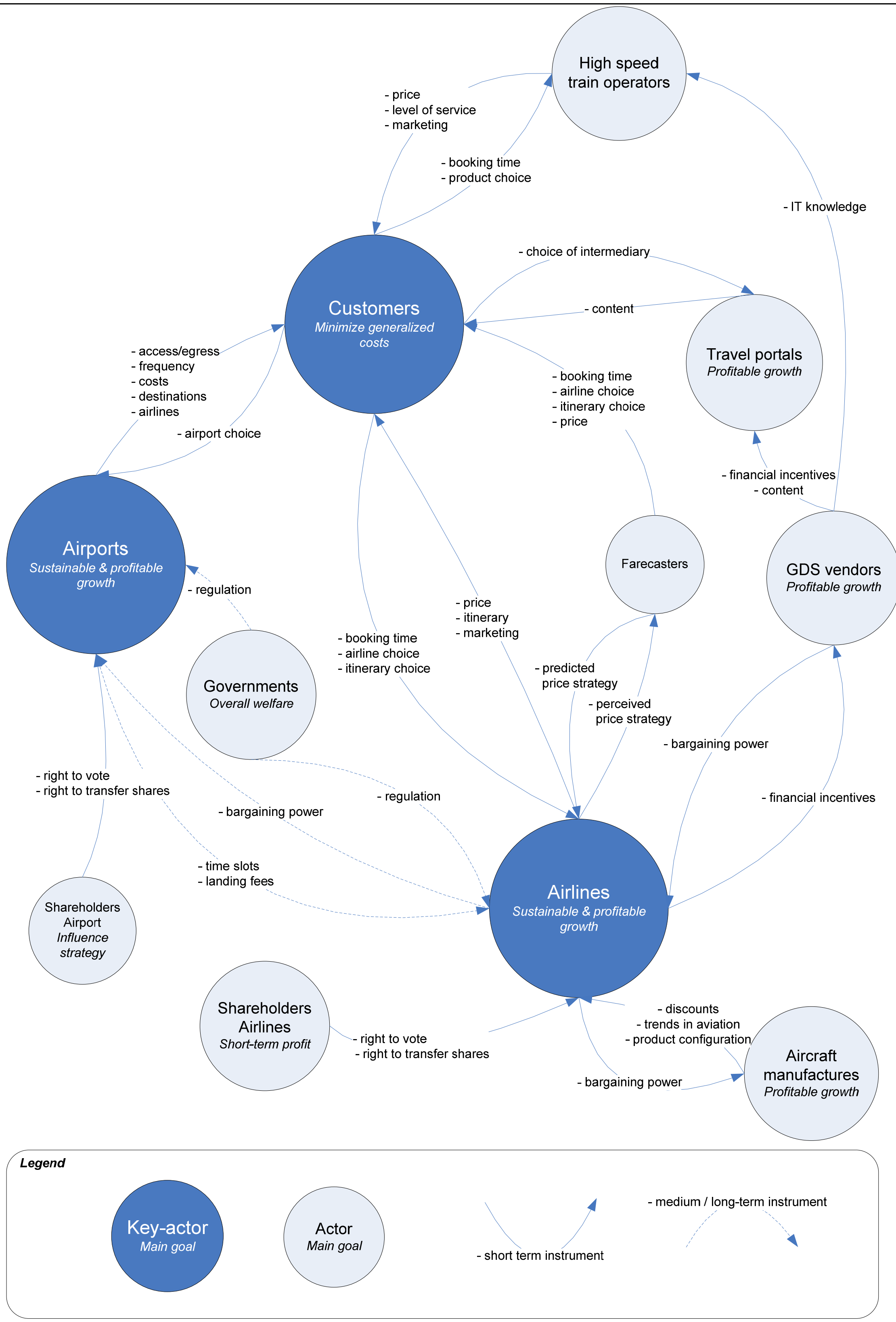
Airlines and high speed train operators are probably not willing to offer a comprehensive travel product on their websites, as they directly compete with each other for passengers. KLM has conducted a pilot project together with the Dutch railways. During this pilot project, it was possible to travel to Schiphol with an airline ticket. Lufthansa offers a joint product with the German Railways. In such a situation, the railways serve as a feeder to the airline's hub-and-spoke network. However, extensive information such as price is not available on the website of Lufthansa or the German railways.

Perhaps most striking in Figure 2-4 is the missing link between high speed train operators and travel portals, where the GDS vendors play a facilitating role. As discussed, they see airlines and hotels as their main suppliers. An opportunity may be the addition of high speed train operators to their supplier base. The American roots of the travel portals may be the reason for the lack of rail information services. On the other hand: railways traditionally have had a national focus; prices for international services are notoriously hard to obtain. Only for international joint-ventures such as Thalys, Citynightline and Eurostar fares are published on the Web. In Appendix A.4 some opinions of online travel agents are summarized. They see further bundling of products, such as a hotel and a flight, as an opportunity. However, the opportunity of offering more or a more comprehensive travel product is not mentioned. The importance of traveler segmentation was discussed in section 2.2.2. However, it is not possible at the moment to find out the trip purpose of a traveler through online booking channels, without extra information from the traveler. It may be possible to circumvent this problem in several ways: obtain booking data from corporate travel portals, either from travel portals or directly at multinationals, or decide on type of passenger through information entered in the online query. The latter option coincides with long-term opportunities as seen by travel agents (Appendix A.4).

If the diagram is viewed as dynamic, i.e. a time dimension would be added, it can be seen that the actions of one actor will influence the other and that behavior will change over time. Several studies have investigated the fare setting of airlines over time. These studies mostly use a graphical analysis to analyze fare setting, Pels and Rietveld (2004) also apply time-series analysis. They also conclude, as Button and Vega (2006) do, that there is no real price competition in the markets and dimensions they observed; each company seems to pursue its own revenue management strategy.

Finally, it would be possible to depict the relative power of the actors in the network. This is not done, as the descriptions of the various actors are given on a high level. The power of Lufthansa towards a travel portal will for instance be much higher than the power of a small carrier. The same holds for large airports and small airports.

Figure 2-4 Actor network



2.8 Three Approaches to Asses Actors and Traveler Decision-making

2.8.1 Actors' Instruments and Traveler Decision-making

This section gives an overview of the different instruments of actors and their relation to traveler decision-making in order to gain insight in the relation of actors to travelers and the influence the different actors can exercise directly on the traveler.

For each actor a number of instruments were listed. In addition, factors influencing the travel choice of a traveler were listed in section 2.2.4. The factors influencing travel choice are listed in Table 2-7 as decision attributes and can be considered necessary for an objective of the problem owner: a better understanding of traveler decision-making in aviation and a European model of travel demand. A differentiation is made between attributes regarding the traveler and the journey. Traveler characteristics include:

- *Trip purpose:* the reason for undertaking the journey, based on traveler segmentation (section 2.2.2)
- *Schedule delay sensitivity*: the preference of a traveler for a certain arrival time (section 2.2.3)
- *Booking time*: the booking time of the ticket, which on the one hand can be influenced by airlines through pricing but on the other hand is dependent on external factors, such as planned meetings.
- *Resident:* a resident is better aware of access alternative to the airport
- *Network knowledge:* the familiarity of the traveler with airports, itineraries and access modes.

Journey characteristics include:

- *Access attributes and origin airport attributes* which comprise of home-end attributes;
- *Transport attributes,* which comprise of service characteristics of the transport;
- *Egress and destination airport attributes*, which comprise of activity-end attributes.

Three actors are listed: airlines, travel portals and airports. Actors do not influence the traveler's characteristics directly, i.e. trip purpose, schedule delay sensitivity, booking time, resident and network knowledge.

From this overview it becomes apparent that the role of travel portals with regard to the actual travel product is fairly limited. However, they influence the travelers' awareness of the travel product. Public transport costs are listed as an instrument of both airports and airlines. Airlines can subsidize public transport; airports can provide easy access and egress. Airlines and airports both have instruments to influence the comfort and waiting time at an airport. Collaboration between airports and airlines already is common here: airports offer waiting lounges for frequent flyers and alliance passengers. Schiphol recently introduced a no-frill terminal for low-cost airlines, with no sanitary facilities and kiosks (Schiphol 2007). This indicates cooperation can go two ways. Unique, operator of Zurich Airport, refuses to do so as they believe that this will influence the airport image for the worse (Unique 2007a).

Frequency and timings is another decision attribute which can be influenced by both airlines and airports. However, where it may be in the interest of an airline to provide a high frequency, an airport may be interested to offer a large set of destinations. On airports with scarce capacity, this may be come an issue. This interaction is also discussed and shown in Appendix A.7.

Table 2-7 Decision attributes and instruments

	Decision attributes	Attribute (levels)	Actor		
			Travel portal	Airline	Airport operator
Traveler characteristics	Trip purpose	Reimbursed business			
		Self-paying business			
		Leisure			
	Schedule delay sensitivity	Low/High			
	Booking time	Days in advance			
	Resident	Yes/No			
	Network knowledge	Low/High	X	X	X
Journey characteristics	Access *Home – end*	Comfort			
		Time		(X)	(X)
		Parking costs			X
		Private transport costs			
		Public transport costs		X	X
	Origin airport *Home – end*	Comfort		X	X
		Waiting time		X	X
		Image of airport			X
	Transport	Comfort		X	
		Airline image		X	
		Travel time		X	
		Equipment		X	
		Ticket fare	X	X	
		Frequency		X	X
		Timings		X	X
		Transfer point		X	
		Airport costs			X
	Destination airport *Activity - end*	Image of airport			X
		Waiting time			X
	Egress *Activity - end*	Comfort			
		Time			(X)
		Costs			

2.8.2 Traveler Decision-Making and Contributions of Science

In this section the listing of decision attributes will be combined with the literature overview given in this chapter. The goal is to assess the current state of research and identify possible contributions for the IVT to current research. The overview is presented in Table 2-8 and excludes mode choice (i.e. high speed train, airline, car). Furthermore, destination airport and egress is not included as no literature was found on this topic. This will be discussed further in this section.

Most studies consider a set of the attributes relevant to decision-making and not consider the entire journey, i.e. door-to-door travel. Most notably, the studies cover three main topics:

- *Access mode and airport choice studies*, with itinerary characteristics included in the airport attributes, e.g. frequency, airline;
- Studies of *itinerary choice models* excluding fare, for the application in network planning models;
- Studies of *fare product choice*, for the application in revenue management.

Not shown in the table is the fact that studies including fare are studies based on stated preference data (Theis*, et al.* 2006, Garrow*, et al.* 2007). These studies make it possible to determine the monetary valuation of service characteristics (e.g. type of aircraft, transfer, seat pitch) of an itinerary. Revealed preference studies of itinerary choice exclude fare in the choice of itinerary, whereas studies regarding the choice for fare product do not include other flight characteristics, an exception being the study carried out by Proussaloglou and Koppelman (1999).

A contribution to current literature could then be found in:

1. The consideration of the combined choice of mode, access mode, airport and itinerary;
2. The addition of fare to itinerary based on revealed preference data and thus having the advantage that actual choices are considered;
3. The influence of network knowledge on decision-making and outcome;
4. An investigation towards destination area and airport choice. For instance, a leisure traveler can choose to go Spain by low-cost carrier and can decide between Barcelona, Madrid and Valencia. In the case of Barcelona, there are two airports, one located in the immediate vicinity, a second located approximately 80 kilometers to the north and which is served by Ryanair.

Two remarks should be made with regard to the overview in Table 2-8:

1. The overview presented is only one of many imaginable cross-sections of state-of-the-art research. Another imaginable cross-section would be by modeling approach. This is not considered appropriate in this chapter, as it is the objective to contribute to insight in decision-making, opposed to extending microeconomic decision-making theory.
2. As already mentioned, mode choice is not considered in this overview and has not been part of the literature review. For further reading in this context, for instance see González-Savignat (2004). It does however represent a current issue in aviation and will therefore be further discussed in the final chapters of this thesis as part of the recommendations.

Table 2-8 Decision attributes and state-of-the-art research

	Decision attributes	Attribute (levels)	Study
Traveler characteristics	Trip purpose	Different segmentations	(Garrow, *et al.* 2007) (Hess & Polak 2006a) (Hess & Polak 2006b) (Tron, *et al.* 2007)
	Schedule delay sensitivity	Low/High	(Koppelman, *et al.* 2007)
	Booking time	Days in advance	(Garrow, *et al.* 2007)
	Resident	Yes/No	(Hess & Polak 2006a) (Hess & Polak 2006b)
	Network knowledge	Low/High	
Journey characteristics	Access *Home – end*	Comfort	(Bondzio 1996) (Pels, *et al.* 2001) (Pels, *et al.* 2003) (Hess & Polak 2006a) (Hess & Polak 2006b) (Tron, *et al.* 2007)
		Time	
		Parking costs	
		Private transport costs	
		Public transport costs	
	Origin airport *Home – end*	Comfort	(Bondzio 1996) (Hess & Polak 2006a) (Pels, *et al.* 2001) (Pels, *et al.* 2003) (Tron, *et al.* 2007)
		Waiting time	
		Image of airport	
	Transport	Comfort	(Proussaloglou & Koppelman 1999) (Coldren, *et al.* 2003, Coldren & Koppelman 2005) (Garrow, *et al.* 2007)
		Airline image	(Proussaloglou & Koppelman 1999) (Coldren, *et al.* 2003, Coldren & Koppelman 2005) (Gramming, *et al.* 2005)
		Travel time	(Garrow, *et al.* 2007) (Coldren, *et al.* 2003, Coldren & Koppelman 2005)
		Equipment	(Proussaloglou & Koppelman 1999) (Coldren, *et al.* 2003, Coldren & Koppelman 2005)
		Fare	(Theis, *et al.* 2006) (Garrow, *et al.* 2007)
		Fare product	(Proussaloglou & Koppelman 1999) (Talluri & van Ryzin 2004) (Carrier 2006)
		Frequency	(Coldren, *et al.* 2003, Coldren & Koppelman 2005) (Hess & Polak 2006a) (Hess & Polak 2006b)
		Timings	(Coldren, *et al.* 2003, Coldren & Koppelman 2005) (Gramming, *et al.* 2005)
		Transfer	(Coldren, *et al.* 2003, Coldren & Koppelman 2005) (Theis, *et al.* 2006)
		Airport costs	

2.8.3 Dedicated and Critical Actors

The previous sections considered travelers' decision attributes, actors' instruments and relevant studies. The overviews presented in Table 2-7 and Table 2-8 represented the point of view of the traveler and also represented the interest of the IVT in decision attributes. However, the overviews did not consider issues of actors in the network. Following the elements of decision-making based on a framework of van Zuylen (2005), decision-attributes are divided in three main categories or choice stages:

- *Mode choice*, i.e. choice for car, aircraft or high speed train;
- *Access mode and airport choice*, i.e. choice for private or public transport, choice of airport in a multi-airport region;
- *Itinerary choice*, where the choice for itinerary includes the choice for fare product.

In this section, a similar approach will be followed, the main difference being that issues of the different actors are assigned to a choice stage in Table 2-9. In addition to the different choice stages, the following fields are listed in the overview:

- A *motivation* followed by a series of *questions*. The motivation serves as an brief introduction to the issue, the question is an example of what the actor might pose the IVT. Both are based on the analysis presented in this chapter;
- An overview is given of the *resources* available to the IVT of the actor;
- An actor can be *dedicated*. If they are dedicated, an actor is willing to use his resources;
- Finally, an assessment is made if an actor is *critical* or *non-critical*. An actor is critical if an actor has resources and is non-replaceable. As here is dealt with groups of actors (i.e. airlines, travel portals), this differentiation is not in place. Critical therefore indicates if the resources or the actor are critical to success of a study with regard to possible resources.

From this overview, the following may be derived:

- A dedicated and critical actor is an attractive collaboration partner from the point of view of the IVT. For instance, cooperation on access mode and airport choice, an airport would be the most obvious partner. However, as becomes apparent from the overview no actor is dedicated to all choice stages.
- A non-dedicated, critical actor may be necessary from the point of view of the IVT, but may not be interested in the mentioned issue. A partner with the potential to be interested in all three choice stages is a travel portal, which also appeared from the actor network presented in Figure 2-4.
- A non-dedicated and critical actor in one arena or, in this case, choice stage and dedicated and critical in another choice stage may be convinced to cooperate in multiple choice stages. However, an airline serving multiple airports in the same region, may also be interested in study of airport choice. An example would be Swiss International Airlines in the Basel/Zurich/Geneva triangle.

Two actors are (partially) excluded from the overview presented in Table 2-9. These are policy-makers and aircraft manufacturers. The first category is believed to believe interested in the aggregated outcome of all choices and perhaps on a lower level of detail as shown. Aircraft manufacturers objectives coincide with those mentioned of airlines regarding itinerary choice, as airlines as their first and foremost customers.

Table 2-9 Network analysis

Actors		Motivation / *Question*	Resource	Dedicated	Critical
Mode Choice					
Airlines		Airlines are in increasing competition with high speed trains on some routes. Therefore, insight into factors that influence traveler choice is required.	Data, Funds		
	Q1	*Which factors influence the choice of mode of travelers on short routes?*		Yes	-/+
High speed rail operators		High speed rail operators are in increasing competition with airlines on some routes. Therefore, insight into factors that influence traveler choice is required.	Data, Funds		
	Q2	*Which factors influence the choice of mode of travelers on short routes?*		Yes	-/+
Travel portals		Travel portals may be interested in mode choice as extension to current services.	Data, Funds		
	Q3a	*Which information regarding door-to-door travel and modes should I present in order to satisfy travelers' needs and increase revenues?*		No	-
	Q3b	*Which information is technically feasible to present within an acceptable search time?*		No	-
Policy makers		Outside the scope of this actor analysis were policy-makers, which can be interested in stimulating a certain mode.	Funds		
	Q4	*In which way should I influence travelers to realize a model shift?*		No	-
Access Mode and Airport Choice					
Airports		As argued in section 2.4.2, airports are faced with either (long-distance) competition or may be interested in redirecting traffic to either reduce the load or increase the number of customers. The first holds for multi-airport regions with a common owner, the latter for competing airports in a multi-airport region.	Data, Funds		
	Q5	*How can I realize a change in access mode?*		Yes	-/+
	Q6	*How can I redirect traffic from my congested airport to less congested airports?*		Yes	+
	Q7	*How can I redirect traffic currently heading for other airports?*		Yes	+

Actors		Motivation / *Question*	Resource	Dedicated	Critical
Travel portals		As argued in section travel portals are used more and more for information purposes and not for bookings, leading to decreasing revenues. Travel portals could offer better information, and increasing customer loyalty, by offering information about multiple airports and in a second stage about modes.	Data Funds		
	Q3a, Q3b			No	-/+
Airlines		Airlines operating hub-and-spoke networks, can be interested in attracting traffic to 'their' hub, by offering public transport to the airport at a reduced price. Airlines operating in multi-airport regions may be interested to see which services they should offer at which airport.	Data Funds	No	-/+
	Q8	*Which services should I offer at which airport?*		No	-/+
	Q9	*Do customers value extra access services (I.e. combined train / aircraft ticket?*		No	-/+
Itinerary Choice					
Airlines		As argued in section 2.3.2 airlines are interested for planning reasons. Therefore, issues are flight timings, frequency, aircraft deployment and network layout. Also, competition based on fare and less on fare products becomes important.	Data Funds		
	Q10	*What is the valuation of travelers for service characteristics (i.e. departure time, transfers, carrier image)?*		Yes	+
	Q11	*Is it possible to define a customer profile (segmentation) based on observable criteria of the traveler?*		Yes	+
	Q12	*What is the influence of fare on traveler decision-making?*		Yes	+
	Q13	*Which pricing strategies do my competitors pursue?*		Yes	+
Travel portals		Travel portals currently perceive low customer loyalty as a problem. Marketing includes the search algorithm of the travel portal.	Data Funds		
	Q14	*What are possible improvements for the listings on the website?*		Yes	-/+
	Q15	*Where can further improvements in search algorithms be made?*		Yes	-/+

2.9 Demarcation of the Research

In this chapter, an extensive overview has been given of current issues in the aviation system, stressing issues that reflect a common ground between the IVT as problem owner and the actor network. It is shown, that actors have different questions and there is no such thing as a single solution.

In this section, a subset of the issues presented in the previous two sections is selected for further investigation. The demarcation is based on the research objectives presented in section 1.5. These research objectives were:

1. *From an internal perspective:*
    a. The research should form a continuation of available knowledge within the IVT;
    b. The research should deliver a proof of concept for current and potential industry partners of the IVT.
2. *From a scientific perspective:*
    a. The research should contribute to current research in the aviation system;
    b. The research should incorporate prevailing issues in the scientific community.
3. *From an industry perspective:*
    a. The research should address current issues in aviation;
    b. The research should take into account objectives and instruments of relevant actors.

Based on these criteria, it is chosen to focus on itinerary choice modeling in the ensuing chapters for the following reasons:

1. *From an internal perspective:*
    a. Frick and Meister (2006) have conducted an analysis of itinerary choice modeling. Fröhlich (2006) developed a VISUM based air demand assignment model, which is available to the IVT and Kissilef (2006) has carried out a preliminary study of pricing practices. Furthermore, Erath (2004) used available data to assign travel demand to the air network;
    b. Cooperation is already taking place between SWISS/Lufthansa and the IVT. In addition, it was argued that an airline such as SWISS International Airlines may be interested in extending models of itinerary towards different choice stages. Furthermore, it was argued that travel portals form an interesting cooperation partner, as these have the potential to offer a door-to-door travel product as an extension to current services;
2. *From a scientific perspective:*
    a. Itinerary choice modeling can be extended by including fare, making it possible to estimate willingness-to-pay and price elasticities.
3. *From an industry perspective:*
    a. Airlines currently perceive willingness-to-pay and pricing as an important topic;
    b. Itinerary choice involves a single actor that influences traveler-decision making, making cooperation less complex and results easy transferable.

2.10 Conclusions

In this chapter, an extensive overview has been presented of recent and current literature on aviation and consumer choice behavior. In addition, several actors have been questioned towards their needs and views. Together with the fact that a large share of literature reflects the opinions of practitioners, a representative overview of issues in the aviation system is obtained.

Starting with travelers, it could be seen that a traveler is interested foremost in a travel product facilitating his journey from origin to destination. The airport is only a stopover in his travel process and might as well not occur in this travel process, as high speed train lines provide an increasingly competitive service. It is assumed that a traveler attempts to maximize his or her trip utility, thereby minimizing the generalized costs.

Prevailing issues, both in the scientific community and industry, are estimation willingness-to-pay and price elasticities. Incorporation of discrete choice models in revenue management (RM), albeit often mentioned, is still a theoretical issue. Furthermore, for the incorporation of discrete choice models in RM specific information is needed on fare products and the field of RM itself. As RM optimize the main source of income for an airline, ticket revenues, the information is probably highly confidential and may lead to conflicts with the goals of the scientific community in general.

Willingness-to-pay and price elasticities are of interest for airlines to set their fares in an evermore competitive environment. For online travel agents, an increased understanding in willingness-to-pay and other factors driving consumer behavior can aid with the listings of air travel tickets and with their search algorithms. The latter can be of central importance when online travel agents will consider offering a more elaborate travel advice to their customers, which either can include more airports or even different modes. To keep the consideration set of a traveler of a reasonable size and reduce search time, this understanding is necessary.

Both airlines and high speed train operators can be interested in surveys directed towards forecasting models demand on routes where they directly compete. In order to fully understand traveler behavior, the exact origin and final destination of a traveler are necessary to know.

Based on the actor analysis presented in this chapter, it is chosen to focus on itinerary choice in the ensuing chapters. This choice follows from the research objectives stated in Chapter 1. On the one hand, itinerary choice forms a continuation of previous research carried out at the IVT. On the other hand, with itinerary choice it is possible to address prevailing issues in the scientific community and active issues in the aviation industry. Table 2-10 shows the research questions relevant to itinerary choice.

Table 2-10 Selected research questions

Itinerary Choice		
Airlines	Q10	*What is the valuation of travelers for service characteristics (i.e. departure time, transfers, carrier image)?*
	Q11	*Is it possible to define a customer profile (segmentation) based on observable criteria of the traveler?*
	Q12	*What is the influence of fare on traveler decision-making?*
	Q13	*Which pricing strategies do my competitors pursue?*
Travel portals	Q14	*What are possible improvements for the listings on the website?*
	Q15	*Where can further improvements in search algorithms be made?*

For the realization of a European model of travel demand, collaboration with an unbiased party may be the best way to proceed. This can for instance be an aircraft producer, such as Boeing or Airbus or a policy-making instance. It should be kept in mind, however, that these parties may lack the resources necessary, especially data.

Chapter 3 will continue with concepts regarding modeling choice behavior in general and in aviation specific, thereby addressing the notion of the choice set and the formation process underlying it. Chapter 3 will conclude with a set of requirements of data necessary for quantifying itinerary choice behavior. Chapter 4 will then discuss the data available for this research and will conclude with a breakdown of the selected research questions.

Chapter 3 Modeling Choice Behavior in Aviation

3.1 Introduction

Chapter 2 presented the traveler as a utility maximizing agent, as is standard in economic consumer theory. In his choice, the agent is constrained by his available budget and the availability of alternatives. These two combined lead to the consumption possibilities, or the choice set. It is assumed that the agent has consistent preferences, i.e. if product A is preferred above product B, and product B preferred above product C, product A will be preferred above product C. If the continuous alternative set of economic theory is replaced with a discrete representation of alternatives, a form is obtained that is suitable for discrete choice analysis.

This discrete choice modeling approach is followed for two reasons. In the first place, it is a widely accepted method in transportation demand and consumer choice analysis. Second, it is amongst the key competences of the IVT.

Amongst others, this chapter describes the features common to discrete choice models, also known as random utility models. First, an introduction to discrete choice models will be given in section 3.2. The most prominent discrete choice model, the multinomial logit (MNL) model and its properties will be discussed.

In section 2.2 it was argued that more insight in the process underlying choice set formation is necessary. Therefore, special attention will be paid to the notion of the choice set and the notion of the choice set in an aviation setting in section 3.3.1 and 3.3.2 respectively. The focus will lie on possible choice sets in itinerary modeling.

3.2 Discrete Choice Models

3.2.1 Introduction

Discrete choice models are also called disaggregate choice models, meaning that the decision maker is assumed to be an individual. The definition of individual depends on the particular application; the individual can be a person or group of persons, such as a household or a family. It can also be a firm or governmental organization. Besides assumptions about the decision-maker, a discrete choice model contains assumptions about alternatives and their availability to the decision-maker and the attributes of an alternative, which represent the costs and benefits of an alternative. Furthermore it is assumed, just as in economic consumer behavior, that the decision-maker has perfect discrimination capability. However, the analyst is assumed to have incomplete information and, therefore, uncertainty must be taken into account. Four sources of uncertainty can be recognized: unobserved alternative attributes unobserved individual characteristics, measurement errors and proxy variables (Ben-Akiva & Bierlaire 1999). In order to reflect this fact, this uncertainty is modeled as a random variable.

With discrete choice models, a decision-makers' choice is described; any choice is made, by definition, from a non-empty set of alternatives (Ben-Akiva & Lerman 1985). The utility U_{iq} of an alternative i for a decision-maker q is defined by:

$$U_{iq} = V_{iq} + \varepsilon_{iq} = f(\beta, x_{iq}) + \varepsilon_{iq} \qquad (3.1)$$

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

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

$$P(i \mid C_q) = P[U_{iq} \geq U_{jq} \forall j \in C_q] = P[U_{iq} \max_{j \in C_q} U_{jq}] \tag{3.2}$$

From formula (3.2) it can be derived that the level of utility is irrelevant both to the decision-maker and the analyst, only differences in utility matter: $P(i \mid C_q) = P[U_{iq} - U_{jq} \geq 0 \forall j \in C_q]$. The same holds for adding a constant to the utility of all alternatives, the alternative with the highest utility doesn't change. If the utility is decomposed into the observed part and the unobserved parts, the following equation is obtained: $P(i \mid C_q) = P[\varepsilon_{jq} - \varepsilon_{iq} \leq U_{iq} - U_{jq} \forall j \in C_q]$. Here again, the utility only depends on the differences.

3.2.2 MNL Model

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

$$P(i \mid C_q) = \frac{e^{V_{iq}}}{\sum_j e^{V_{jq}}} \tag{3.3}$$

In Appendix B the derivation of the MNL-model is given.

3.2.3 IIA-property

The Independence of Irrelevant Alternatives (IIA) property states that the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives (Ben-Akiva & Lerman 1985, p. 33). In the case of the MNL model, this can be illustrated by:

$$\frac{P_q(i)}{P_q(k)} = \frac{e^{V_{iq}} \big/ \sum_j e^{V_{jq}}}{e^{V_{kq}} \big/ \sum_j e^{V_{jq}}} = \frac{e^{V_{iq}}}{e^{V_{kq}}} = e^{V_{iq} - V_{kq}} \tag{3.4}$$

This property stems from the fact that the distribution of the disturbances are assumed to be mutually independent and requires that the sources of errors contributing to the disturbances do so in a way that the total disturbances are independent.

3.2.4 Overcoming the IIA property

An important challenge in the field of transport modeling is how to overcome the IIA (independence of irrelevant alternative) property of the classic Multinomial Logit model (MNL). Recent research has focused on three different general approaches: changing the variance-covariance structure, nesting alternatives, and introducing similarity factors in the deterministic part of the utility function. The main issue is to find a solution that is first flexible and able to represent complex correlations, allows second a more thorough understanding of people's transport behavior and is third easy to compute and applicable to large choice sets. Chapter 7 will go into more detail on this issue.

3.2.5 Interpretation & Application Model Results

Following Louviere *et al.* (2000, p. 51) and Train (2003, p. 72), several model outputs will be discussed.

First, an estimate of β_{ik}, $\hat{\beta}_{ik}$ can be interpreted as the weight of the attribute k in the utility expression V_i of alternative i. With the estimates of the β's, an estimate of V_{iq} can be calculated the by taking β_{iq}'s and the X_{iqk}'s for individual q and using the following formula:

$$\hat{V}_{iq} = \sum_{k=1}^{K} \hat{\beta}_{ik} X_{iqk} \tag{3.5}$$

Instead of using a specific X_{iqk}, it is also possible to use the mean or median value of X_{iqk}. In that way, the level of relative utility U_{iq} of an attribute can be determined.

A t-test shows if an estimated parameter is statistically different from zero. Common is a confidence level of 95%, giving a t-value of 1.96, which gives a 95% confidence interval that the mean is different from zero. Other frequently used t-values are 1.439 (85%) and 1.645 (90%).
The goodness-of-fit can be calculated by:

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)} \tag{3.6}$$

Where $LL(\hat{\beta})$ is the maximized value of the log likelihood function and $LL(0)$ the value of the log likelihood function when all parameters are set equal to zero. The higher the value for ρ^2, the better the model fit.
Two models fitted on the same data can be compared with the following test statistic:

$$-2(LL(\hat{\beta}_c) - LL(\hat{\beta}_{uc})) - \chi^2 \tag{3.7}$$

Where $LL(\hat{\beta}_c)$ is the log likelihood of the constrained model or simple model and $LL(\hat{\beta}_{uc})$ is the log likelihood of the unconstrained model, the model containing more parameters. If the value exceeds the critical value of the chi-squared distribution, the null hypothesis that the added explanatory variables are equal to zero, is rejected.

Discrete choice models can be used to derive estimates of the willingness-to-pay (WTP) or willingness to accept (WTA) of an individual to obtain a benefit or avoid a cost. In a linear model, where each attribute is associated with a single parameter, the ratio of two parameters is the WTP or WTA, holding all other constant. If one of the attributes is measured in monetary units, the ratio can be interpreted as a valuation.

3.3 Choice Set Formation and Generation

3.3.1 Conceptual models of choice set formation

To fit within the discrete choice model framework the set of alternatives needs to exhibit three characteristics. First, the alternatives must be mutually exclusive from the decision maker's perspective. Choosing one of the alternatives implies not choosing any of the other alternatives. Second, the choice set must be exhaustive, in a way that all possible alternatives are included. Third, the number of alternatives must be finite (Train 2003, p. 15).

The environment of the decision maker determines the universal set of alternatives. Any single decision maker considers a subset of this universal set of alternatives, the choice set or consideration set. The identification of the list of alternatives is usually referred to as choice set generation or choice set formation. It is however important to make a clear distinction between choice set generation and choice set formation. In the ensuing, it is assumed that choice set generation is a process performed by the analyst. Choice set generation will be discussed further in section 4.4. Choice set formation is the result of a behavioral process of an individual and results in the consideration set of the individual.

Several approaches are mentioned in literature to determine the choice set which contains the alternatives that were available to the decision maker. On the one hand, Swait (2001) proposes to formulate several choice sets (a set of choice sets) and estimate the probability of a choice set being the true choice set. This work uses the two stage characterization of the choice process of Manski (1977) as basis:

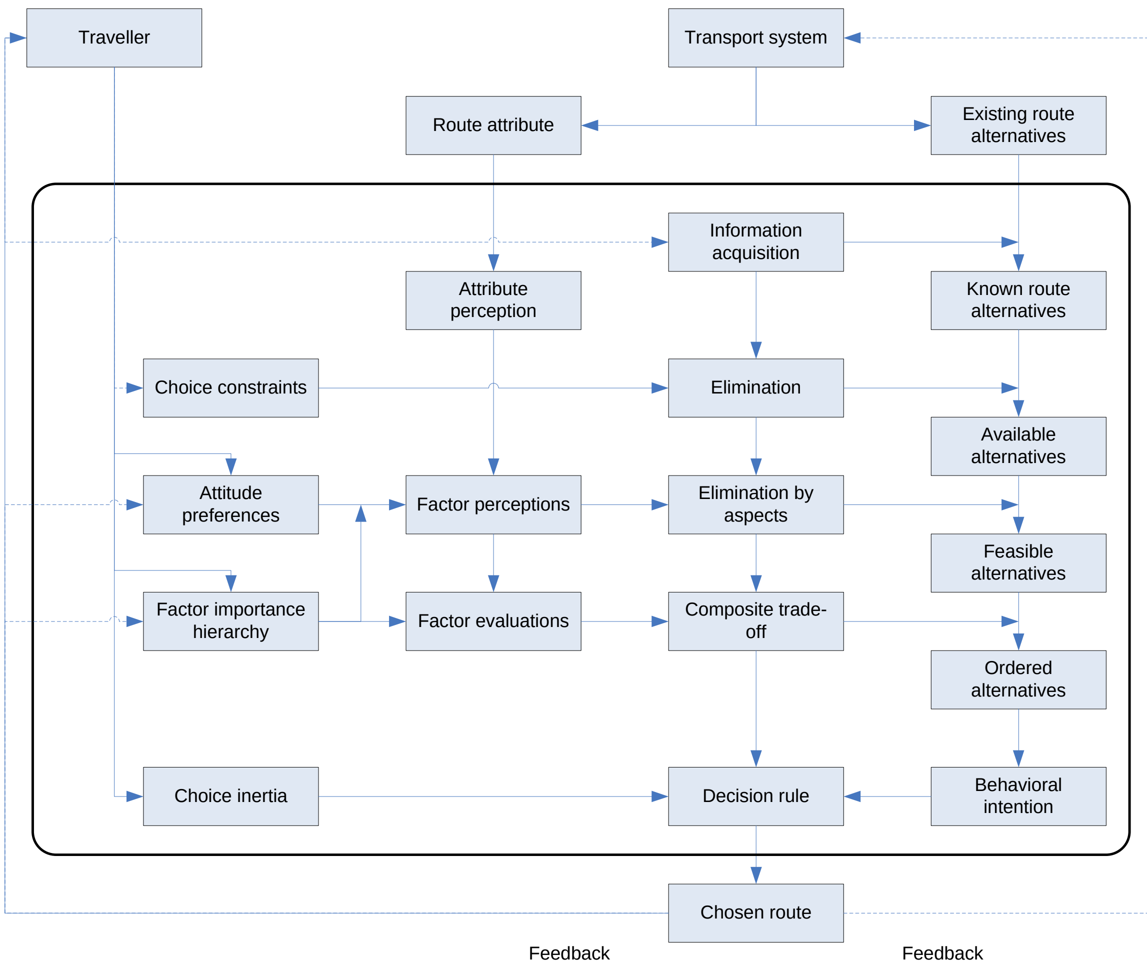
$$P(i) = \sum_{C \subseteq \Delta(M)} P(i \mid C) Q(C) \tag{3.8}$$

Where C is a choice set in $\Delta(\mathrm{M})$, the set of subsets of M, $Q(C)$ is the probability is the true choice set and $P(i \mid C)$ is the conditional probability of choice given set C. In route choice modeling, the set M can be equal to the universal set or the master set. In both cases, the number of subsets $\Delta(\mathrm{M})$ will be very large.

Bovy (1990, 2007), on the other hand, stresses the necessity to make a clear distinction between choice set formation and choice from a choice set in the case of route choice analysis. This necessity stems from specific characteristics of route choice sets: the population of available routes (the universal set) is very large, the subset of feasible and attractive routes is also very large and the identification of relevant alternatives is not a trivial task because of complex patterns of overlap.

In both cases, it is necessary for the researcher to formulate a choice set. Therefore, a closer look will be given to the formation of the choice set from the perspective of a decision-maker. As a guideline, a framework by Bovy and Stern (1990) is taken, which can be seen in Figure 3-1. Both Hoogendoorn-Lanser (2005, p. 21) and Fiorenzo-Catalano (2007) provide a more extensive discussion of choice set formation and generation in a route choice context. The interested reader is referred to their work.

Figure 3-1 Conceptual framework for choice set formation of an individual traveler



Source: Bovy and Stern (1990, p. 31)

The input of the framework consists of a traveler with his or her needs and preferences and a physical environment, with its objective opportunities and fragments. At the right side, the figure shows a series of route sets that follow from a variety of experimental and mental processes, which are shown to the left.

The network, represented by the transport system in the figure, offers a large and complex set of route alternatives for a trip of which the traveler has limited awareness. The traveler's awareness is, amongst others, influenced by previous experiences and his manner of information acquiring. The alternatives, of which the traveler is aware, are the *known alternatives*. Not all known alternatives will be considered as genuine travel alternatives; time sensitive travelers will consider other alternatives than cost sensitive travelers or comfort seeking travelers. From the perspective of the traveler, the known alternatives that satisfy cost, time and comfort constraints form the set of *available alternatives*.

From the set of available alternatives, a limited set of alternatives will be considered feasible. The set of available alternatives is limited through elimination by aspects and depends on the traveler's choice factors, which may be not directly measurable characteristics of the routes. The subjective values of these factors follow from his perception of objective route attributes relevant for his trade-off and choice and the relative importance of the these factors (factor importance hierarchy).

Only a fairly limited set of feasible alternatives remains (*the consideration set*). Between the alternatives in this set, a more elaborate trade-off will be made.

Chorus (2007) provides a useful addition to this framework, which would consist of an extra loop between the decision rule and the information acquisition stage of the decision-making process: not each decision will lead to a chosen route. Instead of choosing a route, an individual may choose to gather more information. Chorus presents a discrete choice modeling approach to describe the full sequences of possibly multiple information acquisitions, followed by a travel choice. In this case, the choice set contains all travel alternatives until the travel choice is made.

As might be clear now, the main distinction between these two approaches for determining the available alternatives to an individual is that the approach proposed by Bovy and Stern totally separates the choice set generation model from the choice model. In addition, it offers valuable insight in the realization of a consideration set. The approach proposed by Swait allows for multiple choice sets. Seen from a modeling perspective, this approach allows the analyst to estimate the probability that a choice set formed by the analyst is the true choice set. All these choice sets originate from the universal set of alternatives and have to be formulated. Thus, the presented framework offers valuable insight in choice set formation. Several questions arise from the framework:

- Which alternatives are known to the decision-maker?
- Which alternatives are available to the decision-maker?
- Which of the known alternatives are chosen?

In the next section, the generic framework will be put in the context of itinerary modeling.

3.3.2 Choice Sets in Itinerary Choice Modeling

In the previous section a generic framework of choice set formation was presented. This framework considered choice set formation from an individual's point of view. In this section, several dimensions specific to itinerary choice modeling influencing the composition of the choice sets will be highlighted.

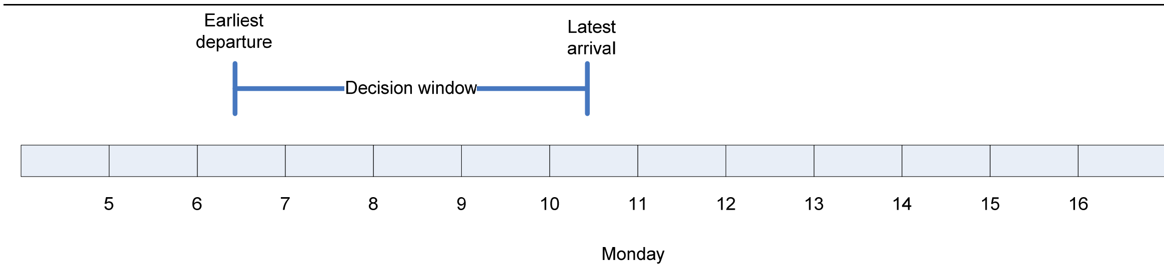
At least four dimensions can be recognized when an individual chooses for an itinerary and which influence the composition of the consideration set. These four dimensions can correspond to the information acquisition stage, elimination stage and elimination by aspects stages, if they were related to the generic framework of Bovy and Stern.

1. The first dimension involves the booking period: an individual can choose to book his ticket any time in the period between the decision to make a trip and the preferred departure time. This concept was also illustrated in section 2.2.3, Figure 2-1. However, a traveler is only aware of itineraries after retrieving information one or multiple times. This dimension will be referred to as the *booking consideration dimension* in the ensuing
2. The second dimension includes the choice of air transportation service provider. An individual may consider all possible transportation service providers for his journey, but it is also very well possible that an individual is bound to a carrier through a loyalty scheme or shows a preference for low-cost carriers. This will be referred to as the *information acquisition dimension*. It should be noted, that a traveler is not likely to be able to consider all possible outcomes of decisions. An individual will gather information until he is convinced that he cannot improve his choice by gathering more information (Simon 1955).
3. The third dimension is the departure time choice: a traveler may make a trade-off between his preferred departure and preferred arrival time and attributes of other known alternatives. This dimension will be referred to as the *preferred arrival time dimension.*
4. The fourth dimension concerns the fare of an itinerary. A traveler might not consider all the fares or fare products offered by an airline. This dimension will be referred to as the *fare dimension.*

Figure 3-4 depicts a set of possible outbound and inbound itineraries and serves as an illustration for dimension two and three. Itinerary 1 and 2 are outbound itineraries departing in the morning; itinerary 3 and 4 are outbound itineraries departing in the evening. Itinerary 5 and 6 are inbound itineraries.
If a traveler prefers to arrive between the arrival time of itinerary 1 and 3 and he may consider only outbound itineraries 1, 2 and 3. A cost sensitive traveler may consider all itineraries, assuming that they are all in the same price range. Finally, a traveler acquiring information from one or multiple carriers may only consider a fairly limited set of the itineraries offered.

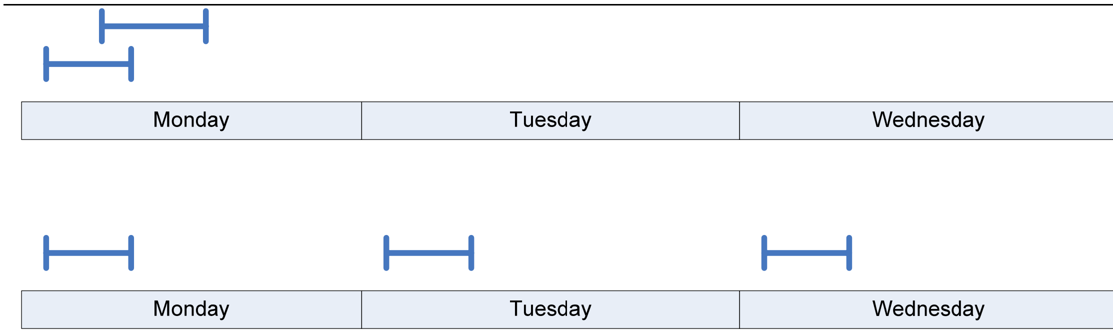
Two examples of the preferred arrival time dimension are shown in Figure 3-2 and Figure 3-3. Figure 3-2 shows the decision window of a traveler, which is bounded by the earliest preferred departure time and the latest preferred arrival time. Figure 3-3 shows an example of day inflexible and day flexible travelers.
The larger the decision window or the flexibility of the traveler, the larger the choice set will become. The same occurs when a traveler considers multiple carriers or gathers information multiple times: the choice set size will become larger. Furthermore, the overlap between choices will increase: it can be that the same itinerary occurs in the choice set, but with different fares or departure days.

Figure 3-2 Decision window



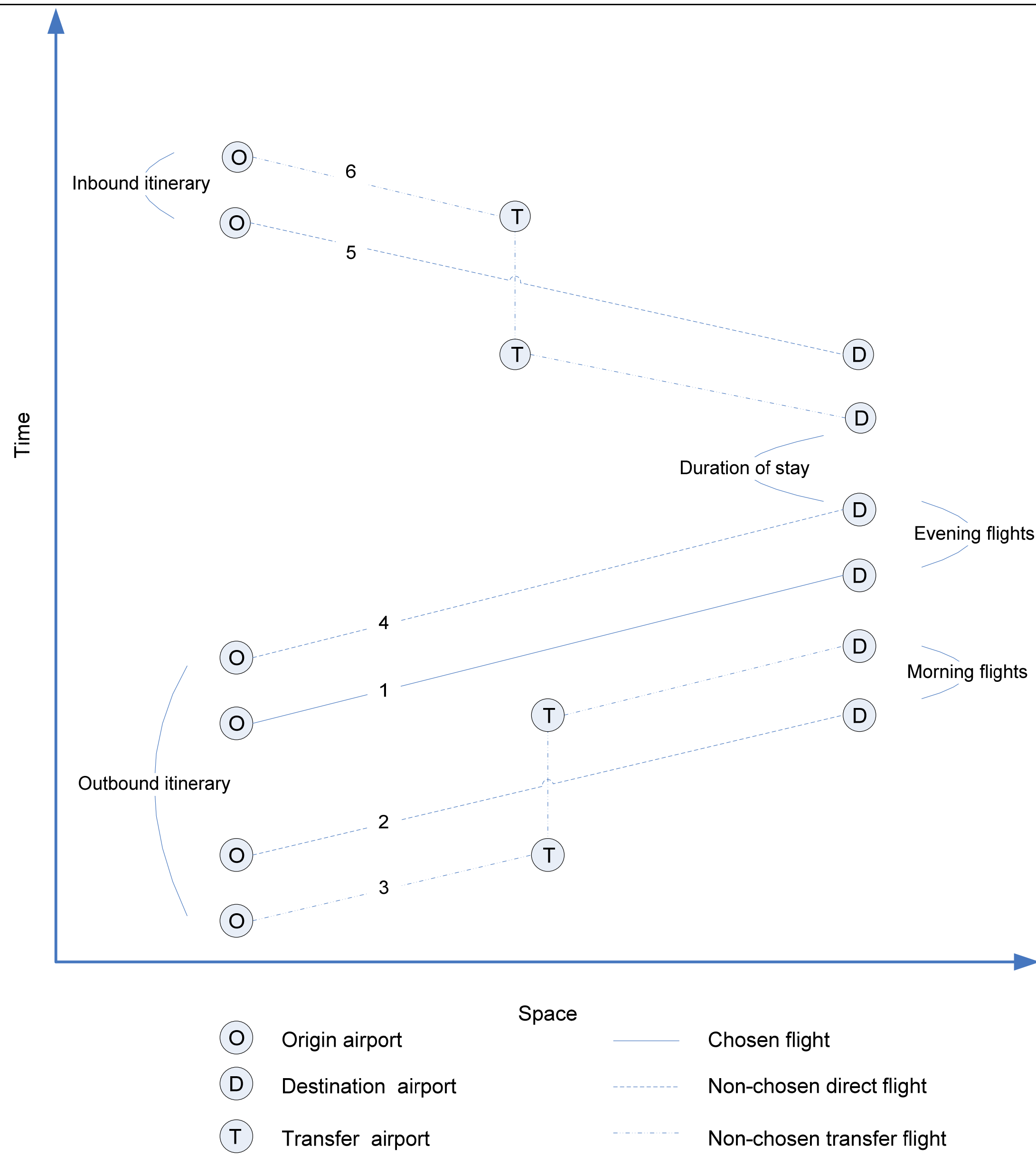
Source: Boeing Commercial Airplanes Group (1993)

Figure 3-3 Day inflexible and day flexible decision windows



Source: Boeing Commercial Airplanes Group (1993)

Figure 3-4 Known outbound and inbound itineraries to a traveler



From the aforementioned it can be concluded that several types of consideration sets can be distinguished when modeling itinerary choice. The consideration set will contain different alternatives, based on the ranges a traveler considers of each dimension. In Table 3-1 two levels are assigned to the booking consideration dimension (single booking consideration/multiple booking consideration), the information acquisition dimension (one carrier/multiple carriers) and preferred arrival time dimension (low preference/high preference).

Table 3-1 Combination of ranges of dimensions

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

3.3.3 Choice Set Terminology

In section 3.3.1 a generic framework of choice set formation was presented. This framework considered choice set formation from an individual's point of view. In section 3.3.2 several dimensions specific to itinerary choice modeling influencing the composition of the choice sets were highlighted. This section will combine the generic framework and the several dimensions, influencing choice set formation. Both the researcher's perspective and a traveler's perspective will be addressed. In addition, several terms regarding choice set formation will be given, following a similar discussion as given by Hoogendoorn-Lanser (2005).

Table 3-2 lists choice set notions from an individual traveler's perspective, together with the definition and relevant dimension, i.e. the dimension influencing the composition of the choice set. The term 'actual' is included as the traveler is aware of the different choice sets.

Table 3-2 Choice set notions from the traveler's perspective

Terminology	Definition	Relevant dimension
Universal choice set	The universal choice set consists of all existing route alternatives and are all itineraries offered by the transportation network between an origin destination pair between the trip making decision and the moment of preferred departure.	Booking consideration dimension
Actual master set	Subset of the universal choice set containing the known alternatives to a traveler after acquiring information; the information acquiring process can consist of a single stage or multiple stages. If a traveler informs himself only on an airline website, all known alternative set will contain itineraries provided by the same carrier. On the other hand, if a traveler informs himself by means of an online travel portal, the known alternative set will contain itineraries offered by a number of carriers. In addition, a traveler can inform himself multiple times.	Information acquisition and booking consideration dimension
Actual subjective choice set	Subset of the universal choice set containing known itineraries and feasible itineraries. Time sensitive and cost sensitive travelers will consider other itineraries and will lead to different sets of available alternatives. A cost sensitive traveler will consider a broader set of itineraries (i.e. itineraries departing over the entire day or even multiple days) than time sensitive travelers.	Arrival time dimension and fare dimension
Actual consideration set	The remaining alternative sets are constructed by traveler's preferences which can include carrier preferences, departure time preferences and aircraft preferences.	
Actual chosen alternative	Itinerary that is chosen and is part of the consideration set	

Table 3-3 lists the different choice set notions from a researcher's perspective. It can be seen that the term 'observed' and 'generated' replaces the term 'actual' in the choice set definitions. Furthermore, the known and feasible alternatives are in the observed / generated subjective choice set, contrary to the definition given by Hoogendoorn-Lanser, who assumes that the observed / generated objective choice set also contains the feasible alternatives. It is argued, that factors influencing an individual's known and feasible alternatives are not known to the researcher and therefore belong in the observed / generated subjective choice set.

Table 3-3 Choice set notions from a researcher's perspective

Terminology	Definition
Universal choice set	The universal choice set consists of the existing route alternatives and are all itineraries offered by the transportation network between an origin destination pair between the observed / generated trip making decision, observed / generated booking time and the moment of observed / generated departure.
Observed / generated objective choice set	Subset of the universal choice set containing itineraries assumed to be logical to individuals, leaving within a preferred departure time interval satisfying their travel needs.
Observed / generated subjective choice set	Subset of the universal choice set assumed to contain the known alternatives and feasible itineraries, leaving within a preferred departure time interval satisfying their travel needs.
Observed / generated consideration set	Subset of the universal choice set assumed to contain considered itineraries.
Observed / generated chosen alternative	Itinerary observed / generated by the researcher and assumed to be chosen by the traveler.

3.3.4 Application of generated choice sets

Choice sets may be used for several applications (Hoogendoorn-Lanser 2005):

- Analysis of available alternatives;
- Estimation of parameters in utility functions;
- Prediction of choice probabilities to determine market shares;
- Data completion

In this thesis, choice sets will be used for analysis and estimation purposes. For the estimation of parameters, choice sets need not to be exhaustive, but may also contain a subset of relevant alternatives. For the prediction of choice probabilities, or market shares, choice sets need to be exhaustive, as the flow on a link, or (part) of an itinerary is the sum of many flows.

3.4 Conclusion

In this chapter, a brief introduction is given to the choice modeling, starting with the general formulation of the utility function and the multinomial logit (MNL) model. Several interpretation methods have been presented, such as sign, significance, relative weight of the parameter estimates.

Every choice is made from a set of alternatives, the choice set. The composition of the choice set is influenced by a behavioral process, the choice set formation process. The choice set formation process results in a consideration set, from which a choice is made.

In itinerary modeling, several dimensions can be recognized that influence the composition of the consideration set, namely the booking consideration dimension, the information acquisition and frequency dimension, the preferred arrival time dimension and the fare dimension. By varying the bounds or levels of these dimensions, the composition of the consideration set varies.

Notions of the choice set differ from a traveler's and researcher's perspective. The traveler is aware of his actual subjective and actual consideration set. To the researcher, these sets are unknown and can only be approximated.

For this research, the discussion presented in this chapter implies the following:

- To fit within the discrete choice framework, it necessary to have a choice or an observed itinerary booking, and:
- It is necessary to have an observed or generated choice set, preferably an:
    - Observed / generated objective choice set, containing itineraries logical to individuals, leaving within a preferred departure time interval satisfying their travel needs, or preferably:
    - Observed / generated subjective choice set, containing the itineraries assumed to be known and available itineraries to the traveler;
    - If the choice set is to be generated, insight is needed in the ranges of dimensions influencing the actual subjective choice set composition.

Chapter 4 will match these requirements with the selected issues in the actor network and will present the available data to meet these requirements.

Part II – Case Study 'Itinerary Choice Modeling'

- This page left blank intentionally -

Chapter 4 Required and Available Data for the Case Study

4.1 Introduction

Chapter 2 concluded with a demarcation of the issues in the actor network and a selection of possible questions an actor can ask the IVT. The selection was based on criteria reflecting the IVT and her environment, the scientific community and the aviation industry. These questions are revisited in Table 4-1. Chapter 3 provided valuable insight in choice set formation and in the characteristics of discrete choice models.

Table 4-1 Research questions - revisited

Itinerary Choice		
Airlines	Q10	*What is the valuation of travelers for service characteristics (i.e. departure time, transfers, carrier image)?*
	Q11	*Is it possible to define a customer profile (segmentation) based on observable criteria of the traveler?*
	Q12	*What is the influence of fare on traveler decision-making?*
	Q13	*Which pricing strategies do my competitors pursue?*
Travel portals	Q14	*What are possible improvements for the listings on the website?*
	Q15	*Where can further improvements in search algorithms be made?*

The research questions and characteristics of discrete choice models lead to a series of requirements to necessary data:

- Each choice is made from a choice set, which leads to two requirements:
    1. An observed choice is necessary, which boils down to a chosen itinerary in the case of itinerary choice;
    2. A choice set is necessary, preferably an observed subjective choice set or the consideration set.
- Each itinerary needs to contain:
    3. A set of service characteristics, such as transfer, carrier, travel time and departure time in order to answer Q10;
    4. An attribute representing fare in order to calculate the valuation of these service characteristics in order to answer Q10 and make it possible to answer Q12.

Three datasets are available to the IVT, namely the *Marketing Information Data Tapes (MIDT),* the *Expedia dataset,* and the *Official Airline Guide (OAG).* As can be seen in Table 4-2., these three datasets are examples of (cross-sectional) revealed preference (RP) datasets. Other possibilities would have been (1) stated preference (SP) data, (2) simulated data or (3) combined revealed preference/stated preference data. For the following reasons it is chosen to work with the given datasets:

- The majority of the studies concerning willingness-to-pay have used SP data (section 2.8.2)., having the advantage that the exact information (e.g. alternatives, attributes) presented to the respondent is known. While offering these advantages, SP data represents choices made in a hypothetical context and not, as is the case with RP data, choices made in real-life situations.

- The fact that the datasets contain revealed preferences and thus actual choices can convince both researchers and practitioners of the results;
- The datasets are readily available and only need processing.

Table 4-2 Requirements for data

Requirement	Dataset	Section
Observed choice	Marketing Information Data Tapes (MIDT) contains bookings of itineraries in November 2006 worldwide.	4.2.1
Choice set	Expedia dataset, which contains fares of itineraries in the month November 2006 on 70 origin-destination pairs collected between September and November 2006.	4.2.2
Service characteristics	Official Airline Guide (OAG) which contains detailed flight timings, code share information, type of aircraft and number of seats for November 2006.	4.2.3

In this chapter the available datasets for this research will be discussed in more detail. After an overview of the available datasets available in section 4.2, the datasets will be related to choice set terminology in section 4.4.2. The processing and matching of these datasets will be discussed in section 4.3. Choice set generation with the available data will be discussed in section 4.4. A further breakdown of the research questions is presented in section 4.5. Conclusions are presented in section 4.6. Chapter 5 will present a descriptive analysis of the choice sets.

4.2 Datasets: MIDT, Expedia and OAG

4.2.1 Marketing Information Data Tapes (MIDT)

The MIDT dataset contains data collected by Computer Reservation Systems (CRS), also known as Global Distribution Systems (GDS). A brief description of GDS-vendors can be found in sections 2.6.2. CRS systems included in the dataset are Amadeus, Abacus, Galileo, Worldspan and Apollo. A rough comparison with Eurostat figures has led to the conclusion that the CRS data cover between the 40% and 90% of the passenger bookings on any one route. Variables included in the CRS dataset are: booking date, trip origin, trip destination, leg origin, leg destination, departure date, return date, departure and arrival times, carrier abbreviation, and flight number per leg. A leg represents a single segment in an itinerary. In Appendix C.2 an extensive overview of the contents of the MIDT dataset is presented, together with the comparison with Eurostat figures. The fare of the booked itinerary is not listed. Also, traveler characteristics, such as age and gender, are not shown.
From the MIDT dataset, it is possible to extract booked itineraries. These will also be referred to as the chosen itineraries or bookings in the ensuing sections.

4.2.2 Expedia Dataset

The second dataset was obtained by webbots (i.e. a pre-programmed query) querying Expedia (http://www.expedia.de) on a nearly daily basis in the period September – November 2006 for flights departing in November 2006 on 70 origin-destination pairs in Europe. Three durations of stay were queried: a trip returning on the same day, a trip returning on the next day and a trip returning in two weeks time. Every origin-destination pair was queried in one direction (e.g. Frankfurt – Istanbul but not Istanbul – Frankfurt). Variables obtained from Expedia include query date, trip origin, trip destination, departure date, return date, departure and arrival times, carrier name(s), flight number(s) and most notably fare. More detailed information about the Expedia dataset can be found in Appendix C.1.

4.2.3 Official Airline Guide (OAG)

The official airline guide (OAG) contains information on all scheduled flights worldwide, including low cost airlines. The information concerns code sharing, type of aircraft operated, number of seats, operating days, departure times and arrival times etc. This information is entered by airlines, is considered to be a neutral source of airline scheduling data and is updated frequently (OAG Worldwide Limited 2006).

4.2.4 Comments on datasets

Two general remarks should be made. First, the available datasets and most importantly the observed bookings are for November. Despite no information being available on passenger type, it can be said that in November the percentage of business travelers is fairly high, as it is off-season.
Second, all available data is scheduled and 'static' data: no information on flight delays and aircraft and no-shows is known. In the yearly statistics bulletin of the German Aviation Authority, special attention was paid to flight delays (Grunewald*, et al.* 2007). Delays can be a consequence of the strategy of an airline: airlines operating hub-and-spoke networks may have more delays as they tend to wait for connecting flights. However, delays were mostly airfield specific and not airline specific and were larger on routes containing a busy airport. It is therefore more likely that delays influence the choice of (transfer) airport directly and the choice of airline indirectly.
The dataset containing the observed bookings is an example of a revealed preferences dataset. In a typical revealed preferences, cross sectional dataset the challenge for the modeler is to determine which alternatives are available to an individual (Ortuzar & Willumsen 2001). This will be discussed to more extent in the next section.

4.3 Processing & Matching Datasets

4.3.1 Adding fare to the MIDT dataset

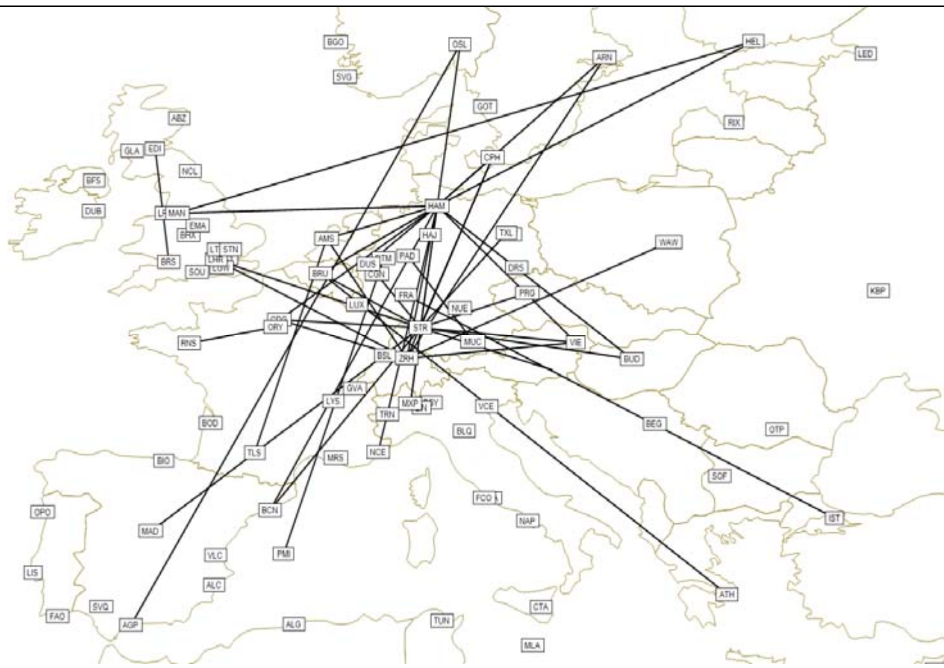
One of the research questions concerns the influence of fare on consumer choice. Therefore, the Expedia dataset, which contains fares, is taken as starting point for the matching of data. Only booked flights on the same origin-destination pairs as observed in the Expedia dataset are extracted from the MIDT dataset.
Again from the Expedia dataset, itineraries offered in November 2006 are extracted. These are combinations of outbound and inbound itineraries. The outbound and inbound itineraries can each consist of one or more flights. Each unique combination of outbound and inbound itinerary is given an ID. To the MIDT dataset, the same ID is assigned. This itinerary ID makes it possible to match the Expedia dataset and the MIDT dataset. The itinerary ID, together with the departure date, booking date and duration of stay makes a definitive match possible. In this way, it is possible to assign a fare

to a booked itinerary. A more elaborate discussion of the matching of the MIDT and Expedia dataset can be found in Appendix C.5.

The result of this exercise is 21.978 observed itinerary choices with fare, of which 18.895 are in the same direction as observed by the webbots. In Appendix C.5, the distribution of the bookings over the origin-destination pairs is shown. Most origins are located in either Hamburg or Stuttgart, with as main destinations Berlin Tegel and Dusseldorf. The percentage of bookings per origin destination pair remains fairly constant by adding constraints (e.g. same booking day, same booking day and departure day). For origin-destinations pairs with a very low number of bookings in the MIDT, no bookings with price are observed. This is because a low number of bookings leads to a low chance of matching exactly the same itinerary. For subsequent research, it should be kept in mind that it the matching chance becomes lower as soon as the number of transfers increases: it is probable that a traveler is offered a different route as listed on Expedia by either an airline or another travel portal.

The matching criteria are fairly rigid; one could argue that the difference between booking time and departure time also can provide ticket fare information. This is not considered here.

Figure 4-1 Origin-destination pairs



With acknowledgements to Fröhlich (2006)

4.3.2 Service characteristics

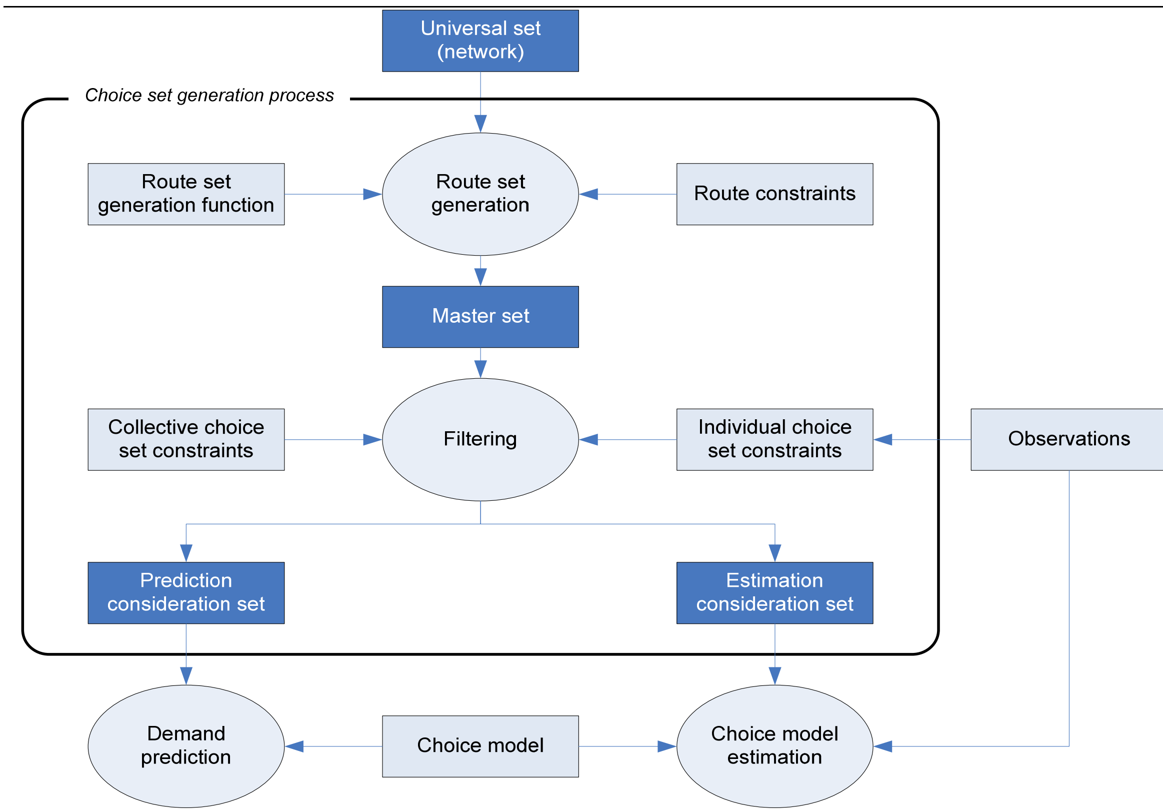
As said, a distinction is made between the outbound part and the inbound of an itinerary. Each outbound part and inbound part of an itinerary can consist of multiple flights. By matching these flights with the OAG dataset on their flight numbers and carrier abbreviations, the data of the OAG dataset can be used. The OAG dataset is based on original flight numbers, whereas the MIDT and Expedia dataset only contain the code share codes. The first step is thus to add the original flight number to the latter datasets share (see Appendix C.3). Following steps then include a match on departure day of week and original flight number. With the OAG dataset, waiting times, in vehicle times, type of aircraft and code share can be added to the characteristics of the outbound and inbound parts of the itinerary.

4.4 Choice Set Generation

4.4.1 Choice Set Generation Process

In the previous chapter, it was discussed that a distinction should be made between choice set generation and choice set formation. Choice set generation is a process carried out by the analyst, whereas choice set formation is the result of behavioral process of the decision-maker. The choice set generation process is visualized in Figure 4-2. The universal set of alternatives is given by the network. However, it is unknown to both the decision-maker and the analyst. The analyst generates a sets of routes based on a series of constraints and the generation function. This can for instance be a branch-and-bound algorithm; Boeing uses an approach that directly takes into account the utility of each route (Parker, *et al.* 2005). The result of the route set generation process is a master set of alternatives. It should be noted, that if a utility based approach is used, the generated set is not equal to the master set. Alternatives from the master set are filtered, based on collective and individual choice set constraints.

Figure 4-2 Choice set generation process



Source: Bovy (2007)

For this research a program was written that follows the same outline. Expedia is used for route set generation. The filtering process is done in a second stage; to this means, the non-chosen alternatives are selected on several criteria from the Expedia dataset based on the ranges of dimensions discussed in the previous chapter. These were the booking time dimension, information acquisition dimension and the preferred arrival time dimension. As discussed in the previous chapter, these dimensions can be combined, which has as result several types of choice sets. It is possible to generate all the types listed in Table 3-1. At the moment, the constraints demarcating the dimensions are deterministic, i.e. each choice set is based on the same ranges of dimensions (it is however possible to vary the constraints of these dimensions per duration of stay). A probabilistic approach would be imaginable, is however not implemented.

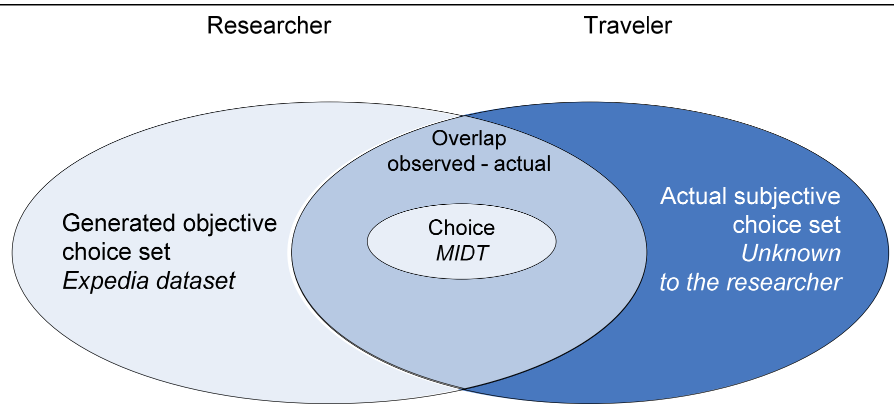
The choice sets are written to a data file format suited for Biogeme and an initial Biogeme model file is created. From each choice set, descriptive statistics (minimum, maximum, mean, median, standard deviation) and frequencies (counts and percentages) are recorded per attribute where a distinction is being made between statistics of a chosen and non-chosen alternative attributes. Statistics are written to an Excel file. This process is also discussed in Appendix C.6.

4.4.2 Relationship datasets – Choice Set Terminology

In this section, the relationship between the datasets and choice set terminology will be revisited in more detail. The MIDT dataset contains the booking of an itinerary and thus the *choice*. The Expedia dataset is used to generate an *objective choice set*.

Figure 4-3 shows the relationship between the datasets and the choice set terminology presented in section 3.3.3. The Expedia dataset contains the *observed objective choice set*. The actual subjective choice set is unknown to the researcher.

Figure 4-3 Relationship between datasets and choice set terminology



4.4.3 Example choice set

Table 4-3 presents an example choice set for the route Hamburg Vienna for an itinerary returning the next day. Of the inbound itinerary, only departure and arrival times are listed. If a the code is not identical to the original code, it concerns a code share flight.

Table 4-3 Example choice set for Hamburg - Vienna for itineraries returning the next day

	Outbound														Inbound		
	Airline 1	Code 1		Aircraft Segment 1	Airline 2	Code 2		Aircraft segment 2	Original flightcode 1		Original flightcode 2		Departure time	Arrival time	Departure time	Arrival time	Fare
1	Lufthansa	LH	35	Airbus A320-100/200	Lufthansa	LH	6326	Fokker 70	LH	35	OS	112	6:10	9:45	17:35	19:15	334.26
2	Czech Airlines	OK	543	Aerospatiale/Alenia ATR 42-300 / 320	Czech Airlines	OK	606	Aerospatiale/Alenia ATR 42-300 / 320	OK	543	OK	606	9:15	19:45	8:55	18:45	735.97
3	Czech Airlines	OK	543	Aerospatiale/Alenia ATR 42-300 / 320	Czech Airlines	OK	606	Aerospatiale/Alenia ATR 42-300 / 320	OK	543	OK	606	9:15	19:45	20:15	8:45	735.97
4	Czech Airlines	OK	545	Aerospatiale/Alenia ATR 72	Czech Airlines	OK	604	Aerospatiale/Alenia ATR 42-300 / 320	OK	545	OK	604	19:15	8:20	8:55	18:45	735.97
5	Czech Airlines	OK	545	Aerospatiale/Alenia ATR 72	Czech Airlines	OK	604	Aerospatiale/Alenia ATR 42-300 / 320	OK	545	OK	604	19:15	8:20	20:15	8:45	735.97
6	Lufthansa	LH	3590	Boeing 737-300 pax					LH	3590			7:00	8:30	19:40	21:10	173.77
7	Lufthansa	LH	3590	Boeing 737-300 pax					LH	3590			7:00	8:30	17:35	19:15	203.77
8	Lufthansa	LH	3592	Boeing 737-500 pax					LH	3592			13:25	14:55	19:40	21:10	173.77
9	Lufthansa	LH	3592	Boeing 737-500 pax					LH	3592			13:25	14:55	17:35	19:15	203.77
10	Lufthansa	LH	3594	Boeing 737-500 pax					LH	3594			17:35	19:05	17:35	19:15	203.77
11	Lufthansa	LH	3594	Boeing 737-500 pax					LH	3594			17:35	19:05	19:40	21:10	173.77
12	Swiss	LX	3645	Boeing 737-500 pax	Swiss	LX	3552	Fokker 70	LH	3645	OS	3552	7:15	12:15	19:40	21:10	244.35
13	Swiss	LX	3699	Boeing 737-500 pax	Swiss	LX	3562	Fokker 70	LH	3699	OS	3562	7:10	12:15	17:35	19:15	269.95
14	Swiss	LX	3699	Boeing 737-500 pax	Swiss	LX	3562	Fokker 70	LH	3699	OS	3562	7:10	12:15	19:40	21:10	239.95
15	Lufthansa	LH	6362	Canadair Regional Jet					OS	172			20:00	21:30	19:40	21:10	228.77
16	Lufthansa	LH	6364	Canadair Regional Jet					OS	176			10:35	12:10	19:40	21:10	228.77
17	Austrian Airlines	OS	7252	Boeing 737-300 pax					LH	3590			7:00	8:30	19:40	21:10	173.77
18	Austrian	OS	7252	Boeing 737-300 pax					LH	3590			7:00	8:30	17:35	19:15	203.77

	Outbound														Inbound		
	Airline 1	Code 1		Aircraft Segment 1	Airline 2	Code 2		Aircraft segment 2	Original flightcode 1		Original flightcode 2		Departure time	Arrival time	Departure time	Arrival time	Fare
	Airlines																
19	Austrian Airlines	OS	7254	Boeing 737-500 pax					LH	3592			13:25	14:55	19:40	21:10	173.77
20	Austrian Airlines	OS	7254	Boeing 737-500 pax					LH	3592			13:25	14:55	17:35	19:15	203.77
21	Austrian Airlines	OS	7254	Boeing 737-500 pax					LH	3592			13:25	14:55	19:40	21:10	173.77
22	Austrian Airlines	OS	7256	Boeing 737-500 pax					LH	3594			17:35	19:05	19:40	21:10	173.77
23	Austrian Airlines	OS	7256	Boeing 737-500 pax					LH	3594			17:35	19:05	17:35	19:15	203.77
24	Austrian Airlines	OS	7256	Boeing 737-500 pax					LH	3594			17:35	19:05	19:40	21:10	173.77
25	Air Berlin	AB	8330	Airbus A319					AB	8330			15:50	17:15	17:55	19:15	147.35
26	Air Berlin	AB	8330	Airbus A319					AB	8330			15:50	17:15	8:30	9:55	154
27	Air Berlin	AB	8330	Airbus A319					AB	8330			15:50	17:15	21:20	22:45	147.35
28	Air Berlin	AB	8468	Boeing 737-800 pax					AB	8468			6:30	7:50	17:55	19:15	162.77
29	Air Berlin	AB	8468	Boeing 737-800 pax					AB	8468			6:30	7:50	18:15	21:15	279.85
30	Air Berlin	AB	8468	Boeing 737-800 pax					AB	8468			6:30	7:50	21:20	22:45	160.55
31	Air Berlin	AB	8682	Boeing 737-700 pax	Air Berlin	AB	8151	Airbus A320-100/200	AB	8682	HG	8151	6:30	10:10	17:55	19:15	195.35
32	Air Berlin	AB	8682	Boeing 737-700 pax	Air Berlin	AB	8151	Airbus A320-100/200	AB	8682	HG	8151	6:30	10:10	18:15	21:15	312.43
33	Air Berlin	AB	8682	Boeing 737-700 pax	Air Berlin	AB	8151	Airbus A320-100/200	AB	8682	HG	8151	6:30	10:10	21:20	22:45	195.35
34	Air Berlin	AB	8846	Airbus A320-100/200					AB	8846			19:20	20:40	17:55	19:15	166.55
36	Air Berlin	AB	8846	Airbus A320-100/200					AB	8846			19:20	20:40	21:20	22:45	168.77

4.5 Research Questions – Breakdown

In section 4.1, a set of research questions relevant for the case study was presented. Based on the available data, a further breakdown (Table 4-4) of these research questions is made:

- Research question Q10 is divided into 11 sub-questions, based on the service attributes extracted from the different datasets, most notably the Official Airline Guide;
- Research question Q11 is divided into 3 sub-questions, based on the information of the traveler at the moment of booking;
- Research question Q12 is not divided into further sub-questions;
- Research question Q13 will be briefly discussed in the next chapter, but is not considered into detail;
- Research question Q14 follows from research question Q10 and Q12, as this the listings contain service characteristics sorted in a certain order.

Table 4-4 Breakdown research questions case study

Itinerary Choice

Q10	What is the valuation of travelers for *service characteristics* (i.e. departure time, transfers, carrier image)?
Q10a	What is the relative valuation of *type of carrier*?
Q10b	What is the relative valuation of *carrier*?
Q10c	What is the relative valuation of *frequency*?
Q10d	What is the relative valuation of *type of aircraft*?
Q10e	What is the relative valuation of *code-share*?
Q10f	What is the relative valuation of *outbound departure time*?
Q10g	What is the relative valuation of *inbound departure time*?
Q10h	What is the relative valuation of *transfers*?
Q10i	What is the relative valuation of *waiting-time*?
Q10j	What is the relative valuation of *in-vehicle time*?
Q10k	What is the relative valuation of *total travel time*?
Q11	Is it possible to define a *customer profile* (segmentation) based on observable criteria of the traveler?
Q11a	What is the effect of *booking period* on traveler decision-making?
Q11b	What is the effect of *day of week* on traveler decision-making?
Q11c	What is the effect of *duration of stay* on traveler decision-making?
Q12	What is the influence of *fare* on traveler decision-making?
Q13	Which *pricing strategies* do my competitors pursue?
Q14	What are *possible improvements* for the listings on the website?

4.6 Conclusions

In this chapter the data available for this research was presented. Choice sets are determined by the available itineraries on the day of booking as listed on Expedia.

In addition to fare, several characteristics are extracted from the datasets. These are: in-vehicle time, waiting time, number of transfers, code share, type of carrier, carrier, departure time, duration of stay (days and minutes), type of aircraft, departure day of week and the difference between booking date and departure date. Put in the context of a traveler's decision attributes as listed in section 2.8, Table 2-7, most of these attributes relate to the transport attributes. In Chapter 2 it was argued that a traveler's trip purpose influenced his choice. However, no information is known on the trip purpose, origin and destination airport choice. Instead information is available on days before booking, duration of stay and departure day of week. This information coincides, not coincidentally, with information given by travelers (or the decision-maker) when booking an itinerary. Together with the fare of an itinerary this is information not considered in previously studies, as discussed in Chapter 2. Therefore, the evaluation of the effect of these variables on itinerary choice can contribute to current studies.

Finally, it should be noted that with the given data, it is also possible to follow a different modeling approach: with the observed bookings in the MIDT dataset and either using routes generated by Expedia or another route generation algorithm, different types of choice sets can be constructed. It is not possible to directly include fare in these choice sets. This would be similar as the approach followed by Coldren (2003). With this approach, network planning needs of airlines can be addressed.

Chapter 5 Descriptive Statistics of Datasets and Choice Sets

5.1 Introduction

The previous chapter presented an overview of the datasets available to answer the set of the research question regarding itinerary choice. In this chapter, an analysis of the datasets is presented. The exploration of the characteristics of the chosen and non-chosen itineraries is carried out to gain insight into variables influencing itinerary choice and answer the sub-questions presented in Table 4-4. Furthermore, the effect of different choice sets is to be evaluated. Three choice set types will be highlighted as shown in Table 5-1. These are all confined to a single booking occasion. Type II and IV will be compared on choice set size and fare. Other types are not considered, because choice sets become very large; Expedia lists 50 or more alternatives on the day of booking. If a passenger acquires information multiple times, again the choice set of a passenger will become very large. It is therefore questionable if passengers will actually consider such large choice sets; more should be known about booking behavior and passenger preferences to add itineraries available on other days than the booking day.

Table 5-1 Considered ranges of dimensions

Choice Set Type	Booking time dimension	Information acquisition dimension	Preferred arrival time dimension	Analysis
III	Single booking consideration	One carrier	Low preference	Choice set size
II	Single booking consideration	Multiple carriers	High preference	Choice set size, fare
IV	Single booking consideration	Multiple carriers	Low preference	Choice set size, fare, and service characteristics.

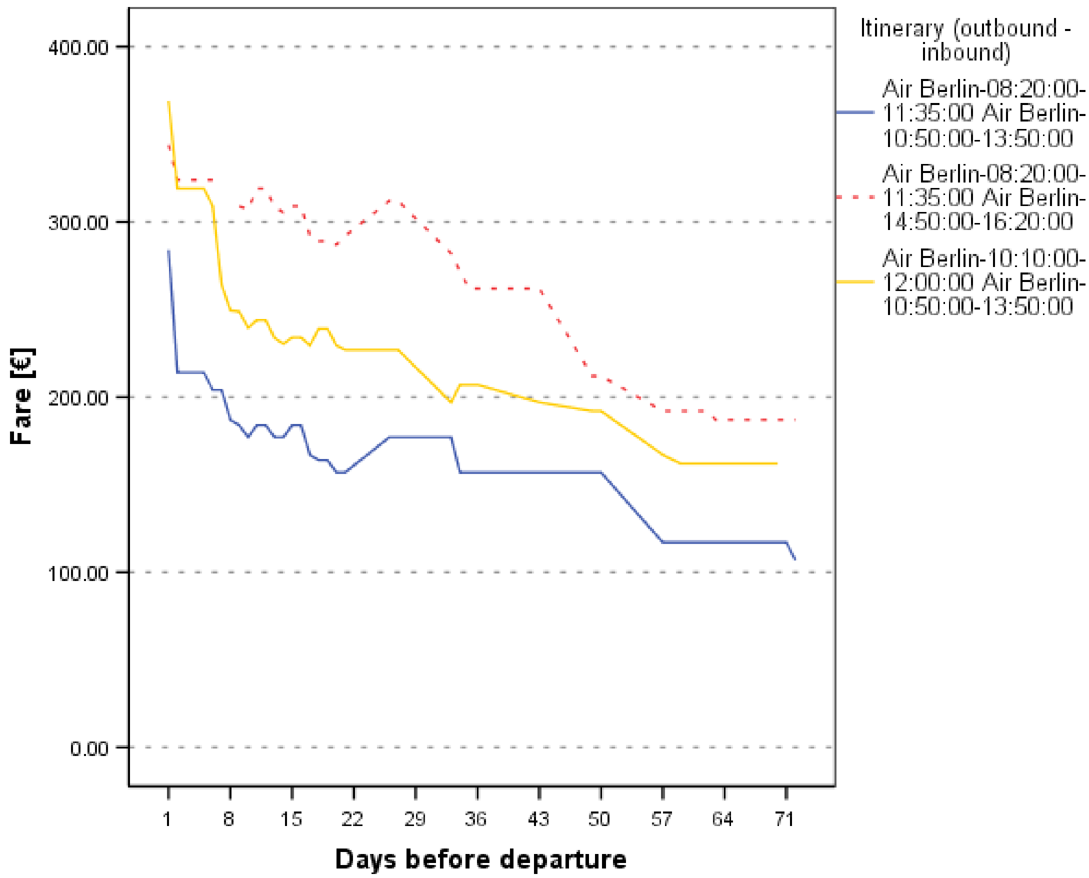
First, attention will be paid to characteristics independent of the choice set, then a comparison will be made between the fare in two types of choice sets. The latter paragraphs contain descriptive statistics about choice set type IV.

5.2 Fare setting of airlines

Because the webbots queried Expedia on a nearly daily basis, it is possible to investigate fare strategies of airlines. Two simple examples are shown in Figure 5-1 and Figure 5-2. Generally speaking, it can be said that fares differ strongly per weekday and duration of stay. In addition, Hüni and Merz (2007) found out that fares differ per code share and type of airline. For instance, the operating airline offers a lower fare as the partner-airline. Furthermore, they found that the variation in price is smaller between flights returning the same day or the next day than the flights returning within two weeks. Also, they made an analysis of booking trends and fares. The fares of Lufthansa, Swiss and British Airways itineraries remain fairly constant up to 25 days before departure, before adjusting their fares. The increase in price differs per origin-destination pair and carrier.

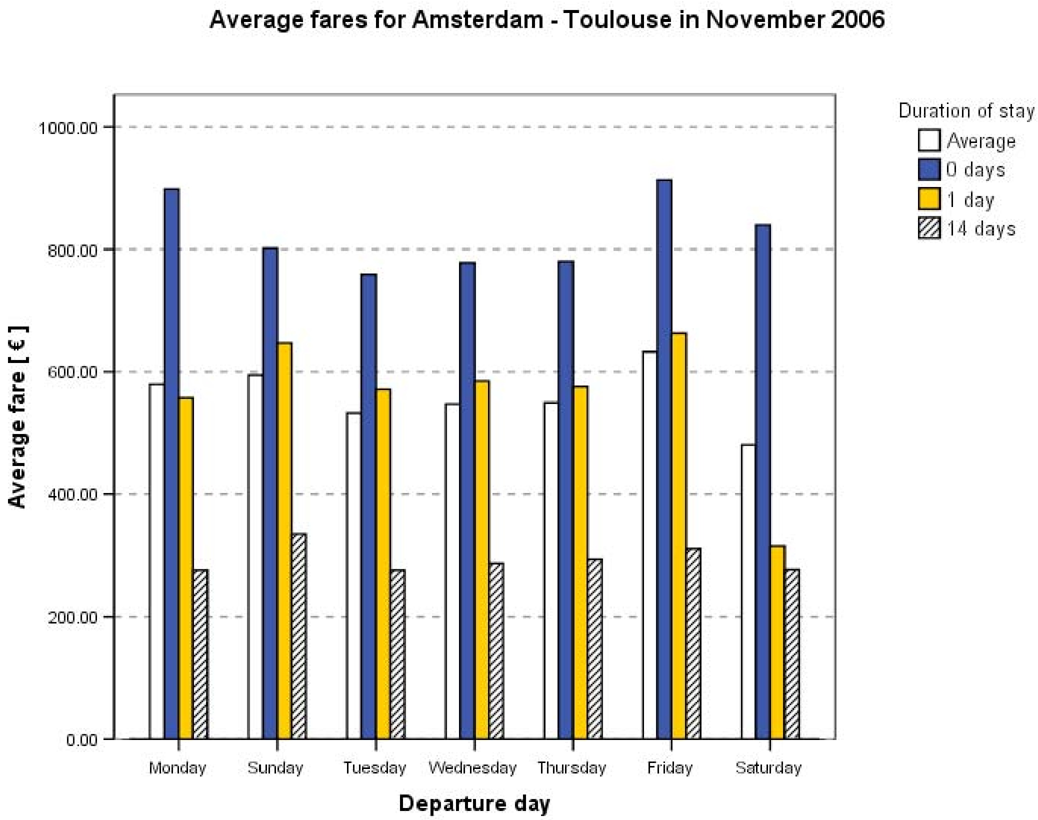
Figure 5-1 shows the fare setting of three Air Berlin itineraries for the route Hamburg-Barcelona. It can be seen that the fares for the itinerary follow approximately the same trend. There is however a difference in the fare-levels. The most expensive flight departs in the early morning and returns in the afternoon of the next day. Second is the itinerary departing in the late morning and returning the next in the late morning of the next day. The third itinerary has the same return flight as the second flight, but departs in the early morning and is less expensive.
The difference in fare is fairly large, being approximately € 70,-.

Figure 5-1 Fare setting Hamburg – Barcelona by Air Berlin flights departing 23-11-2006



A second example of fare setting is shown in Figure 5-2. In the figure, average fares per weekday and stay category are presented. First, it can be seen that the fare differs per duration of stay, returning the same day being more expensive than returning the next day, and returning in a fortnight being less expensive as returning the next day. Travelers departing on Saturday and not returning the same day, pay the lowest fare on average. Returning the same day is cheapest on Sunday.

Figure 5-2 Average fares Amsterdam – Toulouse per weekday and duration of stay



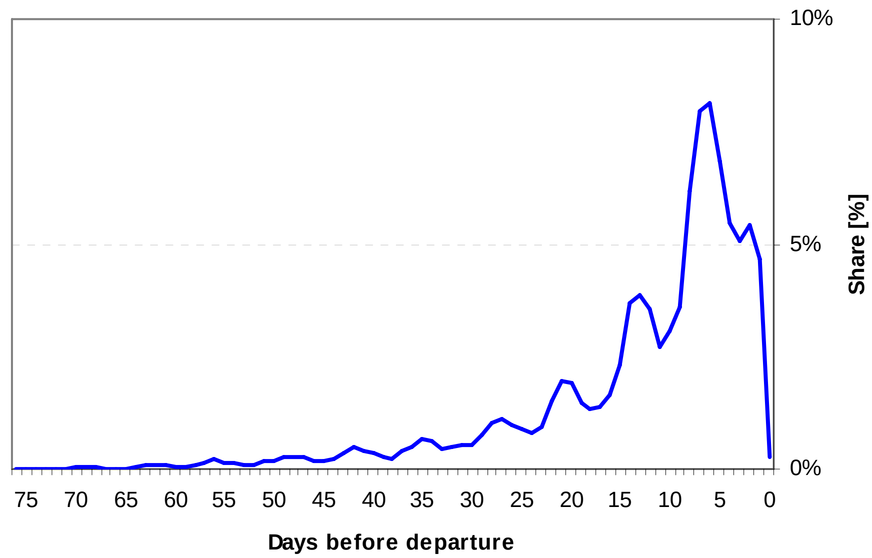
5.3 Duration of Stay, Days in Advance of Booking, and Days of Week

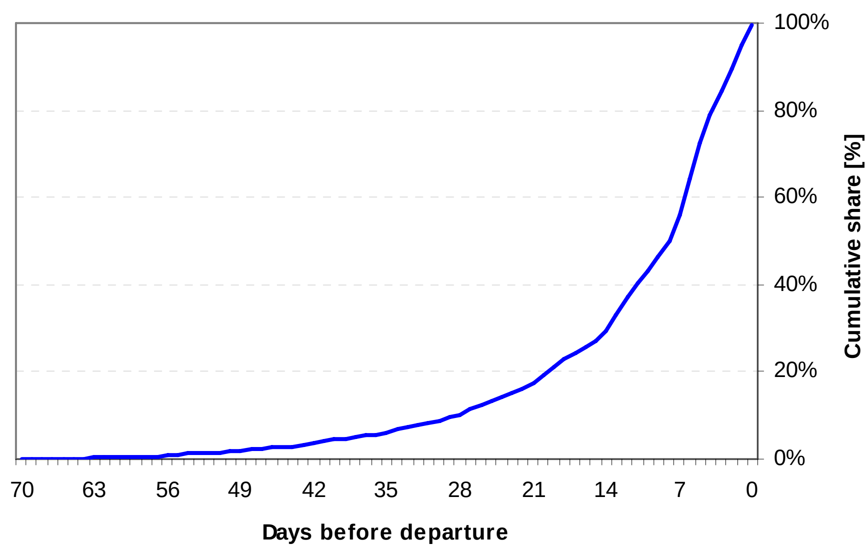
Three durations of stay were queried on Expedia: flights returning the same day, flights returning the next day, and flights returning in two weeks. The observed bookings in the MIDT dataset are added to three stay categories, corresponding with the Expedia queries. In Table 5-2 the number of observations per stay category can be seen. Most passengers return the same day, almost 40% of the passengers return the next day and only 6% stay longer than 6 days. Compared to the number of bookings per stay category (Appendix C.5), the relative number of bookings in stay category 0 and 1 is much higher. This can be because the number of possible itineraries is smaller for stay category 0 and 1, which leads to a higher probability of a similar itinerary occurring on Expedia.

Table 5-2 Bookings per stay category

Observed bookings in MIDT – duration of stay	Count	Cumulative count	Percentage	Cumulative percentage
0 days	10537	10537	55.77%	55.77%
1 day	7300	17837	38.63%	94.40%
> 6 days	1058	18895	5.60%	100.00%

Figure 5-3 Days in Advance of Booking





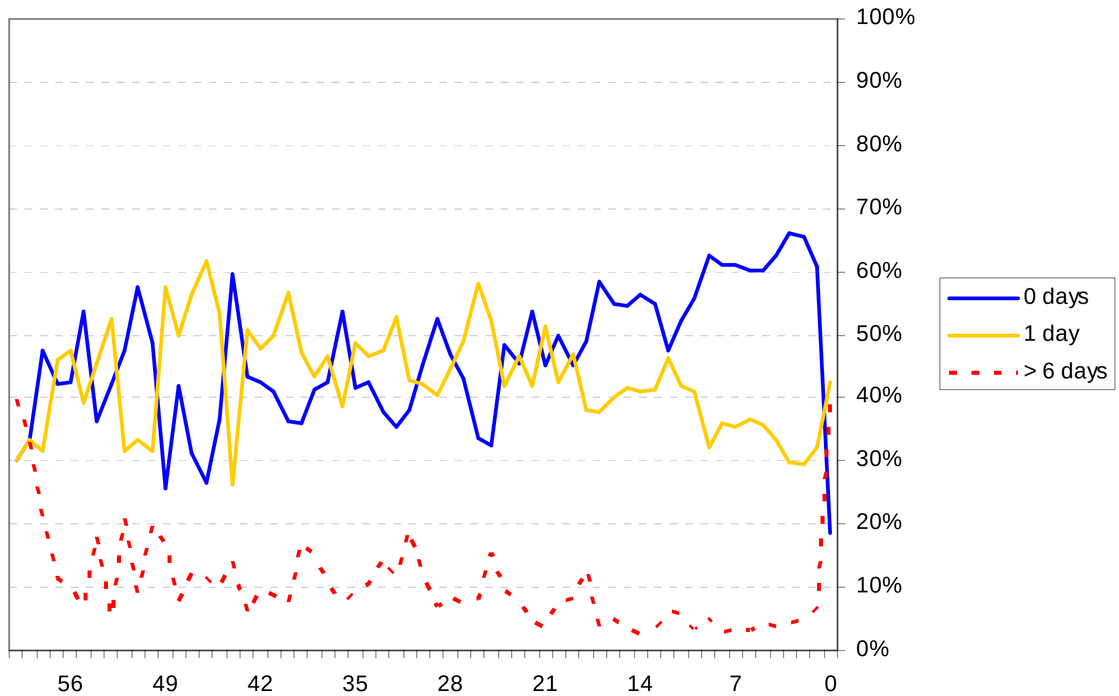
N_{obs}= 18,895

Figure 5-3 shows the percentage of tickets booked per day. The following observations are made:

- 3 days before departure 85% of the tickets were booked.
- 50% of the tickets are booked up to 8 days before departure
- 5% of the tickets more than 36 days in advance.

It can be seen that several peaks occur. No analysis has been carried out on these peaks. It can however be seen that these occur with a frequency of approximately 7 days. In Figure 5-4 the distribution of booked tickets per stay category per number of days before departure is shown. The expected pattern can be observed: travelers staying a short time at their destination book a short time in advance, whereas persons staying at their destination longer book further in advance. Remarkable is the increase in the number of booked tickets just before departure for a longer duration of stay (> 6 days). One explanation might be because that these tickets are actually cheaper to book than tickets returning the same day and can be the result of irrationalities in pricing systems (Garrow, *et al.* 2007) and is also confirmed by a very large number of one-way bookings (or a very long duration of stay) observed in the MIDT-dataset (Appendix C.2.1). These one-way bookings are not considered in the analysis.

Figure 5-4 Share of booked tickets per days before departure per duration of stay



$N_{obs,1}$=10,537, $N_{obs,2}$= 7,300, $N_{obs,3}$= 1,058

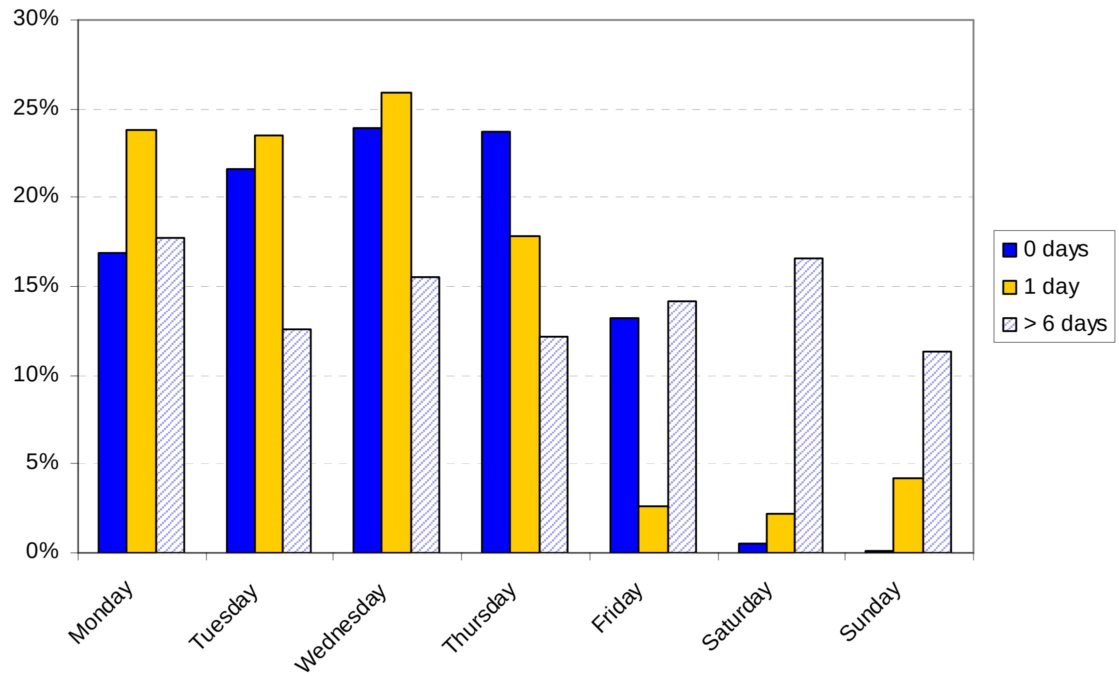
In Figure 5-5 the departure days are depicted per stay category, exact figures are shown in Table 5-3. From the figures presented in the table, the following may be derived:

- Passengers returning the same day, depart on weekdays, thus being home during weekends;
- Passengers returning the next day exhibit a preference for departing from Monday to Thursday, thus being home on Saturday and Sunday.
- Passengers staying longer at their destination do not show a clear preference for a departure day; departures are spread more or less evenly across the week.

Table 5-3 Departure days per week per duration of stay

Day	0 days		1 day		> 6 days	
	Count	Percentage	Count	Percentage	Count	Percentage
Monday	1783	17%	1742	24%	187	18%
Tuesday	2272	22%	1712	23%	133	13%
Wednesday	2524	24%	1892	26%	164	16%
Thursday	2502	24%	1300	18%	129	12%
Friday	1394	13%	189	3%	150	14%
Saturday	50	0%	161	2%	175	17%
Sunday	12	0%	304	4%	120	11%
Totals	10537	100%	7300	100%	1058	100%

Figure 5-5 Departure days of week per duration of stay



$N_{obs,1}$=10,537, $N_{obs,2}$= 7,300, $N_{obs,3}$= 1,058

5.4 Choice Set Size

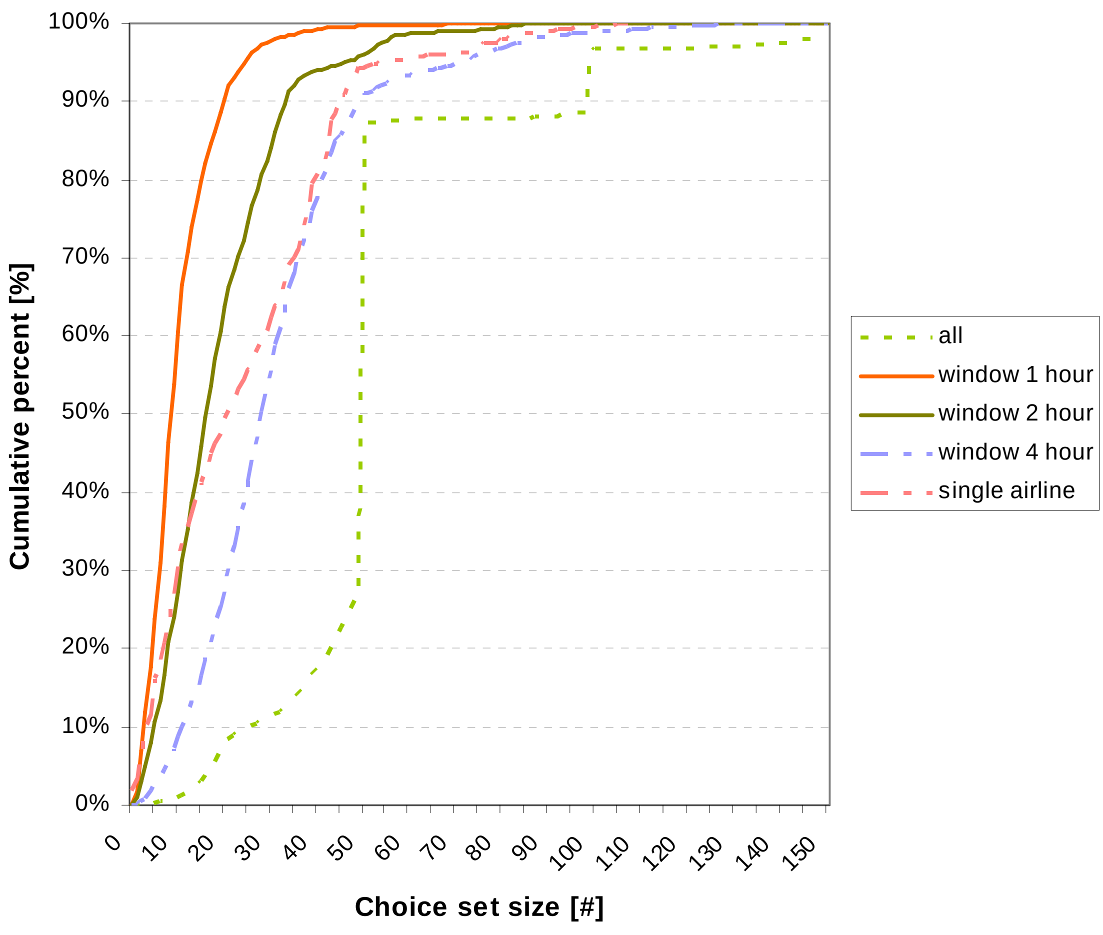
In this section a look will be given to the influence of the arrival time dimension on choice set size and information acquisition dimension:

- The arrival time dimension is defined as a window: around each chosen itinerary an arrival time window is defined, which includes all itineraries and carriers arriving up to n hours earlier or later.
- The information acquisition dimension concerns the number of carriers in the choice set. In this case, a differentiation is made between a single carrier (choice set type II) and multiple carriers (choice set type III and IV).

In Figure 5-6 the choice set size is depicted. It can be seen that the choice set size steadily increases if the window is enlarged. If the window is set to 1 hour, 90% of the choice sets contain 20 alternatives or less, if the window is set to 2 hours this number becomes 30. A window of 4 hours leads to even larger choice sets. If a passenger considers all flights departing on the same day, a choice set can contain up to 150 flights, 60% of the choice sets will contain 50 alternatives. A jump can be observed in the choice set size of latter category. This because in some cases, Expedia returns more itineraries than the 50 it usually does.

With regard to the information acquisition dimension, the following can be derived from Figure 5-6. A large percentage of the choice set consists of a single airline. For instance, 50% of the choice sets have a size of 25 if a single airline is considered.

Figure 5-6 Choice set size as a function of time window considered

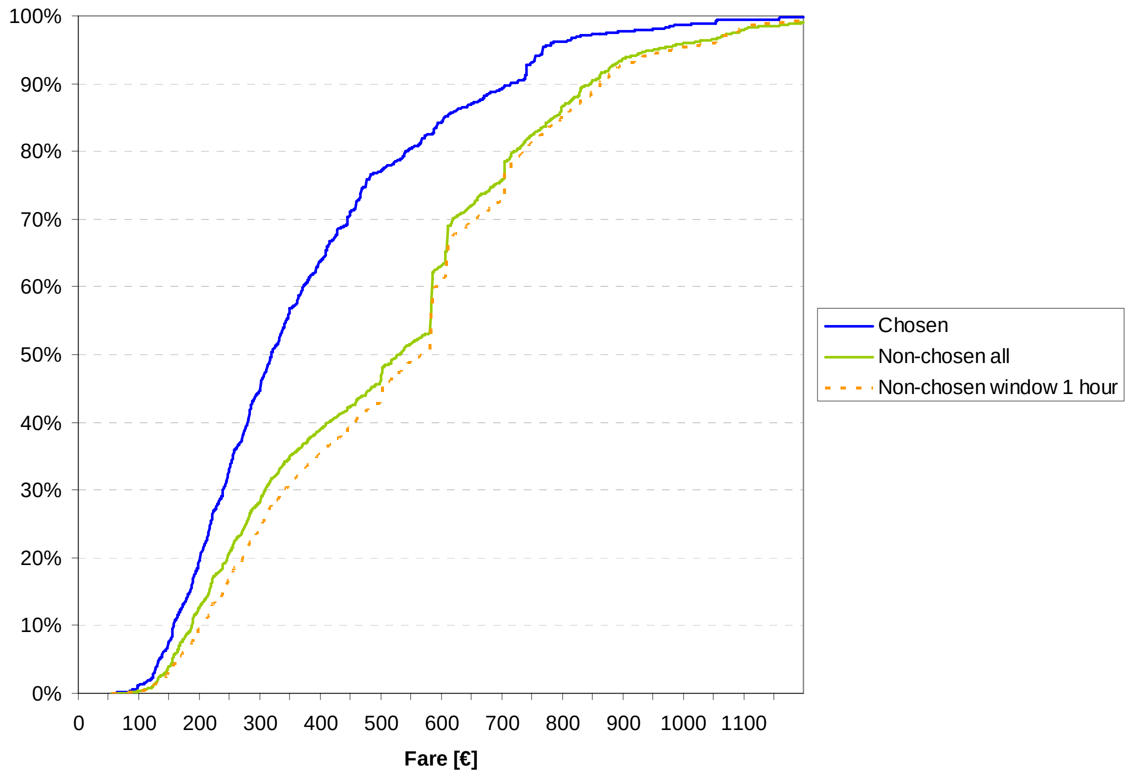


5.5 Fare

In Figure 5-7, the cumulative distribution of the fare of the chosen alternatives versus the fare of non-chosen alternatives can be seen. On first sight, it seems that the chosen alternatives are chosen based on fare, as the cumulative distributions of the fare of the non-chosen alternatives lie lower as the cumulative distribution of the chosen fare. However, the graph only shows that there were higher

fares offered on all routes. Also, the graph shows that the fares offered remain fairly constant, despite the time window, as the lines of the different time windows do not vary much.
Therefore, a closer look is given to the distribution of fares in the choice set.

Figure 5-7 Fare in chosen and non-chosen itineraries



In Figure 5-8 and Figure 5-9, again the chosen fare versus the non-chosen fares is depicted. This time however, two subsets are extracted from each choice set: a subset that contains the alternatives with fares lower as the chosen fare and a subset that contains fares higher than the chosen fare.
Figure 5-8 shows the average fare in the two subsets. For example, if the chosen fare is € 420,-, the average of the lower fares in the choice set is € 307,- and the average of the higher fares in the choice set is € 606,-. Figure 5-9 shows the percentage of the itineraries in the two subsets. For instance, for a chosen fare of € 50,- 100% of the itineraries in the choice set are more expensive than the chosen itinerary.
It can be seen that the average fare in the lower subset indeed equals the chosen fare in the case of the lower chosen fares. As the chosen fare increases however, it can be seen that lower fares are available. Even if the window is limited, chosen fares do not equal the average of the lowest fare. Knowing this, it can be said that fare is not always the decisive criterion for an individual.

Figure 5-8 Fare in chosen and non-chosen itineraries per fare category

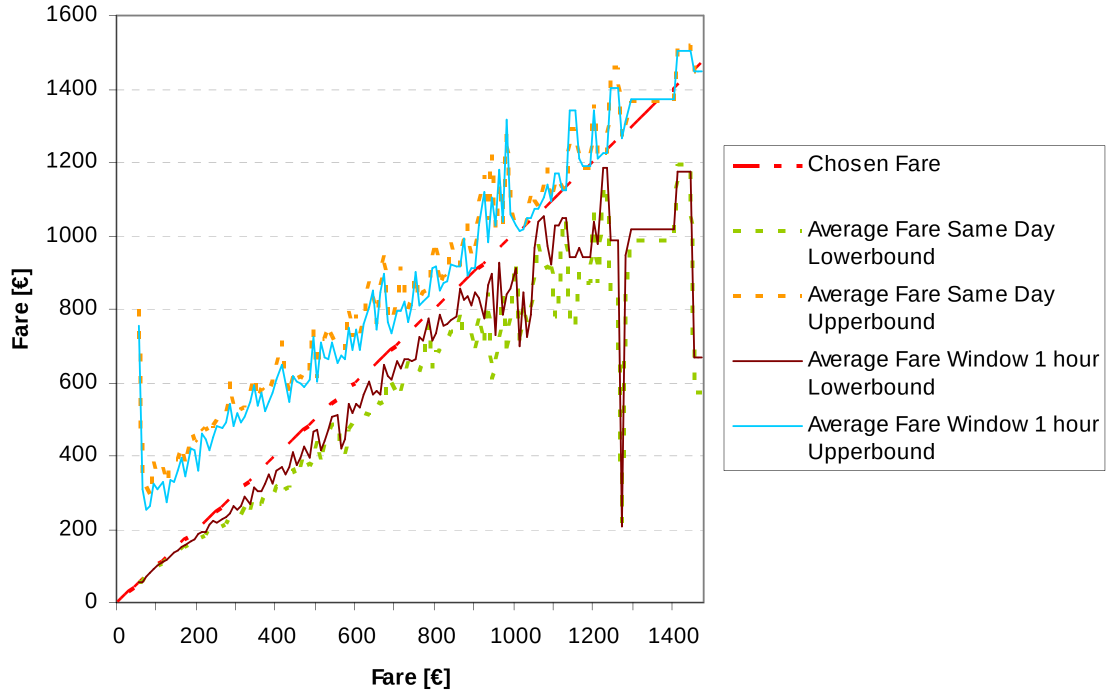
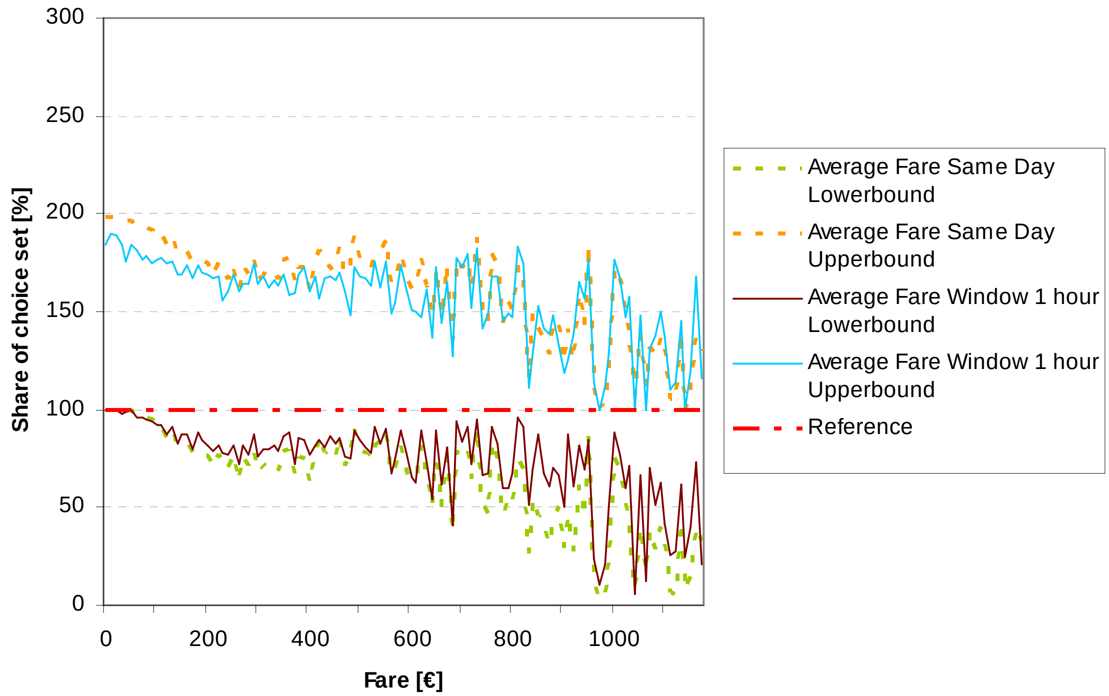


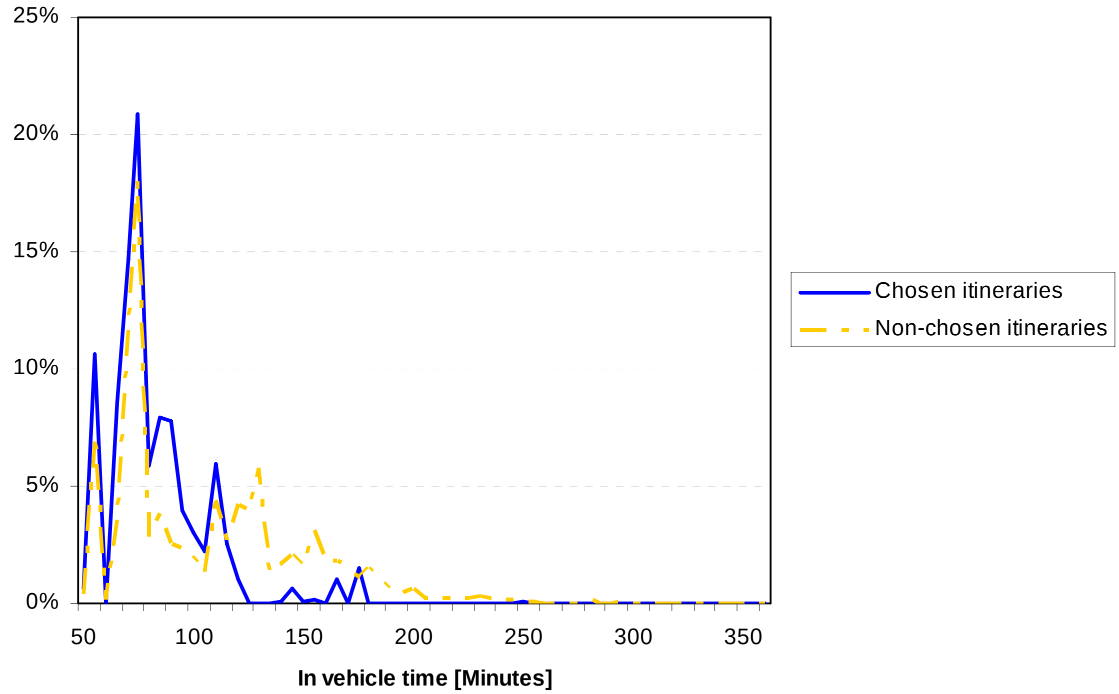
Figure 5-9 Fare in chosen and non-chosen itineraries relative to choice set



5.6 In-Vehicle Time, Waiting-Time & Transfers

In Figure 5-10 the distribution of the in-vehicle time of chosen and non-chosen itineraries can be seen. Up to an in-vehicle time of approximately 130 minutes the distribution is approximately equal. Longer in-vehicle times in the non-chosen itineraries can be explained by the fact that a part of the non-chosen itineraries are itineraries containing a transfer.

Figure 5-10 Frequency of in-vehicle time in chosen and non-chosen itineraries



N_{obs}= 18,895, $N_{non\text{-}chosen}$= 968,352

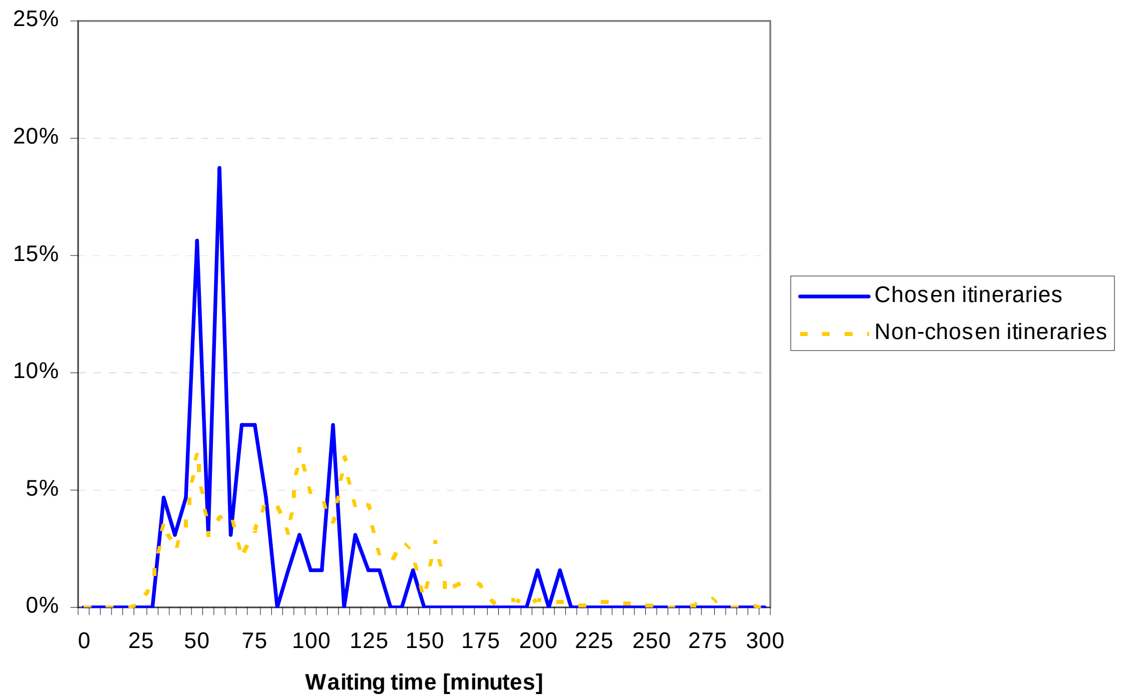
Figure 5-11 shows the waiting time of the itineraries with a transfer is chosen; it can be seen that a shorter waiting time is preferred. It should be mentioned that the number of chosen itineraries that contain a transfer is very low. This is shown in Table 5-4. Only 75 individuals are observed who chose a transfer. However, 42% of the offered itineraries contain a transfer. An initial analysis of the number of transfers on all European bookings in November 2006 revealed that the number of transfers on intra-European flights is higher: approximately 23% of the bookings contain one transfer. This is also shown in Appendix C.2.3.

The correlations between the outbound itineraries are all very low, the only notable exception being the number of transfers, which correlates positively with flag carriers and regional carriers and negatively with low cost carriers. However, these correlations are fairly small (< 0.15).

Table 5-4 Number of Transfers in Chosen and Non-chosen itineraries

	Chosen itineraries		Non-chosen itineraries	
	Count	Percentage	Count	Percentage
0	18820	99.60%	560035	57.83%
1	75	0.40%	408317	42.17%

Figure 5-11 Frequency of waiting time in chosen and non-chosen itineraries

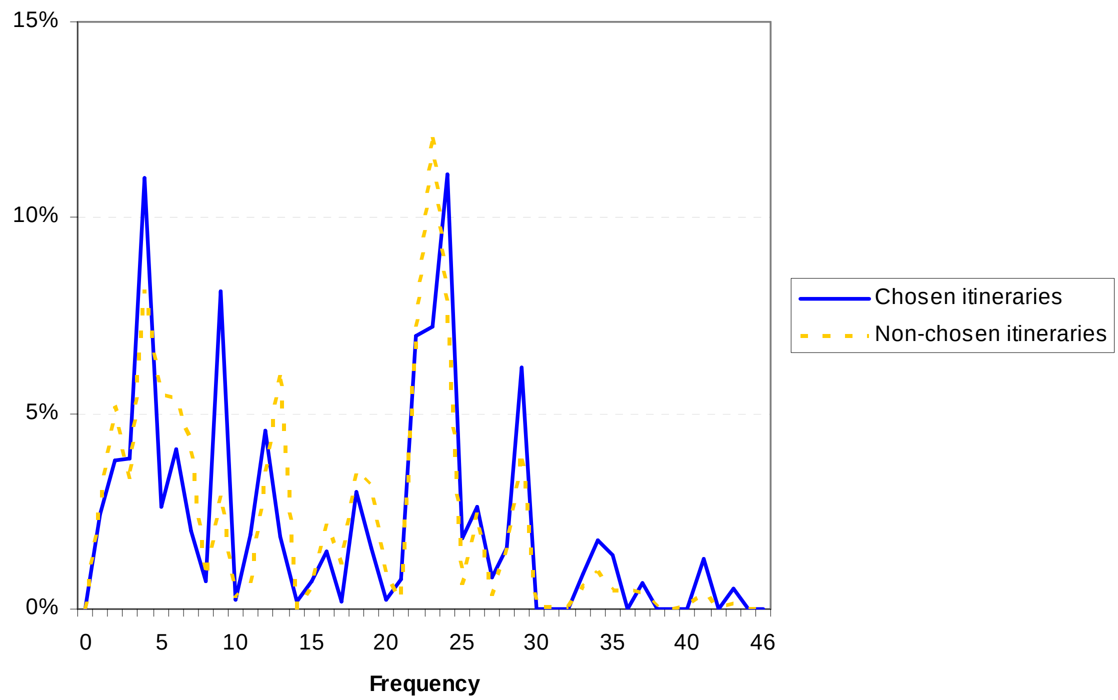


N_{obs}= 75, $N_{non\text{-}chosen}$= 408,317

5.7 Frequency

Frequency is defined as the number of distinct flight numbers of a carrier departing on an origin-destination pair on a certain day. Following this definition, the most remarkable conclusion is that some of the observed frequencies are very high; this is for the following reasons. First, on some origin destination pairs the number of code share flights offered is very high. Second, carriers often offer a direct flight and a flight with a transfer possibility. The distribution of frequency followed for chosen and non-chosen itineraries is approximately the same as can be seen in Figure 5-12.

Figure 5-12 Frequency of carrier in chosen and non-chosen itineraries



N_{obs}=18,895, $N_{non\text{-}chosen}$= 968,352

5.8 Carrier Characteristics

This section presents an analysis of carrier characteristics and carrier. A distinction is made between three types of carriers:

- Flag carrier;
- Regional carrier;
- Low-cost carrier.

Furthermore, a distinction is made between:

- Code-share flights;
- Non-code share flights.

Finally, a differentiation is made between:

- Domestic carrier or home carrier (country of departure is equal to the home country of the carrier);
- Non-home carrier.

In Figure 5-13 these characteristics are displayed, with respect to the frequency of their occurrence in chosen alternatives and non-chosen alternatives.

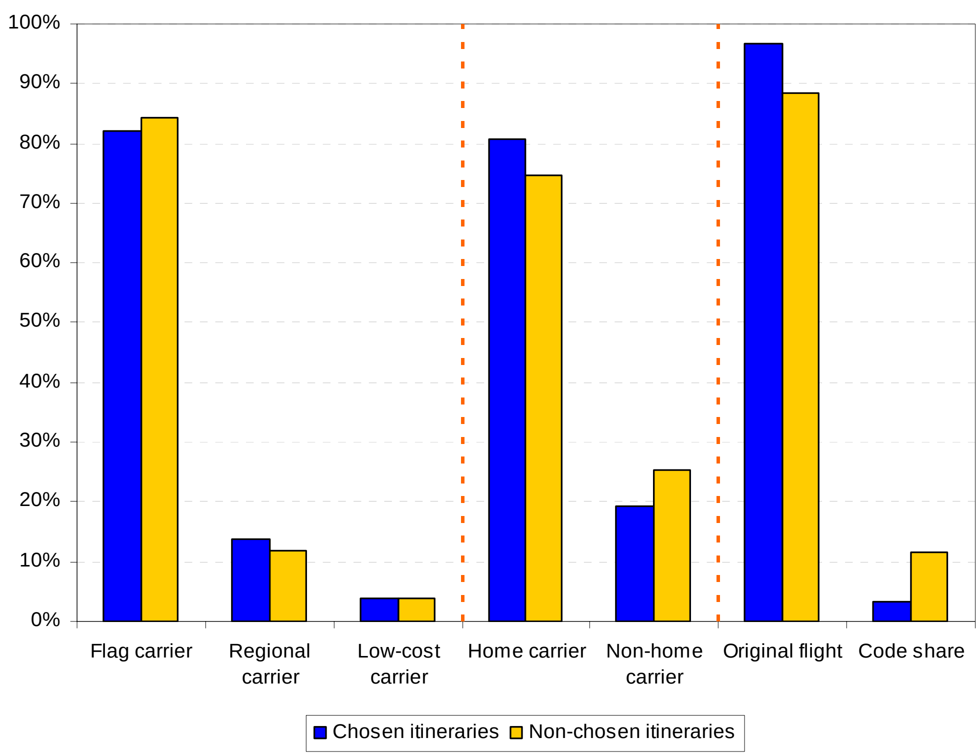
Most flights are operated by a flag carrier, both in the chosen and non-chosen alternatives. Over 80% of the alternatives fall in this category. Regional carriers are chosen slightly more than they are offered; the same holds for low cost carriers. The latter category forms only 4% of the alternatives. The number of low cost carriers might be low, because most of the low cost carriers only offer tickets on their own websites and are thus not represented in this data set. Furthermore, Expedia does not offer all low cost carriers, such as German Wings, which is included in the MIDT dataset.

Domestic carriers are dominant in the chosen alternatives; they are chosen more than they are offered. The opposite holds for non-domestic carriers.

Another preference that can be observed is the preference for flights operated by the original carrier. One explanation can be that the a flight is often operated by the domestic carrier, the carrier of the destination country than offers the same flight under a code share agreement. However, it could also be the case that travelers prefer the original carrier, as this carrier is often cheaper (Hüni & Merz 2007).

Flag carriers show a positive correlation with flights departing at 7:00, for low cost and regional carriers this correlation is negative. Itineraries departing at 8:00 however, have a positive correlation with low cost and regional carriers and a negative correlation with flag carriers.

Figure 5-13 Carrier characteristics of chosen and non-chosen itineraries



N_{obs}= 18,895, $N_{non\text{-}chosen}$= 968,352

Table 5-5 Airline in Chosen and Non-chosen itineraries

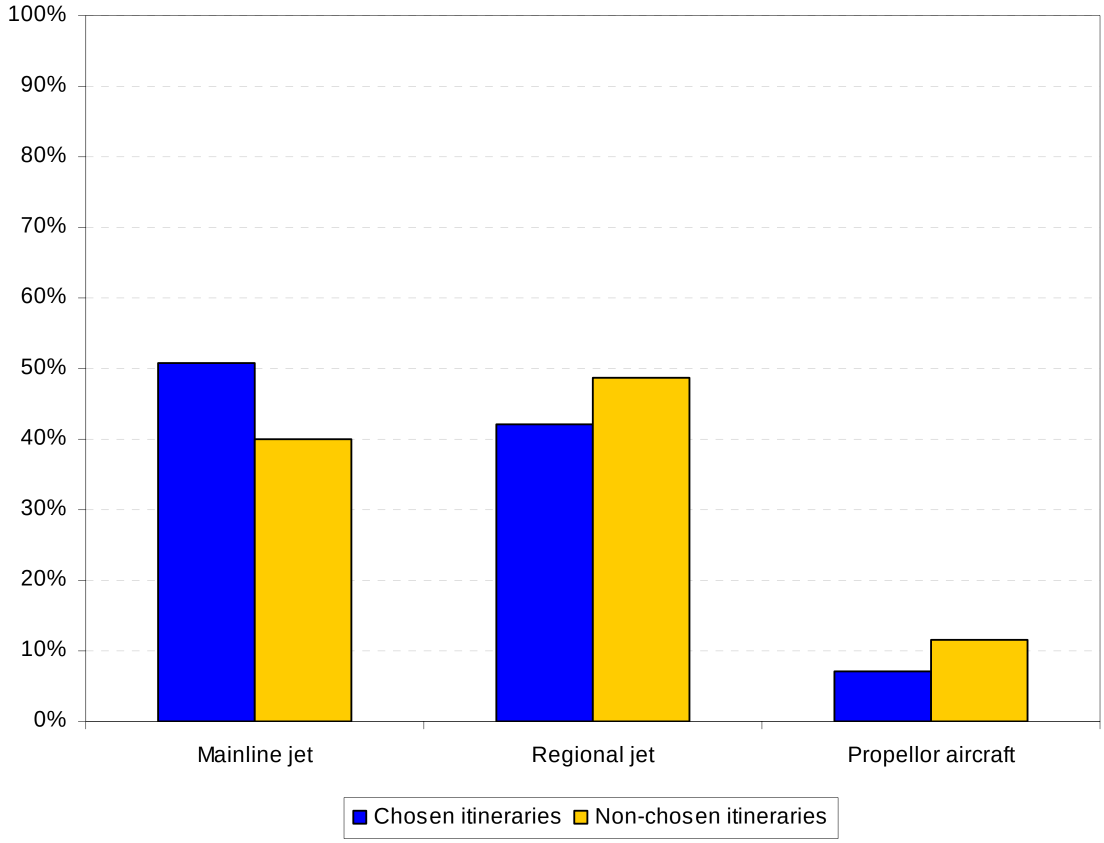
Airline code	Airline name	Chosen itineraries		Non-Chosen itineraries	
		Count	Percentage	Count	Percentage

5.9 Type of Aircraft

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

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

Figure 5-14 Type of aircraft in chosen and non-chosen itineraries



N_{obs} = 18,895, $N_{non\text{-}chosen}$ = 968,352

5.10 Departure time

A further distinction between itineraries is their departure time. In this study, itineraries are aggregated by hour and per stay category, i.e. 5:00 – 5:59, 6:00 – 6:59. A higher level of aggregation can then be made in following steps.

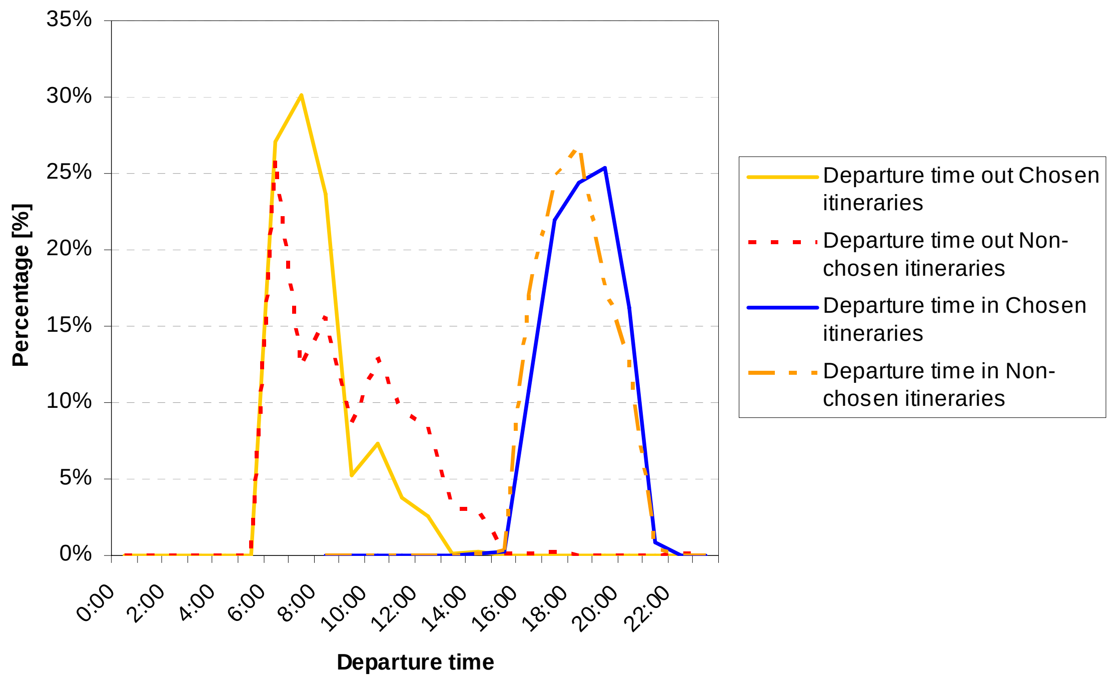
In Figure 5-15, Figure 5-16 and Figure 5-17 the departure times of the chosen itineraries per hour and stay category can be seen.

Figure 5-15 shows the departure times of itineraries returning the same day. Most chosen itineraries returning on the same day depart in the period 6:00 – 9:00 and return between 16:00 and 22:00. It can be seen that non-chosen itineraries are distributed somewhat more evenly than the chosen itineraries.

Figure 5-16 shows the departure time of itineraries returning the next day. Most chosen itineraries returning on the next day depart in the period 6:00 – 9:00. A second peak can be observed during observed in the period 16:00 – 19:00. Again, the non-chosen itineraries are distributed somewhat more evenly than the chosen itineraries.

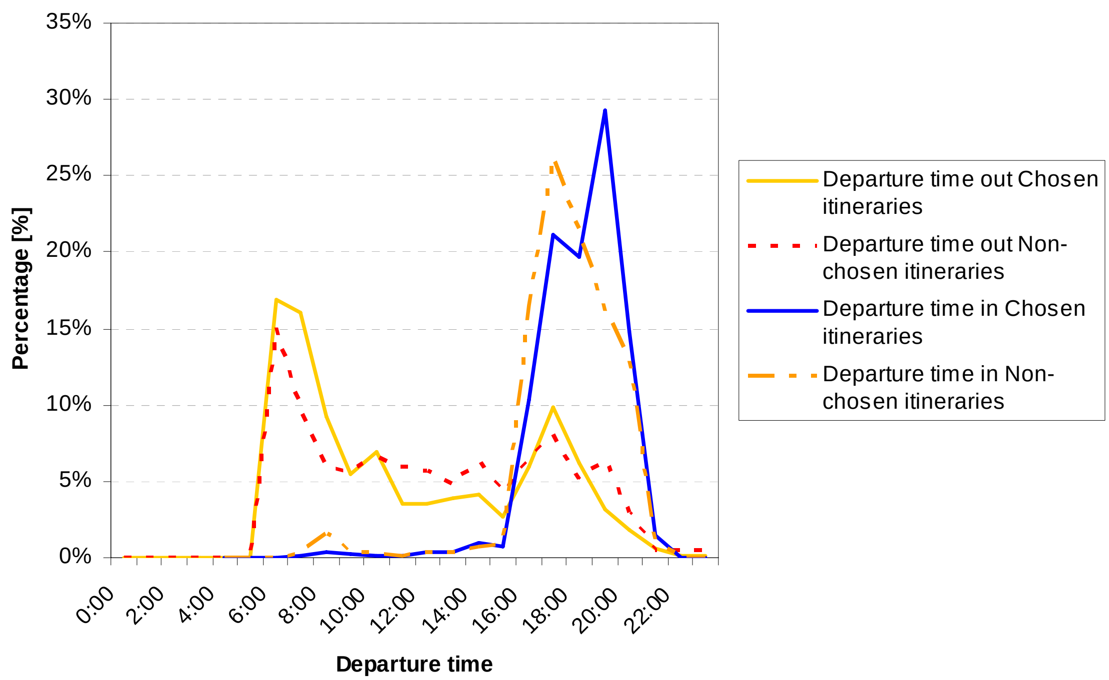
Passengers staying at their destination longer as six days do not show clear preference at first sight for the departure time of the outbound itinerary. However, a morning peak can be observed for the outbound itineraries and a peak in the late afternoon and early afternoon can be observed for the inbound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

Figure 5-15 Departure time of chosen and non-chosen itineraries – duration of stay 0 days



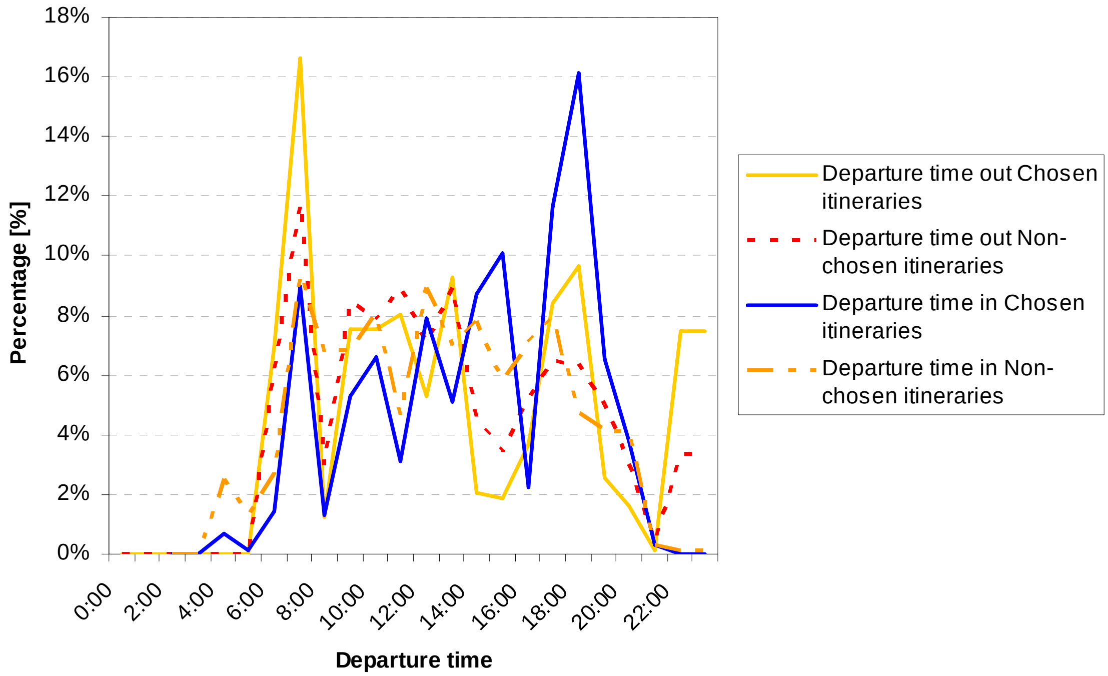
N_{obs}= 10,537, $N_{non\text{-}chosen}$ = 546,939

Figure 5-16 Departure time of chosen and non-chosen itineraries – duration of stay 1 day



N_{obs}= 7,300, $N_{non\text{-}chosen}$ = 374,313

Figure 5-17 Departure time of chosen and non-chosen itineraries – duration of stay > 6 days

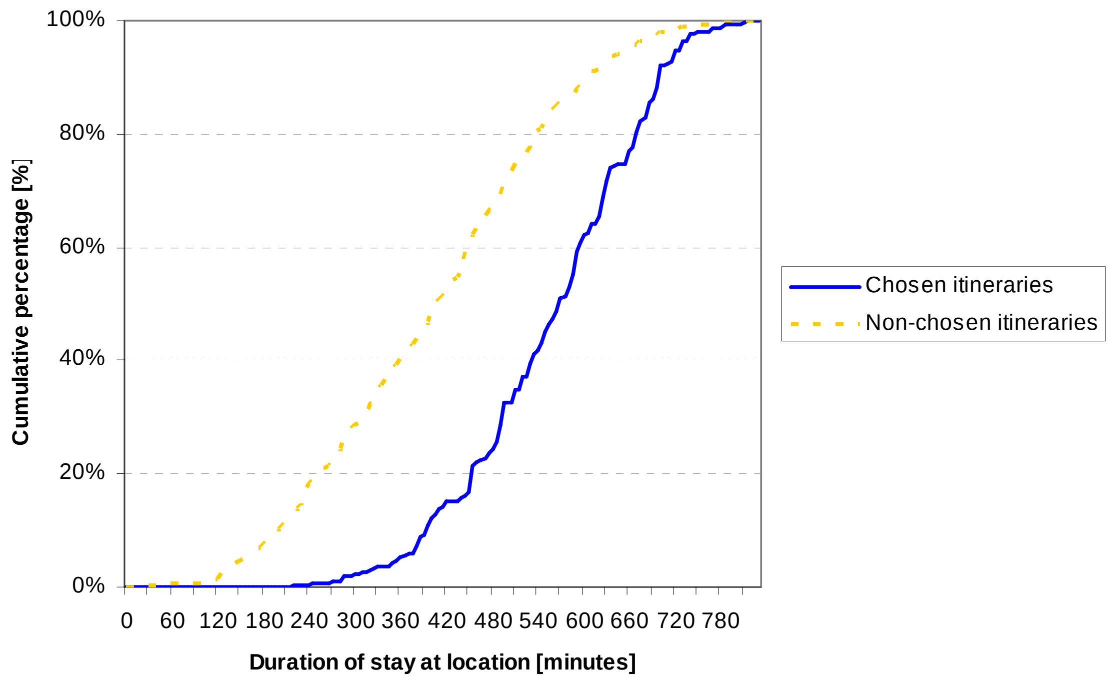


N_{obs}= 1,058, $N_{non\text{-}chosen}$ = 47,100

For itineraries including an overnight stay or a stay of multiple days, it can be argued that departure time is more important. However, for itineraries returning the same day, the time between arrival at the destination airport and departure from the destination airport can influence a traveler's decision more.

In Figure 5-18 the time difference in minutes between the arrival time of the outbound itinerary and the departure time of the inbound itinerary is shown for travelers returning the same day. It can be seen that in the chosen itineraries, the duration of stay is longer than in the non-chosen itineraries. A preliminary conclusion which can be drawn from this is that the search algorithm of Expedia takes into account a minimum stay. This minimum stay is however fairly short as compared to the minimum stay in the chosen alternatives.

Figure 5-18 Difference arrival time outbound itinerary–departure time inbound itinerary s.c. 0



N_{obs}= 10,537, $N_{non\text{-}chosen}$ = 546,939

Itineraries returning the same day correlate positively with the duration of stay if the itineraries depart between 6:00 and 8:00 and negatively if the itinerary departs later. The departure time of the inbound flight correlates negatively for flights departing between 16:00 up to 20:00 and positively after. Both positive correlations are between 0.3 – 0.4; negative correlations are in the range -0.1 - -0.3.

Itineraries returning the next day correlate positively with the duration of stay if the itineraries depart between 6:00 and 8:00 and negatively if the itinerary departs later, most notably between 17:00 and 20:00. The departure time of the inbound flight correlates positively for flights departing between 18:00 up to 20:00 and positively after. Positive correlations are between 0.4 – 0.5 and 0.1 respectively; negative correlations are in the range -0.1 - -0.3, where the correlations with the inbound flight are smaller.

Itineraries returning after 6 days have very small correlations with the duration of stay. Also, correlations between outbound hour of day and inbound hour of day are smaller as the correlations with duration of stay.

5.11 Conclusions

In this chapter, an analysis has been performed of chosen itineraries and non-chosen itineraries. First, the effect of different departure time preferences has been evaluated through the usage of departure time windows. Choice set size is influenced by the time window chosen; choice set size is reduced by the addition of a time constraint. Fare does play a role, as it seems that travelers chose one of the cheaper itineraries. This is however independent of time; departure time choice is not influenced by fare.

Furthermore, insights were gained in certain preference structures in generated objective choice sets, such as type of carrier, carrier and aircraft type and departure time. Some peculiarities in the route generation algorithm of Expedia have been discovered, such as the fact that itineraries are offered which return almost immediately. Furthermore, customers do not seem to choose itineraries with a different outbound and inbound carrier, despite these being offered in 20% of the non-chosen itineraries. In the chosen itineraries, only 1% of the itineraries have a different outbound than inbound carrier.

A transfer seems not to be preferred by travelers occurring in this dataset, with only 0.5% of the traveler opting for a transfer, whilst being offered in 42% of the itineraries. This is not the case for the entire of Europe. For a possible follow-up research, a table is made with origin-destination pairs and the number of transfers.

Moreover, a distinction has been made between three types of travelers based on their duration of stay. Travelers returning the same day depart on weekdays; travelers returning the next day depart from Monday till Thursday. Travelers remaining at their destination longer than six days do not show a clear preference for departure day. These three categories of travelers clearly reveal a different structure with regard to departure time; it is therefore recommended to make this distinction with regard to duration of stay and departure time.

In Chapter 6 and Chapter 7 utility function and estimated using MNL-models will be presented and analyzed, thereby providing more detailed information in preference structures and valuation of itinerary characteristics.

- This page left blank intentionally -

Chapter 6 MNL-model of Itinerary Choice

6.1 Introduction

From the analysis of the data and the generated objective choice sets in Chapter 5, insight has been gained in characteristics of booked itineraries and the alternatives available on the day of booking for the same duration of stay and origin-destination-pair.

In order to gain insight into travel behavior with regard to itinerary choice, the relative valuation of service characteristics and the role of fare, several utility functions have been specified and according choice models are estimated.

With the given data and constructed choice sets, it is possible to estimate itinerary choice models, based on the listing of Expedia on the day of booking and for itineraries available for the same departure and return date.

It will be shown that the relative weight of fare and a transfer is fairly high, as compared to carrier image and code-share. Furthermore, the role of departure time of both the outbound and inbound itinerary is important.

The attributes, their levels and relevant research questions will be presented in section 6.2. Section 6.3 discusses the data used for the estimation of the models. Model building is discussed in 6.4.1. Statistics of the best MNL-model are presented in 6.4.2. The relative weight of the parameter estimates is shown in section 6.4.4. More results of model estimation will be presented in ensuing sections.

6.2 Specification

6.2.1 Attributes and levels

In Table 6-1, the different attributes and their levels are presented. A distinction is made between the levels of chosen itineraries and non-chosen itineraries. Furthermore, the relevant research questions per attribute are listed in the last column. An extensive discussion of the available data is presented in Chapter 5. Section 6.2.2 to section 6.2.5 provide more precise definitions of type of carrier, carrier, aircraft type, code-share and departure time.

6.2.2 Type of Carrier and Carrier

A distinction is made between three types of carrier: low-cost airline, regional carrier and flag carrier. As it is not sure which preference structure travelers have for type of carrier, only the type of the first carrier of the outbound itinerary is incorporated in the utility function as a dummy variable. As a reference, the type 'flag carrier' is chosen. Parameter estimates for low cost carrier and regional carrier thus represent the relative valuation as compared to a flag carrier.

The same line of reasoning holds for the carrier itself. As it is not sure which is the preference structure for carrier, only the carrier offering the first flight is incorporated in the utility function as a dummy variable As a reference, a frequent chosen carrier is taken, namely Lufthansa. Parameter estimates for carrier thus represent the relative valuation compared to Lufthansa.

Table 6-1 Attributes, description and levels

Attribute	Level and definition	Chosen itineraries	Non-chosen itineraries	Research question
Type of carrier	Low-cost airline, regional carrier, mainline jet	Page 78, section 5.8		Q10a
Carrier	First carrier listed in itinerary	Page 78, section 5.8		Q10b
Frequency	Number of itineraries of a carrier departing on per origin-destination pair and day	Page 77, section 5.7		Q10c
Aircraft type	Propeller , regional jet, mainline jet	Page 80, section 5.9		Q10d
Code-share	Dummy variable indicating if the outbound itinerary contains a time share	2.0%	10.1%	Q10e
Departure hour outbound itinerary	Departure hour of the outbound itinerary (local time)	Page 81, section 5.10		Q10f, Q10g
Departure hour inbound itinerary	Departure hour of the inbound itinerary (local time)			
Fare	Fare of the itinerary as listed on Expedia on the booking day	€377,-	€511,-	Q12
Transfer	Indicates if the itinerary contains a transfer	0.40%	42.17%	Q10h
In vehicle time	Flight time in minutes of the itinerary	89 minutes	110 minutes	Q10j
Waiting time in itineraries with a transfer	Waiting time in minutes at transfer airport	89 minutes	112 minutes	Q10i
Total travel time	Sum of in vehicle time and transfer time in minutes	89 minutes	157 minutes	Q10k

6.2.3 Aircraft Type

A distinction is made between three types of aircraft: mainline jet, regional aircraft and propeller aircraft. If an outbound itinerary consists of multiple flights, it is assumed that a traveler will take into account the 'lowest' type of aircraft in the decision-making process. It is assumed, that a mainline jet is preferred above a regional aircraft and a regional aircraft is preferred above a propeller aircraft. Parameter estimates thus represent the relative valuation of a regional aircraft or propeller aircraft compared to a mainline jet.

6.2.4 Code-share

A distinction is made between itineraries containing a code-share flight and a non-code share flight. If an outbound itinerary consists of multiple flights, it is assumed that a traveler will take into account the flight with a code-share in the decision-making process. Parameter estimates for code-share thus represent the relative utility of a code-share compared to a non-code share.

6.2.5 Departure time

The departure time (local time) of the first flight of the outbound itinerary is specified in the utility function. The definitive model contains a dummy variable specification per departure hour. As a reference, the period 8:00 – 8:59 is taken. Parameter estimates thus represent the relative valuation of departure hour compared to 8:00-8:59.

The departure time (local time) of the first flight of the flight of the inbound itinerary is specified in the utility function. The definitive model contains a dummy variable specification per departure hour. As a reference, the period 16:00 – 16:59 is taken. Parameter estimates thus represent the relative valuation of departure hour compared to 16:00-16:59.

6.3 Data Used for Model Estimation

The data used for the model estimation consists of a dataset compiled of the MIDT dataset, the Expedia dataset and the OAG. This dataset contains 18,895 choice sets. For approximately 10% of the choice sets, the choice set sizes exceeds 50 itineraries. In these cases, only the first 50 itineraries are used for model estimation.

Due to the specification of dummy variables (carrier, departure hour), some choice sets were either excluded or itineraries were made unavailable in the definitive MNL-model:

- Itineraries offered by Expedia departing (outbound/inbound) within an hour never chosen were made unavailable;
- Carriers offered by Expedia but never chosen were made unavailable;
- Choice sets not containing a reference category were excluded from model estimation.

The implication of these constraints is that no parameters for several carriers can be estimated. Most notably, United Airlines is offered quite often by Expedia but is never chosen.

In addition, no parameters for certain departure hours can be estimated. For instance, for itineraries returning the same day, no parameters can be estimated for itineraries departing after 16:00 and returning before 14:00.

The definitive MNL-model is estimated on 18,416 observed itinerary bookings, the total number of alternatives being 800,897. The choice set accompanying each booking is the generated objective choice set: the available itineraries listed on Expedia on the day of booking, the same origin destination pair, the same duration of stay and the same departure day.

6.4 Model Results

6.4.1 Model building steps

Prior to the definitive model presented in Table 6-3, a series of model have been estimated and assessed according to criteria presented in section 3.2.5. These were:

- Statistical significance of the parameters;
- Consistency of parameter signs with expectations;
- Relative weights of parameter estimates;
- Final log likelihood and adjusted rho-square;
- Log likelihood test.

A summary of the model building steps can be seen in Table 6-2. All models were estimated using Biogeme (Bierlaire 2003) and the CFSQP solver algorithm.

Table 6-2 Model building steps and final log-likelihood

Model	Final log likelihood
In-vehicle time and waiting time versus total travel time and number of transfers	
The simplest estimated models contained dummy variables for type of carrier and a dummy variable if the carrier was a home-carrier. A comparison was made between models containing in-vehicle time and waiting time and total travel time and number of transfers. The latter model performed better in terms of explanatory power.	
Model with in vehicle time and waiting time	-56347.5
Model with number of transfers and total travel time	-56187.6
Number of transfer, total travel time and fare	
The addition of fare to the model lead to significant increase in terms of explanatory power.	
Model with number of transfers, total travel time and fare	-53153.8
Departure time outbound and inbound	
The inclusion of variables representing the departure time increased model performance significantly. Two approaches to include departure time in the model was experimented with, namely an approach using dummy variables representing the departure hour and a Fourier approximation. The Fourier approximation confirms the parameter estimates of the approach using dummy variables. A discussion of the Fourier approximation and its results can be found in Appendix D.1.	
Model with fare and departure time dummy variables	-46101.7
Model with fare and Fourier approach	-46109.7
Further specifications of fare	
The inclusion of fare in the model led to a significant increase in model performance. Several specifications of fare were tested, the inclusion of fare per stay category increased model performance best. Other specifications included fare per booking period and a fare per day of departure. These results are shown in Appendix D.2	
Model with ln(fare)	-46815.2
Model with fare parameter per booking period	-46091.1
Model with fare parameter per weekday	-46065.1
Model with fare parameter per stay category	-46021.7

Further specifications and considerations are presented in Appendix D.3. The definitive model contains carrier constants, a dummy variable representing if the itinerary contains a code-share, the total travel time, a variable representing a transfer, variables representing the type of aircraft. Furthermore, departure hour variables and a fare variable are included per stay category. With this approach, an explicit choice is made for a segmentation of passengers only with regard to fare and departure time preferences.

A remark should be made with regard to the terminology of the estimated parameters:

- The parameter estimates for carrier attributes, flight attributes and aircraft attributes are *generic* parameter estimates, assuming that the taste parameters for these attributes are similar across all segments;
- The parameter estimates for departure time and fare are segment *specific* attributes. The parameter estimates can be compared *within* a segment and *between* segments. For instance, passengers returning the same day perceive departing between 12:00-12:59 1.65 times worse than passengers returning the next day perceive departing in the same period, as the ratio of these parameter estimates is -2.76/-1.68. The parameter estimates for departure time and fare are not interaction variables. In the case of interaction variables, an extra reference category would be necessary.

A second imaginable approach to would be estimate separate models per duration of stay. Preliminary results show an increase in explanatory power for a duration of stay of 0 days and longer than 6 days. A drop in explanatory power can be observed for passengers returning the next day. However, it should be stressed that these are preliminary results and that the estimations have not been analyzed sufficiently (e.g. availability of carrier, transfer). Parameter estimates for fare and departure times would be the same, however, as the parameter estimates presented in this chapter for departure time and fare.

6.4.2 Statistics of the best MNL-model

Most of the estimated parameters are significant at a 95% confidence level (t-value > 1.95). Among the non-significant parameters are variables representing carrier and departure time for travelers returning after six days. This will be discussed in section 6.4.5 and section 6.4.6 respectively. It is chosen to leave insignificant parameters in the definitive MNL-model, as these variables represent series.

Parameter estimates for some of the variables representing departure hour and the variable representing a transfer is fairly large as compared to the parameter for total travel time and the variable representing regional aircraft. This topic will be further elaborated upon in section 6.4.4. The parameter estimates for travel time, departure time, fare and number of transfers turn out to be stable across different model specifications, as discussed in section 6.4.1.

A MNL-model with an adjusted rho-square between the 0.2 and 0.4 can be said to have a good fit and can be compared to a rho-square between 0.7 and 0.9 in the case of linear regression. Thus, the adjusted rho-square of 0.332 (Table 6-4) is satisfactory.

Table 6-3 Estimated parameters best MNL-model (N_{obs}= 18,416 and N_{cases}= 800,897)

	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Carrier constants				
Flight attributes				
Non-code share	0.0000	-		
Code share	-0.9215	-12.76		
Total travel time out	-0.0116	-5.66		
Transfer	-4.6511	-12.36		
Aircraft attribute				
Mainline jet	0	-		
Regional aircraft	-0.1530	-5.37		
Propeller aircraft	-1.5518	-14.89		
Departure times stay category 0				
Outbound			*Inbound*	
6:00 - 6:59	-0.3543	-10.17	-	-
7:00 - 7:59	0.3012	6.99	-	-
8:00 - 8:59	0	-	-	-
9:00 - 9:59	-1.0473	-17.01	-	-
10:00 - 10:59	-1.4841	-31.06	-	-
11:00 - 11:59	-2.0104	-34.11	-	-
12:00 - 12:59	-2.7596	-37.17	-	-
13:00 - 13:59	-4.8408	-17.27	-	-
14:00 - 14:59	-4.4808	-20.74	-1.6323	-5.08
15:00 - 15:59	-5.5495	-5.45	-0.7659	-2.99
16:00 - 16:59	-	-	0	-
17:00 - 17:59	-	-	0.5991	14.42
18:00 - 18:59	-	-	0.8246	19.54
19:00 - 19:59	-	-	0.8816	22.03
20:00 - 20:59	-	-	0.3325	7.55

	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
21:00 - 21:59	-	-	0.4206	3.09
22:00 - 22:59	-	-	-	-
Departure times stay category 1				
Outbound			Inbound	
6:00 - 6:59	-0.3794	-6.13	-	-
7:00 - 7:59	-0.2656	-3.97	-	-
8:00 - 8:59	0	-	-1.2016	-5.34
9:00 - 9:59	-0.8122	-9.64	-0.7210	-2.29
10:00 - 10:59	-0.8318	-11.89	-	-
11:00 - 11:59	-1.3406	-16.28	-	-
12:00 - 12:59	-1.6706	-19.41	-	-
13:00 - 13:59	-1.3736	-16.25	0.5032	1.73*
14:00 - 14:59	-1.1646	-14.88	-0.0533	-0.29**
15:00 - 15:59	-1.1883	-12.13	0.2588	1.27**
16:00 - 16:59	-1.0239	-14.26	0	-
17:00 - 17:59	-0.9566	-14.63	0.5907	11.21
18:00 - 18:59	-0.9299	-12.82	0.7830	14.02
19:00 - 19:59	-1.6873	-19.47	1.0273	20.76
20:00 - 20:59	-1.6964	-16.29	0.5686	9.96
21:00 - 21:59	-2.9532	-11.47	0.4250	3.02
Departure times stay category 2				
Outbound			*Inbound*	
6:00 - 6:59	-0.9844	-2.11	0.3294	0.90**
7:00 - 7:59	-0.3082	-0.72**	0.8557	3.34
8:00 - 8:59	0	-	-0.2100	-0.58**
9:00 - 9:59	-1.2215	-2.63	0.8315	2.97
10:00 - 10:59	-0.1765	-0.41**	1.0259	4.02
11:00 - 11:59	-0.1006	-0.23**	1.2060	3.78
12:00 - 12:59	-0.7790	-1.74*	0.7797	3.44
13:00 - 13:59	-1.5377	-3.39	1.2317	4.49
14:00 - 14:59	-0.8276	-1.74*	1.2646	4.62
15:00 - 15:59	-0.7482	-1.34*	1.5728	6.05
16:00 - 16:59	-0.6851	-1.46*	0	-
17:00 - 17:59	-0.5866	-1.37**	1.0198	4.75
18:00 - 18:59	-0.4758	-1.08**	1.6875	6.57
19:00 - 19:59	-1.7218	-3.11	1.1499	4.26
20:00 - 20:59	-1.2242	-2.39	1.2464	4.17
21:00 - 21:59	-2.5105	-2.33	0.1766	0.22**
Fare				
Fare duration of stay 0 days	-0.0077	-68.50		
Fare duration of stay 1 day	-0.0056	-41.96		
Fare duration of stay > 6 days	-0.0083	-8.83		

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

Table 6-4 Model performance best MNL-model

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

6.4.3 Valuation of service characteristics

With the parameter estimates it is possible to estimate several ratios, such as the value of time of a traveler and the value of a transfer. As these parameter estimates are estimated with the MNL-model, the values presented in Table 6-5 are obtained by simply dividing a parameter estimate by another parameter estimate. The estimate value of time and the monetary value of a transfer vary per duration of stay, as a separate fare parameter is estimated per duration of stay. The estimated value of a transfer expressed in minutes remains constant, as only a single parameter is estimated for both duration of stay and a transfer.

The estimated value of time varies from 84 €/h to 125 €/h, whereas the estimated value of a transfer varies from €559 to €832. A transfer is valued at 400 minutes. These estimates in the expected range for air travelers. A discussion of the estimated value of departure time can be found in section 8.2.1.

Table 6-5 Estimated value of time and transfers per duration of stay

	0 days	1 day	> 6 days
Value of time [€/h]	90.50	125.09	84.02
Transfer [€]	602.56	832.92	559.41
Transfer [min]	400	400	400

6.4.4 Relative weight of the estimated parameters

A possible way to gain insight into the utility function of the passenger is to depict the relative weight of the parameter estimates. Two utility functions are shown in Figure 6-1 and Figure 6-2.

Figure 6-1 shows a typical chosen and non-chosen itinerary, based on the values as presented in Table 6-6. Both itineraries are assumed to be operated by the same carrier. The chosen itinerary departs between 8:00 and 8:59 and returns between 19:00 and 19:59. The total travel time is 110 minutes and the fare is € 377,-. The non-chosen itinerary departs between 9:00 and 9:59 and returns between 14:00 and 14:59. The total travel time is 190 minutes and the fare is € 300,-.

In the chosen itinerary fare contributes most to utility, namely 57%, followed by total travel time and the departure time of the inbound of the itinerary. In the non-chosen itinerary, the contribution of fare drops to 20%, which is the result of the transfer in the itinerary. The contribution of a transfer to utility is 39%. The combined contribution of departure time is 23%, larger than the contribution of fare. The lowest contribution to utility is the contribution of a regional aircraft, which is only 1%.

Table 6-6 Values for chosen and non-chosen itinerary

	Chosen			Non-chosen		
	Description	Utility	Contribution	Description	Utility	Contribution
Aircraft	Mainline jet	0	0 %	Regional aircraft	-0.15	1 %
Total travel time	110 minutes	-1.28	25 %	190 minutes	-2.21	18 %
Transfer	No	0	0 %	Yes	-4.65	39 %
Departure time outbound	8:00 - 8:59	0	0 %	9:00 - 9:59	-1.05	9 %
Departure time inbound	19:00 -19:59	0.88	17 %	14:00 -14:59	-1.63	14 %
Fare	€ 377,-	-2.91	57 %	€ 300,-	-2.32	19 %
Total		-3.31			-12.01	

Figure 6-1 Relative utility Chosen (U = -3.31) and Non-chosen (U = -12.01)

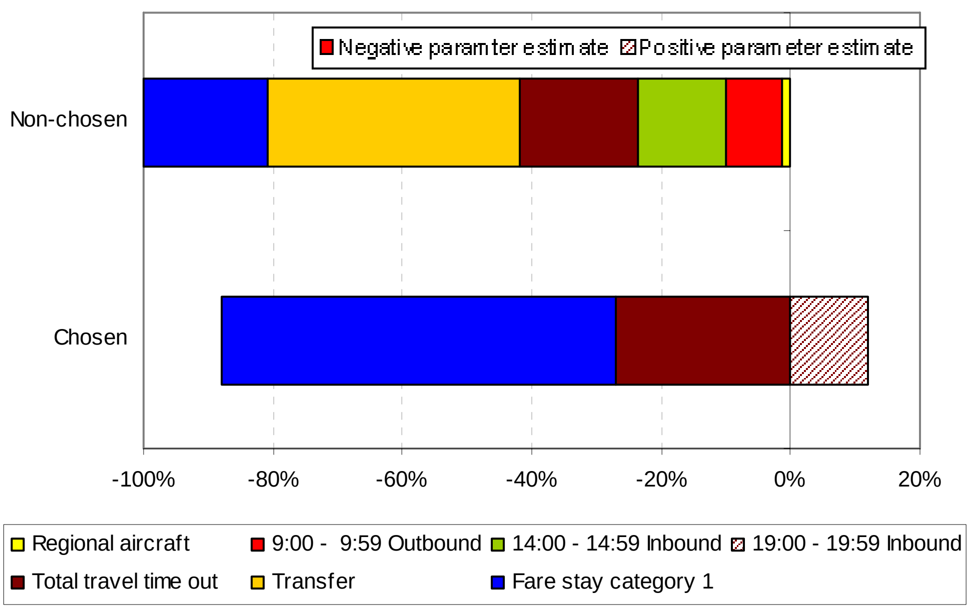


Figure 6-2 Relative utility Air France and KLM

Not depicted

6.4.5 Carrier variables

Not discussed

Figure 6-3 Airline constants

Not depicted

6.4.6 Departure time variables

In Figure 6-4, Figure 6-5 and Figure 6-6 the estimated parameters for outbound and inbound departure hour dummy variables are shown. The reference for the outbound departure time is 8:00, as reference for the inbound departure time 16:00 is chosen.

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

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

In Figure 6-5 the estimated parameters for the outbound and inbound departure time dummy variables for passengers opting for an overnight stay are presented. The estimated parameters for the outbound and inbound departure time dummy variables are all significant at the 95% level. Estimated parameters (except one) for the inbound departure time are significant at the 95% level.

Passengers returning the next day prefer departing at 8:00. Departing after 19:00 is perceived as very negative. More or less indifference can be observed between 14:00 and 18:00, the difference in utility is -0.25, which is valued at €37, and is somewhat less as 10% of the average chosen ticket price.

Passengers prefer returning after 16:00, an exception being departing between 13:00 and 13:59 and perceive returning in the early morning as negative. The preference for 13:00 is thought to be because passengers still can use the morning for other activities, such as meetings.

Figure 6-4 Estimated parameters for stay category 0

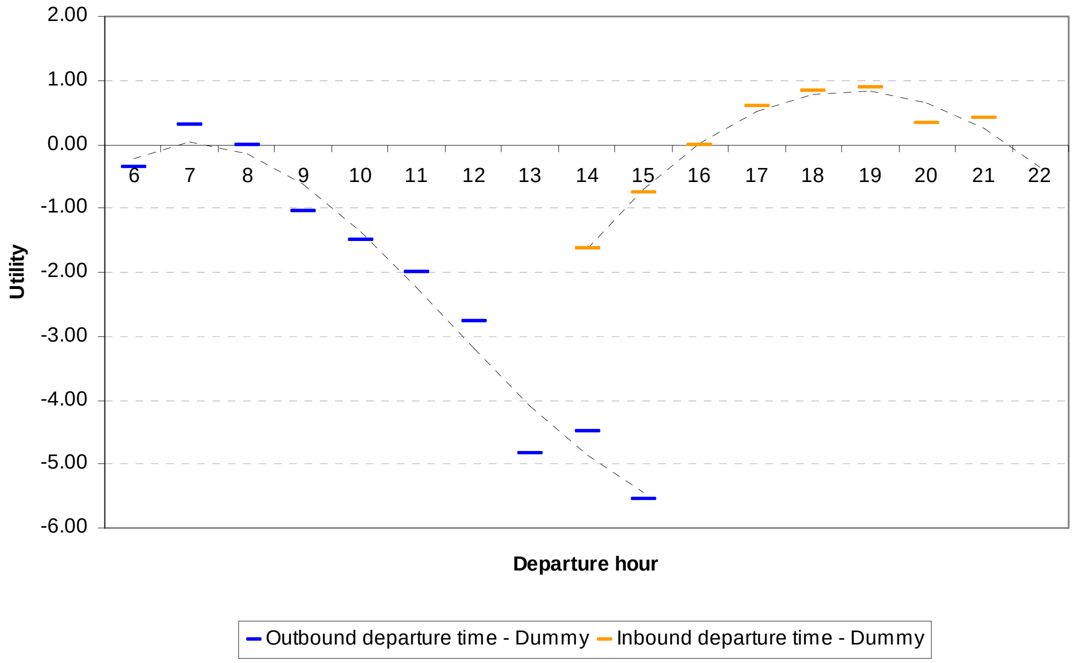


Figure 6-5 Estimated parameters for stay category 1

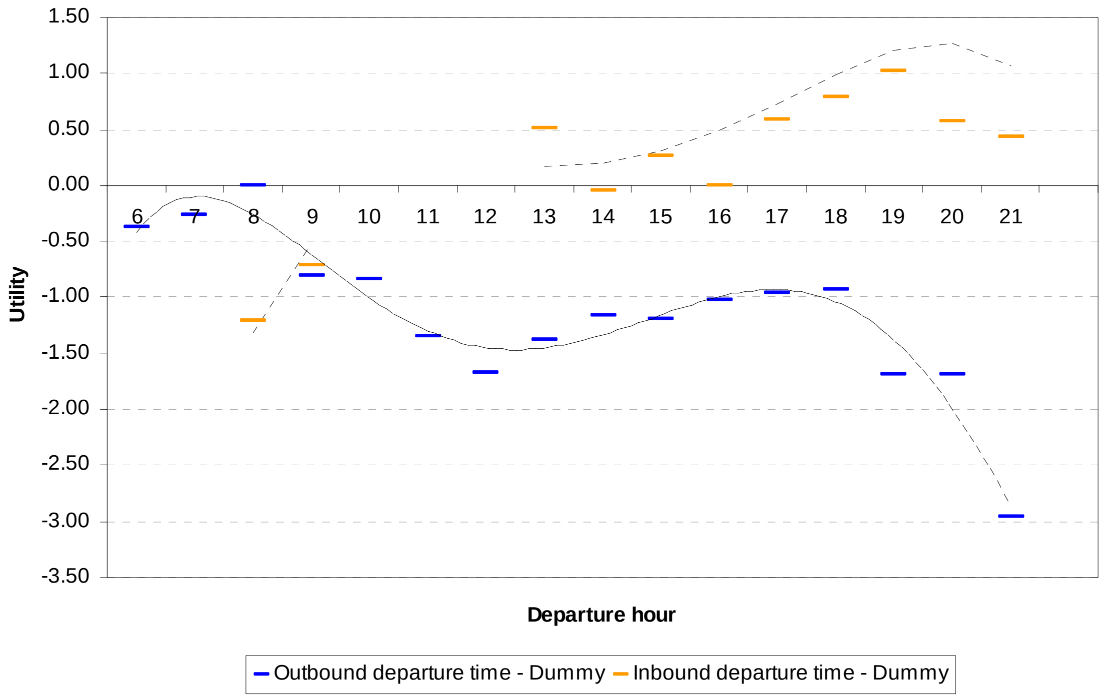
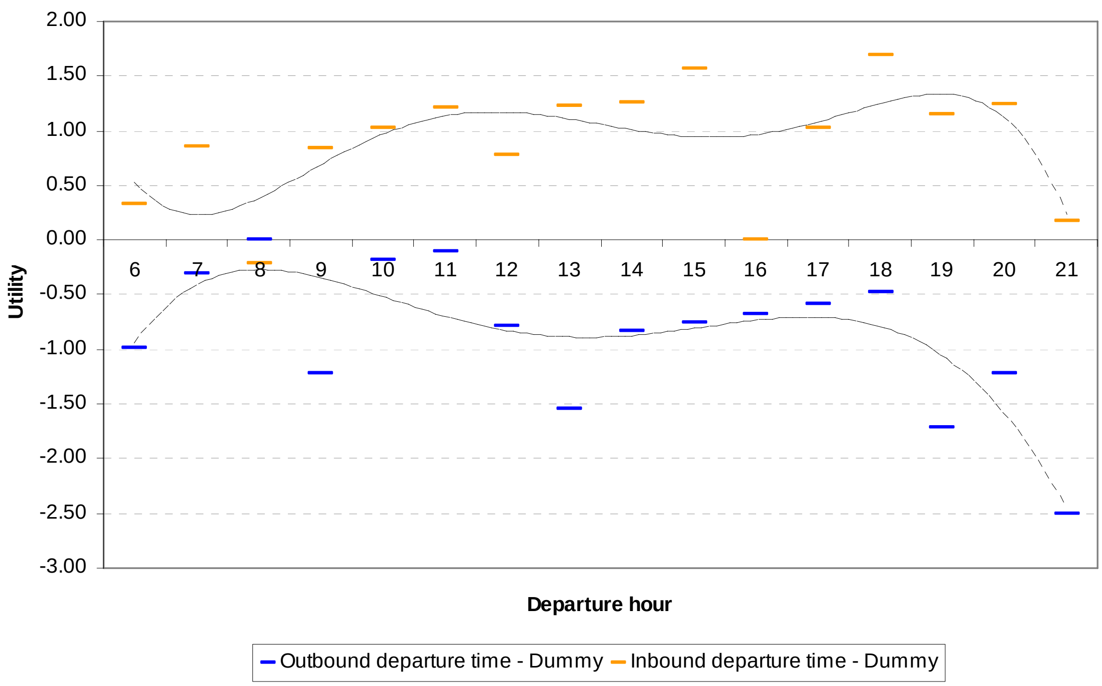


Figure 6-6 shows the estimated parameters for the outbound and inbound departure time dummy variables for passengers staying longer than six days at their destination. The parameters estimated for a departing between 6:00 and 7:00, departing between 13:00 and 14:00 and after 19:00 are significant at a 95% level; parameters estimated for departing between 12:00 and 13:00 and 14:00 and 15:00 are still significant at a 90% level.

From the observed preference structure and significant levels, it can be said that passengers remaining at their destination for a longer time do not reveal a clear preference for a certain departure time. If a conclusion should be drawn, it can be said that passengers prefer departing in the morning or in the late afternoon. The latter leaves passengers their time to find their hotel in the early evening. For their return flight, passengers prefer departing later in the early afternoon or early evening. However, it can well be that other unobservable characteristics have a stronger influence on the choice of passengers, such as the flight being part of a package deal.

Figure 6-6 Estimated parameters for stay category 2



6.4.7 Fare variables

For passengers returning the same day, a lower fare parameter is estimated than for passengers returning the next day. Passengers returning after six days are even more sensitive to fare. It is hypothesized, that passengers retuning the next day are less sensitive as fare only makes up a part of the total costs, which include an overnight stay. A second explanation could be that the fare differences for itineraries returning the same day are larger, as compared to itineraries returning the next day. Finally, an explanation could be found in the used segmentation: it is thought, that in stay category 2 passengers are included, who stay shorter at their destination than 6 weeks. These passengers take advantage of the irrationalities in revenue management systems, which make it possible to book two return journeys for less than one one-way journey. It could also be that these passengers do not know their return date yet, and opt for a cheap ticket, instead of a flexible ticket.

6.5 Conclusions

In order to answer a subset of the research questions and issues in the actor network, MNL-models have been estimated. All MNL-models have proved fairly robust in their parameter estimates, providing trust in the used data set and specification. Model estimation has only been carried out on generated objective choice sets. Furthermore, model estimations are in line with the statistical analysis of the choice sets presented in Chapter 5.

Fare yields a large contribution to utility and explanatory power of the model. In a direct itinerary, the largest part of utility of an alternative is yielded by fare. Arguably more important however, are flight timings. Also, transfers are perceived as very negative. A transfer is valued at 400 minutes or €600,- by passengers returning the same day. Table 6-7 presents further conclusions.

A distinction between different passengers based on their duration of stay is preferred: estimated parameters for departure time differ, as do parameters for fare. The inclusion of fare per stay category yields the highest log-likelihood and is therefore preferred. Furthermore, it is consistent with the segmentation of the departure time variables.

Table 6-7 Conclusions from MNL-itinerary choice model

Question	Keyword	Conclusion
Q10a	Type of carrier	Flag carriers are preferred above regional carriers and low-cost carriers. Regional carriers are preferred above low-cost carriers.
Q10b	Carrier	
Q10d	Type of aircraft	Mainline jets are preferred above regional aircraft, albeit not much. Regional aircraft are preferred above propeller aircraft.
Q10e	Code-share	Itineraries operating under their original flight number are preferred above itineraries containing a code-share.
Q10f	Outbound departure time	Travelers prefer departing in the morning. The preference structure however varies with the duration of stay.
Q10g	Inbound departure time	Travelers prefer departing in the evening. The preference structure however varies with the duration of stay.
Q10h	Transfer	Travelers show an aversion of a transfer in an itinerary.
Q10k	Total travel time	Total travel time forms a relative large part of the utility. It should be noted, that the differences in total travel time are rather small, if it concerns direct itineraries.
Q11c	Duration of stay	The duration of stay influence both the fare parameter and the preference for departure time.
Q12	Fare	Fare yields the largest disutility in direct itineraries.

The MNL-model presented in this chapter contained a segmentation by duration of stay for fare and departure time parameters; taste for carrier, flight attributes and aircraft attributes are assumed to be constant across all segments. From preliminary model results, it can be concluded that separate models per segment yield an increase in explanatory power for passengers returning the same day and after 6 days. It is therefore recommended to look further into this segmentation.

Before continuing with further findings and implications in Chapter 8, as conclusion of the case study, an extension of the MNL-model will be discussed in Chapter 7.

- This page left blank intentionally -

Chapter 7 MNL-Model with Independence of Connection Measure

7.1 Introduction

In the previous chapter, a MNL-model of itinerary choice was presented. The MNL-model is by construction unable to deal with alternatives that are similar, i.e. resulting in correlations between the error terms. In addition, it was discussed in this chapter that travelers are unable to make a distinction with high mutual overlap. Hoogendoorn-Lanser *et al.* (2005) make a distinction between three types of extensions to the MNL-model in order to cope with overlap. These are:

1. Models that allow for a non-zero covariance matrix (e.g. Logit-Kernel);
2. Models using a nested choice structure (e.g. Nested, Cross-Nested);
3. Models including an overlap factor in the utility function (e.g. C-logit, Path-Size Logit).

The latter two will be elaborated upon in the ensuing.
With a nested logit model structure, the set of alternatives j faced by a decision-maker is partitioned into K non-overlapping subsets, denoted B_k, which are called nests. For any two alternatives in one nest, the IIA assumption holds. For any two alternatives in different nests, the assumption usually does not hold. The choice probabilities of the nested logit model can be calculated as:

$$P_{in} = \frac{e^{V_{in}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{jn}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{jn}/\lambda_l}\right)^{\lambda_l - 1}}$$

Coldren *et al.* (2003) apply a nested logit model structure to air-travel itinerary modeling in the United States as an extension to earlier MNL model development at a major U.S. carrier. Two-level nested logit models were estimated. Results showed that itineraries sharing a common time period (morning, afternoon, evening) or common carrier exhibit competition amongst themselves, as the estimated parameters were overall somewhat lower as the estimated MNL model. This indicates that that the competition between itineraries departing in the same period is higher than between itineraries departing in different time periods because they are more alike.
In addition, Coldren estimated three-level nested models, which showed that itineraries sharing time-period and carrier exhibit competition amongst themselves.

Another approach to cope with mutual overlap between routes is to add a utility component to the deterministic part of the utility function. In this way, the simplicity of the MNL-model is maintained. The basis for these approaches is the implicit availability/perception (IAP) model presented by Cascetta *et al.* (1996). They state that a decision-maker is not able to consider all alternatives of the universal choice set because of the individual's imperfect knowledge of the alternatives and limited information processing abilities. The underlying assumption attests that the similarity of an alternative with other alternatives decreases its utility because it decreases its probability to be perceived an alternative. However, recent studies (Hoogendoorn-Lanser & Bovy 2006, Frejinger & Bierlaire 2007) suggest that this assumption might not hold for all choice contexts. In route choice situations for instance, the utility of a route increases with its similarity to other routes because the decision-maker gains the possibility to switch routes while traveling.

Furthermore, Cascetta (2001) proposes to introduce a commonality factor in the utility function, which reduces the systematic utility of a route proportionally to its level of overlapping with other alternative routes. The commonality factor can be calculated in several ways; all methods count in one way or the other the number of overlapping links in a route and transform this.

Ben-Akiva and Bierlaire (1999) introduce the Path-Size Logit factor, which includes the length and number of overlapping links with other alternative routes. Ramming (2001) extends this notion to the Extended Path Size Logit, which corrects for long routes having a low probability of yielding a large amount of overlap. Hoogendoorn-Lanser *et al.* (2005) add the Path-Size to their choice model for multi-modal route choice, an improvement of 0.5% was observed in the log-likelihood of the model. The addition of the Extended Path-Size Factor led to an improvement of 1.4% in the case of a multimodal network and 4.3% in the case of a uni-modal network. They give two reasons for the difference in impact: (1) the multi-modal model has twice as many parameters, which makes it more difficult to improve them, (2) the Extended Path Size factor captures a part of the explanatory factors in the uni-modal model and leads to a drop in parameters related to the Extended Path Size factor.

Most of these applications are mainly applied to road networks. Friedrich *et al.* (2001) suggest a more comprehensive measure, which allows to account for differences in fare, arrival time, departure time and journey time.

The Path-Size Logit and Extended Path-Size Logit both incorporate a length measure; it is questionable if a traveler is aware of the distance he travels when booking an itinerary. A same type of argument holds for the Commonality Factor by Cascetta (2001); a traveler only knows the number of transfers, and mostly not the exact location, thus not knowing the number of overlapping links. The implementation of the Path Size by Hoogendoorn-Lanser is calculated based on the number of legs.

The measure proposed by Friedrich has the advantage that it captures the similarities in the dimensions that play a role when choosing an itinerary: fare, arrival time, departure time and journey time, which are the similarities that occur in itineraries, stressing the multi-dimensional nature of the itinerary choice problem. Therefore, a closer look will be given to the latter measure in the next section.

7.2 Independence of Connection Measure

As already mentioned, the measure of Friedrich, Hofsaess and Weckek assumes that travelers decisions are based on perceived journey time (PJT), the difference between the desired and actual departure time and the differences in fare. By combining these three, the independence of an itinerary, or connection $c \in C$ can be computed. Important to notice is that the independence it is not an attribute of the alternative but a measure how an alternative is viewed in the choice set. The independence of a connection can be calculated as:

$$IND(c) = \frac{1}{\sum_{c' \in C} f_c(c')} = \frac{1}{1 + \sum_{c' \in C; c' \neq c} f_c(c')} \tag{7.1}$$

Where f_c is an appropriate non-negative evaluation function, with $f_c(c)=1$ and $0 \le f_c(c') \le 1$, $c' \in C$. The purpose of f_c is to model the impact of other connections in choice set C on c. $IND(c)$ will become smaller if it is similar to other alternatives, thus f_c will evaluate towards 1 if alternatives are identical. Furthermore, the independence of a connection is dependent of choice set size. f_c is defined as follows:

$$f_c(c') = \max\left(0, \left(1 - \frac{x_c(c')}{s_x}\right)\right) \cdot \left(1 - \gamma \cdot \min\left(1, \frac{s_z\left|y_c(c')\right| + s_y\left|z_c(c')\right|}{s_y s_z}\right)\right) \tag{7.2}$$

Or more intuitively:

$$f_c(c') = \max\left(0, \left(1 - \frac{x_c(c')}{s_x}\right)\right) \cdot \left(1 - \gamma \cdot \min\left(1, \frac{\left|y_c(c')\right|}{s_y} + \frac{\left|z_c(c')\right|}{s_z}\right)\right) \tag{7.3}$$

Where

$$\begin{aligned} x_c(c') &= \frac{\left|DEP(c) - DEP(c')\right| + \left|ARR(c) - ARR(c')\right|}{2} \\ y_c(c') &= PJT(c') - PJT(c) \\ z_c(c') &= Fare(c') - Fare(c) \\ PJT(c') &= IVT(c') + 2WT(c') + 2NT(c') \end{aligned}$$

And

$$s_y = \begin{cases} s_y^{+} & \text{y(c')} \le 0 \\ s_y^{-} & \text{y(c')>0} \end{cases}$$

$$s_z = \begin{cases} s_z^{+} & \text{z(c')} \le 0 \\ s_z^{-} & \text{z(c')>0} \end{cases}$$

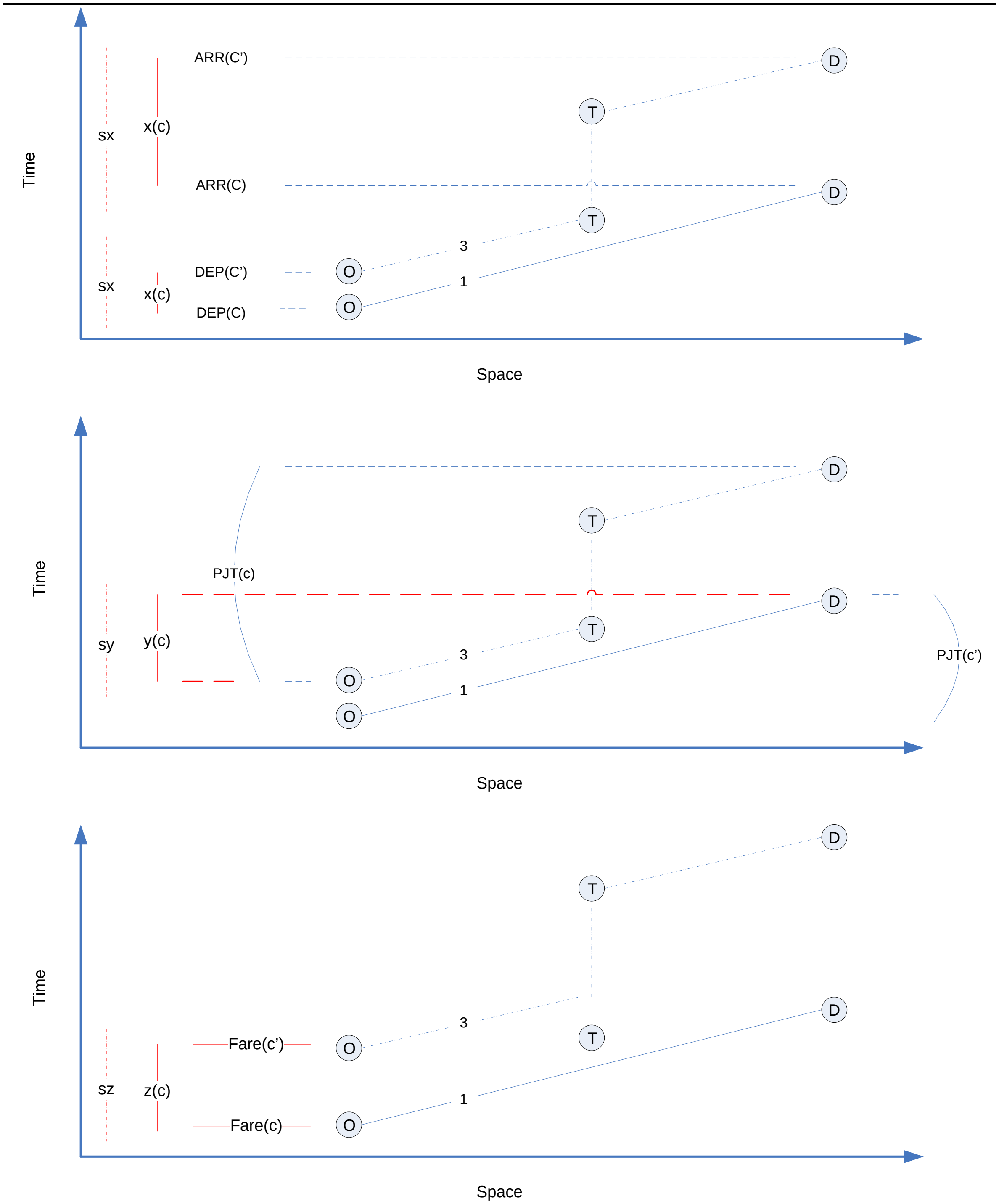
It can be seen from formula (7.3) that s_x, s_y and s_z set the range of influence of $x_c(c')$, $y_c(c')$ and $z_c(c')$. s_y and s_z depend on the sign of $y_c(c')$ and $z_c(c')$ in order to model the asymmetry between connections; if there are difference in terms of the perceived journey time $PJT(c')$, the superior connection will exert a stronger influence on the inferior one and vice versa. The perceived journey time is the sum of the in-vehicle-time IVT , twice the waiting time WT and twice the number of transfers NT .

If the left-hand side of the formula evaluates to a value smaller than zero, no similarity is assumed, as $f(c')$ will evaluate to 0. The same holds for the right-hand side: if the sum between the brackets evaluates larger than 1, no similarity is assumed.

The parameter γ sets the influence of the right-hand side of the formula, thus influencing the relative importance of the combination of fare difference and perceived journey time. A low γ will result in a higher influence of the left-hand side of the formula. In Figure 7-1 the relevant concepts are shown per similarity component.

It should be noted that the formulation of the f_c takes into account three dimensions simultaneously and is an approach to account for the similarity between itineraries in a multi-dimensional way.

Figure 7-1 Similarity components and parameters



For the calculation of the choice probabilities including the connection independence, the MNL-model is changed as following:

$$P_{iq} = \frac{e^{V_{iq}} \cdot IND(i)}{\sum_j e^{V_{jq}} \cdot IND(j)} = \frac{e^{V_{iq} + \beta \ln(IND(i))}}{\sum_j e^{V_{jq} + \beta \ln(IND(j))}}$$

Weis (2006) and Frick and Meister (2006) both add this measure to the utility function and find conflicting results; Weis conducted a study of local public transportation route choice and reports a positive estimated coefficient, which indicates that travelers perceive similar alternatives as negative. Frick and Meister conducted a study with as subject itinerary choice and estimate a negative parameter, indicating that similarity is perceive positive by travelers.

In the next section, a closer look will be given to the setting of the parameters s_x, s_y and s_z.

7.3 Parameter Setting

Several remarks should be made with regard to the setting of the parameters. First, the right part of the formula has to be smaller than 1 on average in order to influence the independence. Thus, the sum of $\frac{|y_c(c')|}{s_y}$ and $\frac{|z_c(c')|}{s_z}$ should be smaller than 1. Second, as already discussed, an asymmetry is modeled by using different values for s_y^+, s_y^-, s_z^+ and s_z^-. However, a certain asymmetry already exists in the dataset since the chosen alternative is usually cheaper than the other alternatives and/or its in-vehicle time is shorter. Finally, from the formula it can be seen that the parameter γ weights the right part of the formula and therefore eventually influences the value of $IND(c)$.

For s_x values between 120 and 720 minutes have been tested, whereas s_y is varied between 60 and 780 minutes for $y_c(c') \leq 0$ and between 30 and 780 minutes for the remaining cases, with an average PJT of 90 minutes in the chosen alternatives. The range of tested values for s_z goes from 15% to 170% percent of the fare for $z_c(c') \leq 0$ and from 10% to 170% for $z_c(c') > 0$. For z_c, a percentage is used as measure because per choice set, the level of fare can differ per choice set and origin-destination pair. The differences in travel time are less large, therefore absolute values are used for the parameters. The parameter settings are partially based on the available data and partially on common sense. To evaluate their effect, different parameter settings have been examined, some of which are presented in Table 7-1.

Table 7-1 Parameter settings for the independence measure

	s_x, s_y, s_z tests		γ tests		Final model
Parameter	Symmetric setting	Extreme values	γ = 0.25	γ = 0.75	γ = 0.5
s_x	720	360	720	720	120
s_y^+	780	180	780	780	780
s_y^-	780	120	540	540	540
s_z^+	1.7	1.1	1.7	1.7	1.7
s_z^-	1.7	0.9	1.5	1.5	1.5
γ	0.5	0.5	0.25	0.75	0.5

The influence of the different parameter settings on the partial similarity components can be seen in Figure 7-2. The figure shows the average values of $x_c(c')/s_x$, $|y_c(c')|/s_y$ and $|z_c(c')|/s_z$ in the chosen and non-chosen alternatives. Generally speaking, they are larger for the chosen alternatives than for the non-chosen alternatives. Due to the above mentioned asymmetry $y_c(c')$ and $z_c(c')$ are usually positive for the comparison between the chosen and a non-chosen alternative. Thus, s_y^- and s_z^- are used resulting in higher partial similarity components. Furthermore, the values of $|y_c(c')|$ and $|z_c(c')|$ themselves may be larger for the comparison between the chosen and non-chosen alternatives than for the one between the non-chosen alternatives. This is illustrated in the case where the parameters are set equal. The difference is most pronounced for the partial similarity component for fare.

The second parameter analysis examined the effect of varying γ and is illustrated in Figure 7-3, that shows the distribution of $IND(c)$ for chosen and non-chosen alternatives. It reduces or increases the influence of the differences in perceived journey time and fare on $IND(c)$. This leads to two effects: First, a higher γ induces a higher total $IND(c)$. Second, the distribution of the independence measure becomes wider for higher γ. Since γ weights the temporal similarity relative to the journey time and fare similarities, this effects shows that the distribution of the temporal similarity in this dataset is wider than the distribution of combined journey time and fare similarities.

Figure 7-2 Influence of the parameter setting on the partial similarity indicator

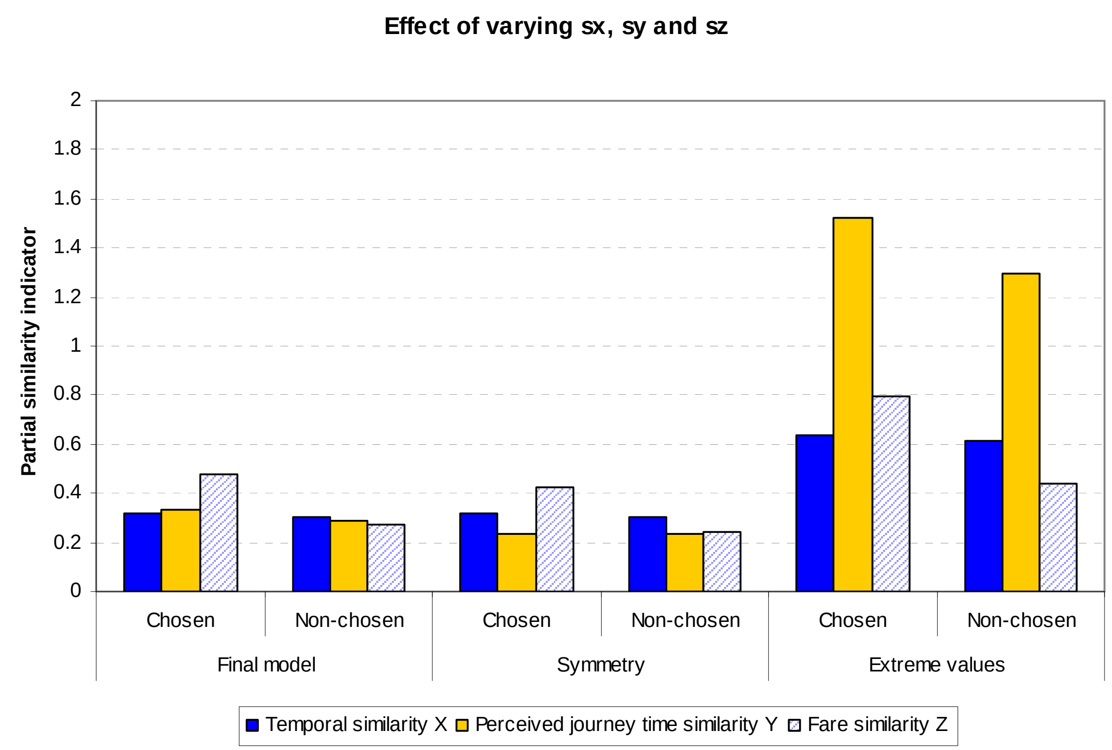


Figure 7-3 Influence of the parameter setting on the independence measure

7.4 Model Results

Model estimation is performed with models containing only a single variable representing fare, opposed to fare variable per duration of stay. In Table 7-2 the parameter estimates for $\ln(IND(c))$ can be seen. All parameter estimates of the MNL-model and the MNL-model with the independence measure can be found in Appendix E.2.

All estimated parameters are positive. Recalling that $IND(c)$ is between 0 and 1, the $\ln(IND(c))$ is negative. Similar connections are thus perceived as negative. This is a confirmation to the implicit availability/perception theory by Cascetta et al. (1996), it is however in contradiction with earlier findings by Hoogendorn-Lanser and Bovy (2005) or Frejinger and Bierlaire (2007). They showed that similarity can have a positive influence on the utility of ground-based transport alternatives. The findings of Hoogendoorn-Lanser and Bovy only hold in the case of a train-leg. Their interpretation refers this to the possibility to switch routes or connections while the passenger is traveling.

The first three specifications presented in Table 7-2 show the effect of varying γ on the parameter estimate for $\ln(IND(c))$. A γ of 0.25 leads to an insignificant parameter estimate for $\ln(IND(c))$. Values of 0.5 and 0.75 increase the significance of $\ln(IND(c))$ and improve model performance, judged by the t-tests, log-likelihood and r-square, indicating that the combination of fare and perceived journey have an influence on the perception of the independence of a connection.

The contribution of $IND(c)$ to the utility of an alternative increases with an increasing γ.

The parameter s_x can be said to set the window of alternatives which are considered similar to the chosen alternative, based on the difference in departure and arrival time. In Table 7-2 it can be seen that a value for s_x of 120 minutes yields the largest log-likelihood. This can indicate that itineraries departing an hour earlier or later are not perceived as similar anymore, as this parameter yields the highest significance.

Table 7-2 Modeling results for different parameter settings

Model	Estimated parameter for $\ln(IND(c))$	Robust t-test	Average value for $IND(c)$ chosen	$\ln(IND(c))$	Utility	Final log-likelihood	Adjusted r-square	Difference in LL compared to MNL
MNL	-		-	-	-	-46101.7	0.3308	-
$s_x = 720, \gamma = 0.25$	0.1608	1.48	0.0388	-3.2545	-0.5234	-46100.5	0.3308	0.00%
$s_x = 720, \gamma = 0.5$	0.2194	2.18	0.0458	-3.0512	-0.6693	-46099.1	0.3308	0.01%
$s_x = 720, \gamma = 0.75$	0.3992	4.72	0.0626	-2.779	-1.1094	-46089.2	0.3309	0.03%
$s_x = 120, \gamma = 0.5$	0.5379	13.59	0.1660	-1.8018	-0.9691	-46004.2	0.3322	0.21%
$s_x = 240, \gamma = 0.5$	0.2210	4.82	0.0983	-2.3187	-0.5123	-46089.6	0.3309	0.03%
$s_x = 360, \gamma = 0.5$	0.1804	3.37	0.0732	-2.6118	-0.4713	-46095.8	0.3308	0.01%

In an earlier stage of this research, itinerary models were estimated with a different utility function. This utility function contained dummy variables per departure period, opposed to a dummy variable per departure hour. The criteria for the outbound time periods were:

- *Morning* : Departure time of outbound connection before 9:00
- *During the day*: Departure time of outbound connection between 9:00 and 16:00
- *Evening*: Departure time of outbound connection after 16:00

For the inbound itineraries the time periods were:

- *Morning*: Departure time of outbound connection before 9:00
- *Afternoon*: Departure time of inbound connection between 16:00 and 18:00
- *Evening*: Departure time of outbound connection after 18:00

Separate parameters for departure time were estimated for each stay category. The same parameter settings were used for the calculation of the independence measure. The parameter estimate for $\ln(IND(c))$ was negative, indicating a preference for similar alternatives. This is the opposite as compared to models containing dummy variables for departure hour, which shows that similar alternatives are perceived as negative.

All parameter estimates are presented in Appendix E.1, Table 7-3 shows the log-likelihood and adjusted rho-square of the best models. It can be seen that the more detailed specification of time is more beneficial in terms of explanatory power than the addition of the $IND(c)$.

Table 7-3 Comparison of IND(c) with different model specifications

	Model per departure period		Model per departure hour	
	Final log likelihood	Adjusted rho-square	Final log likelihood	Adjusted rho-square
Without $IND(c)$	-48816	0.2948	-46101.7	0.3308
With $IND(c)$	-48622	0.2976	-46004.2	0.3322

The question that arises is the following: what is actually captured in the error term of the utility function? It can be argued that in the models containing departure time period variables, the error term contained departure hour preferences, in addition to other, unobservable factors. A positive perception of utility then indicates that itineraries similar to each other compete, which is confirmed by Coldren *et al.* (Coldren & Koppelman 2005), who estimated two-level nested logit models and showed that itineraries sharing a common time period (morning, afternoon, evening) competed with each other.

In models containing a specification per departure hour, the departure time preferences are included in the systematic part of the utility function on an hourly level, with the error term still containing the overlap in departure time, travel time and fare. The independence of connection measure tries to capture this overlap. With the hourly preferences in the systematic part of the utility function, overlap is perceived as negative.

7.5 Conclusion

The main deficit of the MNL model is its IIA characteristic. This characteristic can be overcome in several ways, either by deploying more advanced model structures or by adding a deterministic term to the utility function. For this research, the usage of the independence measure of Friedrich *et al*. (2001) is evaluated.

For the calculation of the independence factor, several parameters are required which set the range of influence. Correct setting of these parameters is required for the independence measure to work properly. Experiments are made with several combinations of parameter settings. These experiments lead to better insights into the degree of similarity between alternatives in a choice set. Perhaps even more important, the experiments show that the independence or similarity of an alternative can be quantified and the relative similarity on the several dimensions can be determined. Especially with revealed preference data, where not much knowledge of the decision-maker's preferences is available, the inclusion of multiple dimensions can lead to more insight in preferences without penalizing other alternatives too strong.

In order to evaluate the effect of the independence measure on the utility of an itinerary, the measure is added to the utility function. Estimated parameters for the independence measure are highly significant and the model performance increases. Thus, it can be shown that the independence measure captures the similarity between alternatives at least to a certain extent without increasing estimation time.

For all models the sign of the parameter indicates that passengers perceive similar alternatives as negative. This however depends on the specification of the utility function: in a utility function specified per departure period of day, passengers perceive similar alternatives as positive.

Chapter 8 Findings and Implications Case Study

8.1 Introduction

In the previous three chapters, a demonstration was given of the possibilities with regard to the given data and statistic analysis as well as MNL-model estimation. This chapter forms the conclusion of this demonstration. First, the general findings will be discussed in the next section. Second, the implications for airlines, travel portals and airports will be sketched.

8.2 Findings

Parameter estimates of the MNL-model are in line with the statistical analysis of the choice sets: a similar preference structure for aircraft type, transfer flights, code sharing and departure time. This gives trust in the used datasets and utility function specification.

In the remainder of this section, the research questions applicable to the case study will be answered. Section 8.2.1 presents the research questions regarding the service characteristics, or non-monetary attributes, of an itinerary. Section 8.2.2 will continue with the questions regarding customer profile.

8.2.1 Service characteristics

What is the relative valuation of carrier and type-of-carrier? - Q10a, Q10b

Not discussed

What is the relative valuation of frequency? - Q10c

Earlier models included a variable for daily frequency, as a proxy for market presence of a carrier. The estimated parameter carried a positive sign, indicating that a traveler perceives an increase in frequency as positive. However, the variable was not significant at an 85% level. Laesser and Wittmer (2007) conducted a survey amongst passengers of Zurich airport and found that frequent travelers preferred an increase of frequency on a certain route above an increase in destinations. Frequency was defined by Laesser and Wittmer as the number of itineraries on an origin-destination pair by all airlines. This indicates that it is a variable to consider if modeling airport and destination choice and as such is a confirmation of the discussion presented in section 2.4.2.

What is the relative valuation of type of aircraft? – Q10d

A traveler prefers a mainline jet above a regional jet and a propeller aircraft. It should be noted that the negative perception for a regional jet is not as pronounced as the negative perception of a propeller aircraft. This is confirmed in both the statistical analysis and the parameter estimates for these variables. A traveler returning the same day values a regional aircraft at approximately -€20,-, whereas a propeller aircraft is valued at -€224,-.

What is the relative valuation of code-share? – Q10e

A code-share flight is valued more negative than a non-code-share flight. Hüni & Merz (2007) conducted an analysis of fare setting on the Expedia dataset and found that itineraries containing a code-share were more expensive than the original flight. However, as fare is already incorporated in the model, code-share should proxy for other unobservable characteristics. A frequent comment in literature on code-share practices is that travelers are not sure which product they buy and therefore have an aversion code-share flights (Hanlon 1999).

What is the relative valuation of outbound departure time and inbound departure time? – Q10f, Q10g

The estimated models show that passengers have a high preference for an early departure time. Figure 8-1 shows the valuation of a fare per departure hour. Passengers returning the same day are willing-to-pay up to -€ 700,- more than compared to an itinerary departing at 8:00. The answer can be found in the used dataset and the observed travel behavior: a large number of passengers return the same day or the next day, passengers who are likely to prefer an arrival in the morning and a departure in the evening. Valuations of the return flight are lower (Figure 8-2).

Figure 8-1 also reveals a similar valuation of early departing itineraries amongst all stay categories and yielded largely insignificant results for parameters representing the departure time of the return flight of passengers staying at their destination for a longer time. It is therefore thought that passengers booking a ticket with a duration of stay of six days, are actually passengers returning earlier, with a second ticket booking. It is concluded that this is the result of the awareness of irrationalities in revenue management systems and probably book two itineraries, both returning in a further moment in the future. Other findings support this hypothesis. First, it was shown that a large number of one-way tickets were booked, which are can be itineraries returning after thirty days. This figure is based on the MIDT dataset, that contains only flights returning in November. Second, an increase of bookings returning after six days is observed close to the departure date. Travelers actually staying long at their destination will need more preparation (hotel), which makes it unlikely that passengers actually return after those six days. Fare listings on the Web have made it easier for travelers to exploit the irrationalities of revenue management systems. Nevertheless, it indicates that passengers returning after six days or more are less sensitive to fare.

Figure 8-1 Valuation per departure hour outbound and stay category

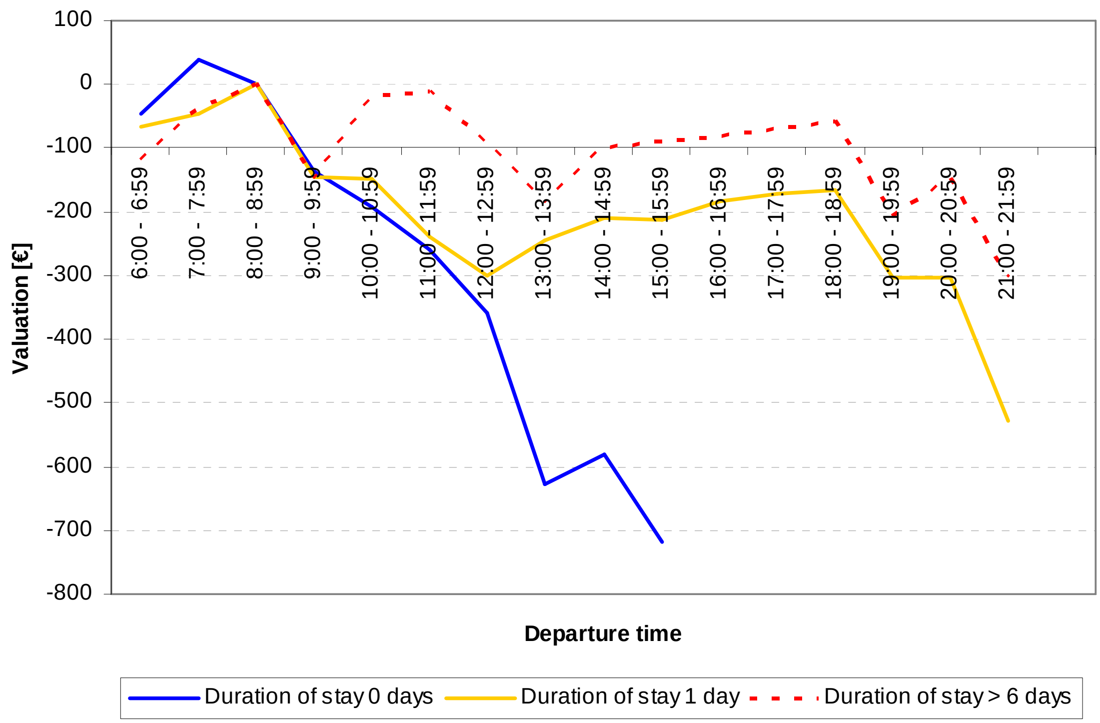
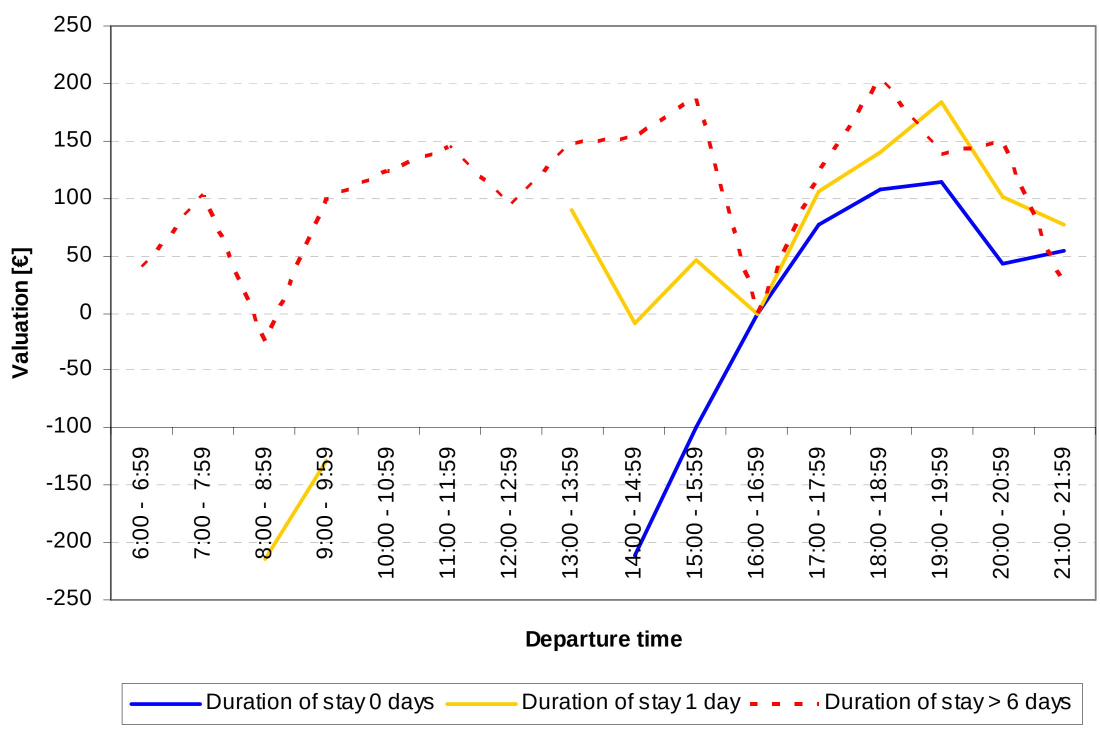


Figure 8-2 Valuation per departure hour inbound and stay category



What is the relative valuation of transfers, waiting-time, in-vehicle time and total travel time on traveler decision-making? – Q10h, Q10i, Q10j and Q10k

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

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

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

8.2.2 Customer profile

What is the effect of booking period, day of week and duration of stay on traveler decision-making? - Q11a, Q11b and Q11c

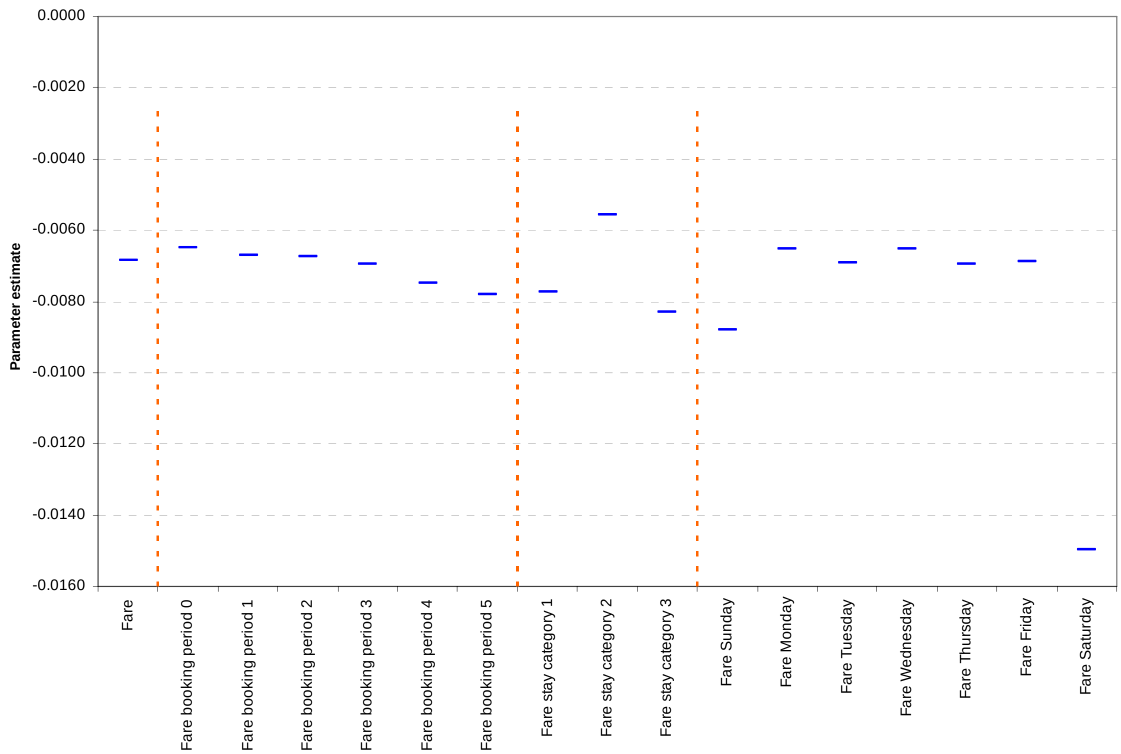
The effect of booking period, day of week and duration of stay has been addressed by specifying the utility function according to these factors. The effect of booking period, day of week and duration of stay has been measured by specifying a separate parameter for fare. Moreover, departure time preferences have been segmented by duration of stay, which has been discussed already in section 8.2.1

From the estimated parameters for fare (see also Figure 8-3, a more extensive discussion is presented in Appendix D.2), the following may be derived:

- A differentiation by duration of stay yields best model performance and shows that travelers returning the next day are less sensitive to fare. It is hypothesized, that this is because itinerary costs only form a part of total costs, such as dinner and a hotel.
- A differentiation by weekday (weekday 2 to 6) and weekend (weekday 1 and 7) is recommended, as passengers departing from Monday to Friday (weekday 2 to 6) yield a similar parameter estimate.
- A differentiation based on booking period can best be based on bookings longer than 3 weeks (booking period 4 and 5) before departure and shorter than 3 weeks (booking period 0 to 3) before departure.

Further research is necessary to determine the combinations of these differentiations. It should be stressed, that before conducting further research, possible application areas of these parameters should be investigated (e.g. forecasting for revenue management, determining real-time willingness-to-pay).

Figure 8-3 Estimated parameters for fare, all parameter significant at a 95% level



8.2.3 Role of Fare

What is the influence of fare on traveler decision-making? – Q12

The previous sections have used parameter estimates for fare to determine a monetary valuation of service characteristics. The relative utility of fare in the utility functions has however not been addressed in this chapter and will be the topic of this section.

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

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

Further research is recommended into the calculation of fare elasticities.

8.3 Implications for airlines

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

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

Airlines not focusing on business passengers but on leisure traffic can avoid airports during congested moments: passengers staying longer at their destination have a low preference for departure time. The results presented advocate strongly against hub-and-spoke systems, at least on intra European flights. Again, this can be due to the used data: most bookings used in the estimation process were on specific city-pairs and for passengers returning the same or next day. This group does however represent 60% of the bookings in Europe. Also, if it is economical to operate smaller aircraft than mainline jets, it is better to operate regional jets than propeller aircraft. With this, a general trend in aviation is followed which predicts an increase usage of regional jets due to their economic characteristics.

8.4 Implications for travel portals

What are possible improvements for the listings on the website? – Q14

Travel portals list itineraries based on information a potential customer enters on the website. Figure 8-4 shows the (partial) result of such a query for an itinerary returning the same day, requested 5 days before departure. No additional information to the arrival, return date and origin and destination is entered on the website. It can be seen that three of these itineraries will probably not be considered, namely the first itinerary and the last two. This hypothesis is based on the number of transfers and/or the price. Especially the last itinerary, with a duration of stay of 3 hours, a flight time of 15 hours and a fare of $ 2571,- is most likely not to be considered. The question arises whether these itineraries are explicitly listed to enforce a choice, or that these are the result of algorithm parameters. If the latter is the case, the estimated choice models can either help to enhance performance of the algorithms by directly including utility in the choice set generation process or calibrating branch-and-bound parameters. Perhaps a more farfetched opportunity is to explicitly include itineraries that will increase the choice probability of a certain itinerary. This could however be an opportunity for airlines instead of travel portals. In this choice set for instance, an itinerary of Air France – KLM will be chosen, if a traveler decides for one of these itineraries.

Figure 8-4 Example choice set from Expedia for Amsterdam – Toulouse

$934.70

9:35 am Depart Amsterdam (AMS) Arrive Toulouse (TLS) **11:35 am**	Mon **8-Oct** Duration: 2hr 0mn	**Air France** 8261 Nonstop flight
8:45 pm Depart Toulouse (TLS) Arrive Amsterdam (AMS) **8:00 am** +1 day	Mon **8-Oct** Duration: 11hr 15mn	**Air France** 7791 / 1140 Connect in Paris (CDG)

Choose this flight

$938.40

9:35 am Depart Amsterdam (AMS) Arrive Toulouse (TLS) **11:35 am**	Mon **8-Oct** Duration: 2hr 0mn	**KLM** 1303 Nonstop flight
5:30 pm Depart Toulouse (TLS) Arrive Amsterdam (AMS) **7:30 pm**	Mon **8-Oct** Duration: 2hr 0mn	**Air France** 8264 Nonstop flight

Choose this flight

$1,542.70

8:00 am Depart Amsterdam (AMS) Arrive Toulouse (TLS) **11:40 am**	Mon **8-Oct** Duration: 3hr 40mn	**KLM** 1229 **Air France** 7782 Connect in Paris (CDG)
6:10 pm Depart Toulouse (TLS) Arrive Amsterdam (AMS) **10:10 pm**	Mon **8-Oct** Duration: 4hr 0mn	**Air France** 7789 **KLM** 1242 Connect in Paris (CDG)

Choose this flight

$2,571.00

8:10 am Depart Amsterdam (AMS) Arrive Toulouse (TLS) **4:25 pm**	Mon **8-Oct** Duration: 8hr 15mn	**Iberia** 3215 / 8768 Connect in Madrid (MAD)
7:40 pm Depart Toulouse (TLS) Arrive Amsterdam (AMS) **11:15 am** +1 day	Mon **8-Oct** Duration: 15hr 35mn	**Iberia** 8773 / 3254 Connect in Madrid (MAD)

Choose this flight

A second recommendation regarding travel portals involves the so-called matrix display shown on travel portals. Two examples of current practices are given in Figure 8-5. Both displays show a matrix, with in the row headers the number of transfers, in the column headers the airline and in the cells the fare. Model results show the importance of departure time of the outbound and inbound itinerary. Therefore, a matrix display containing departure times outbound in the column header and departure times inbound in the row header and fares in the cells can aid a traveler better with his decision.

Figure 8-5 Matrix displays Orbitz and Expedia

With the estimated parameters, it is also possible to determine how much a customer is willing-to-pay. An opportunity for a travel portal is to try to 'skim' the difference between the price listed by the airline and the calculated willingness-to-pay.

8.5 Implications for airports

The role of airports in this case study has remained limited. As such, only a conclusion will be drawn with regard to flight timings and the role of hubs. The role of airport costs and the influence on airport decision-making was outside the scope of this research.

Airports serving a large share of travelers staying short at their destination, such a business travelers, should not expect to be able to spread demand out over the day, as travelers prefer departure time. If this is an issue, airports should consider serving different types of passengers. Airlines can request premium fares from passengers staying short at their destination. Airports could consider adjusting their fees correspondingly, based on the estimation results.

Airports pursuing a strategy to be become a hub for intra-European flights should consider diversifying their business.

8.6 Implications IVT

The datasets collected by the problem owner prove valuable, both for statistical analysis and choice set formation. The inclusion of fare in the choice models leads to an increase in explanatory power of the model. More important however than the increase in explanatory power, is the contribution of this research to current traveler choice behavior research and actual issues existing in industry. Research based on revealed preference data has not included fare directly, research based on stated preference data arguably included fare arbitrary and gives an indication of traveler behavior under simulated conditions. The fact that actual behavior is analyzed can prove both important for convincing both practitioners and researchers of the findings. As such, the author agrees with Chorus (2007).

State-of-the-art research in choice modeling explores the effect of similarity between alternatives and seeks ways to capture this similarity. The independence of connection measure applied in this research requires the setting of several parameters. Several times, it was pointed out that the parameter setting and the measure itself lack guidance. However, other measures for aviation have not been proposed yet.

- This page left blank intentionally -

PART III – Synthesis

- This page left blank intentionally -

Chapter 9 Synthesis, Conclusions and Recommendations

9.1 Introduction

The research presented in this thesis has evolved from a rather straightforward research question posed by the IVT to a research covering a broad range of topics covering the aviation system, traveler-decision making, discrete choice models and itinerary choice modeling.

In the first part of this study, the Institute for Transport Planning and Systems (IVT) was selected as problem owner, who has expressed its interest in the aviation system, in order to extend and expand its knowledge of air traveler behavior specifically and traveler choice behavior in general.
In order to gain insight in the aviation system an actor analysis was conducted. Among the actors included in the analysis were: travelers, airlines, airports, travel portals, high speed train operators and policy-makers. Travel portals cannot influence transport characteristics, but can influence the perception and knowledge of the different attributes relevant to decision-making. Airlines influence itinerary characteristics, whereas airports can influence access-mode and airport choice.
From the analysis in Part I, it was decided to focus on itinerary choice modeling, as this forms a continuation of recent research conducted at the IVT and industry relations, addresses prevailing issues in the scientific community as well as current issues industry such as willingness-to-pay and price elasticities, and offers the opportunity to engage in new industry relationships, most notably with travel portals.
Part II was then dedicated to answering research questions regarding itinerary choice by using data available to the IVT, conducting a statistic analysis of the datasets and the estimation of various choice models.

This chapter starts with a synthesis of Part I and Part II of this thesis in section 9.2, that brings together the findings of the empirical research presented in Part II. This synthesis is then used to formulate general conclusions and recommendations in 9.3 and methodological conclusions and recommendations in section 9.4. This chapter concludes with a reflection in 9.5. In Chapter 10 two recommendations are discussed in more extent.

9.2 Synthesis

9.2.1 Synthesis – Conclusions

This section presents conclusions from the case study presented in Part II, combined with the research questions and concepts formulated in Part I. The case study consisted of a discussion of the available data, a data analysis and the estimation of various choice models.
It is concluded that with the given data, assuming that the traveler is a utility-maximizing agent, and using generated objective choice sets, it is already to possible to answer a series of hypothetical questions posed by the aviation industry to the IVT with regard to itinerary choice. The generated objective choice sets contain itineraries listed by Expedia on the day of booking for the same day of departure and duration of stay. The questions and answers to them are given in Table 9-1.

Table 9-1 Synthesis - research questions answered

Itinerary choice

What is the valuation of travelers for service characteristics (e.g. departure time, transfers, and carrier image)?

For travelers returning the same day, a large part of the utility is contributed by departure time of the outbound itinerary. This group of travelers prefers departing in the morning. Travelers returning the next day prefer departing in the morning as well, but do not perceive departing in the afternoon as very negative. Travelers returning after 6 days do not reveal a clear preference for departure time. It is concluded that the longer the duration stay, the greater the role of unobservable characteristics becomes and the role of the trip in the activity chain.

Is it possible to define a customer profile (segmentation) based on observable criteria of the traveler?

Differences exist in sensitivity to fare between booking periods, duration of stay and departure day. Passengers booking earlier are more sensitive as passengers booking short before departure, passengers departing on weekdays are less sensitive than passengers departing during weekends. In addition, different preferences can be observed for departure time between passenger segments based on the duration of stay. The same conclusion is drawn as in the previous question: the longer the duration of stay, the greater the role of other, non-observable activities becomes.

Therefore, it is concluded that it is possible to segment passengers based on these criteria.

What is the influence of fare on traveler decision-making?

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

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

What are possible improvements for the listings on the website?

Possible improvements are to clearly display itineraries by their departure times and then by fare, preferably in matrix displays. Airlines already do this, travel portals however do not. This conclusion especially holds for intra-European flights and for passengers returning the same or next day.

Where can further improvements in search algorithms be made?

Thought to be most promising however, is the inclusion of different airports and modes in itineraries presented, a travel portal being an impartial party to do so. If search time is an issue, a utility based or branch-and-bound algorithm based on choice model parameters and information entered by a potential customer can reduce search time.

9.2.2 Synthesis on actors' instrument and findings of the case study

This section presents a synthesis of the different instruments of actors and the findings of the case study. Such an overview shows which instruments can be effective to reach the accompanying goal on the hand and gives a concrete example to actors on the other hand. For the sake of readability, this synthesis is presented in a table-like manner.

Table 9-2, Table 9-3 and Table 9-4 list the instruments of airlines, travel portals and airports respectively, assessing the relative effectiveness of each of these instruments based on the findings of the case study.

Table 9-2 Conclusions on airlines' instruments

Instrument	Conclusion
Increase commission costs to third parties	The relative contribution of fare is large as compared to other, non-monetary characteristics of an itinerary. Therefore, it is concluded that a reduction of commission costs to third parties may be in place, if the strategy is to sell as many tickets as possible through all sale channels
Reduce airport costs by flying on off-peak hours	Airlines serving passengers returning the same or next day, should offer their service during peak-hours. Airlines serving passengers staying longer at their destination may cut costs by flying on off-peak hours.
Optimal fleet assignment	Passengers reveal a preference for mainline jets (e.g. Boeing 737), second being regional aircraft. Propeller aircraft are perceived as negative. It is concluded, that if a mainline jet is either too large or expensive, a regional aircraft is the second best option.
Enhance revenue management systems	Preferences vary per passenger per departure day, a segmentation based on weekend and week, together with duration of stay and booking period seems appropriate. No conclusions can be drawn on the role of the fare product. Furthermore, it is thought by the author that travelers exploit current irrationalities in revenue management systems, buying two return tickets instead of a single ticket returning the same or next day. Transparency brought by the Web has made this possible.
Marketing	Not discussed
Increase routes	If an airline would increase the number of routes on intra-European flights, it is concluded that this is an option if it concerns direct routes, i.e. without transfer. Adding routes with a transfer for the sake of extra supply will not attract passengers.
Enlarge perceived network by code-share and franchise agreements	No evidence has been found that passengers prefer extensive networks in the case of intra-European networks. Franchising and code-share in an option to extend the supply basis, whereby the first would be preferred in the case of positive valued airlines, the second would be preferred if a carrier is valued more negative than a code-share
Optimize aircraft deployment	Aircraft deployment has not been investigated. However, a frequent flying traveler may not be pleased if a propeller aircraft is deployed if a regional aircraft was listed.
Optimize alliance networks	Again, in the case of intra-European networks, no evidence has been found that passengers perceive extended networks as positive. Variances amongst parameters for alliance members is observed as well.

Table 9-3 Conclusions on travel portals' instruments

Instrument	Conclusion
List itineraries that will be considered by the traveler;	Travelers have strong preference for departure time. Improvements of listings taken into account these preferences are recommended.
Provide information that the traveler perceives as necessary for decision-making;	Travelers perceive information regarding departure time of the outbound, inbound flight necessary, followed by fare important. An extension towards travel advice in multi-airport regions and door-to-door travel advice is recommended.

Table 9-4 Conclusions on airports' instruments

Instrument	Conclusion
Increase frequency of flights	No evidence has been found that passengers perceive an increase or decrease of frequency of a single airline as positive or negative. However, the number of itineraries on any given origin-destination pair may be relevant to decision-making.
Improve ease of transfer and connections	On intra-European flights, transfer is an exception. Therefore, improving the ease of them may be unnecessary.

9.2.3 Synthesis – Recommendations

This section presents recommendations, based on the results from the case study, data available for it and the concepts presented in Part I of this thesis. These two combined lead to a set of knowledge gaps, which are presented in the form of recommendations in Table 9-5. The first three recommendations directly follow from the case study.

Table 9-5 Synthesis – Recommendations

Keyword	Recommendation
Service characteristics	Two important service characteristics of airlines have not been included in the estimated modes, namely the frequent flyer program and the on-board service level. The first may be of importance for intra-European flights and influence decision-making. The second may be of less importance on intra-European flights, but may be of more importance on longer flights.
Customer profile	A further specification of customer profiles is possible; before continuing with all possible permutations of segmentations however, it is recommended to take a close look to the requirements of actors and possible implementations. In addition, it is recommended to further estimate separate models for several segments. Therefore, a more detailed data-analysis should be carried out.
Fare	The role of fare product has not been addressed in this thesis. Revenue management revolves around fare products. The author recommends a research to the role of fare product and the possibilities of 'insurances' instead of fare product, i.e.: an upgrade for each ticket, starting with a minimum price. An example of these insurances can be found in the simplified fare products of KLM or Swiss.

Keyword	Recommendation
Choice set formation	Part I of this research has shown the composition of the choice set depends on the information offered through interfaces, as these influence network knowledge. Further research over the role of these interfaces is necessary for a better understanding of choice processes, gaining insight in the composition of the subjective choice set and consideration set. In addition, more insight in factors eliminating alternatives should be obtained. Cooperation with either an airline or a travel portal is recommended.
Itinerary choice	In the case study, itinerary choice has been addressed based on booking observed in the MIDT dataset and itineraries offered by Expedia. Two recommendations follow directly from the bookings and itineraries in these datasets: - The MIDT dataset contains a low number of bookings at low cost carriers. It is recommended to look into possibilities to overcome this deficiency; - Expedia offers a low number of low cost carriers. It is recommended to observe fares on sites of low cost carriers who are not listed by Expedia, such as German Wings. Furthermore, only attention was paid to itinerary choice in Europe. Traveler's preferences may very well vary for inter-continental flights, especially with regard to transfers and travel time.
Itinerary, origin (airport) choice and destination (airport) choice.	Itinerary choice can be put in a broader context, for instance by including choice for the origin airport and destination airport. Then trade-offs could be observed for fare and access time. It is recommended to seek cooperation with an airline operating in multi-airport region, for instance Swiss/Lufthansa in the Basel/Geneva/Zurich triangle flying to the London region and then continue step by step. It should be kept in mind, that airport choice may become more (or less) important with the length of the trip. For instance, inter-continental passengers may put more effort in their searches for an itinerary matching their preferences. These trade-offs could be matched. An airport-owner as Schiphol, who thinks of moving passengers from one airport to a second, may be interested in such a study as well: it can show airlines which passengers will opt for the new airport and thus making a move from one airport to the airport more acceptable for an airline.
Assignment of air demand	Assignment of air travel demand can aid airlines with their network and schedule planning. Despite fare yielding a large contribution to utility, it should not be necessary to work with fares directly. Fare ratios between carriers may suffice.

9.3 General conclusions and recommendations

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it". (Simon 1971)

9.3.1 General conclusions

Based on the research conducted in Part I of this thesis, the case study carried out in Part II and the synthesis it is concluded that it is possible for the IVT to contribute substantially to prevailing issues in the scientific community and the aviation industry.

However, it is also concluded that within the aviation industry a wide range of possible application areas exist for the knowledge of the IVT. These application areas all have their own requirements to possible solutions. Furthermore, in a competitive and complex environment, it may be difficult for a research institute to obtain complete and objective data, arguably a requirement for scientific work. Considering beforehand which issues to address and obtaining an overview of requirements of actors may shorten the transition of the results of scientific research into practice.

From an actor and network analysis it was decided to highlight only issue, namely itinerary choice. It was concluded that itinerary choice serve as a jumping point for further research, contributes to current research and forms a continuation of current industry relations.

For the case study, several datasets were used, namely a dataset containing bookings, the MIDT dataset, a dataset containing fares observed on Expedia, and the Official Airline Guide. These three datasets prove to be suited for analyzing itinerary choice, including fare. Further conclusions regarding the datasets are presented in section 9.4. The fact that actual behavior in the case study is analyzed with revealed preference data can prove both important for convincing both practitioners and researchers of the findings.

Information on transport characteristics is provided to or requested by the traveler. Currently, information provision to the traveler is fairly limited: only information on parts of the journey can be obtained at separate locations and actors able to provider complete information are not doing so at the moment.

It is concluded that travelers consider fare important, as well as departure time. Offering a complete overview of journey costs and door-to-door travel time is an opportunity for airports, airlines and most obvious, travel portals. Such an overview can lower transaction costs for passengers and aid a traveler with his decision. The importance of fare has not made revenue management systems obsolete, on the contrary. However, a revenue management system based on fare products and return tickets might need to be adjusted towards simplified, one-way fares.

9.3.2 General recommendations

The aviation system concerns a vast and complex network of actors. Decision-making by a traveler involves numerous dimensions and involves a number of these actors. The first and foremost recommendation to the IVT is therefore to consider beforehand which problems and decisions are to be addressed. Data can be collected subsequently.

An obvious next step would be to consider trade-offs in access time and itinerary fare, for instance from the Basel/Geneva/Zurich region to the London region, based on a combination of stated preference/revealed preference data. An issue remains the incorporation of low-cost carriers and the role of the trip in the activity chain of a traveler. It is recommended to look closer into this.

To actors providing information to travelers, it is recommended to adjust the information displayed per traveler segment. Such a segment can be based on observable characteristics, as duration of stay. In addition, it is recommended to look closer at the combined role of fare and fare product, as a trade-off by travelers will be made between them.

9.4 Methodological conclusions and recommendations

9.4.1 Conclusions and recommendations on the datasets available

With the given data, it is possible to model itinerary choice of passengers. More specifically, it is possible to include fare to itineraries, if it is assumed that a passenger considers the itineraries listed on Expedia on the day of booking or a subset of these itineraries. The used datasets yield stable parameter estimates across different model specifications, giving trust in model results.

Fare can be included only for passengers returning the same day, the next day or after six days. 70% of the travelers do remain at their destination for this period of time. 30% of the travelers remain between 2 and 6 days at their destination. These passengers are not taken into account in model estimation and may have a different utility function, as these will probably present a different market segment. It is recommended to collect fares on Expedia or a second travel portal for the remaining durations of stay and make a distinction with regard to weekend stay and week stay, as these will involve different customer segments.

Furthermore, a large number of *origin-destination pairs* were selected based on supply characteristics and not on demand characteristics. Itineraries listed on a travel portal for a certain origin-destination pair do not indicate the presence of demand. The same holds for itineraries generated on basis of the Official Airline Guide.

It is recommended to select origin-destination pairs based on demand characteristics. Use bookings as observed in the MIDT dataset as basis and select a cross-section similar to the period one plans to observe fares for. A start is made for this: for all origin-destination pairs in Europe the number of booked itineraries with zero, one or two transfers is calculated and stored in a separate table.

The main deficiency of such an approach is that, as soon as transfers are explicitly included, the number of possible itineraries will increase strongly, making a match between booking (e.g. MIDT) and itinerary with fare (e.g. Expedia dataset) less probable.

Possible solution: record the booking and choice sets directly at an actor, i.e. travel portal, airline or GDS system provider.

In addition, data *was collected in November*, a month in which business travelers are most likely to travel. Itineraries serving holiday destinations, which are to be reached with a transfer flight, will not be very popular. This offers the opportunity to analyze a less diffuse group of passengers.

Finally, the quality of MIDT data will be discussed. The number of bookings made on carriers' websites is ever increasing. On the other hand, not all carriers list their itineraries on travel portals, most notably low cost carriers.

A possible solution is to record the booking and choice sets directly at an actor, i.e. travel portal, airline or GDS system provider.

To gain more insight in itinerary choice at least three approaches are imaginable:

1. A revealed preference approach, for instance based on click-and-stream analysis of a website visitor data. More insight in the actual choice sets of a passenger could be gained, if traveler's inputs on a website such as preferred departure time, preferred airline, etc. would be recorded. It would remain uncertain if all itineraries would actually be considered by a traveler. A disadvantage would be the huge amounts of data, which could be tedious to analyze, but could give insight in the choice set formation process. A second disadvantage is the fact that one would always work with a subset of itineraries. Just as the case study presented in this thesis was based on the itineraries as listed on Expedia, other research would be based on a different subset. A second approach, and more feasible, could be the compilation of different data sets. An advantage is that an airliner as SWISS/Lufthansa knows which tickets are booked through their website. Compiling the SWISS/Lufthansa tickets and available SWISS/Lufthansa itineraries with the remaining sold tickets and available travel portal itineraries on a route can lead to more realistic, objective, choice sets.
2. A stated preference approach. A contribution to current research could be made if the respondents would be asked which alternatives they consider and which attributes they find important. In addition, respondents could be asked to order the itineraries on their preferred characteristic, i.e. fare, travel time, number of transfers. A possible location for such a stated-preference research is in the aircraft. On a number of intercontinental flights, a traveler has a screen available, on which question could be presented. In addition, passengers on longer flights can answer paper questionnaires.
3. A possible way to determine and apply the willingness-to-pay would be to compare the results of discrete choice models with the price a consumer paid preferably based on the information a traveler entered. This can serve as a validation and calibration.

9.4.2 Remaining methodological conclusions and recommendations

Independence of Connection Measure

The perception of overlap between itineraries has been investigated with an independence measure that takes into account fare, differences in departure time and arrival time and travel time in a multi-dimensional way.

In order to evaluate the effect of the independence measure on the utility of an itinerary, the measure is added to the utility function. Estimated parameters for the independence measure are highly significant and the model performance increases. Thus, it can be shown that the independence measure captures the similarity between alternatives at least to a certain extent without increasing estimation time.

For all models the sign of the parameter indicates that passengers perceive similar alternatives as negative. This is however dependent on the specification of the utility function: in a utility function specified per departure period of day (i.e. morning, afternoon, evening), passengers perceive similar alternatives as positive.

The independence of connection measure proves easy to calculate, but does require the setting of several parameters. The setting of parameters has been carried out in an experimental way. It is recommended to compare the results with a more advanced discrete choice model and investigate a possible decomposition of the measure. This will be done in the next chapter.

Choice Set Formation

In literature on itinerary choice, the role of choice set formation is neglected in the opinion of the author. This thesis has provided several guidelines by which choice sets are formed by a traveler. Several dimensions have been proposed that influence the composition of the choice set, namely the booking time dimension, the information acquisition (travel portal/ carrier) and frequency dimension, the preferred arrival time dimension and the fare dimension.

Multiple Choice Sets

Choices on some websites comprise of multiple choice stages, a choice for departure day and return day and a choice for departure hour and return hour. These different choice stages represent different choice sets and form an interesting field of research. Furthermore, it can be that a traveler obtains information multiple times. However, within Europe tickets are booked short before departure. Therefore, the role of further looking into the combination information gathering – decision-making may be unnecessary.

Schedule Delay

This research has used departure times of itineraries to approach departure time preferences of travelers. This approach does not reveal the value of schedule delay to the traveler and is also not possible with the available data: a parameter representing the experienced delay of a traveler based on the *chosen* departure will always be highly significant and will carry a negative sign. It is recommended to determine the value of schedule delay either through revealed preferences questionnaires (e.g. travel diaries) or stated preference research.

Variation across travelers' preferences

In this research, a segmentation of travelers has been made based on their duration of stay with regard to departure time and fare. Preliminary results have shown that such a segmentation can be justified.

Mixed Logit models can perhaps better capture variation between travelers. Also, latent class models combined with discrete choice models form an exciting new field and may also be applied to an aviation setting.

9.5 Reflection

"Never work with real data". - Kay W. Axhausen

This section presents a twofold of reflection of the research conducted in this thesis. First, a reflection will be given on the main assumptions. Second, a review of the objectives stated in the research proposal will be given.

9.5.1 Reflection on research proposal

The enumeration presented below contains a copy of the research questions posed in the research proposal. The research presented in this thesis has only considered question 4 and 5 in detail, and considered question 8 and 9 partially. Question 3, which addresses revenue management, has been discussed briefly in an appendix, and it is concluded by the author that this topic is complex enough to dedicate a thesis to. The research proposal adopted a multi-actor perspective, opposed to the single problem owner approach followed in this thesis.
The results of this thesis confirm the importance of fare setting; however does not confirm if customers pick up the price trends. This remains a topic of further research. However, in the context of this research, where travelers book their ticket short before departure and remain short at their destination, it is thought that fare is important at the moment of booking, but might not influence booking time itself. The choice for booking time is influenced by the choice to undertake an activity at a certain location.

1. Do airlines play games with each other and are these games negative/positive for the airlines and for the customer?
2. Do customers pick up the price trends of airlines?
3. Which forms of pricing are in use currently and what is their relation to expected demand?
4. What is the impact of price on choice behavior?
5. What is the impact of other trip characteristics on choice behavior?
6. What are the elasticities of these characteristics?
7. When looking at these elasticities and characteristics, where can policy makers step in?
8. When looking at these elasticities and characteristics, where can airports step in?
9. When looking at these elasticities and characteristics, where can airlines step in?

9.5.2 Reflection on assumptions

This enumeration presents a reflection on a range of assumptions made throughout this thesis:

- The first assumption made was the choice for problem owner and to asses the aviation system through the eyes of the IVT. The author is of the opinion that a multi-actor approach would have led to similar conclusions if the same problem were to be addressed, which in this case was itinerary choice. Furthermore, the author is of the opinion that the choice of problem owner is a valid one, as he believes that it occurs frequently that a problem owner, both in science and industry, has a solution but does not know where and how to 'sell' this solution.

- Second, a systems view of problem solving was adopted. This view could have been used more clearly as leitmotif in this thesis, is however not. On the contrary, a solution was readily available and several conceptual models were available as well. The challenge has been to find the connections between the solution, conceptual models and formulate an accompanying perceived problem. However, the presented systems view of problem solving also shows no end and no beginning, making it possible to start with a solution as well, as long as it is kept in mind that a solution should be adapted to the system.
- Several issues of actors were highlighted, which reflected common ground between the IVT and the actors. However, it may very well be that actors have different priorities at the moment, such as cutting fuel costs and reducing the environmental impact of airports. No assessment has been made of this trade-off.
- It was assumed that a traveler has perfect discriminatory capabilities and has perfect knowledge of all available alternatives. It is highly questionable if this is the case. In the case of air travel of Europe, where over 60% of the passengers return the same or the next day, the amount of effort put in gaining information and making a decision may be limited or may be simply based on elimination or satisfaction instead of making complex trade-offs between different service attributes. This is partially reflected in the conclusions drawn. Fare is important, as is a transfer and departure time. Other characteristics play a minor role.
- Also, the validity of the research holds for intra-European travelers on a select number of routes. It may very well be that differences exist among different geographical areas and traveler segments.
- This research has to large extent been data-driven, which has led to an inevitable trade-off between conceptualization and specification. However, looking back, this data has provided valuable and furthermore gave the author the opportunity to operate in a complex environment of actors, and led to trying to steer a middle course between the different members of the graduation committee. In a next research the author would start with the conceptualization and involving various actors and stakeholders in the process, in order to from them and gain insight in their problems and requirements.
- The fact that this research was based on large datasets and the time-consuming processing of these datasets, has led to the negligence of several trajectories, such as sampling of alternatives and deploying more advanced model structures. Playing with the results such as calculating choice, probabilities, assessing different choice sets, elasticities and relative utilities have proven to give the most valuable experiences.
- Most important, the question is if a step has been made towards solving the problem of the problem owner. This remains a question only the problem owner can answer. The author is of the opinion that this research has led to more insight in air traveler behavior, but also that there lies a challenge ahead to obtain a more complete overview of air traveler behavior across all choice stages. The second challenge lies in convincing involved actors to continue and extend cooperation with the IVT.

- This page left blank intentionally -

Chapter 10 Roadmaps

10.1 Introduction

This chapter presents discussed two recommendations in more depth. First, the application of VISUM in an aviation setting will be discussed. Second, an adjusted version of the applied independence measure will be presented.

10.2 Air Demand Assignment

10.2.1 Roadmap

The application of VISUM is stressed here as it is within the possibilities of the problem owner to apply VISUM on a short term with the given data. Assignment in VISUM will lead to insight in traffic flows, but perhaps for a potential client more interesting, unique visualizations of the distribution of air demand. A possible, future approach, would be an agent-based demand assignment.

Define markets

The MIDT dataset contains all bookings in Europe through CRS systems. Despite this dataset lacking bookings of some low-cost carriers, it can nevertheless prove valuable for model estimation. A first step would be to define markets, for instance markets containing an origin or destination in Eastern Europe, itineraries serving cities of a certain size, or simply making a distinction between national and European flights.

Select a sampling approach

For first model estimations and choice set generation, it would carry to far to use all bookings for model estimation. In Europe alone, approximately 12,000,000 bookings are observed for the month November. A sampling approach could reduce the number of bookings to an acceptable size. For instance, only weekend days and one weekday can be sampled. A requirement is that the sample resembles the population with respect to flight, booking time and duration of stay characteristics on a market level. This will reduce estimation time, especially if nested model structures are used.

Choice set generation

In this research, attention has been paid to choice set generation. Several dimensions (fare, information, time) influencing the choice set formation process of a traveler have been identified. Together with the estimated models presented in Chapter 6 and Chapter 7, more insight is obtained in this behavioral process. Passengers staying short at their destination will not consider transfer flights and evening flights, passengers staying at their destination longer have a less strong preference for departure time. These insights can help in either formulating a choice set generation algorithm or setting parameters in VISUM. Choice set generation can be based on the Official Airline Guide (OAG). The algorithm can either be a branch-and-bound algorithm or can directly incorporate the utility.

Demand matrices

Demand matrices can be distilled from the MIDT dataset. It is however questionable if the extracted demand will suffice for assignment purposes; it only represents a lower bound. Collaboration with a carrier or an alliance, can lead to more insight in the quality of the MIDT data and may help in determining scaling parameters for the matrix. A distinction should be made with regard to duration of stay of passengers, as they have different preferences. It is thought, that the most convenient way to do this is through the usage of multiple matrices per stay category.

10.2.2 Interested Actors

The following actors are thought to be interested in the application of VISUM:

- *Airports:* airports require both landside and airside forecasts. Required forecasts include at least transfer passengers (Schengen, Non-Schengen, Europe and International) and non-transfer passengers and airports are interested in these as they influence processes on the airport, such as security checks and luggage systems.

  Increasingly import is the competition between hubs and between hubs and regional airports. A dual role can be perceived here: airports compete with each other, but also want to redirect traffic. This should be kept in mind when communicating results.
- *Policy-makers:* the accessibility of their region
- *Airlines:* for fleet planning and network planning VISUM can give insight in the distribution of traffic over all carriers on an origin-destination pair, not only the own carrier. Furthermore, the change in certain carrier attributes could be evaluated.

10.3 Overlap in alternatives

10.3.1 Development of a similarity measure

Chapter 7 of this thesis was devoted to the measure of connection independence, as proposed by Friedrich *et al. (2001)*. Other measures were discussed as well. The goal of some of these measures is to overcome the IIA-assumption of the classical MNL-model, whilst maintaining the simplicity of the model. A second goal is to gain insight in the perception of overlap of alternatives. The measure of Friedrich can be said to consist of three parts, concerning departure and arrival time difference, perceived journey time difference and fare difference. At least two aspects of this measure are to be found rather arbitrary by several academics, namely the parameter setting and the multiplications made in the calculation of the measure, which lack a theoretical and behavioral foundation. Furthermore, during this thesis, the interpretation, parameter setting and evaluation of the parameter setting was found to be counterintuitive sometimes. Similarity measures stem from the following line of reasoning:

$$P_{iq} = \frac{e^{V_{iq}} \cdot IND(i)}{\sum_j e^{V_{jq}} \cdot IND(j)} = \frac{e^{V_{iq} + \beta \ln(IND(i))}}{\sum_j e^{V_{jq} + \beta \ln(IND(j))}}$$

A possibility would be to decompose up the measure of connection independence into three parts:

$$IND_{temporal}(c) = \frac{1}{\sum_{c' \in C} f_c(c')} = \frac{1}{1 + \sum_{c' \in C; c' \neq c} f_c(c')}$$

$$IND_{pjt}(c) = \frac{1}{\sum_{c' \in C} g_c(c')} = \frac{1}{1 + \sum_{c' \in C; c' \neq c} g_c(c')}$$

$$IND_{fare}(c) = \frac{1}{\sum_{c' \in C} h_c(c')} = \frac{1}{1 + \sum_{c' \in C; c' \neq c} h_c(c')}$$

Where:

$$f_c(c') = \max\left(0, 1 - \frac{\left|DEP(c) - DEP(c')\right| + \left|ARR(c) - ARR(c')\right|}{2s_x}\right)$$

$$g_c(c') = \max\left(0, 1 - \frac{PJT(c') - PJT(c)}{s_y}\right)$$

$$h_c(c') = \max\left(0, 1 - \frac{Fare(c') - Fare(c)}{s_z}\right)$$

$$PJT(c') = IVT(c') + 2WT(c') + 2NT(c')$$

The implementation could then be incorporated in the choice model as following:

$$P_{iq} = \frac{e^{V_{iq}} \cdot IND_{temporal}(i) \cdot IND_{pjt}(i) \cdot IND_{fare}(i)}{\sum_j e^{V_{jq}} \cdot IND_{temporal}(j) \cdot IND_{pjt}(j) \cdot IND_{fare}(j)} = \frac{e^{V_{iq} + \beta_t \ln(IND_{temporal}(i)) + \beta_f \ln(IND_{fare}(i)) + \beta_p \ln(IND_{pjt}(i))}}{\sum_j e^{V_{jq} + + \beta_t \ln(IND_{temporal}(j)) + \beta_f \ln(IND_{fare}(j)) + \beta_p \ln(IND_{pjt}(j))}}$$

This formulation does not take into account the asymmetry, but could be easily be incorporated.
If the similarity measure would be decomposed into three components, no cancellation between the different forms of overlap will take place, as is the case in the original formulation.

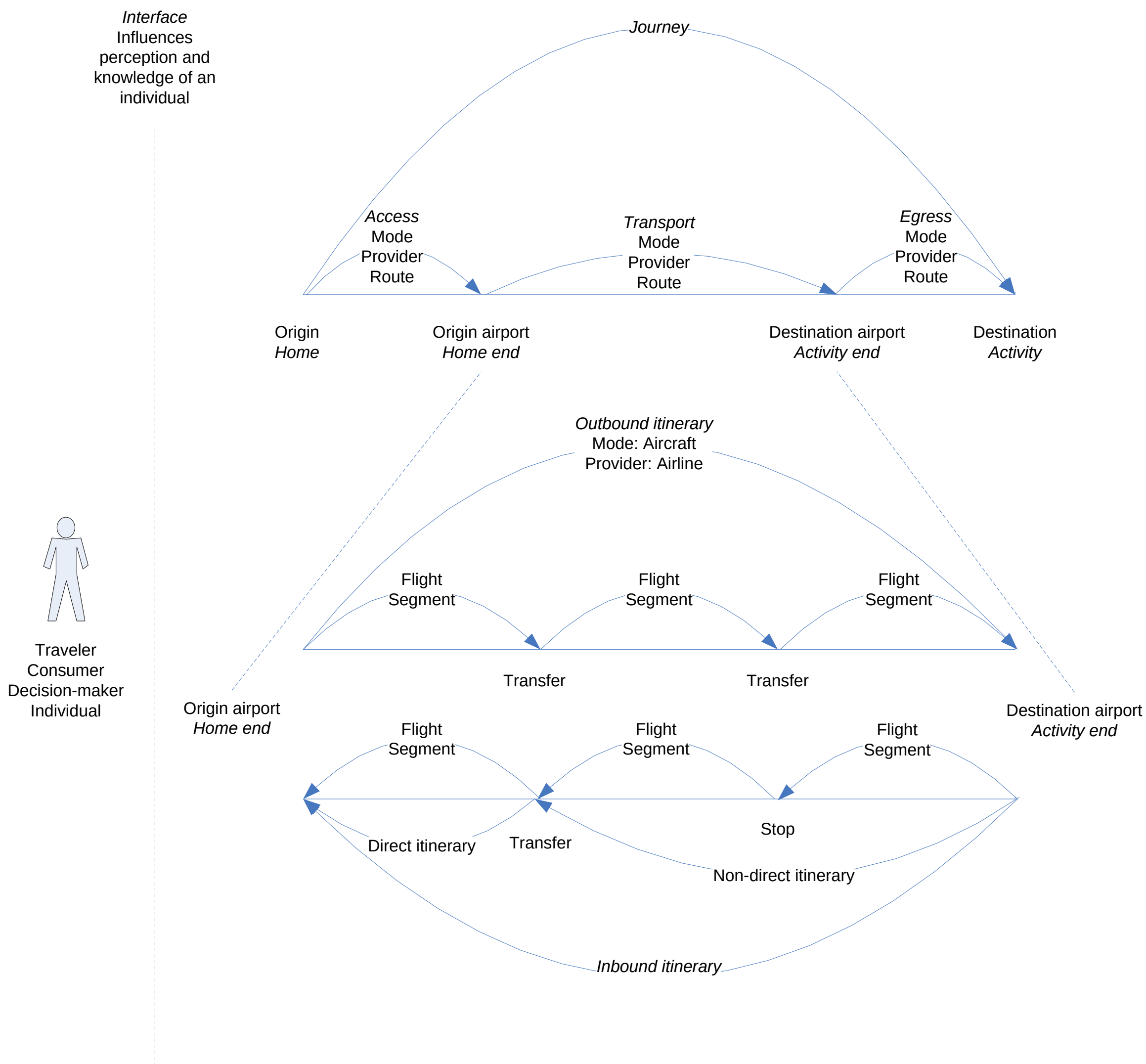
10.3.2 Different model structures

Instead of deploying a similarity measure, it is also possible to estimate nested and cross-nested models. A comparison would be possible if nesting structure would be used similar to the parameter used in the similarity measure, thus travel time, differences in the departure time and arrival time and fare.

- This page is left blank intentionally -

Definitions

This section provides a graphical overview of definitions. In addition, a series of definitions can be found on the next page.



Itinerary
An itinerary is a spatial and temporal constrained route from an origin airport to a destination airport and from the destination airport to the origin airport operated by an airline with as mode an aircraft.

Outbound itinerary
An outbound itinerary is a spatial and temporal constrained route from an origin airport to a destination airport operated by an airline with as mode an aircraft.

Inbound itinerary
An inbound itinerary is a spatial and temporal constrained route from the destination airport to the origin airport operated by an airline with as mode an aircraft.

Example itinerary

Selected flights

LO0270	**Amsterdam** (AMS)	19.12.2006 07:40	**Warsaw** (WAW)	19.12.2006 09:45
LO0269	**Warsaw** (WAW)	27.12.2006 19:55	**Amsterdam** (AMS)	27.12.2006 22:05

Direct itinerary
A direct itinerary is an itinerary without a transfer or stop.

Non-direct itinerary
A non-direct itinerary is an itinerary with a stop, but where passengers remain on the same aircraft.

Flight
A flight or segment is a trip from one airport to a second airport. Each itinerary can contain of one or more flights or segments.

Fare product
Airlines offer fare products based on itineraries; a fare product is based upon rules and restrictions, for instance a number of days in advance a ticket should be booked, or if the itinerary should include an overnight stay or a weekend stay. In addition, a fare product can offer flexibility to a traveler, such as the possibility to reimburse the ticket or change departure time and/or date.

Revenue management
Revenue management (RM) is the practice of enhancing firm revenues while selling essentially the same amount of product. Revenue management revolves around pricing, discount allocation and overbooking.

References

Barrett, S. D. (2000) Airport competition in the deregulated European aviation market, *Journal of Air Transport Management*, **6** (1) 13-27.

Belobaba, P. (2006) Airline Management, *MIT Open courseware*, Massachusetts Institute of Technology, Boston.

Ben-Akiva, M. E. and M. Bierlaire (1999) Discrete choice methods and their application to short term travel decisions, *Handbook of Transportation Science*, 5-35, Kluwer, Dordrecht.

Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Massachusetts.

Bierlaire, M. (2003) *BIOGEME: a free package for the estimation of discrete choice models*, Paper presented at the 3rd Swiss Transport Research Conference, Ascona, May 2003.

Bierlaire, M. (2005) *An introduction to BIOGEME (Version 1.4)*, http://transp-or2.epfl.ch/biogeme/doc/tutorial.pdf.

Boeing Commercial Airplanes Group (1993) Decision Window, Path Preference Methodology, Time Mode, November 1993.

Boland, D., D. Morrison and S. O'Neill (2002) *The future of CRM in the airline industry: A new paradigm for customer management,* IBM Institute for Business Value.

Bondzio, L. (1996) Modelle für den Zugang von Passagieren zu Flughäfen, *PhD Thesis*, Ruhr-Universität Bochum, Bochum.

Bots, P. W. G. (2002) Introduction to Systems Engineering and Policy Analysis, *Reader*, TU Delft, Delft.

Bovy, P. H. L. (2007) *Modeling route choice sets in transportation networks: a preliminary synthesis*, Tristan V Conference, Phuket, June 2007.

Bovy, P. H. L. and E. Stern (1990) *Route choice and way finding in transport networks*, Kluwer Academic Publishers, Dordrecht.

Bray, J. (2007) Consumers turn the tables on revenue managers (with Metasearch 2.0), *Journal of Revenue and Pricing Management*, **5** (4) 324-325.

Brueckner, J. K. (2002) Airport Congestion When Carriers Have Market Power, *The American Economic Review*, **92** (5) 1357-1375.

Burghouwt, G. (2007) *Airline Network Development in Europe and its Implications for Airport Planning*, Ashgate, Hampshire.

Burton, J. and P. Hanlon (1994) Airline alliances: cooperating to compete?, *Journal of Air Transport Management*, **1** (4) 209-227.

Button, K. and H. Vega (2006) Airlines competing with themselves: A note on the temporal pattern of fare setting prior to departure, *International Journal of Transport Economics*, **XXXIII** (3)

Carrier, E. (2006) *Modeling the Joint Choice of an Airline Itinerary and Fare Product: Implications for Airline Pricing Strategies*, 86th Annual Meeting of the Transportation Research Board, Washington D.C.

Cascetta, E. (2001) Transportation Demand Models, *Transportation Systems Engineering: Theory and Methods*, 175-250, Kluwer Academic Publishers, Dordrecht.

Cascetta, E., A. Nuzzola, F. Russo and A. Vitetta (1996) A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks, *Proceedings of the International Symposium on Transportation and Traffic Theory*, 697-711, Edited by J. B. Lesort, Pergamon, Lyon.

Chorus, C. G. (2007) Traveler Response to Information, *TRAIL Thesis Series*, TRAIL Research School, Delft.

Chorus, C. G., J. L. Walker and M. E. Ben-Akiva (2007) A model of travel information response: describing the full sequence of information acquisition and travel choice, *Working paper*,

Coldren, G. and F. Koppelman (2005) Modeling the competition among air-travel itinerary shares: GEV model development, *Transportation Research A*, **39** (4) 345-365.

Coldren, G., F. Koppelman, K. Kasturirangan and A. Mukherjee (2003) Modeling aggregate air-travel itinerary shares: logit model development at a major US airline, *Journal of Air Transport Management*, **9** (6) 361-369.

de Neufville, R. and A. Odoni (2003) *Airport Systems: Planning, Design and Management*, McGraw-Hill, New York.

Dodgson, J. S. (1994) Competition policy and the liberalisation of European aviation, *Transportation*, **21** (4) 355-370.

easyJet (2007) URL: http://www.easyjet.com, Accessed on September 10, 2007.

Enserink, B., J. F. M. Koppenjan and W. A. H. Thissen (2002) Analysis of Complex Environments (Analyse van complexe omgevingen), *Reader*, TU Delft, Delft.

Erath, A. (2004) Nachfrageverteilung im europäischen Luftverkehr, *Semesterarbeit*, IVT, ETH Zurich, Zurich.

Erdem, T. and J. Swait (2004) Brand Credibility, Brand Consideration, and Choice, *Journal of Consumer Research*, **31** (1) 191-198.

Eurostat (2005) *Draft glossary version 6 on air transport statistics*.

Eurostat (2007) *Eurostat News Release - Air transport in the EU25 (19 January 2007)*.

Executive Board of the ETH Zurich (1996) URL: https://www.ethz.ch/about/missionstatement/index_EN, Accessed on August 8, 2006.

Expedia (2007) Annual Report 2006.

Fiorenzo-Catalano, M. S. (2007) Choice Set Generation in Multi-Modal Transportation Networks, *TRAIL Thesis Series*, TRAIL Research School, Delft.

Fox, J., A. Daly and H. Gunn (2003) Review of RAND Europe's Transport Demand Model Systems, RAND Europe, Leiden.

Frejinger, E. and M. Bierlaire (2007) Capturing correlation with subnetworks in route choice models, *Transportation Research Part B: Methodological*, **41** (3) 363-378.

Frick, M. and K. Meister (2006) Route choice in air transport (Routenwahl im Flugverkehr), *Übungsbericht Messung- und Modellierung*, Zurich, IVT, ETH Zurich.

Friedrich, M., I. Hofsaess and S. Wekeck (2001) Timetable-Base Transit Assignment Using Branch and Bound Techniques, *Transportation Research Record*, **1752** 100-107.

Fröhlich, P. (2006) Design and Application Fields for a Worldwide Air Transport Model, *Ingenieurgemeinschaft MODUS*, Zurich.

Garrow, L., M. Ferguson, P. Keskinocak and J. Swann (2006) Expert opinions: Current pricing and revenue management practice across U.S. industries, *Journal of Revenue and Pricing Management*, **5** (3) 237-247.

Garrow, L. A., S. P. Jones and R. A. Parker (2007) How much airline customers are willing to pay: An analysis of price sensitivity in online distribution channels, *Journal of Revenue and Pricing Management*, **5** (4) 271-290.

Garrow, L. A. and F. S. Koppelman (2004a) Multinomial and nested logit models of airline passengers' no-show and standby behaviour, *Journal of Revenue and Pricing Management*, **3** (3) 237-253.

Garrow, L. A. and F. S. Koppelman (2004b) Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information, *Journal of Air Transport Management*, **10** (6) 401-411.

González-Savignat, M. (2004) Competition in Air Transport, *Journal of Transport Economics and Policy*, **38** 77-107.

Good, D. H., L. H. Roller and R. C. Sickles (1993) US Airline Deregulation: Implications for European Transport, *The Economic Journal*, **103** (419) 1028-1041.

Graham, A. (2001) *Managing Airports - An International Perspective*, Butterworth-Heinemann, Oxford.

Grammig, J., R. Hujer and M. Scheidler (2005) Discrete choice modelling in airline network management, *Journal of Applied Econometrics*, **20** (4) 467-486.

Grunewald, E., P. Berster, A. Classen, M. Hepting, H. Pabst and H. Pak (2007) Luftverkehrsbericht 2006, Deutsches Zentrum fuer Luft- und Raumfahrt, Cologne.

Hackney, J. K. (2005) Results of discrete choice models for long-distance travel based on the DATELINE survey, *Arbeitsbericht Verkehrs- und Raumplanung*, **334**, IVT, ETH Zürich, Zürich.

Hanlon, P. (1999) *Global Airlines: competition in a transnational industry - 2nd edition*, Elsevier Butterworth Heinemann, Oxford, United Kingdom.

Hannegan, T. F. and F. P. Mulvey (1995) International airline alliances An analysis of code-sharing's impact on airlines and consumers, *Journal of Air Transport Management*, **2** (2) 131-137.

Heiss, J. (2007) *FareCompare ready to soar with migration to JAVA SE 6*, URL: http://java.sun.com/developer/technicalArticles/J2SE/farecompare, Accessed on 17/2/2007.

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

Hess, S. and J. W. Polak (2006a) Airport, airline and access mode choice in the San Francisco Bay area, *Papers in Regional Science*, **85** (4) 543-567.

Hess, S. and J. W. Polak (2006b) Exploring the potential for cross-nesting structures in airport-choice analysis: A case-study of the Greater London area, *Transportation Research Part E: Logistics and Transportation Review*, **42** (2) 63-81.

Hoogendoorn-Lanser, S. (2005) Modelling travel behavior in multi-modal networks, *TRAIL Thesis Series*, TRAIL Research School, Delft.

Hoogendoorn-Lanser, S. and P. Bovy (2006) *Modeling overlap in multi-modal route choice by inclusion of trip part specific path size factors*, Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington D.C., January 2007.

Hoogendoorn-Lanser, S., R. van Nes and P. H. L. Bovy (2005) Path size and overlap in multi-modal transport networks, *Transportation and traffic theory - Flow, Dynamics and Human Interaction (Proceedings of the 16th International Symposium on Transportation and Traffic Theory)*, Edited by H. Mahmassani, Elsevier, University of Maryland, College Park.

Hubert, J.-P. and F. Potier (2003) What is known?, *Capturing Long-Distance Travel*, Edited by K. W. Axhausen, J.-L. Madre, J. W. Polak and P. L. Toint, 45-70, Research Studies Press, Baldock.

Hüni, C. and M. Merz (2007) Dynamik der Preisstrukturen im Luftverkehr, *Bachelor thesis*, ETH Zurich, Zurich.

Jacobs, P. H. M. (2005) The DSOL Simulation Suite - Enabling Multi-Formalism Simulation in a Distributed Context, *PhD Thesis*, TU Delft, Delft.

Joppien, M. G. (2003) *Strategisches Airline-Management: Evolutionstheorie und Luftverkehrsstrategien*, Verlag Paul Haupt, Bern.

Kahn, A. E. *Airline Deregulation*, The Library of Economics And Liberty, URL: http://www.econlib.org/library/Enc/AirlineDeregulation.html, Accessed on April 7, 2007.

Kisselef, B. (2006) Dynamik der Preisstrukturen im Luftverkehr - Semesterarbeit, *Semesterarbeit*, IVT, ETH Zurich, Zurich.

KLM (2007) URL: http://www.klm.com/travel/corporate_en/mission_people/mission/index.htm, Accessed on August 6, 2007.

Koppelman, F., G. Coldren and L. Garrow (2006) Schedule Planning, *Working paper*, Georgia Institute of Technology, School of Civil and Environmental Engineering.

Koppelman, F., G. Coldren and R. A. Parker (2007) Schedule Delay Impacts on Air-Travel Itinerary Demand, *Paper submitted January 2007*,

Laesser, C. and A. Wittmer (2007) *Customer issues related to limitations in air connectivity*, Paper presented at the 7th Swiss Transport Research Conference, Monte Verita, Ascona, September 2007.

Lohatepanont, M. and C. Barnhart (2004) Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment, *Transportation Science*, **38** (1) 19-32.

Louviere, J. J., D. A. Hensher and J. Swait, D. (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge.

Lufthansa (2007) URL: http://konzern.lufthansa.com/en/html/ueber_uns/management/strategie/index.html, Accessed on August 6, 2007.

Manski, C. F. (1977) The structure of random utility models, *Theory and Decision*, **8** (3) 229-541.

McFadden, D. (1974) Conditional choice analysis of qualitative choice behavior, *Frontiers in Econometrics*, Edited by P. Zarembka, 105-142, Academic Press, New York.

McFadden, D. (1978) Modelling the choice of residential location, *Spatial Interaction Theory and Planning Models*, Edited by A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, 75-96, Amsterdam.

Mercer Management Consulting (2002) *Impact of Low Cost Airlines - Summary of Mercer Study*.

Mitroff, I. I., F. Betz, L. R. Pondy and F. R. Sagasti (1974) On Managing Science in the Systems Age: Two Schemas for the Study of Science as a whole Systems Phenomenon, *Interfaces*, **IV** (3) 46-58.

Morrish, S. C. and R. T. Hamilton (2002) Airline alliances--who benefits?, *Journal of Air Transport Management*, **8** (6) 401-407.

OAG Worldwide Limited (2006) Official Airline Guide 2006, Bedfordshire.

Orbitz (2007) URL: http://www.orbitz.de, Accessed on August 13, 2007.

Ortuzar, J. and L. G. Willumsen (2001) *Modelling Transport, Third Edition*, Wiley, Chicester.

Park, J. H. (1997) The effects of airline alliances on markets and economic welfare, *Transportation Research Part E: Logistics and Transportation Review*, **33** (3) 181-195.

Parker, R. A. (2007) *What Passengers Will Pay To Fly: A Review of Passenger Choice Modeling at Boeing*, Paper presented at the 11th World Congress of Transportation Research (WCTR), Berkeley, July 2007.

Parker, R. A., R. Lonsdale, F. Ervin and Z. Zhengjie (2005) *GMAS: The Boeing Global Market Allocation System*, Presentation for the AGIFORS Scheduling and Strategic Planning Study Group, Seattle, May 2005.

Pels, E., P. Nijkamp and P. Rietveld (2001) Airport and Airline Choice in a Multiple Airport Region: An Empirical Analysis for the San Francisco Bay Area, *Regional Studies*, **35** (1) 1 - 9.

Pels, E., P. Nijkamp and P. Rietveld (2003) Access to and competition between airports: a case study for the San Francisco Bay area, *Transportation Research Part A: Policy and Practice*, **37** (1) 71-83.

Pels, E. and P. Rietveld (2004) Airline pricing behaviour in the London-Paris market, *Journal of Air Transport Management*, **10** (4) 277-281.

Pels, E. and E. T. Verhoef (2004) The economics of airport congestion pricing, *Journal of Urban Economics*, **55** (2) 257-277.

Porter, M. (1985) *Competitive Advantage*, The Free Press, New York.

Proussaloglou, K. and F. S. Koppelman (1999) The choice of air carrier, flight, and fare class, *Journal of Air Transport Management*, **5** (4) 193-201.

Railteam (2007) URL: http:/www.railteam.eu/, Accessed on August 21, 2007.

Ratliff, R. and B. Vinod (2005) Airline pricing and revenue management: A future outlook, *Journal of Revenue and Pricing Management*, **4** (3) 302-307.

Schiphol (2007) URL: http://www.schipholgroup.com/, Accessed on August 6, 2007.

Schiphol, Nederlands Instituut voor Vliegtuigontwikkeling en Ruimtevaart, Technisch Universiteit Delft, KLM, Luchtverkeersleiding, Stork Aerospace and Nationaal Lucht- en Ruimtevaartlaboratorium (2007) Schiphol de meest innovatieve en duurzame luchthaven: Uitvoeringsprogramma, September 2007.

Simon, H. A. (1955) A Behavioral Model of Rational Choice, *Quarterly Journal of Economics*, **69** (1) 99-118.

Simon, H. A. (1971) Designing Organizations for an Information Rich World, *Computers, Communications and the Public Interest*, Edited by M. Greenberger, 38-52, The Johns Hopkins Press, Baltimore, MD.

Swait, J. (2001) Choice set generation within the generalized extreme value family of discrete choice models, *Transportation Research Part B: Methodological*, **35** (7) 643-666.

Talluri, K. and G. van Ryzin (2004) Revenue management under a general discrete choice model of consumer behavior, *Management Science*, **50** (1) 15-33.

Talluri, K. T. and G. J. van Ryzin (2005) *The Theory and Practive of Revenue Management*, Springer, New York.

Tedeschi, B. (2007) An insurance policy for low airfares, *The New York Times*, January 22, 2007.

The Economist (2007) A high speed revolution, *The Economist Print Edition*, July 5, 2007.

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

Train, K. E. (2003) *Discrete Choice Models with Simulation*, Cambridge University Press, Cambridge.

Tron, A., L. McCoomb, M. Nowicki and P. Kowal (2007) *Modeling Air Traveler Choice Behavior in the Southern Ontario Market*, Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington D.C., January 2007.

Unique (2007a) URL: http://www.unique.ch, Accessed on August 15, 2007.

Unique (2007b) Annual Report 2006, Zurich.

United States General Accounting Office (2003) *Airline ticketing: Impact of Changes in the Airline Distribution Industry*, GAO-03-749.

van Ryzin, G. J. (2005) Future of revenue management: Models of demand, *Journal of Revenue and Pricing Management*, **4** (2) 204-210.

van Zuylen, H. J. (2005) *Lecture Notes CT5804, ITS and Dynamic Traffic Management*, TU Delft, Delft.

Weis, C. (2006) Route choice in public transport (Routenwahl im ÖV), *Master Thesis*, IVT, ETH Zurich, Zurich.

Westermann, D. (2006) (Realtime) dynamic pricing in an integrated revenue management environment: An approach to handling undifferentiated fare structures in low-fare markets, *Journal of Revenue and Pricing Management*, **4** (4) 389-405.

Wittwer, J. (2007) Mit Zug statt Flugzeug zum Geschäftstermin, *Tages-Anzeiger*, May 22, 2007.

Zeid, M. A., T. F. Rossi and B. Gardner (2006) *Modeling time Of Day Choice in the Context of Tour and Activity Based Models*, Paper at the 85th Annual Meeting of the Transportation Research Board, Washington D.C., January 2006.

Appendices

List of Figures Appendix

List of Tables Appendix

- This page left blank intentionally -

Appendix A Notes to Chapter 2

A.1 Revenue management

This section is largely based on the book *The Theory and Praxis of Revenue Management* by K.T. Talluri and G.J. van Ryzin (2005).

Revenue management (RM) is concerned with demand-management decisions and the methodology and systems required to make them. Demand management comprises of estimating demand and its characteristics and using price and capacity control to manage demand. It involves managing the firm's interface with the market with the objective of increasing revenues. Synonyms are yield management, pricing and revenue management, pricing and revenue optimization, revenue and process optimization.

A firm's demand has multiple dimensions, including (1) the different products the firm sells, (2) the type of customers it serves and (3) time. All three dimensions are important. However, methodologically often the problem is simplified to develop implementable solutions.

RM is often qualified as either quantity based RM or price based RM if it uses (inventory- or) capacity allocation decisions or prices as the primary tactical control for respectively managing demand. Traditional airlines use quantity based RM, low cost carriers often use price-based RM:

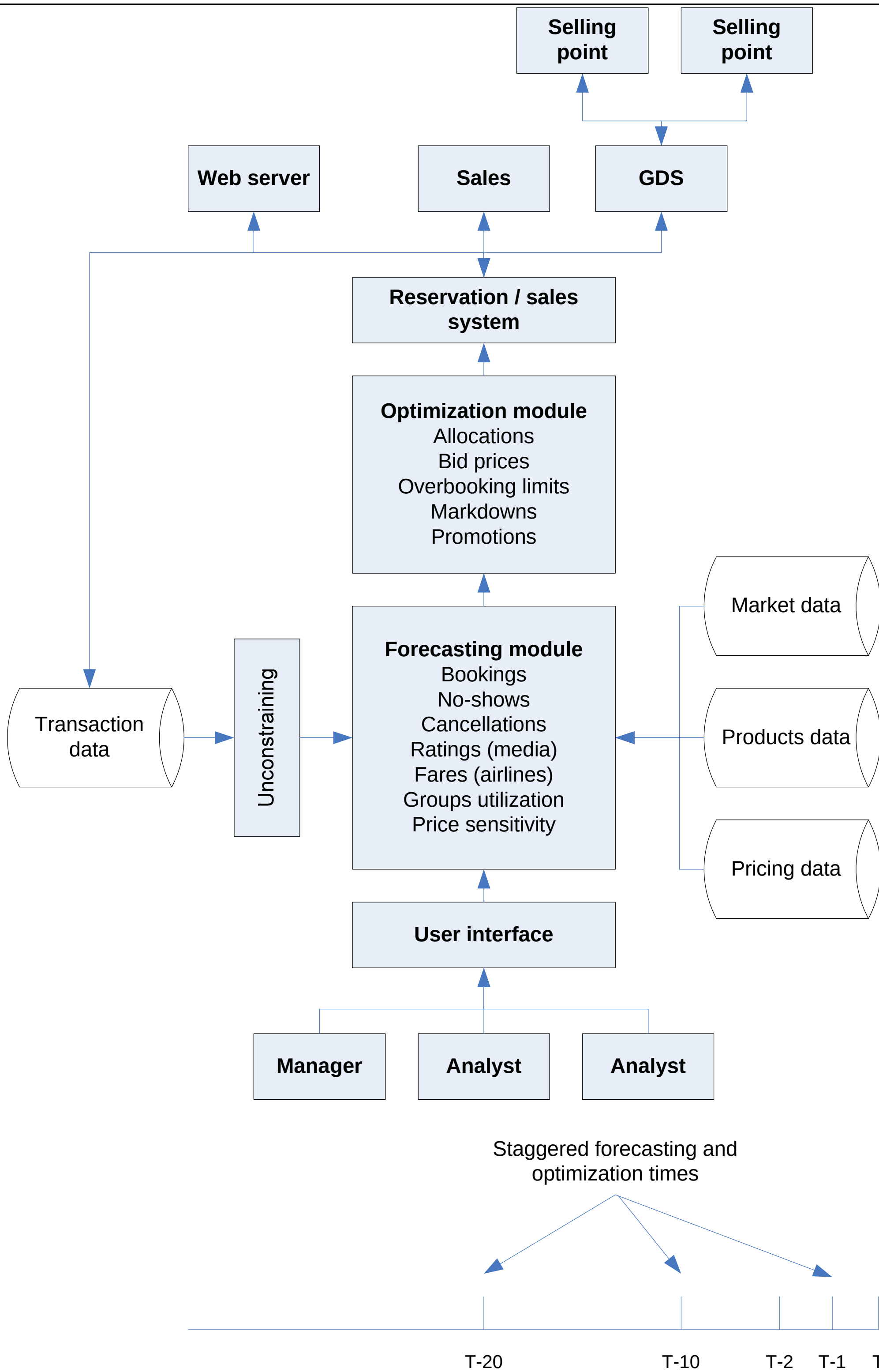
In general, our fares increase as the departure date gets nearer, so to get the best deals make sure you book as early as possible. In addition, the level of demand affects prices, so cheaper fares are often available during less busy periods - for example, on mid-week flights. We recommend that once you see a seat at a price you like you buy it immediately as it may no longer be selling at that fare if you come back later! (easyJet 2007)

RM generally follows four steps:

1. Data collection; Collect and store relevant historical data (prices, demand, causal factors);
2. Estimation and forecasting: Estimate the parameters of the demand model; forecast demand based on these parameters and other relevant quantities like no-show rates, cancellation rates.
3. Optimization: Find the optimal set of controls (allocations, prices, markdowns, discounts, overbooking limits) to apply until the next re-optimization.
4. Control: Control the sale of the inventory using the optimized controls. This is done either through the firm's own transaction processing systems or through shared distribution systems (GDSs).

Figure_APX A-1 shows the aforementioned process. Data is fed to the forecasting module; the forecasts are input to the control optimizer and the controls are uploaded into the control optimizer.

Figure_APX A-1 Forecasting module in a RM system. Periodic reforecasting timeline.



Adjusted from Talluri and van Ryzin *(2005, p. 409)*

If a closer look is taken to the required forecasts, it can be seen that quantity-based RM and price-based RM forecasts have different requirements.

Quantity-based RM forecasts require knowledge of the arrival process of different types of customers. This is also called booking-profile forecasts. In addition, cancellation and no-show rates are required. The first is usually modeled as a function of time, whereas the cancellations are modeled with a certain probability function. Optimization modules require spill and recapture quantities. Spill refers to the amount of demand that is lost when a class is closed; recapture is the amount of spilled demand recaptured by the firm's substitute products.

Price-based RM-forecasts require en estimate of the parameters of the demand function. Cross-price elasticity estimates may also be required when a there are significant substitution effects.

Estimation methods are usually some form of regression (i.e. time series analysis, moving average processes, auto-regressive processes) and is the calibration of a forecasts models. This is done relatively infrequently. Forecasts are carried out on an operational basis.

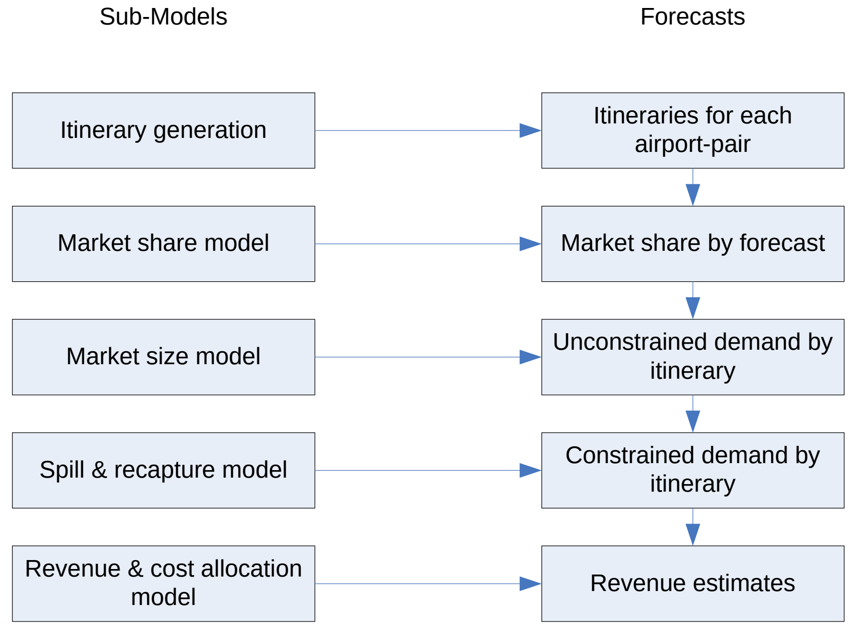
A.2 Network Planning Models

This section is largely based on the working paper "Schedule Planning" by Frank Koppelman, Gregory Coldren and Laurie Garrow (Koppelman*, et al.* 2006) .

The term "Network planning models" refers to a set of sub-models, as shown in Figure_APX A-2. The sub-models are:

- An *itinerary generation model* or algorithm, which is based on the Offical Airline Guide (OAG Worldwide Limited 2006). Itineraries are constructed from the OAG and include distance based circuitry logic, minimum and maximum connection times. They are typically generated for each day of the week to account for day-of-week differences. An exception is the itinerary generation developed by Boeing Commercial Airplanes. Boeing's algorithm integrates discrete choice theory into both the itinerary generation and itinerary selection.
- A *market share model* is used to predict the percentage of travelers that select each itinerary in an airport pair.
- Demand on each itinerary is determined by multiplying the forecasted *market size,* the number of travelers traveling between an airport-pair.
- The demand for certain flights may exceed the available capacity. *Spill and recapture models* are used to reallocate passengers from full flights to flights that have no capacity.
- *Revenue and cost allocation models* are used to determine the profitability of a certain flight or an entire schedule.

Figure_APX A-2 Components of network-planning models



(Koppelman, *et al.* 2006)

A.3 Experiences from travelers

During the course of the research, some acquaintances were questioned about their travel behavior in an informal context. Some excerpts of these talks are given in this section;

- An exchange student from Sweden, living in Zurich, searched for over a week on internet sites for cheap fares. This is something she only did if she planned a trip for a larger party, such as family or a group of friends. When booking her own ticket, she made a decision based on information she obtained from several airline sites. For both groups and herself she did not mind transfers, departure and arrival time and travel time. Also, departure and arrival days were flexible.
- A former consultant of McKinsey considered price to be an important factor, both for holiday and for business purposes. For business purposes, he often booked two itineraries. In both cases, he considered the outbound itinerary most important. The inbound itinerary was selected at a further point in the future. More important as price however, was arrival time. Ticket booking was often done by support staff, but he was aware of the prices as he saw the actual ticket prices on the project expenses bill and was responsible for the project budget. He helped older family members with booking their tickets, as he knew the travel system better. He suggested using corporate booking databases, and perhaps matching these with other databases. Large multinationals have not much to lose and can provide valuable information on choice behavior.
- A second consultant at a leading consultancy firm in the Netherlands delegated ticket bookings to a secretary. Here money was no issue: an airport close to his home in Rotterdam and the destination was most important. For private purposes, he knew most airlines serving the route in question and searched for the cheapest fare. However, he did not consider secondary airports,

which in his case were Eindhoven, Brussels and Girona. The latter airport is marketed by Ryanair as close to Barcelona.

- A lawyer at Heineken N.V. was restricted to using the corporate itinerary search engine and corporate travel agency. At Heineken, corporate policy stated that employees should choose the cheapest fare close to their preferred arrival time. If a transfer flight was cheaper, employees should choose the transfer flight, despite the longer travel time.
- A sales engineer at Philips N.V. was also restricted to using the corporate itinerary search engines. Arrival time was more important than price; he had relative freedom in choosing his tickets. More expensive, direct flights were always preferred. He did not use a travel portal when traveling for private purposes. He used a site that listed all carriers flying to a certain airport from a second region and then searched the carrier sites for the fares. Fare then became a decisive factor. He also calculated the total costs from origin to destination, so included fuel and/or public transport prices.

A.4 Opinions from practitioners

Several persons active in aviation have been questioned, more or less informally, about their ideas and thoughts on the application of discrete choice modeling in the aviation industry. Two persons both active in the academic world and with thorough practical experience with discrete choice models have been asked for their opinion on topics as formulated in 0. In addition, a recent article in the *Journal of Revenue and Pricing Management* summarizes the recent views of experts at the second annual Revenue Management and Price Optimization conference which was held at the Georgia Institute of Technology. The theme of the conference was on how the Internet is changing traditional revenue management and pricing practices (Garrow*, et al.* 2006) . Furthermore, a list of questions was sent to Schiphol Airport, department Aviation/Capacity Planning/Development & Innovation , based on issues as formulated in section 2.4 and as discussed by de Neufville and Odoni (2003) and Graham (2001) .

Laurie Garrow, assistant professor at Georgia Tech and a former analyst at United Airlines' revenue management department was asked for her view on discrete choice models and airlines, giving her a set of ideas in advance.

Determining the willingness-to-pay of potential passengers for service attributes and the itinerary with stated-preference surveys is a promising direction for future research. With regard to revenue management systems, several applications are imaginable, such as spill and recapture models and repurchase intent models. More theoretical applications, such as the implementation of discrete choice into revenue management are still considered to be in the what-if stage.

Philipp Fröhlich, former PhD student at the IVT, ETH Zurich and currently active in his own firm, has developed an application based on the software package VISUM, produced by PTV. Visum is traditionally used by private and public transport planners for strategic planning. Mr. Fröhlich adjusted the program in such a way, that it may be used for aviation purposes. The application is recently awarded with the third price for most-innovative application of VISUM.

Mr. Fröhlich sees several applications of his model, some of which are discussed in the paper "Design and Application Fields for a Worldwide Air Transport Model". First, the application serves as a powerful addition to the Official Airline Guide. It is possible to visualize competition on routes and airports that can be reached within a certain travel time of transfers. Second, the application can be used to assign transport demand to the aviation network. For the assignment, the minimum requirements are an origin-destination matrix and a choice model. Choice sets can be generated by Visum. These choice sets can be used for choice model estimation. Mr. Fröhlich is of the opinion that a dataset comparable to the *Origin Destination Data Bank 1A or Data Bank 1B* of the U.S. Department Transportation could be of aid for modeling air transport in Europe. The data are based on a 10 percent sample of flown tickets collected from the passengers as they board aircraft operated by U.S. airlines.

At the second annual Revenue Management and Price optimization conference, several panelists mentioned the ability to protect brad value while effectively selling off distressed inventory as a motivation for working with on-line travel agents. Hotels may offer highly discounted rates to Travelocity, Expedia and Orbitz for use in travel packages. The majority of panelists agree that the internet has increased price transparency and blurred traditional segmentation lines. Some panelists mention the awkward supplier relation between travel portals and their suppliers. People will shop online but then go to the supplier and shop direct. This especially holds when the supplier lures a customer with award points. Travelocity and Expedia view customized packages as a future direction, while hotels view the Internet as an opportunity to differentiate themselves. In addition, hotel representatives feel that they have a strong brand presence and thus may need not to rely heavily on on-line travel agents to be in a consumer's consideration set. Other related opportunities provided by the opportunity Internet related to customization of offers, screen presentations or prices. The ability to set price appropriately depends partly on the ability to monitor sales in real-time. Within the broader perspective, the ability to further customize offers, screen presentations, prices, etc. to individual consumers will be driven by the ability to analyze click-stream data. A lot of this data is unique to the on-line travel agents and not available to merchants such as airlines.

Prognoses

Wat voor typen prognoses met betrekking tot passagiers (transfer, Europese vluchten, intercontinentale vluchten) worden gemaakt?

Periodiek (circa 1 per jaar) worden er zgn vliegschema's gemaakt, met informatie over het aantal vluchten tijdens een piekweek (zomerweek). Deze omvat info als intercontinentaal/Europese vluchten, toesteltypes, airlines etc. O.b.v. van historie worden bezettingsgraden bepaald, om tot het aantal passagiers te komen. Van groot belang zijn de piekuren te bepalen.

Wordt er onderscheid gemaakt tussen andere type passagiers, bijv. business of vakantie?

Onderscheid tussen KLM (en haar partners), leisure en low costs (de Easyjets etc).

Wordt er onderscheid gemaakt tussen bestemmingen en herkomst van passagiers?

Ja, wat betreft de herkomst en bestemming is een onderscheid van groot belang: er wordt een onderscheid gemaakt tussen Schengen en niet-Schngen passagiers, plus een onderscheid tussen Europese en Intercontinentale passagiers. De eerste is van belang voor de passagiers stromen en de 2e voor de bagage stromen

Is het belangrijk om te weten wat voor type vliegtuigmaatschappij (charter, low-cost carrier, allianties) van Schiphol gebruik maken in de toekomst voor de prognoses?

Absoluut, KLM (plus partners) leveren veel transfererende passagiers (dus; maken overstap via schiphol naar andere luchthaven) en kennen dus andere processen dan de andere airlines met m.n. passagiers die van/naar Schiphol vliegen.

Indien er onderscheid (bijv in bestemming) wordt gemaakt, welk belang heeft dit voor de prognoses?

Zie eerder; van groot belang. Een verkeerde inschatting van airlines-segment en herkomst/bestemming levert een totaal andere passagiers-segment op en dus een verkeerde vraag naar de verschillende processoorten.

Hoe hangen deze prognoses met de prognoses voor het aantal vliegbewegingen?

Is 1 gezamenlijke prognose.

Op welke methoden zijn deze prognoses gebaseerd (bijv. regressie, zwevend gemiddelde, trend extrapolatie)?

Enerzijds o.b.v. trends, maar belangrijkste zijn de marktonderzoeken en marktverkenningen; wat verwachten de airlines te gaan doen (nieuwe bestemmingen, vlootontwikkelingen etc)

Geven deze methoden inzicht in de relatieve waardering van verschillende aspecten van het vliegveld, de afweging welke de consumenten maken?

Niet in dit proces; waardering wordt apart via tevredenheidsonderzoeken en benchmarks uitgevoerd.

Instrumenten

Hebt u inzicht in hoe de afweging tot toegang tot het vliegveld, auto of openbaar vervoer, gemaakt wordt?

Via marktonderzoeken/enquêtes wordt deze informatie inderdaad onderzocht

Hebt u inzicht in waarom reizigers expliciet voor Schiphol kiezen? Denkt u deze keuze te kunnen beïnvloeden?

Wederom, via marktonderzoeken/enquêtes. Er worden inderdaad studies en projecten uitgevoerd om ons marktaandeel te optimaliseren, zoals betere/snellere bereikbaarheid van passagiers op grotere afstand.

Schiphol is aandeelhouder in meerdere vliegvelden in Nederland. In de toekomst kan het zijn, dat Schiphol passagiers en vluchten wil verplaatsen naar andere vliegvelden in Nederland.
Welke instrumenten zou Schiphol tot haar beschikking hebben om dit te realiseren?

In principe hebben wij een vaste capaciteit beschikbaar die nog beperkt kan groeien. Als de vraag groter is/wordt dan de capaciteit, zullen airlines automatisch de keuze moeten maken of ze op een andere luchthaven gaan vliegen .

Denkt u dat het mogelijk is vluchten/vliegtuigmaatschappijen naar een ander vliegveld te dirigeren?

In theorie kunnen we dit niet. Iedereen die wil, kan een 'kaartje' kopen voor een slot; dus hoe laat willen ze aankomen en vertrekken. Als zij het afgelopen jaar een slot hadden voor dit tijdstip en ze hebben er het afgelopen jaar minimaal 80% van de dagen gebruik van gemaakt, dan hebben zij het recht om dit zelfde slot het komende jaar weer te gebruik. We kunnen dus niet weigeren.

Slot allocatie

Hoe worden slots op dit moment vergeven op Schiphol? Wordt deze vergeven met het oog op frequentie en bestemming van de vluchten?

Zie vorige vraag. In principe kunnen wij dit niet sturen (Internationale luchthaven wetgeving). Deze wetgeving kent wel vele voorwaarden met criteria, maar daar hebben wij geen invloed op.

Vind er prijsdifferentiatie plaats bij deze slots? Zo ja, waar is deze op gebaseerd?

Enige verschil maken wij voor de airlines met korte omdraaitijden, zoals easyjet. Deze worden op een 'simpele' pier afgewerkt met minder luxe en service. Zij krijgen iets lagere tarieven.

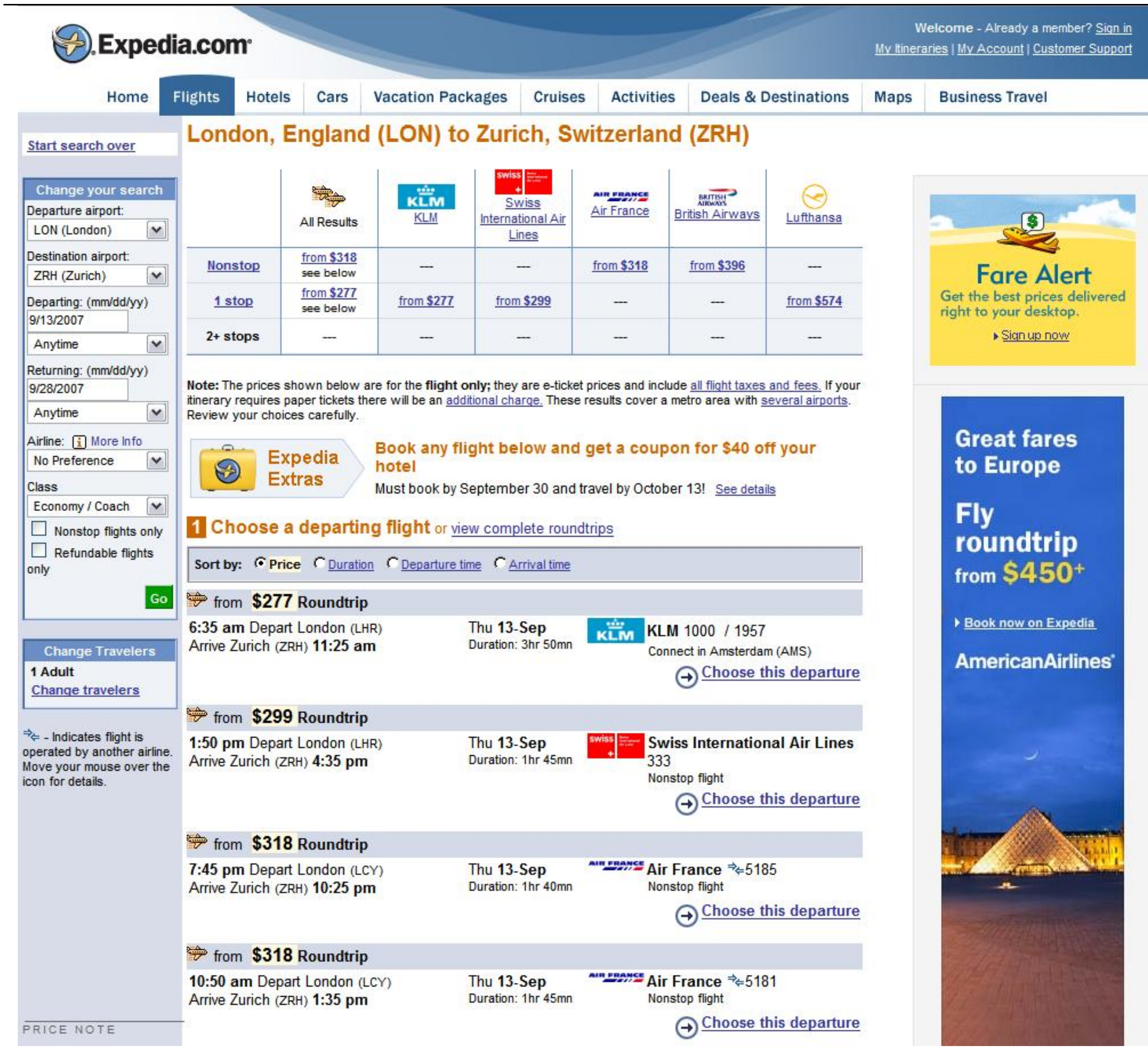
A.5 Interfaces

Figure_APX A-3 Example of Expedia query screen

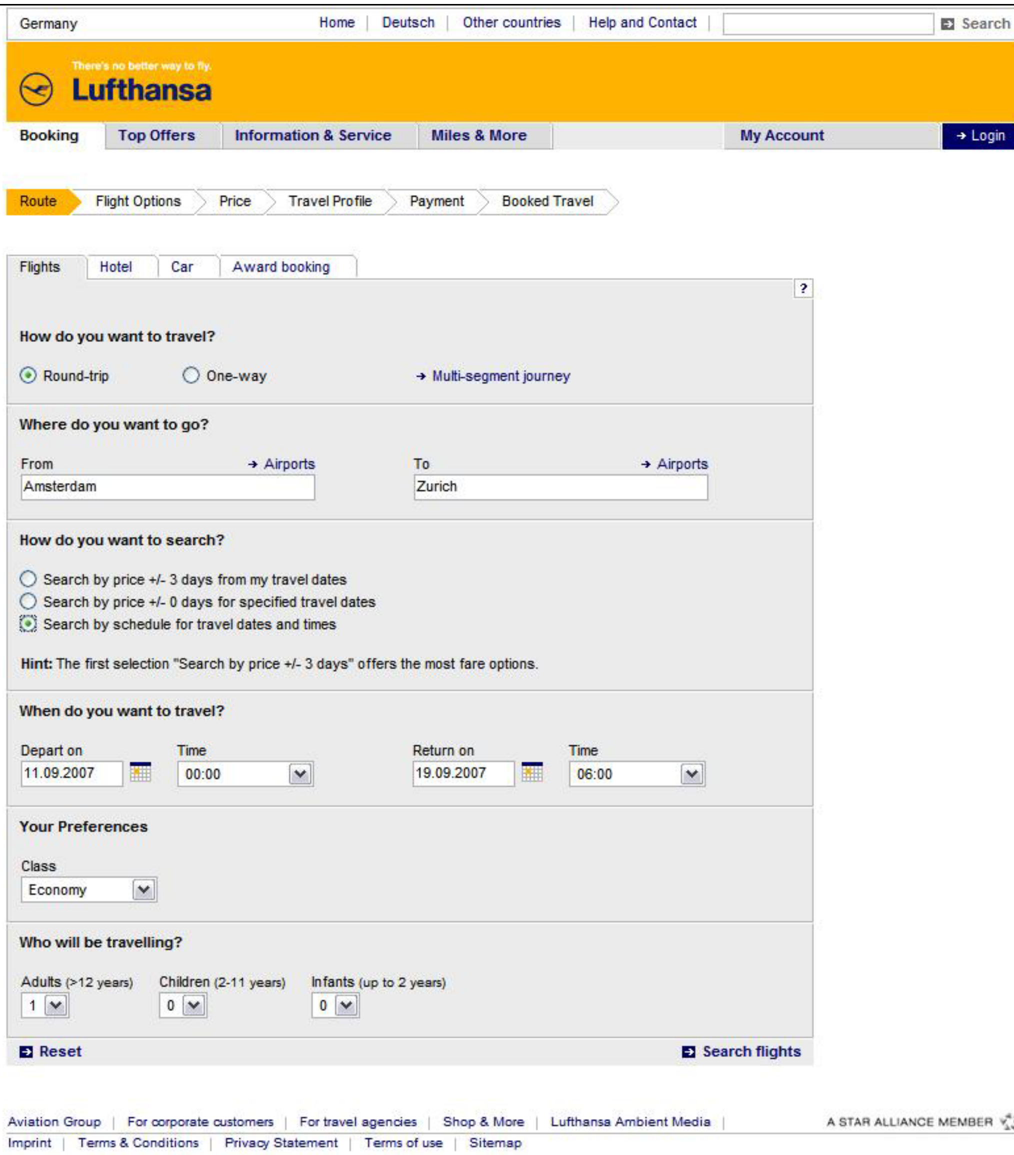
Figure_APX A-3 shows an example of a typical Expedia query screen. A potential customer can enter the origin and destination airport, departure and return data and can limit the search results by entering an airline preference, departure time outbound, return time inbound, class and number of transfers.

The results of the query can be seen in Figure_APX A-4. A potential customer can sort the itineraries by price, duration, departure time and arrival time. Furthermore, a separate frame is available, which shows a matrix with airline, number of transfers and fare.

Figure_APX A-4 Example of Expedia result screen



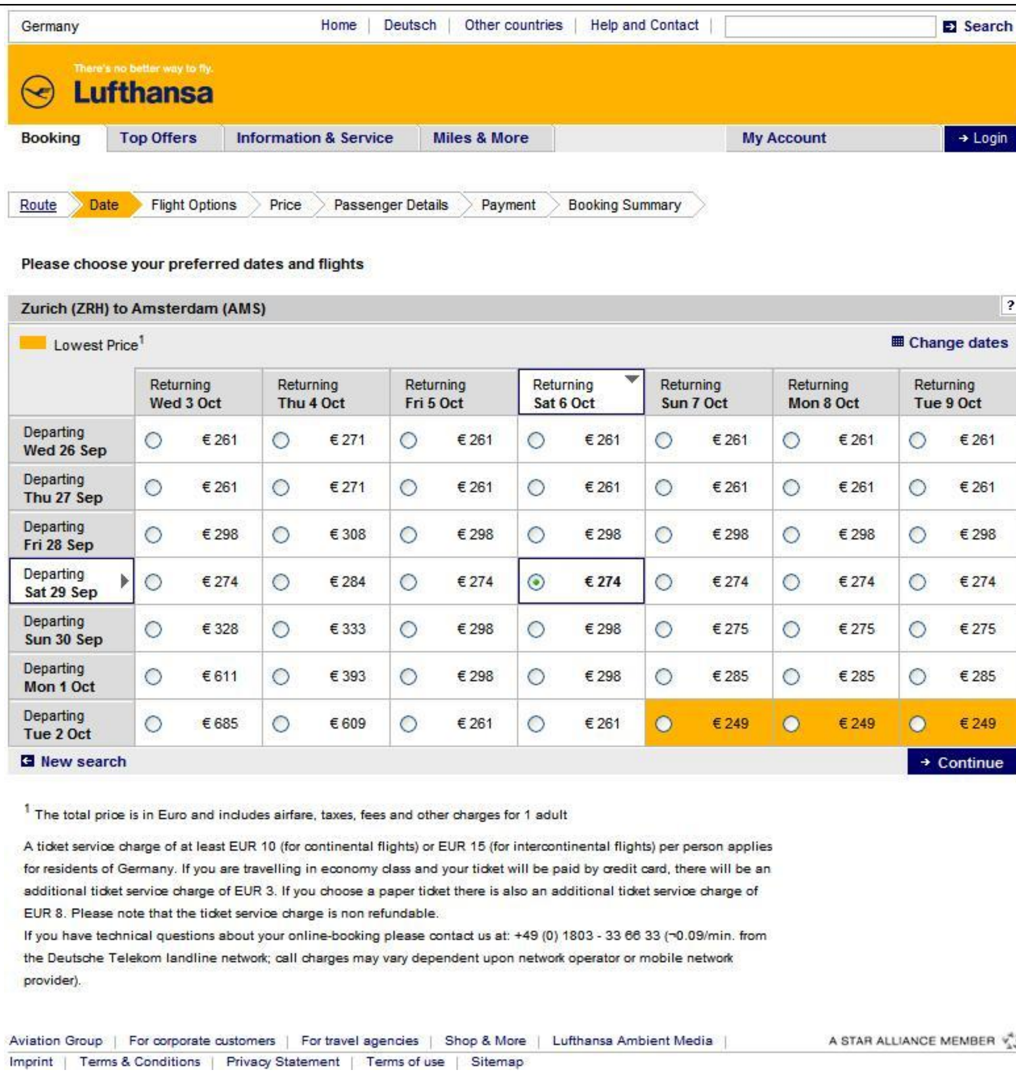
Figure_APX A-5 Example of Lufthansa query screen



A similar query screen can be found on the Lufthansa website, as shown in Figure_APX A-5. The results of the query differ: first, a screen is shown with possible combinations of outbound and inbound date. The prices listed in this screen are the cheapest available combination of outbound and inbound itinerary. Figure_APX A-7 shows the follow-up screen. In this screen, the cheapest itinerary is highlighted yellow. Other combinations vary, dependent on departure time and fare class. Interesting to see here is the two-stage choice process, where first a choice is made between departure date, followed by a choice for departure time.

In both cases, it is reasonable to assume that it would be possible to record the choice set(s) belonging to a booking or reconstruct the choice set.

Figure_APX A-6 Example of Lufthansa result screen – Part 1



Germany | Home | Deutsch | Other countries | Help and Contact | Search

There's no better way to fly.
Lufthansa

Booking | Top Offers | Information & Service | Miles & More | My Account | → Login

Route > Date > Flight Options > Price > Passenger Details > Payment > Booking Summary

Please choose your preferred dates and flights

Zurich (ZRH) to Amsterdam (AMS)

Lowest Price[1] | Change dates

	Returning Wed 3 Oct	Returning Thu 4 Oct	Returning Fri 5 Oct	Returning Sat 6 Oct	Returning Sun 7 Oct	Returning Mon 8 Oct	Returning Tue 9 Oct
Departing Wed 26 Sep	€ 261	€ 271	€ 261	€ 261	€ 261	€ 261	€ 261
Departing Thu 27 Sep	€ 261	€ 271	€ 261	€ 261	€ 261	€ 261	€ 261
Departing Fri 28 Sep	€ 298	€ 308	€ 298	€ 298	€ 298	€ 298	€ 298
Departing Sat 29 Sep	€ 274	€ 284	€ 274	€ 274	€ 274	€ 274	€ 274
Departing Sun 30 Sep	€ 328	€ 333	€ 298	€ 298	€ 275	€ 275	€ 275
Departing Mon 1 Oct	€ 611	€ 393	€ 298	€ 298	€ 285	€ 285	€ 285
Departing Tue 2 Oct	€ 685	€ 609	€ 261	€ 261	€ 249	€ 249	€ 249

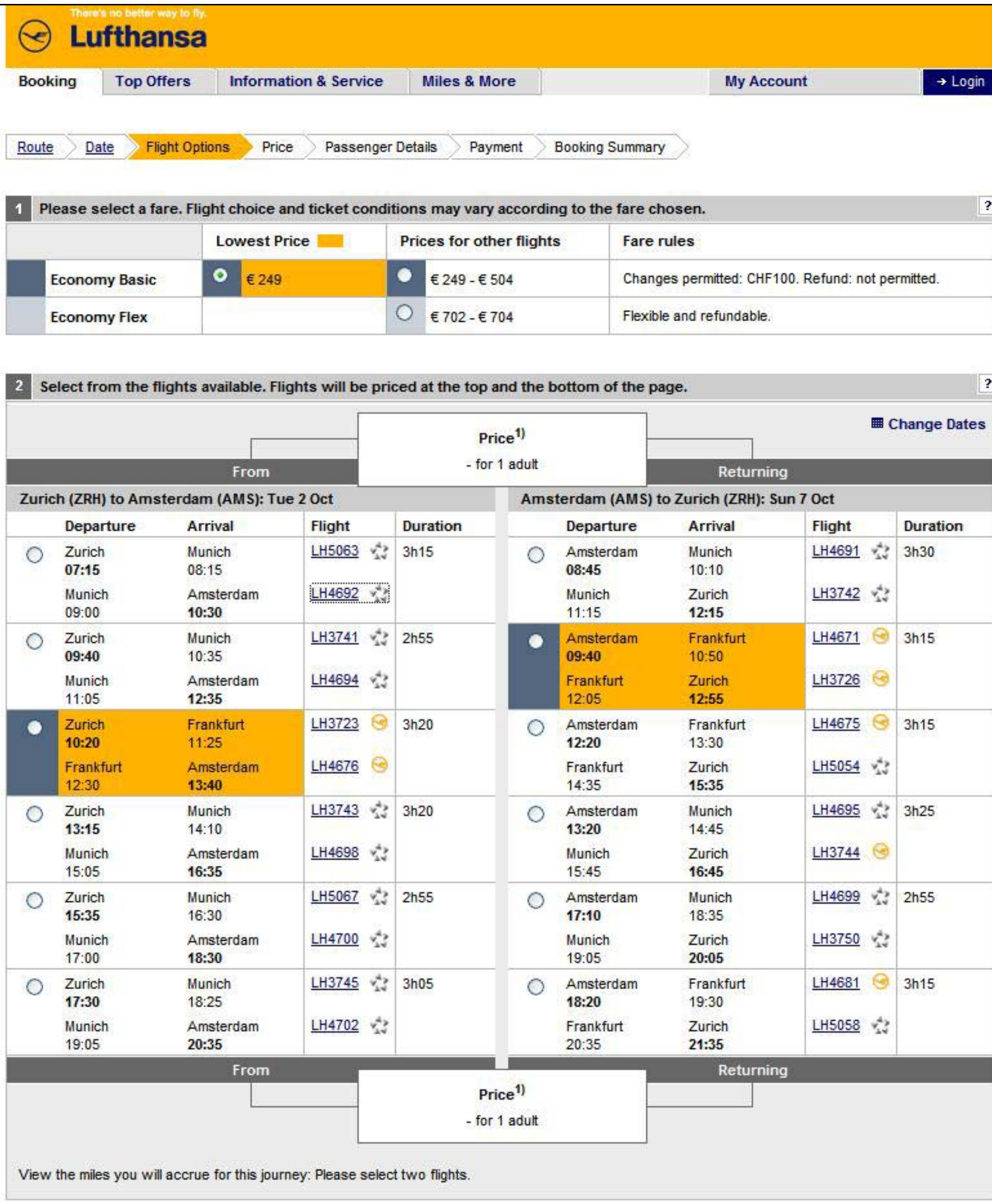
New search | → Continue

[1] The total price is in Euro and includes airfare, taxes, fees and other charges for 1 adult

A ticket service charge of at least EUR 10 (for continental flights) or EUR 15 (for intercontinental flights) per person applies for residents of Germany. If you are travelling in economy class and your ticket will be paid by credit card, there will be an additional ticket service charge of EUR 3. If you choose a paper ticket there is also an additional ticket service charge of EUR 8. Please note that the ticket service charge is non refundable.
If you have technical questions about your online-booking please contact us at: +49 (0) 1803 - 33 66 33 (~0.09/min. from the Deutsche Telekom landline network; call charges may vary dependent upon network operator or mobile network provider).

Aviation Group | For corporate customers | For travel agencies | Shop & More | Lufthansa Ambient Media | A STAR ALLIANCE MEMBER
Imprint | Terms & Conditions | Privacy Statement | Terms of use | Sitemap

Figure_APX A-7 Example of Lufthansa result screen – Part 2


There's no better way to fly.

Lufthansa

Booking | Top Offers | Information & Service | Miles & More | My Account | → Login

Route > Date > Flight Options > Price > Passenger Details > Payment > Booking Summary

1 Please select a fare. Flight choice and ticket conditions may vary according to the fare chosen. ?

	Lowest Price	Prices for other flights	Fare rules
Economy Basic	€ 249	€ 249 - € 504	Changes permitted: CHF100. Refund: not permitted.
Economy Flex		€ 702 - € 704	Flexible and refundable.

2 Select from the flights available. Flights will be priced at the top and the bottom of the page. ?

Change Dates

Price[1)] - for 1 adult

From

Zurich (ZRH) to Amsterdam (AMS): Tue 2 Oct

Departure	Arrival	Flight	Duration
Zurich 07:15	Munich 08:15	LH5063	3h15
Munich 09:00	Amsterdam 10:30	LH4692	
Zurich 09:40	Munich 10:35	LH3741	2h55
Munich 11:05	Amsterdam 12:35	LH4694	
Zurich 10:20	Frankfurt 11:25	LH3723	3h20
Frankfurt 12:30	Amsterdam 13:40	LH4676	
Zurich 13:15	Munich 14:10	LH3743	3h20
Munich 15:05	Amsterdam 16:35	LH4698	
Zurich 15:35	Munich 16:30	LH5067	2h55
Munich 17:00	Amsterdam 18:30	LH4700	
Zurich 17:30	Munich 18:25	LH3745	3h05
Munich 19:05	Amsterdam 20:35	LH4702	

Returning

Amsterdam (AMS) to Zurich (ZRH): Sun 7 Oct

Departure	Arrival	Flight	Duration
Amsterdam 08:45	Munich 10:10	LH4691	3h30
Munich 11:15	Zurich 12:15	LH3742	
Amsterdam 09:40	Frankfurt 10:50	LH4671	3h15
Frankfurt 12:05	Zurich 12:55	LH3726	
Amsterdam 12:20	Frankfurt 13:30	LH4675	3h15
Frankfurt 14:35	Zurich 15:35	LH5054	
Amsterdam 13:20	Munich 14:45	LH4695	3h25
Munich 15:45	Zurich 16:45	LH3744	
Amsterdam 17:10	Munich 18:35	LH4699	2h55
Munich 19:05	Zurich 20:05	LH3750	
Amsterdam 18:20	Frankfurt 19:30	LH4681	3h15
Frankfurt 20:35	Zurich 21:35	LH5058	

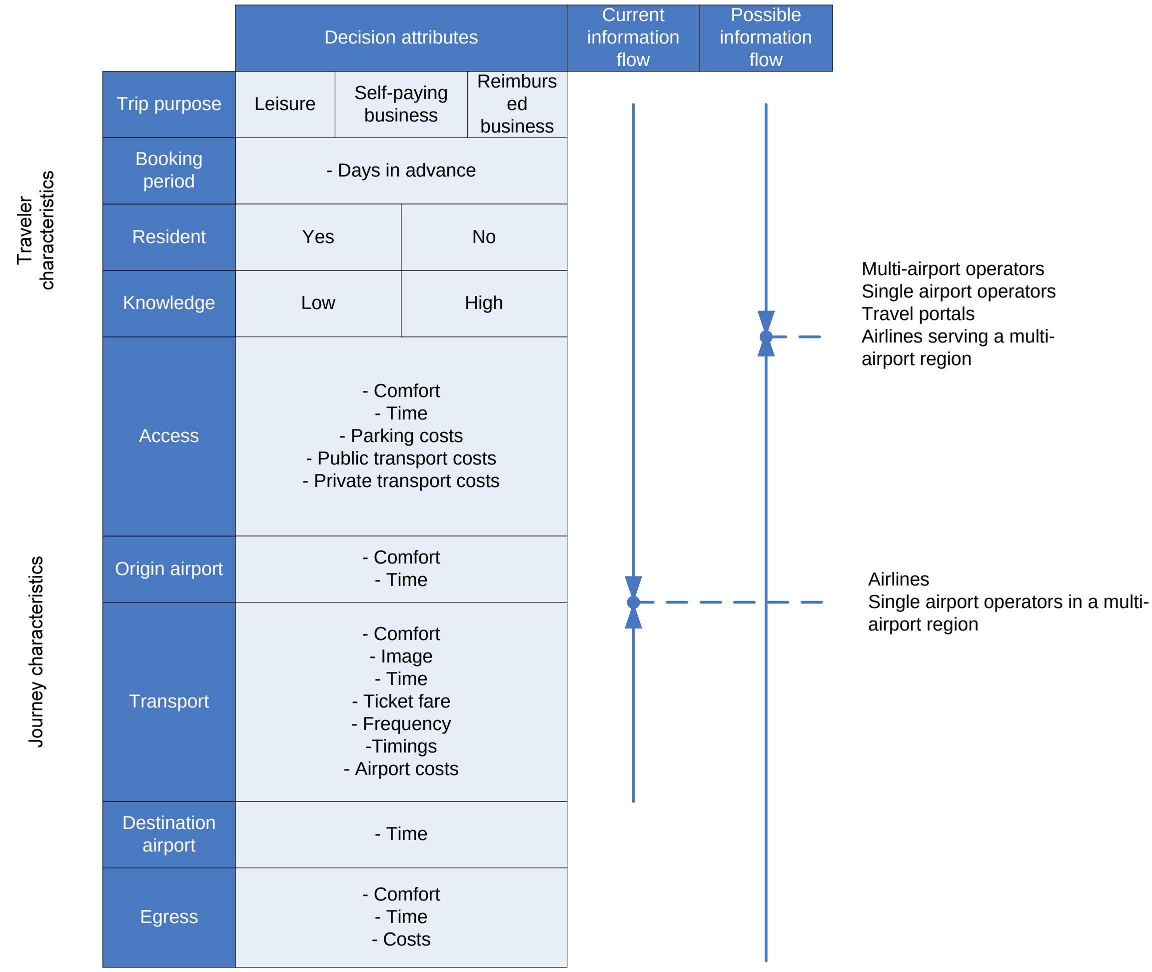
From | Returning

Price[1)] - for 1 adult

View the miles you will accrue for this journey: Please select two flights.

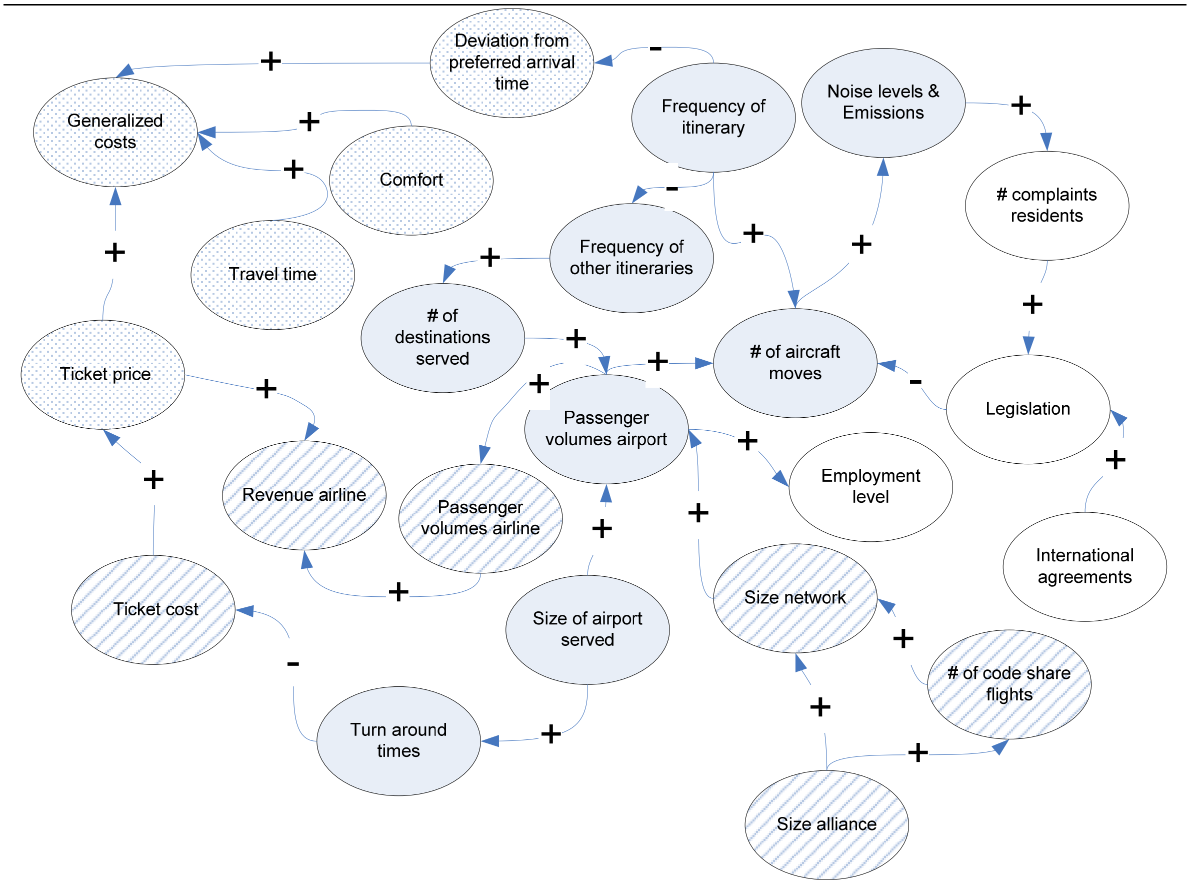
A.6 Information flows

Figure_APX A-8 Current information and possible future information flow



A.7 Causal Diagrams

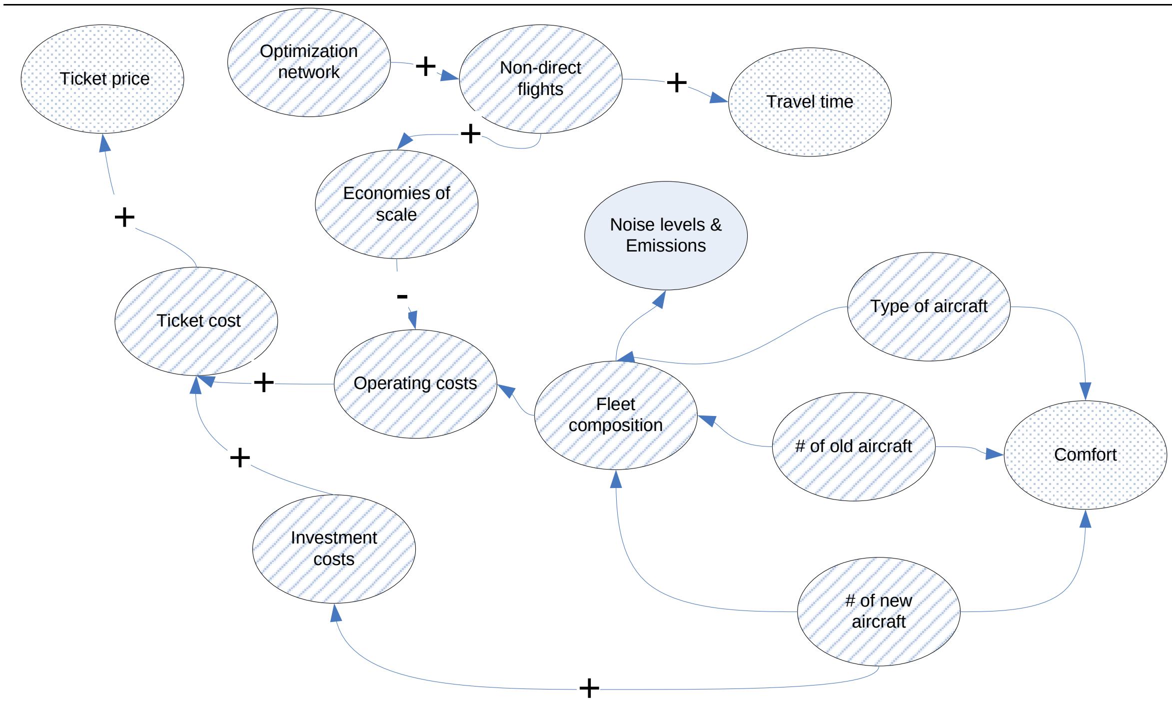
Figure_APX A-9 Global causal diagram



Legenda:
Dotted: passenger specific
Dashed: airline specific
Gray: airfield specific
White: societal environment

In the diagram, a negative feedback loop can be recognized, between the variables legislation, # of aircraft moves, noise levels & emission levels and # complaint residents. The effects can be diminished by decreasing the number of itineraries departing or lowering passenger volumes. The former will lead the airfield to be less attractive for passengers, as already noticed by Graham (2001).

Figure_APX A-10 Causal diagram airline



Appendix B Notes to Chapter 3

The discussion in this Appendix follows the derivation of the MNL-model as documented by Train (2003, p. 38).

B.1 Derivation

The logit model is obtained by assuming that each ε_{iq} is independently, identically distributed extreme value. The distribution is also known Gumbel and type I extreme value. The variance-covariance matrix of the error terms ε_{iq} thus has the following form:

$$\Omega = \sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The density for each unobserved part of the utility is:

$$f(\varepsilon_{jq}) = e^{-\varepsilon_{jq}} e^{e^{-\varepsilon_{jq}}} \quad (1.4)$$

And the cumulative distribution function is:

$$F(\varepsilon_{jq}) = e^{-\varepsilon_{jq}} \quad (1.5)$$

Following McFadden (1974) and knowing that the probability that a decision-maker q chooses alternative i is:

$$P(i \mid C_q) = P[\varepsilon_{jq} \le \varepsilon_{iq} + U_{iq} - U_{jq} \forall j \in C_q] \quad (1.6)$$

If ε_{iq} is given, this expression is the cumulative distribution for each ε_{jq} evaluated at $\varepsilon_{iq} + U_{iq} - U_{jq}$. With expression (1.5) and the assumption that all $\varepsilon's$ are independent, this is the product of the individual cumulative distributions:

$$P_{iq} \mid \varepsilon_{iq} = \prod_{j \ne i} e^{-e^{-(\varepsilon_{iq}+U_{iq}-U_{jq})}} \quad (1.7)$$

As ε_{iq} is not given, the choice probability is the integral of $P_{qi} \mid \varepsilon_{iq}$ over all values of ε_{iq} weighted by its density (1.4):

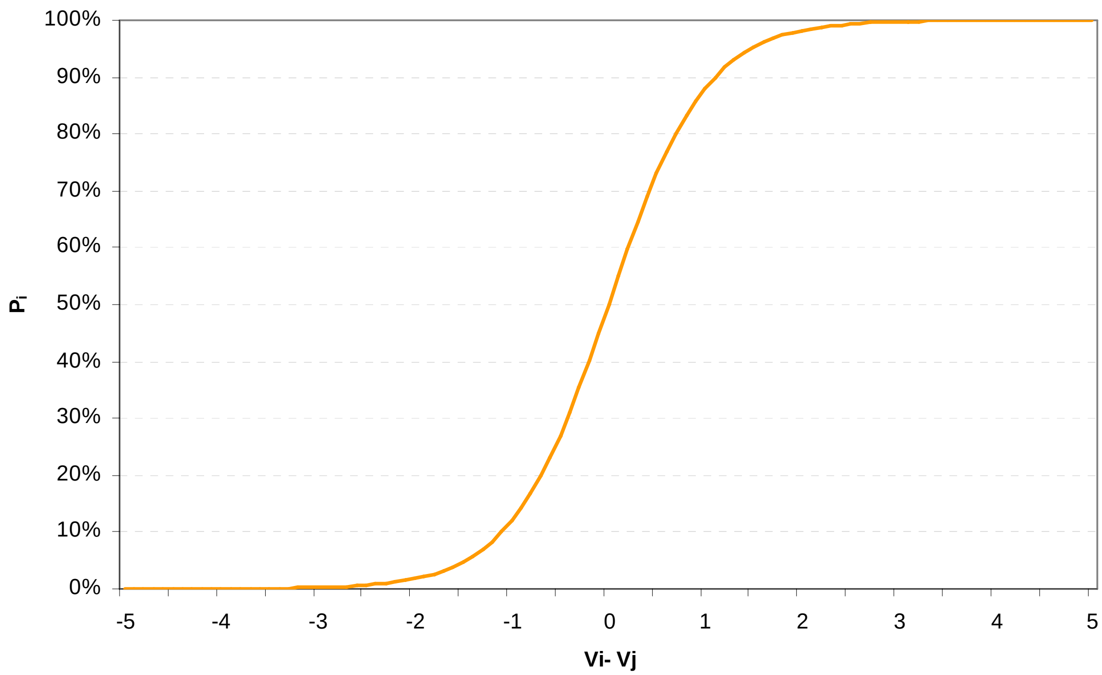
$$P_{iq} = \int \left(\prod_{j \ne i} e^{-e^{-(\varepsilon_{iq}+U_{iq}-U_{jq})}} \right) e^{-\varepsilon_{jq}} e^{e^{-\varepsilon_{jq}}} \, d\varepsilon_{iq} \quad (1.8)$$

By solving this integral, the expression of the logit model is obtained:

$$P_{iq} = \frac{e^{V_{iq}}}{\sum_j e^{V_{jq}}} \quad (1.9)$$

The shape of the model is visualized in Figure_APX B-1.

Figure_APX B-1 Graph of logit curve where V_i = -V_j



B.2 Estimation

The parameters of the logit model are estimated by means of the maximum likelihood estimation method. The likelihood function takes the following form:

$$LL(\beta_1, \beta_2, .., \beta_k) = \prod_{q=1}^{Q} P_q(i)^{y_{iq}} P_q(j)^{y_{qn}} \quad (1.10)$$

Where y_{iq} is 1 if person *n* chose alternative i and 0 if person q chose alternative j and $P_q(i)$ is a vector of $\beta_1, \beta_2, .., \beta_k$. The likelihood function is transformed into the log-likelihood function:

$$LL(\beta_1, \beta_2, .., \beta_k) = \sum_{q=1}^{Q} [y_{iq} \log P_q(i) + y_{jq} \log P_q(j)] \quad (1.11)$$

The objective is to find the maximum for this function by differentiating it with respect to the β's and setting the partial derivatives equal to zero.

Appendix C Notes to Chapter 4 and Chapter 5

This appendix provides an extensive overview of the data sets used for this research. Special attention is paid to steps undertaken to generate the final data file as used for model estimation and the statistics as presented in the Chapter 5. The goal of this overview is to provide transparency and ensure the data and generation tools can be used by other users.

C.1 Expedia Dataset

The data collected by the webbots covers 70 origin-destination pairs and three types of trips: return trip on the same day, return trip on the next day and return trip 2 weeks later. Table_APX C-1 provides an overview of some of the information collected by the webbots. In Table_APX C-2 an overview is given from all the fields collected from the Expedia site. It may be important to mention that all journey information is one entry in the database, opposite to what is the case with the MIDT dataset.

The 70 od-pairs were selected based upon travel time, number of transfers and frequency. For each possible combination at least 2 journeys were chosen. The characteristics were derived from the IVT air transport model for the year 2003.

Table_APX C-1 Visualization of collected Expedia data

Origin	*Destination*	*Query date*	*Departure date*	*Return date*	*Departure time*	*Airline out*	*Fare*
Amsterdam	Toulouse	12-09-2006	1-11-2006	1-11-2006	07:00	KLM	167.73
Amsterdam	Toulouse	12-09-2006	1-11-2006	2-11-2006	07:00	KLM	167.73
Amsterdam	Toulouse	12-09-2006	1-11-2006	15-11-2006	07:00	KLM	167.73
Amsterdam	Toulouse	12-09-2006	1-11-2006	1-11-2006	07:40	Lufthansa	198.59
Amsterdam	Toulouse	12-09-2006	2-11-2006	2-11-2006	07:00	KLM	167.73
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
Zurich	Warsaw	29-11-2006	30-11-2006	14-12-2006	07:55	LOT	329.28

Table_APX C-2 Information collected from Expedia

Name	*Label*
Orgn	Original Airport (IATA Code)
Dstn	Destination Airport (IATA Code)
DepDate	Departure Date
RetDate	Return Date
ID	ID
TEXT	Text (all columns, divided by semicolon)
TIMESTAMP	Timestamp (time of the query)
CURRENCY	Currency of the fare
FARE	Amount of the fare

Name	*Label*
TAXCURRENCY	Currency of the tax
TAX	Amount of the tax
ORI_ARP_OUT	Original airport outbound flight
DES_ARP_OUT	Destination airport outbound flight
DEP_TIME_OUT	Departure time outbound flight
ARR_TIME_OUT	Arrival time outbound flight
TRIPTIME_OUT	Duration of the trip (way out)
ALN_OUT	Airlines outbound flight
CNX_OUT	Connections outbound flight
ALN1_OUT	Airline first leg outbound flight
ALN2_OUT	Airline second leg outbound flight
ALN3_OUT	Airline third leg outbound flight
FLTNUM1_OUT	Flightnumber first leg outbound flight
FLTNUM2_OUT	Flightnumber second leg outbound flight
FLTNUM3_OUT	Flightnumber third leg outbound flight
CNX1_OUT	First connection outbound flight
CNX2_OUT	Second connection outbound flight
ORI_ARP_IN	Original airport inbound flight
DES_ARP_IN	Destination airport inbound flight
DEP_TIME_IN	Departure time inbound flight
ARR_TIME_IN	Arrival time inbound flight
TRIPTIME_IN	Duration of the trip (way in)
ALN_IN	Airlines inbound flight
CNX_IN	Connections inbound flight
ALN1_IN	Airline first leg inbound flight
ALN2_IN	Airline second leg inbound flight
ALN3_IN	Airline third leg inbound flight
FLTNUM1_IN	Flightnumber first leg inbound flight
FLTNUM2_IN	Flightnumber second leg inbound flight
FLTNUM3_IN	Flightnumber third leg inbound flight
CNX1_IN	First connection inbound flight
CNX2_IN	Second connection inbound flight

From time-to-time, the webbots were "down" and collected no price information for November 2006. If we look at the table above, this would mean that several query dates for several od-pairs are lacking.

For several reasons it is necessary to complete the set of records: a complete overview of the prices is needed for a complete analysis of the dynamics of the prices, and more important, it could be that a set of alternatives is needed on the query date because a booking took place on that date. The latter can be determined with the MIDT dataset, which will be discussed in the next paragraph.

Most important is the method used for determining the missing alternatives and prices. The approach followed will be discussed below.

For each query date and od-pair it is determined when the webbots were down. This is simply done by determining if there are records available on a certain date for an od-pair. If no records are available, a lower bound date and an upper-bound date are determined. The lower bound is the last date the webbot was up, the upper bound is the first date the webbot is up again. For instance, for Amsterdam-Toulouse the webbot was down on 16-09-2006. The last date the webbot was running, was on the 15-09-2006, the first date the webbot is up again is on 19-09-2006.

It can only said with certainty that the alternatives for each combination of departure day and return day are available on the missing date if the alternative is available on the lower bound date and is available on the upper bound date. The missing fare can then be calculated as follows:

$$P = P_{LB} + (D_{MISSING} - D_{LB})\frac{P_{UB} - P_{LB}}{D_{UB} - D_{LB}}$$

Where:

P = Price

D = Date collection id number

$MISSING$ = Date on which observation is missing

LB = Lower bound

UB = Upper bound

These records are inserted in a new table (new_webbot_records), which is merged with the table containing original webbot records. The result is a table containing all price observations (price_observations).

Number of records collected by the webbots and number of records created are listed in Table_APX C-3.

Table_APX C-3 Number of Records Expedia Dataset

Table	*Number of records*
Webbot records	7.926.392
“New” records	2.494.924

C.2 MIDT Dataset

C.2.1 General description

The MIDT dataset obtained from SWISS/Lufthansa contains information on all airline tickets booked for November 2006 in the world through CRS systems. In contrary to the Expedia dataset, each entry represents a segment. In total, 55 fields are listed in the dataset. Table_APX C-5 shows a few of these fields, in Table_APX C-7 Fields MIDT Datasetall fields are listed. For each segment, departure time, arrival time and flight number is listed.

Table_APX C-4 Number of Records MIDT Datasets

Table	Table name in database	Number of records
MIDT	midt	37,671,470
MIDT Europe	midt_eur	12,838,271
MIDT Selected od-pairs	midt_sel_od	323,990
Segments selected od-pairs	segments_midt	323,990
Trips selected od-pairs	trips_midt	314,672
Journeys selected od-pairs	journeys_midt	213,188
Outbound + inbound journeys	journeys_midt	101,601
Outbound journeys	journeys_midt	111,587

Journeys can be distilled from the dataset by matching some fields, such as CRS (shown), trip origin and destination, booking date, IATA, record locator etc. This approach is confirmed by the Operations Research/Revenue Management department of SWISS/Lufthansa. In order to ensure the right sequence of segments and trips, a small program was written to extract trips and journeys and insert them in new tables. The number of records per table is listed in Table_APX C-6.

Four tables are made of the original dataset; a table containing journeys, a table containing trips, a table containing segments and a table containing flights. The latter could be included with the trips table. This set-up makes it easier to extract statistics.

Table_APX C-5 Example of MIDT dataset

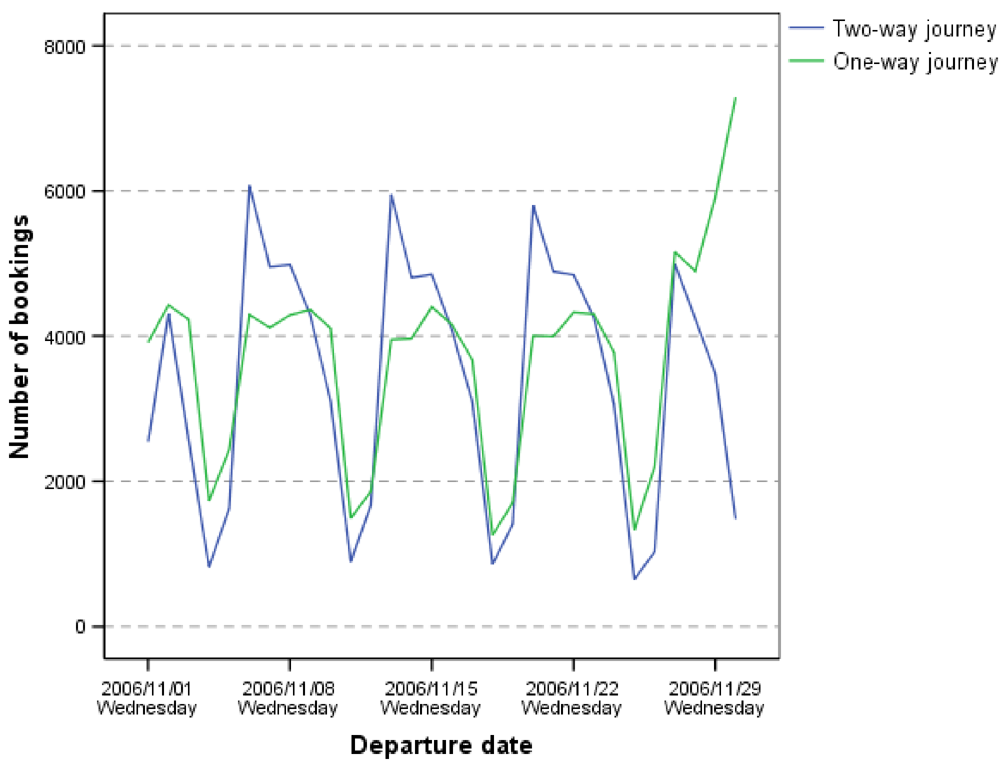
CRS	*Booking date*	*Trip departure date*	*Trip origin*	*Trip destination*	*Trip ID*	*Segment origin*	*Segment destination*
A	12-09-2006	1-11-2006	Amsterdam	Toulouse	100058863	Amsterdam	Paris
A	12-09-2006	1-11-2006	Amsterdam	Toulouse	100058863	Paris	Toulouse
A	12-09-2006	1-11-2006	Toulouse	Amsterdam	100058864	Toulouse	Paris
A	12-09-2006	1-11-2006	Toulouse	Amsterdam	100058864	Paris	Amsterdam
…	…		…	…		…	…
…	…		…	…		…	…

Table_APX C-6 Number of Records in MIDT Datasets

Table	*Table name in database*	*Number of records*
MIDT	midt	37,671,470
MIDT Europe	midt_eur	12,838,271
MIDT Selected od-pairs	midt_sel_od	323,990
Segments selected od-pairs	segments_midt	323,990
Trips selected od-pairs	trips_midt	314,672
Journeys selected od-pairs	journeys_midt	213,188
Outbound + inbound journeys (complete itineraries)	journeys_midt	101,601
Outbound journeys	journeys_midt	111,587

Figure_APX C-1 shows the number of bookings for two-way journeys and one-way journeys. Travelers strongly prefer departing on week-days, where travelers departing in the beginning of the week show a preference for departing in the beginning of the week. Furthermore, the number of two-way journeys is exceeded by the number of one-way journeys. This is also shown in Table_APX C-6.

Figure_APX C-1 Share of number of segments in trip



Table_APX C-7 Fields MIDT Dataset

Fieldname	*Description*
CRS	CRS source of booking
RL	Record Locator
IATA	IATA Number of Agency
DEP_DATE	Departure Date
SEG_DEP	origin of segment
SEG_ARR	destination of segment
SEG_AL	segment airline
SEG_FLTNR	segment flight number
SEG_BKG_CLS	segment booking class
TRIP_PAX	number of passengers
ADD_CANCEL	Booking/Cancellation Identifier
PCC	Pseudo City Code
BKG-DATE	Booking Date
NOT_USED_1	
TRIP_ACCTID	account ID
SEG_TRANS_CLS	segment translation class
NOT_USED_2	
ARR-DATE	Local Arrival Date of segment
UTC_ARR_DATE	utc arrival date of segment
UTC_DEP_DATE	utc departure date segment
UTC_DEP_TIME	UTC departure time
SEG_DEP_TIME	Local departure time segment
UTC-DEP-VAR	UTC Time Variation at Departure
UTC_ARR_TIME	utc arrival time of segment
SEG_ARR_TIME	Segment Local Arrival Time
UTC_ARR_VAR	UTC Time Variation at Arrival
DATE_VAR	Date Variation between Departure and Arrival
STOPS	Numbe of Stops in Segment
AIRCRAFT	Aircraft Type
NOT_USED_3	
NOT_USED_4	
GCD	Great Circle Distance of Segment
DEP_CITY_CODE	departure city code
ARR_CITY_CODE	arrival city code
DEP_COUNTRY_CODE	departure country code
ARR_COUTNRY_CODE	arrival country code
DEP_CONTINENT_CODE	departure continent code
ARR_CONTINENT_CODE	arrival continent code
TRIP_ID	unique trip identifier
TRIP_ORG	Origin
TRIP_DST	Destination

Fieldname	Description
TRIP_SEG_CNT	number of segments in trip
TRIP_BEG_DATE	begin date
TRIP_END_DATE	end date
TRIP_STOPS	number of stops
TRIP_PNR_ORG	First Trip Origin in PNR
TRIP_AL	Trip Airline - Dominant Carrier Calculation Model
TRIP_OP_AL	Trip Operating Airline - Dominant Carrier Calculation Model (optionally)
TRIP_AL_ALL	all trip airlines (max3) concatenated
TRIP_TRANS_CLS	cabin class
TRIP_ELAPS_TIME	elapsed time (of whole trip)
SEG_SEQ_NO	Position of Segment in Trip
SEG_PRO_PAX	Prorated Passenger of Segment
SEG_CONNEX_TIME	connection time to next segment
SEG_ONLINE_CONNECT	next segment on same airline (y/n)
NOT_USED_5	

C.2.2 Comparison with EUROSTAT Figures

Eurostat provides detailed figures on the number of passengers per airport and per route. The regulation foresees to collect monthly detailed data for airports handling more than 150 000 passengers per year, for airports with less than 150 000 but more than 15 000 passengers only aggregated annual data are requested, while for minor airports there is no data provision obligation (Eurostat 2005).

Taking the above into account, the Eurostat databases are consulted for passenger figures on the 70 selected OD-pairs. The database contains route information per country per airport.

From the route information, arriving and departing passengers are retrieved for 2005, as this is the most recent year for which information is available. If possible, a comparison was made if the figures were consistent: over a year, the number of arriving passengers on airport X from airport Y should be equal to the departing from airport Y to airport X. In most cases, this check holds, an exception being the route Paris Charles de Gaulle – Hamburg.

On 30 of the 70 OD-pairs, Eurostat has no route information or figures available. A closer look to the definitions used by Eurostat, learns us that the passengers counted are per flight(number) (Bierlaire 2005), in our case this means by segment; thus no transfer passengers are included.

In Table_APX C-8 the passenger figures from Eurostat and the MIDT database are compared. The passenger figures for Eurostat are computed by taking the average number of passengers per segment and divide the result by 12 months.

Several columns are added to the passenger figures:

- Share MIDT, which is calculated by dividing the MIDT figure by the Eurostat figure.
- Share MIDT 5% No-show: Eurostat only records passengers that actually were on the aircraft, MIDT counts bookings. Therefore, a 5% no-show is taken into account in the calculation of the share.
- Share MIDT 10% No-show: here, a 10% no-show is taken into account in the calculation of the share.

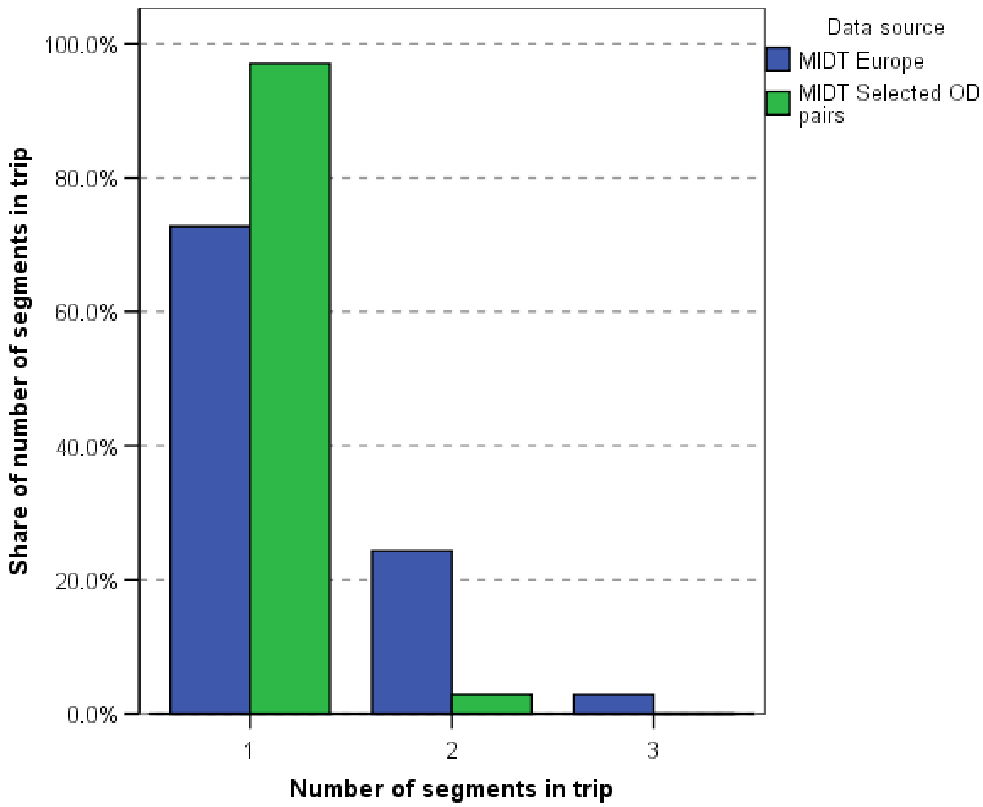
Table_APX C-8 Market share MIDT bookings (compared to Eurostat figures)

City-pair	*Average passengers per month 2005*	*MIDT November 2006*	*Share MIDT*	*Share MIDT No-show 5%*	*Share MIDT No-show 10%*
STR - ZRH	8636	10841	125.50%	132.10%	139.50%
STR - CPH	8839	9341	105.70%	111.20%	117.40%
HAM - CPH	9462	9435	99.70%	105.00%	110.80%
HAM - DUS	30679	29930	97.60%	102.70%	108.40%
ZRH - WAW	10253	9812	95.70%	100.70%	106.30%
HAM - BRU	9269	8233	88.80%	93.50%	98.70%
HAM - AMS	20057	17572	87.60%	92.20%	97.30%
HAM - CDG	26713	21085	78.90%	83.10%	87.70%
STR - MXP	9345	7311	78.20%	82.40%	86.90%
STR - CDG	23994	18430	76.80%	80.90%	85.30%
PAD - MUC	10624	8121	76.40%	80.50%	84.90%
ZRH - CDG	52509	39848	75.90%	79.90%	84.30%
ZRH - ARN	15649	11569	73.90%	77.80%	82.10%
STR - DUS	19455	13480	69.30%	72.90%	77.00%
ZRH - AMS	45750	29112	63.60%	67.00%	70.70%
FRA - IST	55487	34975	63.00%	66.40%	70.00%
ZRH - CPH	31053	19102	61.50%	64.80%	68.30%
ZRH - VIE	49995	29061	58.10%	61.20%	64.60%
HAM - HEL	9956	5346	53.70%	56.50%	59.70%
RNS - ORY	3967	2126	53.60%	56.40%	59.50%
STR - LHR	20746	11103	53.50%	56.30%	59.50%
STR - TXL	64468	34162	53.00%	55.80%	58.90%
AMS - TLS	13253	6994	52.80%	55.60%	58.60%
HAM - BUD	10172	5255	51.70%	54.40%	57.40%
STR - VIE	21647	11197	51.70%	54.40%	57.50%
HAM - ZRH	34817	17374	49.90%	52.50%	55.40%
HAM - VIE	28612	14196	49.60%	52.20%	55.10%
HAM - STR	60439	28800	47.70%	50.20%	52.90%
ZRH - LHR	75125	33208	44.20%	46.50%	49.10%
HAM - MAN	6845	2875	42.00%	44.20%	46.70%
STR - BCN	11734	4590	39.10%	41.20%	43.50%
ATH - BRU	23556	8069	34.30%	36.10%	38.10%
STR - BUD	8412	2686	31.90%	33.60%	35.50%
HAM - BCN	10589	3338	31.50%	33.20%	35.00%
STR - MAD	6678	1497	22.40%	23.60%	24.90%
EDI - BRS	27489	3160	11.50%	12.10%	12.80%
OSL - AGP	13844	1343	9.70%	10.20%	10.80%
DTM - PMI	23412	225	1.00%	1.00%	1.10%
BCN - BFS	5389		0.00%	0.00%	0.00%

C.2.3 Number of transfers

In Figure_APX C-2 the number of segments per trip is compared. It can be seen that the number of segments on the selected OD pairs is much lower than the number of segments in Europe per trip. In both cases, two segments per trip is already uncommon, three being very unusual.

Figure_APX C-2 Share of number of segments in trip



C.3 OAG

From the Official Airline Guide (OAG) it is possible to extract information on departure time, arrival time, aircraft type and code share. In the OAG, flights are stored by their original flight number and operator. Therefore it is necessary to first add the original flight number to the flight alternatives extracted from the Expedia and OAG dataset. Once this is done, it is possible to add the departure times and arrival times of a flight alternative and subsequently calculate the waiting time, in vehicle time and total travel time.
The fields used from the OAG are listed in Table_APX C-9.

Table_APX C-9 Used fields OAG

Fields	*Explanation*
Carrier1	Carrier abbreviation
Carrier1Name	Carrier name
FlightNo1	Flight number
DepAirport	Departure airport code
ArrAirport	Arrival airport code
LocalDepTime	Local departure time
LocalArrTime	Local arrival time
GeneralAcft	General aircraft code
GeneralAcftName	General aircraft name
SpecificAcft	Specific aircraft code
SpecificAcftName	Specific aircraft name
DupCar1	Code share carrier code
DupCar2	Code share carrier code
...	
DupCar8	Code share carrier code
DupFlightNo1	Code share flight number
DupFlightNo2	Code share flight number
...	
DupFlightNo8	Code share flight number

By matching DupCar1, DupCar2 … DupCar8 and DupFlightNo1, DupFlightNo2 … DupFlightNo8 on the carrier abbreviations and flight numbers in the flight alternatives dataset, it is possible to assign the original flight number to the code share flights.

C.4 Other Data Sources

Departure and arrival times were mostly stored in local times. For calculation purposes, these were converted to UMT times. The time differences per airport are stored in the table *airports_europe*. This table indicates in which country an airport is located.
A table with carrier name and their abbreviation was created, because the Expedia dataset does not contain the abbreviations. This table is called *carriers.* Each carrier listed in Expedia is added to a

category. The categories are regional carrier, flag carrier and low cost carrier. This category is based on the marketing of the carrier and common sense.

Finally, a table containing aircraft types and their category was created. This table is called *aircraft_types*. A distinction is made between three type of aircraft: propeller aircraft, regional jet and mainline jet. The categories are based on the listing of the producer's website.

C.5 Matching Expedia and MIDT Datasets

With the Expedia dataset, it is possible to add fares to the records of the MIDT dataset, the bookings. A fare is added to a booking when the following criteria are met:

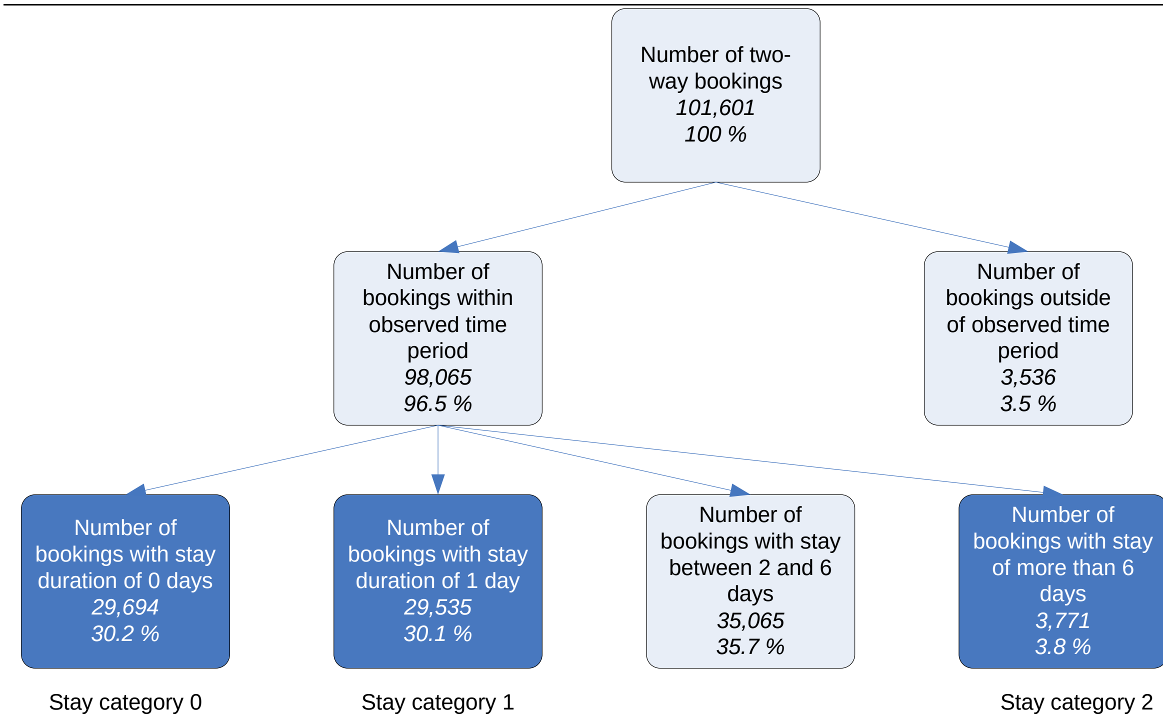
- Same booking date as Expedia query date;
- Same departure date;
- Same flight alternative (combination of inbound and outbound flight);
- Same duration of stay.

The following approach is followed:

- Add id for booking date;
- Add category variable for duration of stay:
    - 0 days = category '0';
    - 1 day = category '1';
    - > 6 days = category '2';
- Match the datasets on booking date, departure date, flight alternative and stay category.

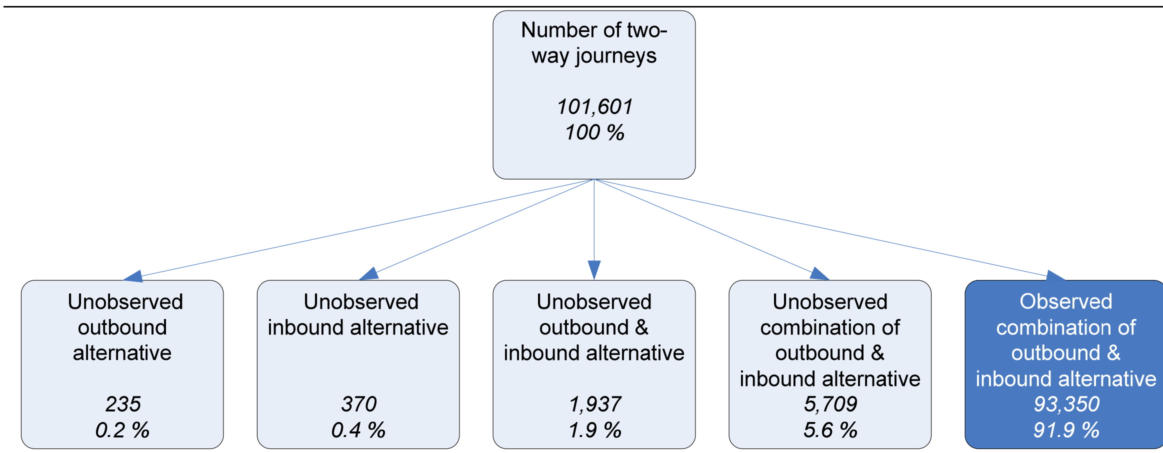
Before making a complete match, an analysis is made to determine which criteria lead to potential record loss. In Figure_APX C-3 the number of bookings within the time period observed by the webbots can be seen. 96.5 % of the bookings are within the observed time period, 60.3% are in stay category 0 or 1.

Figure_APX C-3 Filtering on booking period



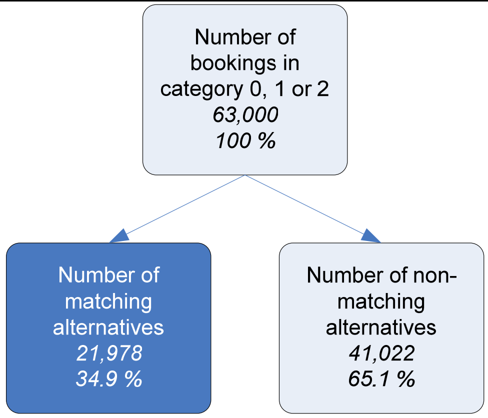
In Figure_APX C-4 the number of records is shown when a match is made on flight alternative. Most records are lost because the combination of outbound and inbound alternative does not occur in the Expedia dataset. This is for instance the case when a carrier is not listed on Expedia, such as Germanwings.

Figure_APX C-4 Matching on flight alternative



Finally, a match is made on all four criteria. The result is visualized in Figure_APX C-5. It can be seen that the match on booking date and departure date leads to a loss of 41,022 records or 65.1% if compared to the number of records in one of the three stay categories. This loss is contributed to the fact that Expedia does consequently list the same alternatives on their website on a daily basis.

Figure_APX C-5 Matching on all four criteria



Table_APX C-10 shows the number of bookings per origin-destination pair per matching step. In the third column, the total number of bookings on the origin-destination pair is shown as observed in the MIDT dataset. The fourth column shows the number of bookings within each stay category. The fifth column shows the number of bookings with fare. The last columns show the amount of bookings relative to the total number of remaining bookings during each step. It can be seen that on origin-destination pairs with a high number of bookings, the chance of matching a booking is higher.

Table_APX C-10 Number of bookings with fare on OD-pairs

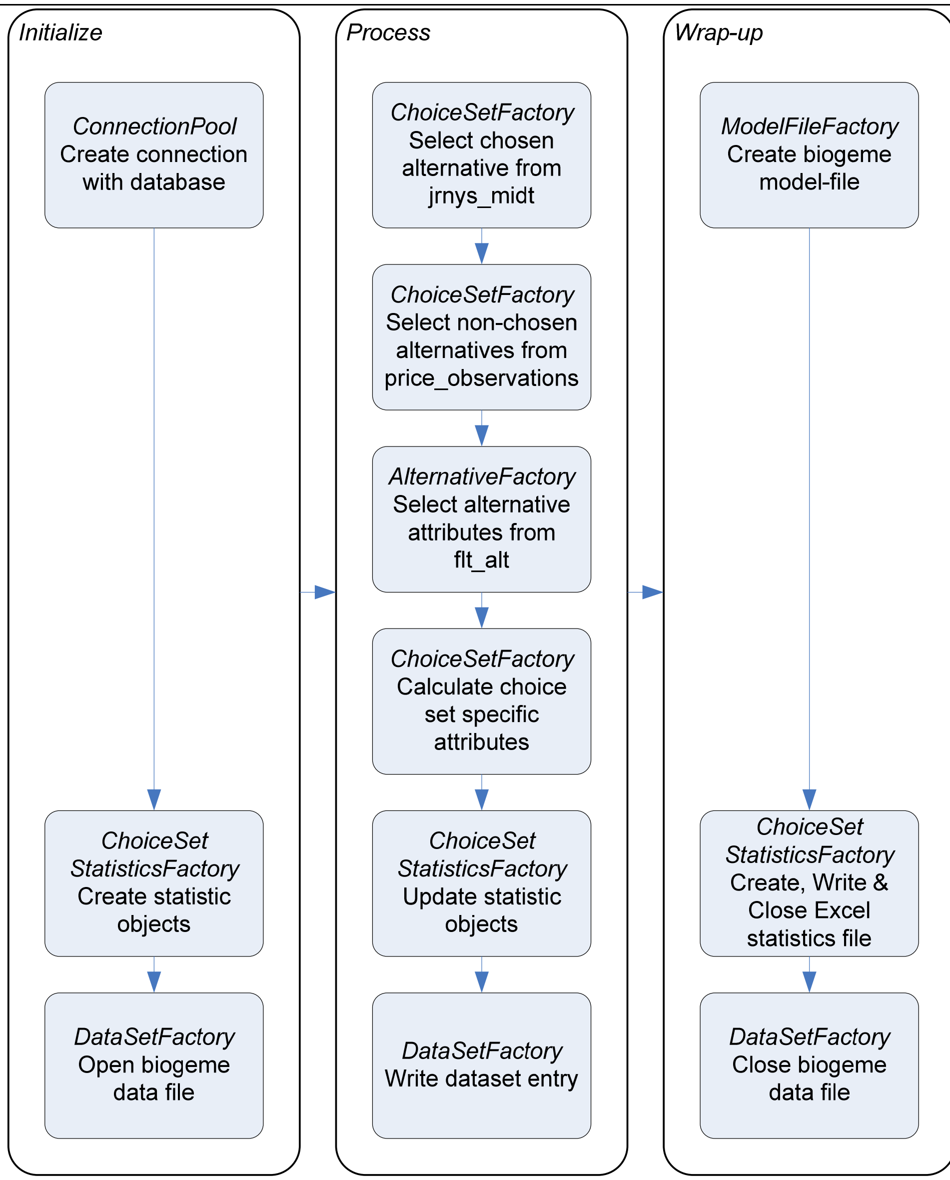
Origin airport	Destination airport	Absolute			Relative to total remaining bookings		
		all duration of stay	within stay category	with price	all duration of stay	within stay category	with price
AMS	TLS	310	125	64	0.57%	0.37%	0.34%
ARN	SCN	1	1	0	0.00%	0.00%	0.00%
ATH	BRU	1011	333	5	1.86%	0.98%	0.03%
BHX	TLL	10	0	0	0.02%	0.00%	0.00%
DTM	PMI	3	1	1	0.01%	0.00%	0.01%
EDI	BRS	457	343	19	0.84%	1.01%	0.10%
EDI	LCG	3	1	0	0.01%	0.00%	0.00%
FRA	IST	2284	922	534	4.20%	2.71%	2.82%
GDN	INN	1	0	0	0.00%	0.00%	0.00%
GLA	RTM	3	0	0	0.01%	0.00%	0.00%
HAM	AMS	1157	825	354	2.13%	2.43%	1.87%
HAM	ARN	498	280	129	0.92%	0.82%	0.68%
HAM	BCN	498	172	59	0.92%	0.51%	0.31%
HAM	BRU	1232	898	376	2.27%	2.64%	1.98%
HAM	BUD	669	281	186	1.23%	0.83%	0.98%
HAM	CDG	2119	1189	771	3.90%	3.50%	4.07%
HAM	CPH	442	332	187	0.81%	0.98%	0.99%
HAM	DUS	4412	3714	2020	8.12%	10.92%	10.65%
HAM	HEL	414	164	97	0.76%	0.48%	0.51%
HAM	MAN	449	272	207	0.83%	0.80%	1.09%
HAM	MXP	636	381	146	1.17%	1.12%	0.77%
HAM	NCE	202	68	31	0.37%	0.20%	0.16%
HAM	OSL	281	177	108	0.52%	0.52%	0.57%
HAM	STR	4390	3304	2223	8.08%	9.71%	11.72%
HAM	VIE	1584	940	530	2.91%	2.76%	2.80%
HAM	ZRH	1815	1116	554	3.34%	3.28%	2.92%
HEL	FDH	7	4	3	0.01%	0.01%	0.02%
MAN	HEL	311	71	38	0.57%	0.21%	0.20%
MSQ	SOF	3	0	0	0.01%	0.00%	0.00%
NRK	AAL	3	2	1	0.01%	0.01%	0.01%
NRK	BSL	10	3	0	0.02%	0.01%	0.00%
NRK	STR	6	2	0	0.01%	0.01%	0.00%
OSL	AGP	106	46	23	0.20%	0.14%	0.12%
PAD	MUC	1209	912	673	2.22%	2.68%	3.55%
PRG	AOI	6	2	0	0.01%	0.01%	0.00%
PRG	JKG	4	0	0	0.01%	0.00%	0.00%
RNS	ORY	45	31	24	0.08%	0.09%	0.13%
SJJ	ALC	1	0	0	0.00%	0.00%	0.00%
SOU	CPH	4	2	0	0.01%	0.01%	0.00%

STR	BCN	789	420	173	1.45%	1.23%	0.91%
STR	BRU	794	603	297	1.46%	1.77%	1.57%
STR	BUD	694	356	290	1.28%	1.05%	1.53%
STR	CDG	1744	1135	746	3.21%	3.34%	3.93%
STR	CPH	574	313	161	1.06%	0.92%	0.85%
STR	DUS	1597	1360	956	2.94%	4.00%	5.04%
STR	GRZ	392	278	206	0.72%	0.82%	1.09%
STR	LHR	1121	648	508	2.06%	1.90%	2.68%
STR	MAD	343	174	17	0.63%	0.51%	0.09%
STR	MXP	658	447	316	1.21%	1.31%	1.67%
STR	PRG	455	239	208	0.84%	0.70%	1.10%
STR	TXL	5881	4140	3127	10.82%	12.17%	16.49%
STR	VIE	1283	759	185	2.36%	2.23%	0.98%
STR	ZRH	78	56	0	0.14%	0.16%	0.00%
ZRH	AMS	2076	1135	634	3.82%	3.34%	3.34%
ZRH	ARN	626	303	104	1.15%	0.89%	0.55%
ZRH	CDG	2852	1679	465	5.25%	4.94%	2.45%
ZRH	CPH	889	508	153	1.64%	1.49%	0.81%
ZRH	LHR	2128	1088	432	3.92%	3.20%	2.28%
ZRH	VIE	2352	1254	530	4.33%	3.69%	2.80%
ZRH	WAW	419	208	90	0.77%	0.61%	0.47%
Totals		54341	34017	18961	100%	100%	100%

C.6 Choice Set Generation

Two bachelor students (Hüni & Merz 2007) conducted an investigation towards the possibility of the imputation of new records in an earlier stage of this research. The goal of the imputation was to increase the price observations in the Expedia dataset, with the objective to match more records with the MIDT dataset. For this reason, a flexible approach was followed towards the generation of choice sets, which is visualized in Figure_APX C-6. With the followed approach, the id of a booking with price is selected from a table containing the bookings with a price observation. From a second table, containing all bookings, the booking characteristics (OD-pair, booking date, departure date, flight alternative, stay category) are selected and stored in an object, the *BookingCharacteristicsDAO*. With the information stored in this object other, non-chosen, alternatives are selected. In this stage, it is possible to add criteria to the selection of non-chosen alternative such as booking time window, departure time window etc. This can be done per chosen alternative. Each alternative has a *BookingCharacteristicsDAO.* With the information stored in the *BookingCharacteristicsDAO,* alternatives are created (*AlternativeDAO*), which consist of a number of attributes (*AttributeDAO*). Alternative flight characteristics are selected from the tables expedia_flt_alt_out_weekday, expedia_flt_alt_in_weekday and price_observations. The alternatives are then added to a choice set (*ChoiceSetDAO*). Each choice set is then handed over the *ChoiceSetStatisticsFactory*, where a distinction is made between chosen alternative statistics and non chosen alternative statistics. Two types of statistic objects are used, which are provided by the JAVA commons math project (http://commons.apache.org/math/). The objects are the frequency object and the descriptive statistics object. The first makes it possible to record the frequency and extract statistics such as frequency, percentage, cumulative frequency and cumulative percentage. The latter makes it possible to extract statistics such as mean, median, minimum, maximum and skewness. Each choice set is written to a data file, according to the format necessary for Biogeme.

Figure_APX C-6 Choice set generation program flow



C.7 Biogeme Data and Model File

Examples of Biogeme data files and their format can be found on the Biogeme website and in the tutorial (Bierlaire 2005). The fields listed in Table_APX C-11 are in the Biogeme data files used.

Table_APX C-11 Fields Biogeme Data File

Fields	*Explanation*
A_A3	Carrier abbreviation
A_AA	Carrier abbreviation
....	
A_ZB	Carrier abbreviation
codeShareInItineraryOut	1 if the outbound itinerary contains a code share
dayOfWeek	1 on Sunday, 2 on Monday, 7 on Saturday
daysBeforeDeparture	Number of days before departure
depHourIn	Hour of the departure time inbound itinerary
depHourInCosI	Used for continuous departure time modeling
depHourInCosII	Used for continuous departure time modeling
depHourInCosIII	Used for continuous departure time modeling
depHourInSinI	Used for continuous departure time modeling
depHourInSinII	Used for continuous departure time modeling
depHourInSinIII	Used for continuous departure time modeling
depHourOut	Hour of the departure time outbound itinerary
depHourOutCosI	Used for continuous departure time modeling
depHourOutCosII	Used for continuous departure time modeling
depHourOutCosIII	Used for continuous departure time modeling
depHourOutSinI	Used for continuous departure time modeling
depHourOutSinII	Used for continuous departure time modeling
depHourOutSinIII	Used for continuous departure time modeling
dummyPropellorAircraftOut	1 if the outbound itinerary contains a propellor aircraft
dummyRegionalAircraftOut	1 if the outbound itinerary contains a regional aircraft
durationOfStay	Duration of stay in minutes
Fare	The fare of the itinerary in euro's.
Frequency	The number of flights offered by the carrier on the OD-pair
homeCarrier	1 if the carrier is the domestic carrier
inVehicleTimeOut	In vehicle time in minutes
lowCostCarrierInItineraryOut	1 if the itinerary contains a low cost carrier
numTransfersOut	Number of transfers in the outbound itinerary
odId	OD-pair ID
outboundCarrierIsInboundCarrier	1 if the carrier outbound carrier equals the inbound carrier
perceivedJourneyTime	Perceived journey time
regionalCarrierInItineraryOut	1 if the itinerary contains a regional carrier
stayCategory	Stay category of the flight
Tax	Tax of the itinerary
totalTravelTime	Total travel time of the outbound itinerary
waitingTimeOut	Waiting time in the outbound itinerary
Av1 …. Av150	1 Indicates if the alternative is available
Choice	Number indicates which alternative is chosen

As can be seen, not all variables are included which are shown in the model estimation results. This is because Biogeme offers the possibility to calculate variables. The calculation should be specified in the model file. For instance, the fare per booking period can be computed by:

$LOOP(XXX 1 150 1) bookingPeriodFare1_XXX = (daysBeforeDepartureXXX >= 4 && daysBeforeDepartureXXX < 8) * fareXXX

And the departure per stay category can be computed by:

$LOOP(XXX 1 150 1) StayCat0_Hour14_XXX = (stayCategoryXXX == 0) * (depHourOutXXX == 14)

The same can be done for departure hour, fare per stay category, etc. The computed values should be specified in the Utilities or GeneralizedUtilities section of the model file.

C.8 Correlation tables

Table_APX C-12 Correlation stay category 0 - departure hour - duration of stay

	Pearson Correlation	Sig. (2-tailed)
duration_of_stay	1.00	
dep_hour_out6	0.39	0.00
dep_hour_out7	0.34	0.00
dep_hour_out8	-0.23	0.00
dep_hour_out9	-0.17	0.00
dep_hour_out10	-0.32	0.00
dep_hour_out11	-0.30	0.00
dep_hour_out12	-0.32	0.00
dep_hour_out13	-0.09	0.00
dep_hour_out14	-0.12	0.00
dep_hour_out15	-0.02	0.02
dep_hour_in13	0.01	0.51
dep_hour_in14	-0.09	0.00
dep_hour_in15	-0.06	0.00
dep_hour_in16	-0.36	0.00
dep_hour_in17	-0.32	0.00
dep_hour_in18	-0.01	0.22
dep_hour_in19	0.22	0.00
dep_hour_in20	0.36	0.00
dep_hour_in21	0.17	0.00

Table_APX C-13 Correlation stay category 1 - departure hour - duration of stay

	Pearson Correlation	Sig. (2-tailed)
duration_of_stay	1.00	
dep_hour_out6	0.46	0.00
dep_hour_out7	0.44	0.00
dep_hour_out8	0.19	0.00
dep_hour_out9	0.06	0.00
dep_hour_out10	0.05	0.00
dep_hour_out11	0.00	0.92
dep_hour_out12	-0.04	0.00
dep_hour_out13	-0.06	0.00
dep_hour_out14	-0.12	0.00
dep_hour_out15	-0.12	0.00
dep_hour_out16	-0.23	0.00
dep_hour_out17	-0.37	0.00
dep_hour_out18	-0.30	0.00
dep_hour_out19	-0.28	0.00
dep_hour_out20	-0.24	0.00
dep_hour_out21	-0.24	0.00
dep_hour_out22	-0.08	0.00
dep_hour_in6	-0.03	0.01
dep_hour_in7	-0.12	0.00
dep_hour_in8	-0.08	0.00
dep_hour_in9	-0.07	0.00
dep_hour_in10	-0.04	0.00
dep_hour_in11	-0.02	0.08
dep_hour_in12	-0.08	0.00
dep_hour_in13	-0.03	0.02
dep_hour_in14	-0.10	0.00
dep_hour_in15	-0.09	0.00
dep_hour_in16	-0.14	0.00
dep_hour_in17	-0.09	0.00
dep_hour_in18	0.13	0.00
dep_hour_in19	0.03	0.00
dep_hour_in20	0.14	0.00
dep_hour_in21	0.06	0.00

Table_APX C-14 Correlation type of carrier - departure hour

	rc_in_itinerary		fc_in_itinerary		lcc_in_itinerary	
	Pearson Correlation	Sig. (2-tailed)	Pearson Correlation	Sig. (2-tailed)	Pearson Correlation	Sig. (2-tailed)
dep_hour_out6	0.10	0.00	-0.11	0.00	0.04	0.00
dep_hour_out7	-0.12	0.00	0.18	0.00	-0.12	0.00
dep_hour_out8	0.16	0.00	-0.21	0.00	0.13	0.00
dep_hour_out9	-0.09	0.00	0.10	0.00	-0.03	0.00
dep_hour_out10	0.05	0.00	-0.01	0.26	-0.06	0.00
dep_hour_out11	0.03	0.00	0.00	0.78	-0.04	0.00
dep_hour_out12	-0.07	0.00	0.09	0.00	-0.04	0.00
dep_hour_out13	-0.03	0.00	0.04	0.00	-0.03	0.00
dep_hour_out14	0.04	0.00	-0.02	0.00	-0.03	0.00
dep_hour_out15	-0.03	0.00	-0.06	0.00	0.14	0.00
dep_hour_out16	-0.04	0.00	0.04	0.00	-0.02	0.02
dep_hour_out17	-0.03	0.00	0.02	0.00	0.01	0.12
dep_hour_out18	-0.02	0.02	0.04	0.00	-0.04	0.00
dep_hour_out19	-0.04	0.00	0.01	0.31	0.05	0.00
dep_hour_out20	-0.01	0.39	0.02	0.01	-0.02	0.00
dep_hour_out21	-0.03	0.00	-0.16	0.00	0.32	0.00
dep_hour_out22	-0.02	0.00	0.03	0.00	-0.01	0.03

Appendix D Notes to Chapter 6

D.1 Departure Time Models

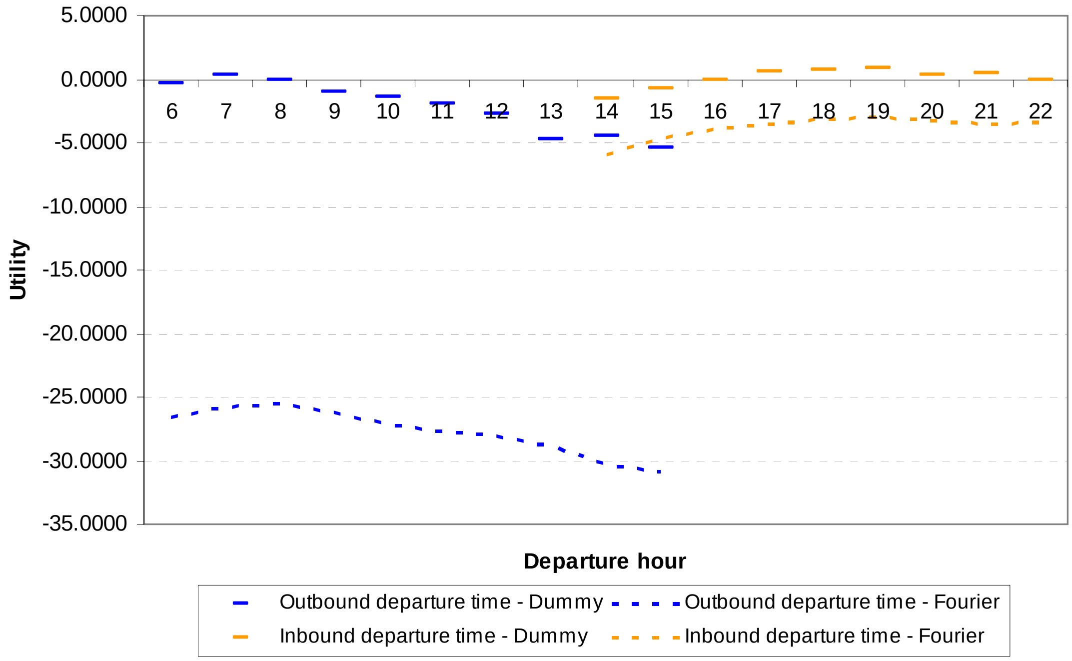
An alternate way to enter the departure time in the utility function will be discussed as the discretization of departure time might give strange changes in choice probabilities. Koppelman *et al.* *(Koppelman, et al. 2007)* propose an approach which is adopted from Zeid *et al.* (2006) to overcome this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

$$\begin{aligned} U(t) = \beta_{\sin 2}\sin\left(\frac{2\pi t}{1440}\right) + \beta_{\sin 4}\sin\left(\frac{4\pi t}{1440}\right) + \beta_{\sin 6}\sin\left(\frac{6\pi t}{1440}\right) \\ +\beta_{\cos 2}\cos\left(\frac{2\pi t}{1440}\right) + \beta_{\cos 4}\cos\left(\frac{4\pi t}{1440}\right) + \beta_{\cos 6}\cos\left(\frac{6\pi t}{1440}\right) \end{aligned} \quad (1.12)$$

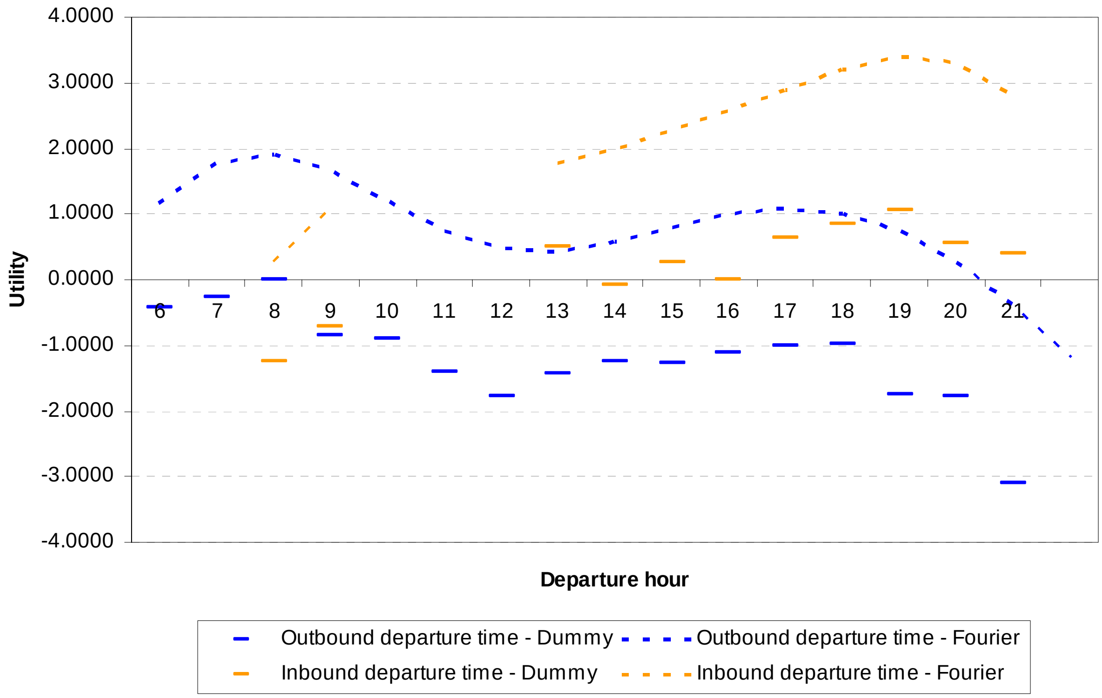
Where t is the departure time in minutes and 1440 the number of minutes per day. As can be seen, this a Fourier series approach. Gramming *et al.* (2005) model departure time preference with a similar approach, namely $\sum_{q=1}^{3}\gamma_q \sin\left(\frac{2\pi q}{T}t_j + \phi_q\right)$, and estimate the parameters γ_1, γ_2, γ_3, ϕ_1, ϕ_2, ϕ_3. As can be seen, this is equal to $\sum_{q=1}^{3}\gamma_q \sin\left(\frac{2\pi q}{T}t_j\right)\cos\phi_q + \sum_{q=1}^{3}\gamma_q \cos\left(\frac{2\pi q}{T}t_j\right)\sin\phi_q$ and therefore equal to formula (1.12). Formula (1.12) is more convenient to implement in a software package such as BIOGEME. Both approaches should yield the same results and require six parameters to be estimated. For the models presented this results in 37 parameters less to be estimated. It may be important to notice that this concerns a dimensionless attribute level. Model performance and parameter estimates are presented in Table_APX D-1 and Table_APX D-2 respectively.

Figure_APX D-1, Figure_APX D-2 and Figure_APX D-3 allow for a visual comparison between the estimates of the model containing a dummy variable per departure hour and the parameter estimates of the Fourier series. The same preference structure can be observed, one notable exception being the estimates for the itineraries returning the following day between 13:00 and 14:00. Here, the Fourier series results in a linear interpolation, between 13:00 and 9:00, as opposed to the dummy variable for this time period. Therefore, it is recommended to use the Fourier series only if a continuous series is observed.

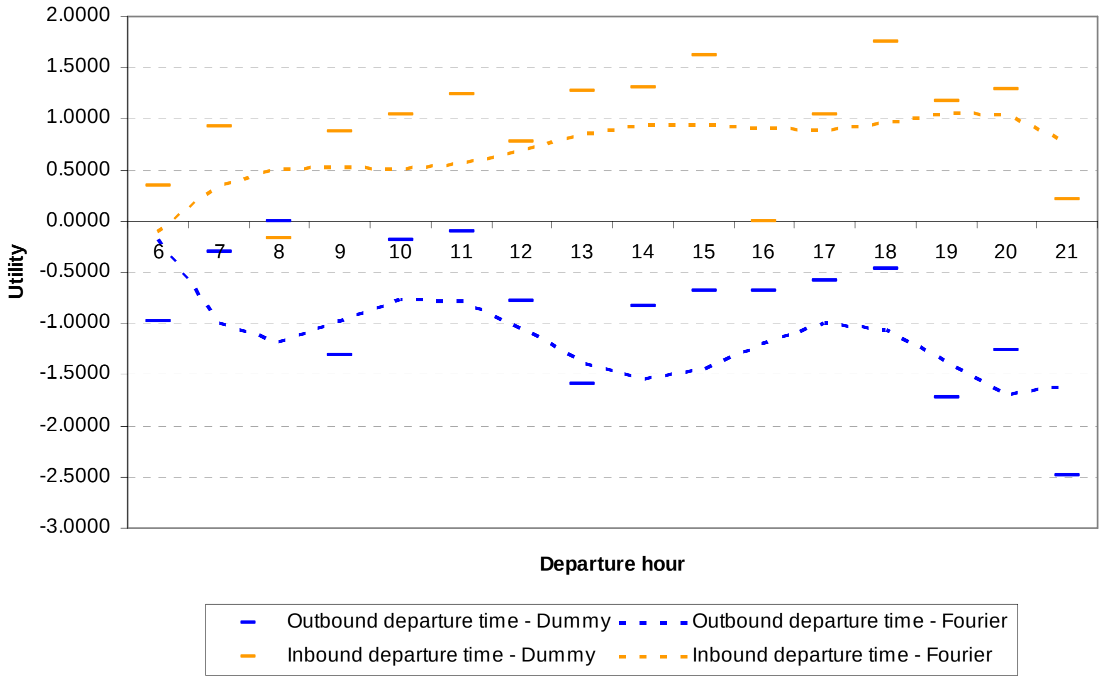
Figure_APX D-1 Estimated parameters for stay category 0



Figure_APX D-2 Estimated parameters for stay category 1



Figure_APX D-3 Estimated parameters for stay category 2



Table_APX D-1 Model performance dummy variables and Fourier series

	Model with dummy variables	Model with Fourier serier
Number of estimated parameters	98	61
Number of observations	18416	18416
Init log-likelihood	-69032.6	-68956.1
Final log-likelihood	-46113.1	-46109.7
Likelihood ratio test	45838.9	45692.9
Rho-square	0.332	0.331
Adjusted rho-square	0.331	0.330

Table_APX D-2 Parameter estimate fare model and Fourier model

	Fare Estimated parameter	Fare Robust t-test	Fare Fourier Estimated parameter	Fare Fourier Robust t-test
Carrier constants				
Not presented				
Flight attributes				
Non-code share	0	-	0	-
Code share	-0.9247	-12.84	-0.9722	-13.68
Total travel time out	-0.0114	-5.47	-0.0005	-6.40
Number of transfers	-4.6923	-12.42	-6.2267	-33.70
Aircraft attribute				
Mainline jet	0	-	0	-
Regional aircraft	-0.1328	-4.68	-0.1720	-6.19
Propellor aircraft	-1.5388	-14.79	-1.4533	-13.73
Departure times stay category 0				
Outbound				
6:00 - 6:59	-0.3136	-9.14		
7:00 - 7:59	0.2992	7.05		
8:00 - 8:59	0.0000	0.00		
9:00 - 9:59	-0.9985	-16.36		
10:00 - 10:59	-1.4108	-30.12		
11:00 - 11:59	-1.9384	-33.38		
12:00 - 12:59	-2.6684	-36.62		
13:00 - 13:59	-4.7459	-16.96		
14:00 - 14:59	-4.4016	-20.40		

Carrier constants		Fare Estimated parameter	Fare Robust t-test	Fare Fourier Estimated parameter	Fare Fourier Robust t-test
	15:00 - 15:59	-5.4656	-5.38		
	16:00 - 16:59	0.0000	0.00		
	17:00 - 17:59	0.0000	0.00		
	18:00 - 18:59	0.0000	0.00		
	19:00 - 19:59	0.0000	0.00		
	20:00 - 20:59	0.0000	0.00		
	21:00 - 21:59	0.0000	0.00		
	22:00 - 22:59	0.0000	0.00		
Inbound					
	6:00 - 6:59	0.0000	0.00		
	7:00 - 7:59	0.0000	0.00		
	8:00 - 8:59	0.0000	0.00		
	9:00 - 9:59	0.0000	0.00		
	10:00 - 10:59	0.0000	0.00		
	11:00 - 11:59	0.0000	0.00		
	12:00 - 12:59	0.0000	0.00		
	13:00 - 13:59	0.0000	0.00		
	14:00 - 14:59	-1.5407	-4.85		
	15:00 - 15:59	-0.7206	-2.91		
	16:00 - 16:59	0.0000	0.00		
	17:00 - 17:59	0.5819	14.14		
	18:00 - 18:59	0.7976	19.14		
	19:00 - 19:59	0.8835	22.27		
	20:00 - 20:59	0.3501	8.02		
	21:00 - 21:59	0.4767	3.46		
	22:00 - 22:59	0.0000	0.00		
Departure times stay category 1					
Outbound					
	6:00 - 6:59	-0.4136	-6.60		
	7:00 - 7:59	-0.2658	-3.90		
	8:00 - 8:59	0.0000	0.00		
	9:00 - 9:59	-0.8599	-10.09		
	10:00 - 10:59	-0.9051	-12.93		
	11:00 - 11:59	-1.4101	-16.93		
	12:00 - 12:59	-1.7703	-20.45		
	13:00 - 13:59	-1.4399	-16.83		
	14:00 - 14:59	-1.2574	-16.15		
	15:00 - 15:59	-1.2753	-12.87		
	16:00 - 16:59	-1.1040	-15.21		
	17:00 - 17:59	-1.0016	-15.20		
	18:00 - 18:59	-0.9707	-13.25		
	19:00 - 19:59	-1.7444	-20.00		
	20:00 - 20:59	-1.7759	-17.05		
	21:00 - 21:59	-3.0921	-11.95		

Carrier constants		Fare Estimated parameter	Fare Robust t-test	Fare Fourier Estimated parameter	Fare Fourier Robust t-test
	22:00 - 22:59	12.2296	4.06		
Inbound					
	6:00 - 6:59	0.0000	0.00		
	7:00 - 7:59	0.0000	0.00		
	8:00 - 8:59	-1.2439	-5.45		
	9:00 - 9:59	-0.7251	-2.25		
	10:00 - 10:59	0.0000	0.00		
	11:00 - 11:59	0.0000	0.00		
	12:00 - 12:59	0.0000	0.00		
	13:00 - 13:59	0.5014	1.71		
	14:00 - 14:59	-0.0744	-0.39		
	15:00 - 15:59	0.2567	1.21		
	16:00 - 16:59	0.0000	0.00		
	17:00 - 17:59	0.6366	11.88		
	18:00 - 18:59	0.8423	14.80		
	19:00 - 19:59	1.0495	20.79		
	20:00 - 20:59	0.5464	9.43		
	21:00 - 21:59	0.4057	2.89		
	22:00 - 22:59	0.0000	0.00		
Departure times stay category 2					
Outbound					
	6:00 - 6:59	-0.9810	-2.14		
	7:00 - 7:59	-0.3001	-0.71		
	8:00 - 8:59	0.0000	0.00		
	9:00 - 9:59	-1.3147	-2.84		
	10:00 - 10:59	-0.1935	-0.45		
	11:00 - 11:59	-0.1046	-0.25		
	12:00 - 12:59	-0.7849	-1.77		
	13:00 - 13:59	-1.5906	-3.53		
	14:00 - 14:59	-0.8375	-1.78		
	15:00 - 15:59	-0.6890	-1.27		
	16:00 - 16:59	-0.6803	-1.47		
	17:00 - 17:59	-0.5908	-1.40		
	18:00 - 18:59	-0.4697	-1.09		
	19:00 - 19:59	-1.7251	-3.17		
	20:00 - 20:59	-1.2552	-2.47		
	21:00 - 21:59	-2.4812	-2.33		
	22:00 - 22:59	15.0059	4.89		
Inbound					
	6:00 - 6:59	0.3493	0.96		
	7:00 - 7:59	0.9234	3.59		
	8:00 - 8:59	-0.1608	-0.45		
	9:00 - 9:59	0.8688	3.12		
	10:00 - 10:59	1.0404	4.08		

Carrier constants		Fare Estimated parameter	Fare Robust t-test	Fare Fourier Estimated parameter	Fare Fourier Robust t-test
	11:00 - 11:59	1.2377	3.87		
	12:00 - 12:59	0.7829	3.47		
	13:00 - 13:59	1.2703	4.60		
	14:00 - 14:59	1.2978	4.73		
	15:00 - 15:59	1.6210	6.23		
	16:00 - 16:59	0.0000	0.00		
	17:00 - 17:59	1.0398	4.87		
	18:00 - 18:59	1.7544	6.80		
	19:00 - 19:59	1.1785	4.36		
	20:00 - 20:59	1.2961	4.34		
	21:00 - 21:59	0.2066	0.25		
	22:00 - 22:59	0.0000	0.00		
Stay category 0 - Outbound					
	Cos 1			41.0632	5.29
	Cos 2			13.6912	4.42
	Cos 3			0.6718	1.36
	Sin 1			-19.6297	-6.71
	Sin 2			-20.9360	-7.61
	Sin 3			-6.7286	-7.75
Stay category 1 - Outbound					
	Cos 1			-1.6512	-6.40
	Cos 2			-1.0834	-8.22
	Cos 3			0.0909	1.91
	Sin 1			-0.1002	-0.92
	Sin 2			-0.6311	-4.97
	Sin 3			-0.1836	-2.60
Stay category 2 - Outbound					
	Cos 1			1.6967	3.15
	Cos 2			0.6295	2.51
	Cos 3			-0.0183	-0.15
	Sin 1			1.2979	4.09
	Sin 2			1.2906	3.51
	Sin 3			0.8590	4.45
Stay category 0 - Inbound					
	Cos 1			8.7991	1.77
	Cos 2			-4.6747	-1.98
	Cos 3			-2.3216	-1.83
	Sin 1			6.8049	1.59
	Sin 2			7.3711	1.69
	Sin 3			-1.0393	-1.59
Stay category 1 - Inbound					
	Cos 1			-2.1254	-4.12
	Cos 2			-0.2559	-1.39
	Cos 3			0.2265	1.69
	Sin 1			-3.2568	-7.61

Carrier constants		Fare Estimated parameter	Fare Robust t-test	Fare Fourier Estimated parameter	Fare Fourier Robust t-test
	Sin 2			-1.9156	-4.59
	Sin 3			-0.3156	-2.18
Stay category 2 - Inbound					
	Cos 1			-1.0801	-2.64
	Cos 2			-0.4304	-1.73
	Cos 3			-0.0489	-0.43
	Sin 1			-0.8756	-5.20
	Sin 2			-0.6509	-3.31
	Sin 3			-0.3420	-2.84
Fare					
Fare		-0.0068	-75.79	-0.0069	-75.54

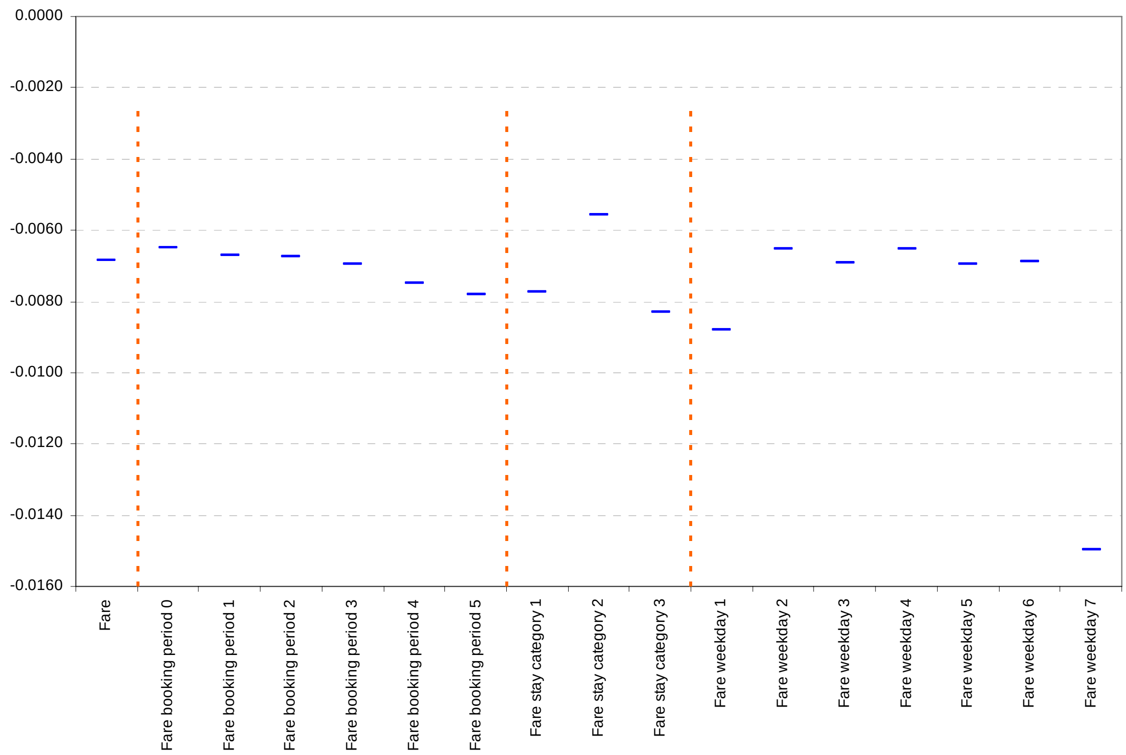
D.2 Fare Models

Several approaches can be recognized with regard to the specification of fare. On the one hand, fare can be a part of the choice set formation process and will lead to a reduction of the consideration set. It can be also possible that a passenger gathers information over several days and that these itineraries are in the consideration, despite not being available anymore. Again, this is a part of the choice set formation. On the other hand, it is possible to distinguish between passengers according to their revealed preferences. Passengers chose to return the same day or the next day, chose a certain departure day and chose to book their ticket in a certain period. These preferences are revealed at the moment they request fare information. The fare parameter is specified in the following ways:

- It is hypothesized that passengers staying longer at their destination will have a higher sensitivity to fare (and a lower sensitivity to departure time). Instead of estimating a single β_{fare}, the parameter for fare is replaced by $\sum_{k=1}^{K}\beta_{fare\_k}D_{stay-category\_k}Fare$, where $D_{stay-category\_k}$ indicates if the itinerary is for period n.
- The same approach is followed for booking period, where the parameter β_{fare} is replaced by $\sum_{m=1}^{M}\beta_{fare\_m}D_{booking-period\_m}Fare$ and $D_{booking-period\_m}$ indicates if the itinerary is booked in period m. A further specification of the fare parameter can be made by defining a fare parameter per weekday.
- The parameter β_{fare} is replaced by $\sum_{n=1}^{N}\beta_{fare\_n}D_{weekday\_n}Fare$ where $D_{weekday\_n}$ indicates if the itinerary is booked for weekday n.
- Finally, fare was treated as a log-transform, depicting the decreasing marginal returns of fare.

Figure_APX D-4 and Table_APX D-3 Summary estimated parameters and t-tests show the different estimated parameters for fare. All estimated parameters are highly significant; other estimated parameters remain approximately the same as compared to the base model with dummy variables for outbound and inbound departure time, except for the model treating fare as a log-transform.

Figure_APX D-4 Estimated parameters for fare



Six booking periods are identified based on the number of forecasts that are made prior to departure of an itinerary. Booking period 0 is up to 3 days before departure, booking period 1 indicates between 4 and 7 days before departure, booking period 2 indicates between 7 and 14 days before departure, booking period 3 indicates between 14 and 21 days before departure, booking period 4 indicates between 21 and 28 days before departure and booking period 5 indicates longer as 28 days before departure. The estimated parameters for fare follow the anticipated preference structure: travelers booking further in advance are more sensitive to fare as travelers booking close to departure. The difference between the estimated parameters for booking period 0 and 5 is approximately 20%.

For passengers returning the same day, a lower parameter is estimated than for passengers returning the next day. Passengers returning after six days are even more sensitive to fare. It is hypothesized, that passengers retuning the next day are less sensitive as fare only makes up a part of the total costs, which include an overnight stay. A second explanation could be that the fare differences for itineraries returning the same day are larger, as compared to itineraries returning the next day.

Table_APX D-3 Summary estimated parameters and t-tests

Fare	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Fare	-0.0068	-75.79						
Ln(Fare)	-2.3042	-83.04						
Fare booking period 0			-0.0065	-27.16				
Fare booking period 1			-0.0067	-43.25				
Fare booking period 2			-0.0067	-49.08				
Fare booking period 3			-0.0069	-31.74				
Fare booking period 4			-0.0075	-22.86				
Fare booking period 5			-0.0078	-26.21				
Fare stay category 1					-0.0077	-68.50		
Fare stay category 2					-0.0056	-41.96		
Fare stay category 3					-0.0083	-8.83		
Fare Sunday							-0.0088	-13.93
Fare Monday							-0.0065	-35.87
Fare Tuesday							-0.0069	-41.89
Fare Wednesday							-0.0065	-39.47
Fare Thursday							-0.0070	-38.09
Fare Friday							-0.0069	-29.25
Fare Saturday							-0.0150	-7.30

Model performance indicators are shown in Table_APX D-4. The models including either a fare variable per stay category or a fare variable per weekday slightly outperform the models with a fare variable per booking period. The model containing a log transformed fare is outperformed by all other models.

Table_APX D-4 Model performance of fare models

	Model with fare	Model with fare booking period	Model with fare per stay category	Model with fare per weekday	Model with ln(fare)
Number of estimated parameters	98	103	100	104	98
Number of observations	18416	18416	18416	18416	18416
Init log-likelihood	-69032.6	-69032.6	-69032.6	-69032.6	-69032.6
Final log-likelihood	-46101.7	-46091.1	-46021.7	-46065.1	-46815.2
Likelihood ratio test	45861.9	45883.1	46021.9	45935.1	44434.9
Rho-square	0.332	0.332	0.333	0.333	0.322
Adjusted rho-square	0.331	0.331	0.333	0.331	0.320

In this section several cross sections of the fare parameter in the choice set were presented, namely with regard to booking period, stay category and per weekday. With this specification, an attempt is made to segment customers based on their revealed preferences at the time of booking. A further specification of the fare parameter is very well imaginable, such as fare per booking period and stay category of fare per weekday and stay category. It should be kept in mind however, that this leads to a further segmentation of customers and perhaps separate models per stay category would capture traveler behavior better.

Table_APX D-5 Estimated parameters fare models

Models	Fare		Ln(Fare)		Fare per booking period		Fare per stay category		Fare per weekday	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Carrier constants										
Flight attributes										
Non-code share	0.0000	1.00	0.0000	1.00	0.0000	1.00	0.0000	1.00	0.0000	1.00
Code share	-0.9247	-12.84	-0.8851	-12.76	-0.9224	-12.83	-0.9215	-12.76	-0.9180	-12.75
Total travel time out	-0.0114	-5.47	0.0000	0.00	-0.0114	-5.51	-0.0116	-5.66	-0.0114	-5.51
Number of transfers	-4.6923	-12.42	-5.0535	-13.45	-4.6866	-12.41	-4.6511	-12.36	-4.6894	-12.36
Aircraft attribute										
Mainline jet	0.0000	1.00	0.0000	1.00	0.0000	1.00	0.0000	1.00	0.0000	1.00
Regional aircraft	-0.1328	-4.68	-0.1195	-4.28	-0.1315	-4.64	-0.1530	-5.37	-0.1312	-4.62
Propellor aircraft	-1.5388	-14.79	-1.5527	-14.39	-1.5426	-14.81	-1.5518	-14.89	-1.5394	-14.78
Departure times stay category 0										
Outbound										

Models	Fare		Ln(Fare)		Fare per booking period		Fare per stay category		Fare per weekday	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
6:00 - 6:59	-0.3136	-9.14	-0.3763	-10.72	-0.3132	-9.12	-0.3543	-10.17	-0.3110	-9.06
7:00 - 7:59	0.2992	7.05	0.2826	6.63	0.3008	7.07	0.3012	6.99	0.2984	7.03
8:00 - 8:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00
9:00 - 9:59	-0.9985	-16.36	-0.9699	-16.49	-0.9986	-16.37	-1.0473	-17.01	-0.9957	-16.28
10:00 - 10:59	-1.4108	-30.12	-1.4890	-31.51	-1.4087	-30.04	-1.4841	-31.06	-1.4030	-29.96
11:00 - 11:59	-1.9384	-33.38	-2.0285	-34.69	-1.9375	-33.31	-2.0104	-34.11	-1.9296	-33.22
12:00 - 12:59	-2.6684	-36.62	-2.7454	-38.03	-2.6645	-36.60	-2.7596	-37.17	-2.6557	-36.51
13:00 - 13:59	-4.7459	-16.96	-4.7556	-16.95	-4.7357	-16.91	-4.8408	-17.27	-4.7302	-16.91
14:00 - 14:59	-4.4016	-20.40	-4.5611	-21.08	-4.3988	-20.38	-4.4808	-20.74	-4.3986	-20.43
15:00 - 15:59	-5.4656	-5.38	-5.5579	-5.45	-5.4801	-5.39	-5.5495	-5.45	-5.4517	-5.37
16:00 - 16:59	-	-	-	-	-	-	-	-	-	-
17:00 - 17:59	-	-	-	-	-	-	-	-	-	-
18:00 - 18:59	-	-	-	-	-	-	-	-	-	-
19:00 - 19:59	-	-	-	-	-	-	-	-	-	-
20:00 - 20:59	-	-	-	-	-	-	-	-	-	-
21:00 - 21:59	-	-	-	-	-	-	-	-	-	-
22:00 - 22:59	-	-	-	-	-	-	-	-	-	-
Inbound										
6:00 - 6:59	-	-	-	-	-	-	-	-	-	-
7:00 - 7:59	-	-	-	-	-	-	-	-	-	-
8:00 - 8:59	-	-	-	-	-	-	-	-	-	-
9:00 - 9:59	-	-	-	-	-	-	-	-	-	-
10:00 - 10:59	-	-	-	-	-	-	-	-	-	-
11:00 - 11:59	-	-	-	-	-	-	-	-	-	-
12:00 - 12:59	-	-	-	-	-	-	-	-	-	-
13:00 - 13:59	-	-	-	-	-	-	-	-	-	-
14:00 - 14:59	-1.5407	-4.85	-1.2965	-4.39	-1.5462	-4.91	-1.6323	-5.08	-1.5350	-4.84
15:00 - 15:59	-0.7206	-2.91	-0.7186	-2.91	-0.7368	-2.97	-0.7659	-2.99	-0.7217	-2.91
16:00 - 16:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00
17:00 - 17:59	0.5819	14.14	0.6063	14.61	0.5814	14.13	0.5991	14.42	0.5819	14.14
18:00 - 18:59	0.7976	19.14	0.8597	20.58	0.7952	19.08	0.8246	19.54	0.7961	19.10
19:00 - 19:59	0.8835	22.27	0.9523	23.87	0.8795	22.17	0.8816	22.03	0.8842	22.27
20:00 - 20:59	0.3501	8.02	0.3755	8.64	0.3501	8.01	0.3325	7.55	0.3528	8.07
21:00 - 21:59	0.4767	3.46	0.5081	3.67	0.4797	3.49	0.4206	3.09	0.4751	3.45
22:00 - 22:59	-	-	-	-	-	-	-	-	-	-
Departure times stay category 1										
Outbound										
6:00 - 6:59	-0.4136	-6.60	-0.5081	-7.90	-0.4170	-6.65	-0.3794	-6.13	-0.4120	-6.59
7:00 - 7:59	-0.2658	-3.90	-0.3936	-5.65	-0.2718	-3.97	-0.2656	-3.97	-0.2658	-3.90
8:00 - 8:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00

Models	Fare		Ln(Fare)		Fare per booking period		Fare per stay category		Fare per weekday	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
9:00 - 9:59	-0.8599	-10.09	-0.8999	-10.67	-0.8668	-10.16	-0.8122	-9.64	-0.8644	-10.17
10:00 - 10:59	-0.9051	-12.93	-1.0126	-14.32	-0.9108	-13.01	-0.8318	-11.89	-0.8987	-12.84
11:00 - 11:59	-1.4101	-16.93	-1.4720	-17.84	-1.4172	-16.99	-1.3406	-16.28	-1.4024	-16.86
12:00 - 12:59	-1.7703	-20.45	-1.8635	-21.33	-1.7761	-20.50	-1.6706	-19.41	-1.7602	-20.34
13:00 - 13:59	-1.4399	-16.83	-1.4906	-17.35	-1.4416	-16.83	-1.3736	-16.25	-1.4346	-16.76
14:00 - 14:59	-1.2574	-16.15	-1.4385	-18.23	-1.2640	-16.22	-1.1646	-14.88	-1.2525	-16.06
15:00 - 15:59	-1.2753	-12.87	-1.3401	-13.58	-1.2806	-12.92	-1.1883	-12.13	-1.2850	-12.86
16:00 - 16:59	-1.1040	-15.21	-1.2064	-16.32	-1.1073	-15.25	-1.0239	-14.26	-1.0915	-15.06
17:00 - 17:59	-1.0016	-15.20	-1.0982	-16.25	-1.0066	-15.26	-0.9566	-14.63	-0.9946	-15.08
18:00 - 18:59	-0.9707	-13.25	-1.0644	-14.35	-0.9774	-13.32	-0.9299	-12.82	-0.9603	-13.12
19:00 - 19:59	-1.7444	-20.00	-1.8322	-20.82	-1.7499	-20.05	-1.6873	-19.47	-1.7373	-19.92
20:00 - 20:59	-1.7759	-17.05	-1.9056	-18.13	-1.7827	-17.10	-1.6964	-16.29	-1.7728	-17.00
21:00 - 21:59	-3.0921	-11.95	-3.1358	-12.09	-3.1116	-11.99	-2.9532	-11.47	-3.0765	-11.89
22:00 - 22:59	12.2296	4.06	-4.1584	-10.09	12.2970	4.09	12.6153	4.23	12.2747	4.08
Inbound										
6:00 - 6:59	-	-	-	-	-	-	-	-	-	-
7:00 - 7:59	-	-	-	-	-	-	-	-	-	-
8:00 - 8:59	-1.2439	-5.45	-1.3285	-5.71	-1.2419	-5.44	-1.2016	-5.34	-1.2574	-5.50
9:00 - 9:59	-0.7251	-2.25	-0.7423	-2.24	-0.7252	-2.25	-0.7210	-2.29	-0.7218	-2.25
10:00 - 10:59	-	-	-	-	-	-	-	-	-	-
11:00 - 11:59	-	-	-	-	-	-	-	-	-	-
12:00 - 12:59	-	-	-	-	-	-	-	-	-	-
13:00 - 13:59	0.5014	1.71	0.4832	1.72	0.5004	1.71	0.5032	1.73	0.4981	1.70
14:00 - 14:59	-0.0744	-0.39	0.0212	0.11	-0.0751	-0.39	-0.0533	-0.29	-0.0772	-0.41
15:00 - 15:59	0.2567	1.21	0.2630	1.26	0.2606	1.23	0.2588	1.27	0.2495	1.18
16:00 - 16:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00
17:00 - 17:59	0.6366	11.88	0.6565	12.27	0.6412	11.96	0.5907	11.21	0.6298	11.73
18:00 - 18:59	0.8423	14.80	0.8782	15.68	0.8441	14.84	0.7830	14.02	0.8371	14.68
19:00 - 19:59	1.0495	20.79	1.1029	21.96	1.0485	20.78	1.0273	20.76	1.0425	20.62
20:00 - 20:59	0.5464	9.43	0.5475	9.59	0.5433	9.37	0.5686	9.96	0.5421	9.34
21:00 - 21:59	0.4057	2.89	0.3273	2.37	0.4079	2.91	0.4250	3.02	0.4004	2.85
22:00 - 22:59	-	-	-	-	-	-	-	-	-	-
Departure times stay category 2										
Outbound										
6:00 - 6:59	-0.9810	-2.14	-0.9931	-2.08	-0.9827	-2.13	-0.9844	-2.11	-1.0027	-2.16
7:00 - 7:59	-0.3001	-0.71	-0.3610	-0.82	-0.3045	-0.72	-0.3082	-0.72	-0.3398	-0.80
8:00 - 8:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00
9:00 - 9:59	-1.3147	-2.84	-1.0511	-2.24	-1.3112	-2.83	-1.2215	-2.63	-1.2666	-2.74
10:00 - 10:59	-0.1935	-0.45	-0.2101	-0.47	-0.1992	-0.46	-0.1765	-0.41	-0.2067	-0.48
11:00 - 11:59	-0.1046	-0.25	-0.1807	-0.41	-0.1118	-0.26	-0.1006	-0.23	-0.1185	-0.28

Models	Fare		Ln(Fare)		Fare per booking period		Fare per stay category		Fare per weekday	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
12:00 - 12:59	-0.7849	-1.77	-0.8237	-1.80	-0.7863	-1.77	-0.7790	-1.74	-0.7838	-1.76
13:00 - 13:59	-1.5906	-3.53	-1.4312	-3.13	-1.5824	-3.50	-1.5377	-3.39	-1.5895	-3.51
14:00 - 14:59	-0.8375	-1.78	-0.8896	-1.83	-0.8394	-1.78	-0.8276	-1.74	-0.8608	-1.81
15:00 - 15:59	-0.6890	-1.27	-0.6598	-1.17	-0.6886	-1.26	-0.7482	-1.34	-0.7230	-1.32
16:00 - 16:59	-0.6803	-1.47	-0.7036	-1.45	-0.6819	-1.47	-0.6851	-1.46	-0.6770	-1.44
17:00 - 17:59	-0.5908	-1.40	-0.5609	-1.28	-0.5925	-1.40	-0.5866	-1.37	-0.6122	-1.44
18:00 - 18:59	-0.4697	-1.09	-0.4848	-1.08	-0.4727	-1.09	-0.4758	-1.08	-0.5019	-1.15
19:00 - 19:59	-1.7251	-3.17	-1.7505	-3.07	-1.7281	-3.18	-1.7218	-3.11	-1.7025	-3.16
20:00 - 20:59	-1.2552	-2.47	-1.2202	-2.35	-1.2557	-2.47	-1.2242	-2.39	-1.2489	-2.45
21:00 - 21:59	-2.4812	-2.33	-2.5037	-2.29	-2.4935	-2.34	-2.5105	-2.33	-2.4707	-2.32
22:00 - 22:59	-	-	-	-	-	-	-	-	-	-
Inbound										
6:00 - 6:59	0.3493	0.96	0.2393	0.65	0.3484	0.96	0.3294	0.90	0.3396	0.93
7:00 - 7:59	0.9234	3.59	0.7313	2.95	0.9235	3.60	0.8557	3.34	0.8709	3.38
8:00 - 8:59	-0.1608	-0.45	-0.3170	-0.89	-0.1641	-0.46	-0.2100	-0.58	-0.2283	-0.63
9:00 - 9:59	0.8688	3.12	0.7085	2.59	0.8657	3.11	0.8315	2.97	0.8203	2.93
10:00 - 10:59	1.0404	4.08	0.8934	3.56	1.0444	4.10	1.0259	4.02	1.0322	4.02
11:00 - 11:59	1.2377	3.87	1.0867	3.44	1.2345	3.86	1.2060	3.78	1.1969	3.73
12:00 - 12:59	0.7829	3.47	0.8326	3.70	0.7839	3.48	0.7797	3.44	0.7790	3.43
13:00 - 13:59	1.2703	4.60	1.1288	4.20	1.2744	4.63	1.2317	4.49	1.2744	4.63
14:00 - 14:59	1.2978	4.73	1.1960	4.48	1.2966	4.73	1.2646	4.62	1.2682	4.62
15:00 - 15:59	1.6210	6.23	1.4299	5.61	1.6156	6.22	1.5728	6.05	1.5749	6.04
16:00 - 16:59	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00	0.0000	0.00
17:00 - 17:59	1.0398	4.87	1.0773	4.98	1.0376	4.86	1.0198	4.75	1.0010	4.64
18:00 - 18:59	1.7544	6.80	1.5782	6.34	1.7553	6.82	1.6875	6.57	1.7173	6.64
19:00 - 19:59	1.1785	4.36	1.0843	4.09	1.1779	4.36	1.1499	4.26	1.1428	4.22
20:00 - 20:59	1.2961	4.34	1.1225	3.84	1.2925	4.34	1.2464	4.17	1.2659	4.23
21:00 - 21:59	0.2066	0.25	0.1488	0.19	0.2078	0.25	0.1766	0.22	0.0407	0.05
22:00 - 22:59	-	-	-	-	-	-	-	-	-	-
Fare										
Fare	-0.0068	-75.79								
Fare booking period 0					-0.0065	-27.16				
Fare booking period 1					-0.0067	-43.25				
Fare booking period 2					-0.0067	-49.08				
Fare booking period 3					-0.0069	-31.74				
Fare booking period 4					-0.0075	-22.86				
Fare booking period 5					-0.0078	-26.21				
Fare stay category 1							-0.0077	-68.50		
Fare stay category 2							-0.0056	-41.96		
Fare stay category 3							-0.0083	-8.83		

Models	Fare		Ln(Fare)		Fare per booking period		Fare per stay category		Fare per weekday	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Fare weekday 1									-0.0088	-13.93
Fare weekday 2									-0.0065	-35.87
Fare weekday 3			1.2036						-0.0069	-41.89
Fare weekday 4									-0.0065	-39.47
Fare weekday 5									-0.0070	-38.09
Fare weekday 6									-0.0069	-29.25
Fare weekday 7									-0.0150	-7.30
ln fare			-2.3042	-83.04						

D.3 Possible further specifications

In the previous sections, several choice models were discussed. All these models were estimated on more or less the same data. The dataset was for instance adjusted when specifying fare parameters per booking period or per weekday. This was however a breakdown of a single parameter in several parameters and is therefore considered as being the same dataset.

It is however possible to use the same utility function specification on different datasets. For instance, it is possible to generate different choice sets, estimate models for a single origin destination pair or estimate models for a group of origin-destination pairs. Also, the used dataset only contained observed bookings in the same direction (directional OD's) as the webbots queried prices.

First, the results of including non-directional OD's in the dataset will be discussed. It is possible to extend the used dataset with bookings observed in the opposite direction as the webbots, which would lead to an increase of 3000 choices, making the total number of observed choices nearly 22,000. In Table_APX D-6 Directional OD's versus Non-directional OD's the exact figures are shown. It can be seen that the usage of non-directional OD's will increase the number of observations in stay category 1 and 2. No increase is observed in stay category 0; Expedia does not list early return flights, which makes matching of booked itineraries and observations on Expedia not possible. Finally, it remains questionable how realistic the formulated choice sets are, as no information is known availability of itineraries and the distribution of itinerary characteristics is likely to change (i.e. departure times).

Table_APX D-6 Directional OD's versus Non-directional OD's

Stay category		Directional OD's		Non-directional OD's	
		Absolute	Relative	Absolute	Relative
	0	10537	48%	10537	56%
	1	7300	44%	9752	39%
	2	1058	8%	1689	6%
Total		18895	100%	21978	100%

Taking the aforementioned in account, the model results will be discussed briefly. First, a basic model with a dummy variable indicating if the carrier is domestic is estimated. This variable remained in the same range. Second, a more advanced model is estimated. This model contains carrier and departure time variables. Departure time variables for stay category 0 remain approximately the same, as do variables for stay category 1. For stay category 2 however, changes occur in the sign and significance of the departure time variables. It is thought, that this is due to the earlier mentioned irrationalities in pricing systems, which influences consumer behavior: passengers actually staying at their destination for a longer period and passengers booking a return ticket are in the same stay category.

Second, different choice sets can be used for the estimation. Choice sets are limited in size for passengers returning the same day and the next day. For the first category, only choice sets containing flights departing within a window of 1 hour of the arrival time of the chosen itinerary, for the second category itinerary departing within a window of 2 hours of the arrival time of the chosen itinerary are generated. For chosen itineraries returning after 6 days, all available itineraries on the booking day are considered. However, the reduction of choice set sizes will have consequences for the earlier specified utility function: a dummy variable for departure hour will not be valid anymore as the reference category is not available in each choice set. This could be overcome by specifying a relative reference category. The main deficiency of a relative reference category lies in the fact that a parameter for a deviation from a chosen category will be estimated, which would be highly negative and significant but would not reveal much about passenger behavior. It can be argued that a Fourier approximation will be able to function. Therefore, a model is estimated. Of the six parameters required, only two are significant, indicating not enough variance. Nevertheless, if a judgment should be made, the same preference structure can be observed as with using the original choice sets.
Choice set size will become much smaller: 50% of the choice sets will contain 12 itineraries at most and 95% of the choice sets will contain 40 itineraries at most, compared to the choice set size of nearly 50 if no arrival time windows are considered.

It would also be possible to estimate different models for OD pairs or groups of OD pairs. With a revealed preference data, it remains a challenge to formulate these markets. On the one hand, it would be possible to group them geographically or by direction. It is also possible to classify them by the number of passengers heading for their destination a certain period of time. No further attention is paid to this issue.

Appendix E Notes to Chapter 7

E.1 Previous results

Table_APX E-1 shows a summary of the model results based on an utility function containing departure time period variables, opposed to departure hour variables.

Table_APX E-1 Modeling results for different parameter settings from a model specified with departure time periods

Model	Estimated parameter for $IND(c)$	Average for $IND(c)$	value chosen $\ln(IND(c))$	Utility	final log-likelihood	r-square	Difference in r-square compared to MNL
MNL	-	-	-	-	-48816	0.2948	-
$s_x = 720, \gamma = 0.25$	-1.1966	0.0386	-3.2545	3.8943	-48712	0.2963	0.5%
$s_x = 720, \gamma = 0.5$	-1.0362	0.0473	-3.0512	3.1617	-48684	0.2967	0.6%
$s_x = 720, \gamma = 0.75$	-0.7210	0.0621	-2.7790	2.0037	-48752	0.2957	0.3%
$s_x = 120, \gamma = 0.5$	-0.1835	0.1650	-1.8018	0.3306	-48798	0.2950	0.1%
$s_x = 240, \gamma = 0.5$	-0.6776	0.0984	-2.3187	1.5712	-48622	0.2976	0.9%
$s_x = 360, \gamma = 0.5$	-0.7666	0.0734	-2.6118	2.0022	-48640	0.2973	0.8%

E.2 Model results

Table_APX E-2 Best MNL-model versus best measure of independence model

Models	Basis		sim5	
	Estimated parameter	Robust t-test	Estimated parameter	Robust t-test
Carrier constants				
Flight attributes				
Non-code share	0-		0-	
Code share	-0.9247	-12.84	-0.9674	-13.54
Total travel time out	-0.0114	-5.47	-0.0105	-5.13
Number of transfers	-4.6923	-12.42	-4.6494	-12.19
Aircraft attribute				
Mainline jet	0-		0-	
Regional aircraft	-0.1328	-4.68	-0.1236	-4.31
Propellor aircraft	-1.5388	-14.79	-1.4719	-14.11
Departure times stay category 0				
6:00 - 6:59	-0.3136	-9.14	-0.3546	-10.39
7:00 - 7:59	0.2992	7.05	0.2837	6.67
8:00 - 8:59	0-		0-	
9:00 - 9:59	-0.9985	-16.36	-1.0621	-17.29
10:00 - 10:59	-1.4108	-30.12	-1.5480	-32.36
11:00 - 11:59	-1.9384	-33.38	-2.0152	-34.80

12:00 - 12:59	-2.6684	-36.62	-2.8529	-39.05
13:00 - 13:59	-4.7459	-16.96	-5.0783	-18.45
14:00 - 14:59	-4.4016	-20.40	-4.9103	-22.34
15:00 - 15:59	-5.4656	-5.38	-5.7791	-5.75
16:00 - 16:59	-	-	-	-
17:00 - 17:59	-	-	-	-
18:00 - 18:59	-	-	-	-
19:00 - 19:59	-	-	-	-
20:00 - 20:59	-	-	-	-
21:00 - 21:59	-	-	-	-
22:00 - 22:59	-	-	-	-
6:00 - 6:59	-	-	-	-
7:00 - 7:59	-	-	-	-
8:00 - 8:59	-	-	-	-
9:00 - 9:59	-	-	-	-
10:00 - 10:59	-	-	-	-
11:00 - 11:59	-	-	-	-
12:00 - 12:59	-	-	-	-
13:00 - 13:59	-	-	-	-
14:00 - 14:59	-1.5407	-4.85	-1.4793	-4.72
15:00 - 15:59	-0.7206	-2.91	-0.6910	-2.68
16:00 - 16:59	0-		0-	
17:00 - 17:59	0.5819	14.14	0.5837	14.13
18:00 - 18:59	0.7976	19.14	0.8039	19.24
19:00 - 19:59	0.8835	22.27	0.8896	22.36
20:00 - 20:59	0.3501	8.02	0.3279	7.46
21:00 - 21:59	0.4767	3.46	0.4388	3.20
22:00 - 22:59	-	-	-	-
Departure times stay category 1				
6:00 - 6:59	-0.4136	-6.60	-0.3615	-5.88
7:00 - 7:59	-0.2658	-3.90	-0.2433	-3.60
8:00 - 8:59	0-		0-	
9:00 - 9:59	-0.8599	-10.09	-0.8874	-10.48
10:00 - 10:59	-0.9051	-12.93	-1.0368	-14.96
11:00 - 11:59	-1.4101	-16.93	-1.4438	-17.78
12:00 - 12:59	-1.7703	-20.45	-1.9097	-22.20
13:00 - 13:59	-1.4399	-16.83	-1.5397	-18.07
14:00 - 14:59	-1.2574	-16.15	-1.3318	-17.35
15:00 - 15:59	-1.2753	-12.87	-1.3385	-13.73
16:00 - 16:59	-1.1040	-15.21	-1.1600	-16.15
17:00 - 17:59	-1.0016	-15.20	-1.0470	-16.03
18:00 - 18:59	-0.9707	-13.25	-1.0503	-14.40
19:00 - 19:59	-1.7444	-20.00	-1.8548	-21.45
20:00 - 20:59	-1.7759	-17.05	-1.9800	-18.78
21:00 - 21:59	-3.0921	-11.95	-3.1597	-12.28
22:00 - 22:59	-	-	-	-

6:00 - 6:59	-	-	-	-
7:00 - 7:59	-	-	-	-
8:00 - 8:59	-1.2439	-5.45	-1.2310	-5.40
9:00 - 9:59	-0.7251	-2.25	-0.7074	-2.16
10:00 - 10:59	-	-	-	-
11:00 - 11:59	-	-	-	-
12:00 - 12:59	-	-	-	-
13:00 - 13:59	0.5014	1.71	0.5269	1.80
14:00 - 14:59	-0.0744	-0.39	-0.0161	-0.08
15:00 - 15:59	0.2567	1.21	0.2995	1.41
16:00 - 16:59	0-		0-	
17:00 - 17:59	0.6366	11.88	0.6481	12.04
18:00 - 18:59	0.8423	14.80	0.8630	15.08
19:00 - 19:59	1.0495	20.79	1.0534	20.81
20:00 - 20:59	0.5464	9.43	0.5551	9.53
21:00 - 21:59	0.4057	2.89	0.4380	3.10
22:00 - 22:59	-	-	-	-
Departure times stay category 2				
6:00 - 6:59	-0.9810	-2.14	-0.9436	-2.12
7:00 - 7:59	-0.3001	-0.71	-0.2225	-0.54
8:00 - 8:59	0-		0-	
9:00 - 9:59	-1.3147	-2.84	-1.2523	-2.78
10:00 - 10:59	-0.1935	-0.45	-0.1693	-0.41
11:00 - 11:59	-0.1046	-0.25	0.1209	0.29
12:00 - 12:59	-0.7849	-1.77	-0.6327	-1.47
13:00 - 13:59	-1.5906	-3.53	-1.4261	-3.24
14:00 - 14:59	-0.8375	-1.78	-0.7165	-1.56
15:00 - 15:59	-0.6890	-1.27	-0.5369	-1.02
16:00 - 16:59	-0.6803	-1.47	-0.5947	-1.32
17:00 - 17:59	-0.5908	-1.40	-0.4940	-1.21
18:00 - 18:59	-0.4697	-1.09	-0.4295	-1.02
19:00 - 19:59	-1.7251	-3.17	-1.6595	-3.13
20:00 - 20:59	-1.2552	-2.47	-1.3648	-2.71
21:00 - 21:59	-2.4812	-2.33	-2.2422	-2.10
22:00 - 22:59	-	-	-	-
6:00 - 6:59	0.3493	0.96	0.3430	0.93
7:00 - 7:59	0.9234	3.59	0.9367	3.61
8:00 - 8:59	-0.1608	-0.45	-0.1743	-0.48
9:00 - 9:59	0.8688	3.12	0.8656	3.09
10:00 - 10:59	1.0404	4.08	1.0384	4.05
11:00 - 11:59	1.2377	3.87	1.2477	3.88
12:00 - 12:59	0.7829	3.47	0.7477	3.29
13:00 - 13:59	1.2703	4.60	1.2978	4.66
14:00 - 14:59	1.2978	4.73	1.3109	4.75
15:00 - 15:59	1.6210	6.23	1.6375	6.24

	16:00 - 16:59	0.0000	0.00	0.0000	0.00
	17:00 - 17:59	1.0398	4.87	1.0101	4.69
	18:00 - 18:59	1.7544	6.80	1.7886	6.89
	19:00 - 19:59	1.1785	4.36	1.1944	4.38
	20:00 - 20:59	1.2961	4.34	1.2976	4.32
	21:00 - 21:59	0.2066	0.25	0.2161	0.26
	22:00 - 22:59	-	-	-	-
Fare					
Fare		-0.0068	-75.79	-0.0066	-70.67
Similarity					
Similarity				0.5379	13.59
Number of estimated parameters		98.0		99	
Number of observations		18416.0		18416	
Number of individuals		18416.0		18416	
Null log-likelihood		-69032.6		-69032.6	
Init log-likelihood		-69032.6		-69032.6	
Final log-likelihood		-46101.7		-46004.2	
Likelihood ratio test		45861.9		46056.8	
Rho-square		0.3322		0.333587	
Adjusted rho-square		0.3308		0.332153	
Final gradient norm		0.0363		0.003382	