NORTHWESTERN UNIVERSITY

Modeling the Competitive Dynamic among Air-travel Itineraries with Generalized Extreme Value Models

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Civil and Environmental Engineering

By

Gregory M. Coldren

EVANSTON, ILLINOIS

JUNE 2005

ABSTRACT

Modeling the Competitive Dynamic among Air-travel Itineraries with Generalized Extreme Value Models

Gregory M. Coldren

This dissertation develops models that forecast air-travel itinerary passenger (market) shares between airport-pairs. This is the first study to model aviation demand at the itinerary level and air-carriers can use these models to assist their strategic decision-making. The motivation for developing these models is both to understand the impact of different air-carrier service attributes on itinerary share, and to understand the underlying competitive dynamic between itineraries.

Aggregate multinomial logit models with independent variables measuring air-carrier and itinerary service attributes such as level-of-service, connection quality, carrier attributes, aircraft size and type, and departure time are estimated. These model estimations describe the importance of these attributes on itinerary share. The results of these models are intuitive and offer new perspectives on the impacts of changing various service attributes on itinerary and carrier market share. Additionally, validation results from the implementation of these models by a major U.S. air-carrier are presented (the logit-based models improved the carrier's forecasting accuracy compared to its previous QSI-based model).

These multinomial logit models cannot measure the underlying competitive dynamic (if any) between itineraries due to constraints inherent in the model derivation. This competition is hypothesized to be differentiated by proximity in departure time, carrier and/or level-of-service. This hypothesis is tested by estimation of incrementally more complex generalized extreme value models. Results indicate that inter-itinerary competition is indeed differentiated by proximity in departure time, by carrier, by carrier within departure time periods and (to a lesser extent) by level-of-service within departure time periods. The advanced models estimated (extending to multi-level generalized nested logit and ordered generalized extreme value models) are shown to outperform the more basic specifications with regard to statistical tests and behavioral interpretations. Additionally, these models offer clear insights into air-traveler behavior and capture the underlying competitive dynamic among air-travel itineraries.

DEDICATION

To my wife Amy,

who understood. Every step of the way. Thank you.

ACKNOWLEDGMENTS

After eight years at Northwestern University, I have finally achieved my dream of earning a Ph.D. This was by far the most difficult and challenging experience of my life, and it would not have been possible without the help, guidance and support from many people.

I would like to begin by thanking the professors at James Madison University who were extremely influential to me in my undergraduate years. These include Carter Lyons and Peter Kohn of the Mathematics department, William Wood of the Economics department, and John Sweigart, Bill O'Meara, Richard Lippke, William Knorpp and Anne Wiles of the Philosophy and Religion department.

An early portion of this research was supported in part by a grant from United Airlines. A later potion was supported with a Dissertation Year Fellowship awarded to me from Northwestern University's Transportation Center. Additionally, John Gliebe, a former Northwestern University graduate student and current employee of PBConsult, helped introduce me to the GAUSS software system (I used GAUSS to estimate all of the models in this dissertation). Finally, Laurie Garrow, a recent graduate of the Civil and Environmental Engineering department at Northwestern University and now Assistant Professor of Civil Engineering at the Georgia Institute of Technology, gave me a tremendous amount of help and advice throughout this project. I am very thankful for this support I received.

I also want to thank the incredibly helpful staff at Northwestern University's Transportation Library (especially Joe Ellison). They helped me tremendously with my literature review, making a very difficult task easier by helping me locate the most obscure references (among other things).

My dissertation committee was instrumental in making my dissertation what it is today. Committee members Joseph Schofer, Joseph Schwieterman and Pablo Durango-Cohen were generous with their time, and their thoughtful suggestions really improved this dissertation. I enthusiastically thank them!

There is not enough space in this document to thank my Ph.D. committee chair, advisor, mentor and friend Frank Koppelman. Frank gave me the opportunity to earn my Ph.D. by taking a big chance on me. Frank is one of the most generous and compassionate people I know, never asking for anything in return. Everything I have achieved here I owe to Frank. Thank you Frank.

I would like to thank my friends for seeing me through this project. In particular, I need to thank two of my best friends: Bill McDaniel and Jason Mott. Their advice and friendship has been invaluable to me throughout this project.

Finally, I want to thank my parents, Joy and Lee Coldren, who always put my interests ahead of theirs (I did not completely appreciate them until I became a parent!); my wife, Amy, whose love has been unwavering and who has sacrificed more than me for this degree; and my beautiful children, Devon and Ivy, for bringing me an infinite amount of happiness. I love and thank you all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The airline industry has experienced drastic changes over the last thirty years. Prior to the Airline Deregulation Act of 1978, a small set of airlines dominated the United States market. Entry by other carriers was essentially blocked. During this era, the Civil Aeronautics Board (CAB) was responsible for establishing prices and determining which airline(s) would provide service between each airport-pair. Additionally, the CAB tended to discourage airlines from competing with respect to onboard amenities. Nevertheless, the CAB did not control the schedules of the airlines. Airlines could determine their flight frequencies, the time of day these flights operated, and the type(s) of equipment used. Therefore, scheduling (including equipment assignment) was the primary mechanism for airline competition.

Deregulation was supposed to usher an era of competitiveness into the U.S. aviation industry. Ironically, during the first fifteen years of the deregulated era, the industry consolidated and the growth of hub-and-spoke networks accelerated. Economies of scale, aircraft technologies, the ability of carriers to insulate themselves from competition, and marketing opportunities are the most cited reasons for this phenomenon (De Vany and Garges 1972; Gordon and De Neufville 1973; Kanafani 1981; Kanafani and Ghobrial 1982; Kanafani and Ghobrial 1985).

Beginning in the early 1990's and continuing through today, low-cost airlines (many of which operate point-to-point service networks) have been aggressively challenging the traditional network-based (legacy) carriers. These low-cost airlines have ushered in an era of price competition and have steadily gained market share. Some traditional carriers have launched "low-cost" airlines of their own in response to this assault. Windle and Dresner (1995a) show low-cost carrier entry into a market significantly reduces prices and increases traffic for the market (the entry of legacy carriers usually has no significant effect on prices and traffic) and Dresner *et al.* (1996) show low-cost carrier entry on a route spurs lower prices and higher traffic on non-competing routes out of the airport of entry as well as competing routes from nearby airports[1]. Simultaneous with the rise of low-cost carriers, legacy carriers have had to deal with the decline of their "bread and butter" business travelers, labor strife, and an overall diminishing of demand due to geo-political concerns and a weak economy. Regardless of the fact that the landscape of the U.S. airline industry is changed in a variety of dimensions, it is still a vital element of the transportation system in particular, and the national economy in general, with the quality of air-carrier facilities and services affecting the accessibility and economic growth prospects of major geographic regions.

Though passenger demand for air-carrier services has fluctuated greatly in response to new technologies, evolving consumer preferences and macroeconomic changes, it remains the foundation of the aviation industry. Understanding this demand is crucial for air-carriers, government policy makers, airport operators, air-carrier suppliers and researchers.

Many studies of aviation demand forecast air-travel volumes for given levels of aggregation such as system (Brown and Watkins 1968; English and Kernan 1976;

[1] That is, if low-cost airline XYZ were to enter the BWI – Cleveland market, this would have price and traffic effects on the BWI – Buffalo market, as well as the Washington National – Cleveland and Washington Dulles – Cleveland markets.

Transportation Research Circular #348 1989), metropolitan region (Mumayiz and Pulling 1992), airport (Skinner 1976; Augustinus and Demakopoulos 1978; Harvey 1987; Ashford and Benchemam 1987; Furuichi and Koppelman 1994; Windle and Dresner 1995b; Suzuki *et al.* 2003; Basar and Bhat 2004; Hess and Polak 2005) or airport (city) pair (Brown and Watkins 1968; Brown and Watkins 1971; Verleger 1972; De Vany and Garges 1972; Douglas and Miller 1974a; De Vany 1974; Kanafani and Fan 1974; De Vany 1975; Ippolito 1981; Anderson and Kraus 1981; Abrahams 1983; Reiss and Spiller 1989; Dresner *et al.* 1996; Corsi *et al.* 1997).

Other studies allocate air-travel volumes to air-carriers at a given level of aggregation. These air-carrier allocation studies typically identify a relationship between airline service attributes and the allocation of air-travel volumes. Additionally, these studies estimate the relative importance of different carrier attributes (*e.g.* market presence, fare-levels, service quality) on demand share and quantify the tradeoffs that passengers make among these in their air-travel purchasing decisions. The most aggregate allocation studies explore methods of distributing system-wide airline demand to individual carriers (Nason 1981; Morash and Ozment 1996; Suzuki *et al.* 2001).

While predicting system-wide market shares is interesting, it does not give researchers or managers enough planning information or behavioral insight due to its lack of detail on the impact of carrier service attributes on demand in different markets. For example, a given carrier's market share can vary greatly across airport-pairs due to its differing levels of market presence, fares and service quality in the different markets. Detailed market-level information about carrier service attributes is necessary if market-level forecasts are desired. Studies focusing on the allocation of airport-pair demand to individual carriers employ richer and more disaggregate data with respect to airport-pair airline-specific quality of service measures than the studies referenced in the preceding paragraph (Ghobrial and Soliman 1992; Nako 1992; Proussaloglou and Koppelman 1995).

Airport-pair allocation studies have contributed to the understanding of the relationship between carriers' airport-pair service attributes and their market share. However, even though some carrier airport-pair characteristics are important in determining demand share (*e.g.* overall presence in market), the service provided by a carrier in a market is typically not homogeneous. For a given day and airport-pair, a carrier can offer service varying by departure time, equipment type, number of stops, or any number of other attributes, each of which can have a profound effect on demand. Thus, it is not realistic to only use airport-pair characteristics in order to describe a carrier's market share.

Some studies in the literature take a step in the right direction by modeling flight-level allocation using survey methodologies to gauge the tradeoffs that passengers make among flight attributes in their flight-choice behavior (Yoo and Ashford 1996; Proussaloglou and Koppelman 1999; Bruning and Rueda 2000; Algers and Beser 2001). However, even though these studies present richer data and more realistic air-passenger choice scenarios than the airport-pair carrier allocation studies, they are still inadequate due to the fact that they don't systematically model all of the air-travel options available to travelers within a given airport-pair.

All the above-mentioned carrier allocation studies fall into at least one of the following categories: 1) studies based on data with a high level of geographic aggregation, 2) studies employing surveys with a very limited range of airport-pairs or 3) studies based on stated preference data which may be subject to bias (Ben-Akiva *et al.* 1989; Morrison 2000; Murphy *et al.* 2005). Additionally, a major limitation of these studies is their failure to model air-travel demand at the itinerary level. An itinerary, as used here, is a leg (flight number) or sequence of legs connecting a given airport-pair (a leg may be included in several itineraries linking a given airport-pair as well as being included in itineraries linking many different airport-pairs). Itineraries are either nonstop, direct (a connecting itinerary not involving an airplane change), single-connect (a connecting itinerary with an airplane change) or double-connect (an itinerary with two connections)[2]. For a given day, an airport-pair may be served by hundreds of itineraries, each of which offers passengers a potential way to travel between the airports.

Itineraries are the products that are ultimately purchased by passengers, and hence it is their characteristics that influence demand. In making their itinerary choices, travelers make tradeoffs among the characteristics that define each itinerary (*e.g.* departure time, equipment type(s), number of stops, route, carrier). Modeling these itinerary-level tradeoffs is essential to truly understanding air-travel demand.

Using comprehensive and official schedule and bookings data, this dissertation builds on a framework established by Coldren *et al.* (2003), Coldren and Koppelman (2005a) and Coldren and Koppelman (2005b), and focuses on the problem of allocating airport-pair demands to individual itineraries linking the airport-pairs. Because this is the first study to model itinerary-level demand, this dissertation provides the most comprehensive and thorough explanation of air-carrier demand to-date.

[2] These four classifications are referred to as an itinerary's "level-of-service" in this dissertation.

CHAPTER 1: PRELIMINARY DISCUSSION

1.1 AIR-TRAVEL ITINERARY SHARE MODELS

As mentioned in the Introduction, it is essential to model the allocation of airport-pair demand at the itinerary level in order to truly understand passenger behavior and hence to accurately forecast passenger demand. Models that address this problem are referred to as air-travel itinerary share models[3]. Itinerary share models forecast the share of passengers expected to travel on each itinerary between any airport-pair. This is done by relating each itinerary's value (quality) to the total value of all itineraries in its respective airport-pair. The value of an itinerary is usually modeled as a function of its service characteristics[4] (independent variables) and their corresponding preference weights (parameter estimates). Though various air-carriers (and aviation consulting firms) own proprietary itinerary share models, no published study currently exists addressing the itinerary share problem except those developed in conjunction with this dissertation[5].

Though itinerary share models vary with respect to the service characteristics that are included in their specifications, they all describe the impact that these service characteristics have on itinerary share. Additionally, advanced itinerary share models (like those presented in this dissertation) allow the underlying competitive dynamic, if any, among air-travel itineraries to be measured.

---

[3] In the aviation industry, these models are also referred to as network-planning or network-simulation models.

[4] Such as number of stops, equipment type(s), departure time, elapsed time, fare, etc.

[5] These are Coldren *et al.* (2003), Coldren and Koppelman (2005a) and Coldren and Koppelman (2005b).

Once an itinerary share model is estimated, these shares (probabilities) can be applied to airport-pair volume forecasts obtained from a separate model[6], allowing for itinerary-level passenger forecasts to be obtained. These itinerary-level forecasts can then be assigned to flight legs to obtain carrier market share at the flight-leg, airport-pair, region, system or any other level of aggregation.

When passenger forecasts from an itinerary share model are combined with fare and cost data, it becomes a powerful tool allowing an airline to evaluate its current schedule and network, as well as proposed network changes, alliance (merger) evaluations and "what-if" scheduling decisions (including modeling the potential actions of competitors) with respect to passengers, revenue and profitability. Thus, these models can be invaluable to a passenger airline's tactical decision-making and strategic planning. Some common situations where these models aid carriers are: merger and acquisition scenarios, route schedule analysis, codeshare scenarios, equipment assignment scenarios, minimum connection time studies, price-elasticity studies, hub location studies, and equipment purchasing decisions.

Because itinerary share models guide critical decisions, it is paramount that they be as accurate as possible since, for a given carrier, improving the forecasting ability of its itinerary share model will translate to improvements in its revenue management, schedule efficiency and profitability.

[6] Airport-pair demand forecasting is beyond the scope of this dissertation. However, these forecasts are relevant since the accuracy of itinerary share forecasts depends on the accuracy of the total airport-pair forecast. See Chapter 2 for a review of studies concentrating on airport-pair (and city-pair) demand forecasting.

1.2 QUALITY OF SERVICE INDEX ITINERARY SHARE MODELS

Most of the proprietary itinerary share models mentioned in Section 1.1 employ a demand allocation methodology referred to as quality of service index (QSI). QSI models[7] relate an itinerary's passenger share to its "quality" (and the quality of all other itineraries in its airport-pair), where quality is defined as a function of various itinerary service attributes and their corresponding preference weights. For a given QSI model, these preference weights are obtained using statistical techniques and/or analyst intuition. Once the preference weights are obtained, the final itinerary quality of service index (QSI) for a given itinerary is usually expressed as a linear or multiplicative function of its service characteristics and preference weights[8]. For itinerary i, its passenger share is then determined by:

$$S_i = \frac{QSI_i}{\sum_{j \in J} QSI_j} \tag{1.1}$$

where S_i is the passenger share assigned to itinerary i,
QSI_i is the quality of service index for itinerary i and
the summation is over all itineraries in the airport-pair.

QSI models are problematic on two fronts. First, a distinguishing characteristic of these models is that their preference weights (or sometimes subsets of these weights) are

---

[7] Information on QSI models obtained from TRB Transportation Research E-Circular E-C040 (2002) as well as the author's personal experience with the models.

[8] For example, suppose a given QSI model measures itinerary quality along four service characteristics (*e.g.* number of stops, fare, carrier, equipment type) represented by independent variables X_1, X_2, X_3, X_4 and their corresponding preference weights β_1, β_2, β_3, β_4. The QSI for itinerary i, QSI_i, can be expressed as $QSI_i = (\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$ or

usually obtained independently from the other preference weights in the model. Thus, QSI models do not capture interactions existing among itinerary service characteristics (e.g. elapsed itinerary trip time and equipment, elapsed itinerary trip time and number of stops). Second, QSI models are not able to measure the underlying competitive dynamic that may exist among air-travel itineraries. This second inadequacy in QSI models can be seen by examining the cross-elasticity equation for the change in the share of itinerary j due to changes in the QSI of itinerary i:

$$\eta_{QSI_i}^{S_j} = \frac{\partial S_j}{\partial QSI_i}\frac{QSI_i}{S_j} = -S_i \tag{1.2}$$

The expression on the right side of Equation (1.2) is not a function of j. That is, changing the QSI (quality) of itinerary i will affect the passenger share of all other itineraries in its airport-pair in the same proportion. This is not realistic since, for example, if a given itinerary (linking a given airport-pair) that departs in the morning improves in quality, it is likely to attract more passengers away from the other morning itineraries than the afternoon or evening itineraries. Since QSI models are not able to capture the interactions between itinerary service characteristics or the underlying competitive dynamic among itineraries, these models are not adequate for modeling air-travel itinerary shares.

$QSI_i = (\beta_1 X_1)(\beta_2 X_2)(\beta_3 X_3)(\beta_4 X_4)$. Other functional forms for the calculation of QSI's are also possible.

1.3 MOTIVATION FOR RESEARCH

Due to the critical importance of air-travel itinerary share models in guiding air-carrier strategy, the inadequacies of QSI-based models currently in use by some carriers, and the lack of itinerary share models in the literature, the research presented in this dissertation fills a gap in the airline forecasting literature by providing a framework for itinerary-level demand allocation modeling.

The itinerary share models developed in this dissertation use aggregate logit-based share techniques[9]. In these models, a value function, the aggregate analog of utility, represents the relative desirability of each itinerary connecting an airport-pair for each day of the week. In specifying these value functions, itinerary-level and carrier airport-pair service-characteristic variables are used to relate each itinerary with its airport-pair demand share.

The motivation for the research contained in this dissertation is twofold given that the overarching goal is to model itinerary shares: 1) to understand the impact of different air-carrier service attributes on itinerary share and 2) to understand the underlying competitive dynamic between itineraries.

1.4 CONTRIBUTIONS OF RESEARCH

The itinerary share models presented in this dissertation are state-of-the-art with respect to itinerary share accuracy, behavioral insights, and modeling flexibility.

---

[9] Aggregate logit shares are used (despite the availability of individual itinerary-level booking information) as the data do not contain information about traveler or trip characteristics. This is discussed in more detail in Chapter 3.

Specifically, the research presented herein contributes to the literature in three important ways.

First, this is the first study to systematically and comprehensively model aviation demand at the itinerary level. Second, unlike QSI-based itinerary share models, the logit-based models developed in this dissertation use advanced econometric techniques to simultaneously estimate parameters for the independent variables used in the model specifications. Explanatory variables used in this study such as level-of-service indicators differing by market type, connection quality variables, departure time variables, and equipment size and type variables have previously not been reported in the literature. Use of these variables yields valuable insights into air-traveler behavior. Additionally, implementation of the most basic itinerary share model presented in this dissertation by a major U.S. carrier yielded substantial improvements in the carrier's forecasting accuracy compared with its previous QSI-based model[10].

Third, generalized extreme value (GEV) (McFadden 1978) specifications are used to model the underlying competition among air-travel itineraries. These models provide flexibility in capturing inter-itinerary competition dynamics along a variety of dimensions. Most of these structures have never been used in the aviation demand literature and some are new to the logit share literature in general. Some of these models are likely to be applicable to areas outside of aviation demand as well.

[10] This improvement is quantified in Chapter 4.

1.5 OUTLINE OF DISSERTATION

Chapter 2 of this dissertation provides an overview of the literature pertinent to the presented research. This includes background information on the forecasting of air-travel volumes at different levels of aggregation, the allocation of air-travel volumes to air-carriers at different levels of aggregation, the impact of different service attributes on airline demand, and airport-pair itinerary share modeling and logit share techniques for modeling itinerary share.

Chapter 3 outlines the conceptual and modeling framework of this dissertation. This includes an introduction to discrete choice analysis and its techniques, as well as a description of the data used.

Chapter 4 details the impact of various itinerary service characteristics on itinerary share via the use of multinomial logit (MNL) models. In particular, the impact of level-of-service, connection quality, carrier presence, fares, carrier, aircraft size and type, and departure time on itinerary share are detailed. Additionally, validation results from the implementation of these models by a major U.S. carrier are presented.

Chapter 5 develops several variations of the nested logit (NL) model in order to capture the underlying competitive dynamic among air-travel itineraries along the dimensions of departure time, carrier and level-of-service. Chapter 6 advances these NL models by estimating ordered generalized extreme value (OGEV) and hybrid-OGEV models allowing for complicated inter-itinerary competition relationships to be measured.

Finally, Chapter 7 summarizes the findings, contributions and conclusions of this research, as well as its limitations and future directions.

CHAPTER 2: LITERATURE REVIEW

Studies in the air-travel demand modeling literature typically either forecast volumes at a high level of aggregation or allocate these volumes to a somewhat more disaggregate level. The research presented in this study belongs in the second category. Until now, the most disaggregate allocation studies have allocated airport-pair volumes to flights (nonstop and direct) linking airport-pairs. These allocation studies are not ideal since they do not consider all potential travel options (that is, they do not consider connections) between the airport-pairs in their respective samples. Additionally, most airport-pairs do not have nonstop (or direct) service. Thus, for a given airport-pair, these studies do not accurately capture air-traveler behavior. The allocation of airport-pair demand should be modeled at the itinerary level (independent of the type of service offered in the airport-pair) since the itineraries linking an airport-pair represent all possible ways to travel between the airports. The itinerary share models developed in this study accomplish this. Though this is the only study to-date that models the allocation of airport-pair demand to the itinerary level, it is beneficial to review air-travel demand allocation studies at a higher level of aggregation. Additionally, because the final quality of any allocation model depends on the accuracy of the volumes it is distributing, it is also worthwhile to review the literature on forecasting total air-travel volumes.

Section 2.1 of this chapter reviews literature detailing the forecasting of air-travel volumes for given levels of aggregation, with emphasis on the forecasting of system, metropolitan region, airport, and city (airport) pair volumes. Section 2.2 reviews studies that allocate system-wide demand to air-carriers, airport-pair (city-pair) demand to carriers, and airport-pair demand to flights linking the airport-pair.

As mentioned in Section 1.3, one motivation for developing the itinerary share models of this dissertation was to better understand the impact of different air-carrier service characteristics on itinerary share. Because no previous studies have modeled air-travel demand at the itinerary level, no information on the link between itinerary-level attributes and itinerary demand share is available. However, many aviation demand studies (both volume forecasting and allocation) do detail relationships between service attributes and airline demand (measured at a higher level of aggregation than itineraries). Section 2.3 reviews these studies. In particular, studies measuring the impact of carrier presence, fares, route, number of stops (connections), equipment, departure time, elapsed trip time, frequent flyer programs and/or service quality on airline demand are reviewed.

Section 2.4 explains that itinerary share modeling (the central problem of this dissertation) has not been adequately addressed in the aviation literature. Further, techniques are reviewed from the logit share literature that can potentially be used to solve the research problem of this study. These techniques are relevant for measuring the underlying competitive dynamic among air-travel itineraries, another primary motivation of this dissertation.

2.1 FORECASTING AIR-TRAVEL VOLUMES FOR GIVEN LEVELS OF AGGREGATION

2.1.1 FORECASTING SYSTEM VOLUME

The broadest air-travel forecasting studies forecast system demand (either world-wide, continent-wide or U.S. domestic). These typically link demand to time or seasonal trends, or broad demographic or economic variables such as gross national product, per-capita income, the consumer price index and/or population. For example, Brown and Watkins (1968) forecast U.S. domestic system demand using aggregate annual data, while English and Kernan (1976) use the Delphi method[11] to forecast world air-traffic.

Typically, public aviation authorities forecast this system demand. The Federal Aviation Administration (FAA) does forecasting for the United States and most studies employing U.S. system demand rely on FAA forecasts rather than developing their own (Gosling 1994). TRB Transportation Research Circular #348 (1989) provides an overview of the methodology employed by the FAA to forecast system demand. Obviously, system-wide forecasts are important to national governmental authorities (*e.g.* planning air-traffic control capacities). However, these forecasts shed no light on individual travel behavior, regional differences in air-travel demand or on the research problem of this dissertation.

2.1.2 FORECASTING METROPOLITAN REGIONAL VOLUMES AND AIRPORT VOLUMES IN MULTI-AIRPORT METROPOLITAN REGIONS

Studies in the literature forecasting air-travel volumes for metropolitan regions generally employ a top-down forecasting methodology where the FAA national forecast is allocated to regions of interest. Regional (for the same reasons as national) forecasts give little insight into air-travel behavior. However, regional air-travel forecasts are important for regional planning agencies when considering airport capacity and access issues, as well as the anticipated effects of changes to the fleet mix. An introduction to regional forecasting methods can be found in Mumayiz and Pulling (1992).

Airport demand in multi-airport regions can be obtained from regional demand via the use of airport distribution or choice models. Originally, the allocation of air-travel demand to individual airports in multi-airport regions was modeled in the literature using aggregate demand distribution models, but has recently been treated using discrete choice analysis. The motivation for these studies was the realization that the distribution of air-travel passengers to airports in multi-airport regions is based on passenger tradeoffs among the attributes of each airport. The majority of the studies reviewed for this dissertation find that some measure of access time to each airport and some measure of air-carrier level-of-service from each airport to the travelers' destination are the most important determinants of airport choice. Not surprisingly, these studies have found that higher levels-of-service and lower access times make an airport more attractive and hence increase its market share within a region.

The following is a brief review of some representative articles on airport choice. Augustinus and Demakopoulos (1978) develop a model that distributes the demand from each aviation-planning zone in a metropolitan area to each airport according to a

[11] The Delphi method is an iterative judgmental forecasting technique that employs a panel of experts to achieve consensus on a given question.

deterministic function that is proportional to the relative access costs of those airports to each zone. Skinner (1976) gauges the tradeoff passengers make between access times and air-carrier levels-of-service. It is believed that he is the first author to model the tradeoffs passengers make between airside and landside attributes in their airport choice process.

Harvey (1987) includes non-linear specifications of access time and flight frequency in his utility functions. He finds that there is decreasing marginal utility for additional flight frequencies from an airport, and there appears to be a saturation point where additional frequencies do not add to passenger utility. Additionally, he finds that the marginal disutility of access time decreases with total trip time. Ashford and Benchemam (1987) incorporate an aggregate fare measure from each airport to the travelers' destination.

Furuichi and Koppelman (1994)[12] estimate a nested logit model measuring the joint destination and departure airport choice of Japanese international air-travelers. The authors incorporate a relative measure of air-carrier level-of-service from each airport[13] that is believed to more accurately capture the impact of level-of-service on passenger airport choice behavior than absolute measures[14]. Windle and Dresner (1995b) include an airport experience measure into some of their specifications. They find passengers who

---

[12] This paper differs from the other reviewed papers on airport choice because it considers departure airports within a country (Japan), not just a metropolitan region.
[13] The share of nonstop flights serving the travelers' destination from the airport with respect to all nonstop flights out of Japan to the travelers' destination.
[14] However, their definition of destination includes countries, continents and even groups of continents, rather than a specific metropolitan region.

have chosen an airport in the past will feel more comfortable with that airport and will tend to choose it again (all else being equal).

Hess and Polak (2005) estimate a mixed multinomial logit (MMNL) model of airport choice. They find random taste variation across and within segments of the population with respect to the impact of access time on airport choice. Basar and Bhat (2004) develop a parameterized probabilistic choice set multinomial logit (PCMNL) model of airport choice. This allows for a random distribution of choice sets across the sample, since even though each airport may be feasible for a traveler, he/she may not consider a given airport.

Suzuki *et al.* (2003) model the phenomena of "airport leakage", occurring when travelers in small metropolitan regions drive long distances to depart from a larger metropolitan airport due to its lower fares or more convenient schedule. The authors employ two unique variables in their specification. First, they employ a quality of experience variable. They find travelers are more likely to choose an airport where they had a good experience over an airport where they had a bad experience. They also find that a traveler is more likely to choose an airport where he/she had a bad experience over an airport where he/she has no experience. Second, they include a variable measuring the impact of an airport offering service from an air-carrier of which the traveler is a frequent-flyer program member.

Common findings of the above studies are:

- Business passengers are more sensitive to access time in their airport choice process than leisure passengers (Harvey 1987; Windle and Dresner 1995b; Suzuki *et al.* 2003; Hess and Polak 2005).

- Business passengers are more sensitive to flight frequencies (schedule convenience) in their airport choice process than leisure passengers (Ashford and Benchemam 1987; Harvey 1987; Windle and Dresner 1995b; Hess and Polak 2005).

- Access time is more important to air-travelers making short trips than to air-travelers making long trips. This is because access time for short trips makes up a higher proportion of the total trip time (Augustinus and Demakopoulos 1978; Harvey 1987).

- Passengers experience non-increasing marginal utility to increases in flight frequencies from an airport to their destination (Harvey 1987; Hess and Polak 2005).

- The general fare level at airports in multi-airport regions is inversely related to the probability of airport choice (Ashford and Benchemam 1987; Suzuki *et al.* 2003; Hess and Polak 2005).

- Leisure travelers are more sensitive to fares then business travelers with respect to their choice of airport (Ashford and Benchemam 1987; Suzuki *et al.* 2003; Hess and Polak 2005).

The above-mentioned studies have improved our understanding of the factors that influence airport choice in multi-airport regions. However, some of the variable specifications used in these studies are problematic. For example, many of the studies use daily (or even weekly) nonstop flight frequencies to the travelers' destination from a given airport to measure its level-of-service. In addition to only looking at nonstop flights, this aggregate approach of combining flight frequencies across time periods and carriers over-simplifies the air-traveler choice process. A measure of the amount and type (ideally by departure time, carrier, routing, etc.) of itineraries linking airports would better quantify the level-of-service offered by the airport. Similarly, the studies that include a fare variable generally use an extremely aggregate measure (across time periods, carriers, classes of service, etc.). These measures do not accurately represent the effect that the distribution of fares has on airport choice. Finally, numerous other variables (not explicitly included in the above studies) may be relevant to airport choice such as the quality of parking facilities, types and quality of transit access, airline lounge availability, terminal congestion and propensity for delays. Clearly, the limited availability of data precludes the estimation of models including many of these variables. However, if possible, these variables (in addition to better fare and level-of-service variables) should be considered in future airport-choice models.

2.1.3 FORECASTING CITY-PAIR (AIRPORT-PAIR) VOLUMES

City-pair forecasts are commonly based on city-pair attributes, measures of air-carrier service quality between the cities[15] and average fare levels. It has been found that city-pair air-travel demand is related to the demographic and geographic properties of the city-pair. For example, demand has been shown to be positively related to income measures of the city-pair (Brown and Watkins 1968; Brown and Watkins 1971; Verleger 1972; De Vany and Garges 1972; Douglas and Miller 1974a; De Vany 1974; Kanafani and Fan 1974; De Vany 1975; Ippolito 1981; Anderson and Kraus 1981; Abrahams 1983; Dresner *et al.* 1996; Corsi *et al.* 1997), to be positively related to population measures of the city-pair (De Vany and Garges 1972; De Vany 1974; Kanafani and Fan 1974; De Vany 1975; Ippolito 1981; Abrahams 1983; Reiss and Spiller 1989; Dresner *et al.* 1996; Corsi *et al.* 1997) and to be related to the distance between the cities which can positively (Ippolito 1981; Corsi *et al.* 1997) or negatively (Brown and Watkins 1968; Brown and

Watkins 1971; De Vany and Garges 1972; De Vany 1974) affect the amount of air-travel between them[16].

Brown and Watkins (1968, 1971) and Abrahams (1983) use variables that are unique in the aviation demand literature to model city-pair demand. For example, Brown and Watkins (1968, 1971) use the number of phone calls between cities as a proxy for the city-pairs' "community of interest". This variable is shown to be positively related to air-travel demand (although the causality is not known). Brown and Watkins (1968) relate a competition index[17] to city-pair demand. Abrahams (1983) incorporates an air-car relative-cost modal split variable into his specification. Finally, Reiss and Spiller (1989) separately model city-pair nonstop and connecting air-travel demand. However, it is doubtful that passengers *a priori* decide (as is assumed in their model) whether to fly nonstop or not. Rather, they decide to travel between the city-pair and other factors (*e.g.* fares, airline preference) determine whether they fly nonstop[18].

The conclusions of many of the above-mentioned studies are weakened by the use of very aggregate variables in their specifications (*e.g.* average city-pair fares). No consideration is given to independent variable differences across itineraries, carriers, departure times, etc. They also do not recognize regional or travel-characteristic differences amongst the city-pairs. Thus, due to the fact that these studies ignore demand

---

[15] Overall nonstop flight frequency (across carriers) is usually used as a proxy for service quality in a city-pair.

[16] Distance may be positively related to city-pair air-travel demand since air-travel may be the only viable mode of transportation between long-distance city-pairs. However, long distances also imply high elapsed travel times which tend to depress demand.

property differences between markets and their use of aggregate variables, their city-pair demand functions may have little connection to the underlying micro-travel demand functions (Verleger 1972). Potentially, city-pair air-travel volume forecasts can be improved by employing measures of city-pair service extracted from the itinerary share models estimated in this study.

2.2 MODELING THE ALLOCATION OF AIR-TRAVEL VOLUMES FOR GIVEN LEVELS OF AGGREGATION TO AIR-CARRIERS OR FLIGHTS

2.2.1 MODELING SYSTEM-WIDE CARRIER SHARE

The broadest allocation studies distribute system-level demand to individual carriers. Nason (1981) develops a model of system-wide airline choice using stated preference survey data by giving respondents a series of questions and asking them to choose which airline has the "best" service for various service attributes. His model predicts airline choice as a function of socioeconomic passenger characteristics as well as these reported service attribute ratings. Morash and Ozment (1996) link airlines' "external time advantages" (*e.g.* on-time reliability, frequency of service, denied boardings, percentage of flights cancelled) and "internal time advantages" (*e.g.* network size and density) to customer perception of quality and ultimately system market share. Finally, Suzuki *et al.* (2001) link airline customer service quality (and other attributes) with system market share using a loss-aversion[19] framework.

---

[17] This variable measures the percentage of the second largest carriers' traffic in a market to that of the largest carriers' traffic in the market. The parameter estimates for this variable were positive indicating that increased competition within a market is a positive factor affecting market demand.
[18] See Subsection 2.3.3 on air-travel route choice for a more detailed discussion of this topic.
[19] Loss aversion theory states that consumers evaluate product attributes relative to a reference point and react more strongly to losses (negative deviations from the reference point) than to equivalent-sized gains (Suzuki *et al.* 2001).

2.2.2 MODELING CITY-PAIR (AIRPORT-PAIR) CARRIER SHARE

Using revealed preference data, a few studies in the literature model the allocation of city-pair demand to air-carriers as a function of the carriers' city-pair characteristics (Ghobrial and Soliman 1992; Nako 1992; Proussaloglou and Koppelman 1995). These studies employ "better" variables (with respect to behavioral interpretations and managerial implications) in their specifications than the system-level allocation studies such as peak and off-peak flight frequency (Ghobrial and Soliman 1992), itinerary quality (number of stops) (Ghobrial and Soliman 1992), commuter airline indicators (Ghobrial and Soliman 1992), frequent flyer program membership variables (Nako 1992; Proussaloglou and Koppelman 1995), percentage of direct flights offered (Nako 1992) and a nonlinear specification for flight share (Proussaloglou and Koppelman 1995).

2.2.3 MODELING AIRPORT-PAIR FLIGHT SHARE

Using both revealed preference (RP) and stated preference (SP) survey methodologies, recent studies have modeled flight-level allocation by gauging the tradeoffs that passengers make among flight attributes in their flight-choice behavior (Yoo and Ashford 1996; Proussaloglou and Koppelman 1999; Bruning and Rueda 2000; Algers and Beser 2001). Yoo and Ashford (1996) develop RP and SP models for the international flight choice behavior of Korean air-travelers. Their specifications include a dummy variable indicating whether a flight was operated by a Korean airline. They find (all things being equal) Korean airlines were viewed more positively by Korean travelers than non-Korean airlines. Algers and Beser (2001) using a blend of RP and SP techniques, and Proussaloglou and Koppelman (1999) using SP techniques, show empirically that schedule delay[20] has a negative impact on flight choice probabilities with the impact being greater for business travelers than leisure travelers (Proussaloglou and Koppelman 1999). Finally, Bruning and Rueda (2000) using SP methods gauge the importance of frequent flyer program membership on flight choice.

2.3 THE IMPACT OF SERVICE ATTRIBUTES ON AIRLINE DEMAND

This section describes the relationships between different service attributes and airline demand that have been discovered in the literature. Note that even though each service characteristic is treated separately, most of them are interrelated. For example, airline presence is positively correlated with airline fares, connection quality is related to elapsed travel time, etc.

2.3.1 THE IMPACT OF AIR-CARRIER PRESENCE ON AIRLINE DEMAND

Loosely defined, air-carrier presence refers to the amount of scheduled service offered between an airport-pair (industry-wide or carrier specific) or the amount of service supplied out of an airport (industry-wide or carrier specific). Early studies (using data from the regulatory period) framed the airport-pair scheduling decisions of the airlines as a proxy for the amount of service quality offered in a market (Douglas 1971; Douglas and Miller 1974a; Douglas and Miller 1974b; De Vany 1975; Ippolito 1981; Anderson and Kraus 1981; Abrahams 1983). In these studies, airport-pair service quality is typically defined as the expected amount of schedule delay[21] an average passenger

[20] The difference in time between a passengers' desired departure time and the departure time of a given scheduled flight.

[21] Schedule delay as defined here is the sum of frequency and stochastic delays. Frequency delay is the time difference between a passengers' desired departure time and the time of the closest scheduled departure. A

incurs from the air-carrier schedule (across carriers) in that market. Using this definition, airport-pair service quality is positively related to flight frequency and equipment size since more frequencies lessen frequency delays and larger airplanes lessen stochastic delays. De Vany (1975), Ippolito (1981) and Abrahams (1983) show that city-pair demand is positively related to increases in market frequency and/or capacity.

The prevalence of hub-and-spoke networks after deregulation resulted in tradeoffs to consumers. There is evidence that hubbing has allowed airlines to exhibit market power at their hubs. For a given route, studies have shown that prices are higher when operated by a carrier who has a hub on either (or both) ends of the route compared with carriers who do not have a hub at either end, as well as being higher for carriers who dominate service on that particular route[22] (Huston and Butler 1988; Borenstein 1989; Berry 1990; Borenstein 1990; Dresner and Windle 1992). However, these price premiums have been modest since (potentially) airlines have passed some of the cost savings achieved from hubbing to consumers as well as the fact that airline markets are contestable[23]. Finally, research has shown the characteristics of the carriers competing on a route determine its price more than the number of carriers. That is, a route dominated

---

passenger incurs stochastic delay if the flight he/she wants to take is booked to capacity and the passenger has to take a different flight.

[22] Airport domination does not necessarily imply route domination for a given route, although they are positively related. In this discussion, they will be treated interchangeably.

[23] Due to the mobile nature of its capital and the low costs of establishing service, it is easy for an airline to enter a given market (assuming gate and slot availability). Thus, there is an incentive for a carrier who dominates a route not to charge too high a fare lest they attract competition (Huston and Butler 1988).

by a low-cost carrier may have lower prices than a less concentrated route served by several legacy carriers (Windle and Dresner 1995a).

Clearly, the growth of hub-and-spoke networks after deregulation resulted in the loss of nonstop service between many small city-pairs. However, this was compensated for by more frequent (though with longer elapsed times) service between these city-pairs via connections through a hub. See Douglas (1971), De Vany and Garges (1972), Kanafani and Chang (1979), Kanafani and Ghobrial (1982), Kanafani and Ghobrial (1985) and Ghobrial and Soliman (1992) for a discussion of this tradeoff. Additionally, hub-and-spoke networks resulted in more frequent service from hubs to spoke cities (since the demand for each flight to a spoke was being fed from all over the system) and, for spoke cities, gave passengers better access to more destinations. Thus, one result of hub-and-spoke networks was that passengers got more frequencies to their destinations but had less direct routings and longer elapsed travel times[24].

In summary, carriers have been able to charge premium prices on routes to and from their hubs while simultaneously offering frequent service to their spoke airports. They can achieve this since carriers who dominate a route will capture a disproportionate share of the market and be able to charge a higher fare than their competitors (Douglas and Miller 1974b; Kanafani and Ghobrial 1985; Borenstein 1989; Berry 1990). Reasons for airlines being able to charge a premium for flights to and from their hubs include hub-

---

[24] Recently, low-cost carriers offering point-to-point service between many airports have tempered these results. Simultaneously, many legacy carriers have dramatically increased their use of regional jets. This regional jet service has provided many small and medium-sized cities with convenient access to several different air-carrier hubs which previously were not directly accessible because of technological and/or economic constraints.

city "hometown" marketing advantages, local travel agents steering passengers to the dominant airline so that the agent can receive higher commissions[25], the ability to offer customers convenient flight schedules, the ability to offer customers in the hub cities more direct flights and (most importantly) the ability to induce customer loyalty by offering effective frequent flyer programs (Kanafani and Ghobrial 1985; Borenstein 1989; Berry 1990; Borenstein 1990; Dresner and Windle 1992).

Recent studies have linked air-carrier presence to airline choice using discrete choice analysis where the impact of presence is estimated relative to other carrier attributes. For example, using survey data, studies show indirect measures of air-carrier presence (such as perceived schedule convenience[26]) influence carrier choice probabilities (Nason 1981; Proussaloglou and Koppelman 1995; Proussaloglou and Koppelman 1999; Algers and Beser 2001). Additionally, direct measures of presence such as the number of nonstop frequencies to a travelers' destination (Ghobrial and Soliman 1992; Nako 1992; Yoo and Ashford 1996), airport dominance (Ghobrial and Soliman 1992; Nako 1992; Proussaloglou and Koppelman 1995; Proussaloglou and Koppelman 1999) and percentage of nonstop frequencies to a travelers' destination (Nako 1992; Proussaloglou and Koppelman 1995[27]) are shown to positively influence carrier choice.

---

[25] Historically, this was one reason dominant airlines could charge a premium. However, since the latest industry downturn magnified by the September 11th, 2001 terrorist attacks, commissions have been severely scaled back or eliminated.

[26] Presumably, this is a function of flight frequencies, number of destinations served, direct flight offerings, etc.

[27] Proussaloglou and Koppelman (1995) use a non-linear specification of airport-pair flight share.

Some of these discrete choice studies relate air-carrier presence with other characteristics via the use of innovative specifications. For example, Ghobrial and Soliman (1992) link the value of a carrier's service offerings with departure time by separately modeling the effects of peak and off-peak flight frequencies on airport-pair airline choice. All things being equal, they find that passengers view a flight during peak periods as 2.18 times more favorable than a flight during the off-peak periods. Nako (1992) finds that the effectiveness of a carrier's frequent flyer program in attracting passengers from an originating airport is dependent upon its market share at the airport. Additionally, he finds that business travelers are willing to pay approximately \$12 for an additional flight frequency to their destination.

Finally, carrier network characteristics such as average flight departures per airport, average market share per city and number of airports served influences carrier choice at the system level (Morash and Ozment 1996; Suzuki *et al.* 2001) and as was shown in Subsection 2.1.2, the level of air-carrier presence at airports in multi-airport regions strongly influences airport choice.

2.3.2 THE IMPACT OF FARES ON AIRLINE DEMAND

Fare is an important attribute influencing air-travel especially amongst leisure travelers. Unfortunately, due to the overall unavailability (to researchers) of industry-wide disaggregate fare data, detailed relationships between fares and demand are not well understood.

Using national aggregate measures, Brown and Watkins (1968) and Kanafani (1980) show that system-wide air-travel demand and fares are inversely related (all things being equal). Total city-pair demand is also negatively related to market fare levels (Brown and Watkins 1968; Brown and Watkins 1971; De Vany and Garges 1972; De Vany 1974; De Vany 1975; Jung and Fujii 1976; Ippolito 1981; Abrahams 1983; Dresner *et al.* 1996; Corsi *et al.* 1997).

Some studies have concluded that the (overall) demand for air-travel is price elastic (Brown and Watkins 1968; Brown and Watkins 1971; De Vany and Garges 1972; De Vany 1975; Jung and Fujii 1976) while others have found that the degree of price-elasticity depends on factors such as the response period (short vs. long-term elasticities), the distance of the trip or the type of traveler (business vs. leisure). For example, it is theorized that long-term price-elasticities of demand are larger in magnitude than short-term elasticities since the effects of price changes take time to be absorbed by the traveling public. Brown and Watkins (1968), Kanafani (1980) and Abrahams (1983) show this empirically.

The relationship between price-elasticities and trip distance has gotten contradictory results in the literature. One theory is that passengers on short-haul trips (who are usually business travelers) exhibit less sensitivity to fares than long-haul passengers since their values-of-time are larger and they may fly regardless of the fare level. A competing theory is that short-haul air-trips have more competition with other transportation modes; thus, it can be expected that the price-elasticities of demand for short trips would be greater than long trips. Abrahams (1983) shows evidence of the former theory while Douglas and Miller (1974a) and Jung and Fujii (1976) show evidence of the latter. Brown and Watkins (1968) find demand to be fare-inelastic for short and long-distance city-pairs but elastic for moderate distance city-pairs (these city-pairs have the highest competition with ground modes). Finally, Brown and Watkins (1971) show no evidence that the price elasticity of demand for air-travel depends on trip distance.

Brown and Watkins (1971) find that first-class passengers are not only sensitive to the price of first-class tickets but are also sensitive to the price of coach tickets (with a strong cross-elasticity of 1.3). Coach passengers, on the other hand, are only sensitive to the price of coach seats (with a direct elasticity of –1.3). Kanafani (1980) in his study of non-business air-travelers finds that leisure air-travelers are inelastic with respect to fares but highly elastic with respect to per-capita recreational expenditures (with a long-run recreational expenditures elasticity of 1.9). Finally, as expected, Abrahams (1983) shows leisure air-travel to be more price-elastic than business air-travel.

Other authors recognize that demand is not only a function of fares (out-of-pocket cost) but also time cost. De Vany (1974) derives a "full-price" elasticity of demand incorporating price and time costs. Using this measure, he finds city-pair air-travel demand is only slightly price-elastic using the traditional measure but much more elastic when incorporating time costs. Similarly, Anderson and Kraus (1981) find city-pair demand is inversely related to the origin city consumer price index plus time costs.

Most of the above-referenced studies use average fare paid at the city-pair (or even national) level or the average published fare to calculate elasticities. Generally, no consideration is given to differences in airline, fare-class, departure time, etc. Since accurate elasticities can be determined only by knowing the actual fares available to disaggregate customers and their corresponding choice of carrier, flight and fare-class, the conclusions of the above studies cannot be wholly trusted. In his convincing study, Verleger (1972) concludes that studies employing aggregate measures of fare do not yield proper demand functions and hence give incorrect price elasticity estimates. He finds air-travel demand to be much more income elastic than price elastic.

Recent studies using discrete choice analysis with stated preference (Nason 1981; Bruning and Rueda 2000), revealed preference (Ghobrial and Soliman 1992; Nako 1992; Proussaloglou and Koppelman 1995) or a blend of SP and RP (Yoo and Ashford 1996; Proussaloglou and Koppelman 1999; Algers and Beser 2001) data to model the carrier or flight choice behavior of passengers have also (not surprisingly) found that higher fare-levels negatively influence choice. In particular, Proussaloglou and Koppelman (1995, 1999) determine that leisure passengers are more sensitive to fares than business passengers. Bruning and Rueda (2000) find that fare is the most important attribute of flight choice, more important than the number of stops, frequent flyer programs and on-time performance[28]. Finally, Suzuki *et al.* (2001) find system-wide carrier share is inversely related to fare levels in a loss aversion fashion. Most of these discrete choice studies also suffer from the use of aggregate fare data. Thus, their conclusions with respect to the relative importance of fare (compared to other attributes) on airline demand are limited.

[28] This conclusion holds for their full sample. Some segments of the sample had different rankings for the importance of these attributes.

2.3.3 THE IMPACT OF ROUTE ON AIRLINE DEMAND

Air-travel route choice[29] models assign passengers to paths between airport-pairs. These passengers can then be "rolled up" to determine flows on the links in the network. Kanafani and Fan (1974), Kanafani and Ghobrial (1982) and Kanafani and Ghobrial (1985) embed a route choice model within a larger model of air-travel demand. Kanafani and Chang (1979) however, explicitly describe and develop a multinomial logit route choice model that assigns quarterly airport-pair demand to the paths between the airports. Aggregate independent variables are used such as average route travel time, frequency, fare, a dummy variable for whether the route is direct, and a dummy variable for whether a connecting route is online or offline. The authors find that a model employing only route time and frequency performs as well as models with more advanced specifications. Ndoh *et al.* (1990) also model route choice with frequency, average journey time, average connecting time (if a connection) and average number of seats supplied as independent variables. One of their findings is that passengers have a seven-fold preference for direct routings (flights) over indirect routings (flights).

Because these studies do not consider disaggregate itinerary-level attributes such as carrier, departure time, connection quality or equipment, they do not adequately explain air-traveler behavior. Modeling itinerary shares and then summing these forecasts up to the route (path) level is the best way to model route shares.

[29] Route choice is not the same as itinerary choice. Route choice (as used here) simply means the choice of a given path between airports (independent of departure time, equipment used, carrier, connecting time, etc.).

2.3.4 THE IMPACT OF THE NUMBER OF STOPS (CONNECTIONS) ON AIRLINE DEMAND

Few studies have modeled the impact of the number of stops on airline demand. The studies that have have used either aggregate measures (*e.g.* average number of stops between a given city-pair) or stated preference survey data. The "schedule convenience" survey measure employed by Nason (1981) and Proussaloglou and Koppelman (1995, 1999) indirectly measures the availability of nonstop flights to a passengers' destination. Bruning and Rueda (2000) show an increase in the number of stops negatively influences flight choice for all travelers, with business and/or frequent travelers rating the number of stops as the most important attribute when making their flight decisions. Nako (1992) calculates that business travelers on a connecting flight would be willing to pay approximately $56 for direct service.

Brown and Watkins (1968) show city-pair demand to be positively related to the "best" air-travel path between them. That is, all things being equal, city-pairs with nonstop service will have more demand than city-pairs with one-stop service as the best service, which in turn will have more demand than city-pairs with two-stop service as the best service[30]. They show that the effect on demand of going from no stops to one stop is approximately the same as the effect of going from one stop to two stops. Ghobrial and Soliman (1992) show that as the average number of stops a carrier has in a given market increases, its carrier choice probability decreases. Finally, in their route choice study,

[30] Of course, it is likely that city-pairs with nonstop service have the service because of high demand rather than the other way around.

Ndoh *et al.* (1990) show nonstop flights are significantly preferred over flights involving a connection.

2.3.5 THE IMPACT OF EQUIPMENT ON AIRLINE DEMAND

The impact of equipment type and size on airline demand has not gotten much attention in the literature. In his survey, Nason (1981) asked his respondents to gauge which airline (in general) offered the greatest number of wide-body flights, while Ghobrial and Soliman (1992) measure the average number of seats offered by a carrier in a city-pair. However, equipment size and type can only be accurately measured at the itinerary level.

2.3.6 THE IMPACT OF DEPARTURE TIME ON AIRLINE DEMAND

The impact of carrier time of day schedule offerings has also not been modeled extensively in the literature. The schedule convenience measure employed by some authors in their survey studies captures the extent to which airlines offer flight services near the desired departure times of their customers (Nason 1981; Proussaloglou and Koppelman 1995; Proussaloglou and Koppelman 1999; Algers and Beser 2001). However, use of these models requires a basis for determining departure time preferences. The specifications employed by these studies assume travelers regard a departure time that is earlier than their desired departure time just as negatively as a departure time that is later than their desired departure time. Proussaloglou and Koppelman (1999) demonstrate that business travelers are more sensitive to deviations from their desired departure times than leisure travelers. Finally, using revealed preference data, Ghobrial and Soliman (1992) include peak and off-peak flight-frequency in their model of city-pair airline choice.

2.3.7 THE IMPACT OF ELAPSED TRIP TIME ON AIRLINE DEMAND

The impact of elapsed travel time on air-travel demand is treated indirectly in several other sections of this literature review. For example, Subsection 2.3.2 mentions that travel time can be thought of as a cost affecting the demand for air transportation, Subsection 2.3.1 relates air-carrier schedules with the potential for frequency and/or stochastic delays, Subsection 2.1.2 mentions airport access time (which in that subsection is used to model airport choice in multi-airport regions), Subsection 2.3.4 relates travel time to the number of connections and Subsection 2.3.5 relates travel time to equipment flown. The following discussion examines the direct relationship between elapsed travel time (defined by flight times plus any connecting ground time) and demand.

Brown and Watkins (1968, 1971) and Kanafani and Fan (1974) show average elapsed travel time in a city-pair is inversely related to city-pair demand, while De Vany and Garges (1972) and De Vany (1974) find that travel time (represented by the shortest elapsed travel time in a city-pair) is inversely related to city-pair demand. Yoo and Ashford (1996) show elapsed time negatively influences flight choice. Additionally, they conclude from one of their model estimations that business travelers have a value of time of approximately $56 per hour, while leisure travelers have a value of time of approximately $48 per hour. Nako (1992) shows average airline travel times in a city-pair negatively influences carrier choice probabilities in that city-pair (he finds business travelers value time at an average of $25 per hour). Finally, Kanafani and Chang (1979) and Ndoh *et al.* (1990) use route choice models to show the elapsed time of a route negatively influences its probability of being chosen.

Because the ultimate air-traveler choice is itinerary choice, where each itinerary has an associated elapsed time (flight time plus ground time if a connecting itinerary), modeling itinerary demand shares as a function of itinerary elapsed time is the appropriate way to gauge the true relationship between demand and elapsed times. The studies mentioned above use measures (such as average elapsed travel time in a city-pair) whose level of aggregation does not allow for true micro-level relationships to be established.

2.3.8 THE IMPACT OF FREQUENT FLYER PROGRAMS ON AIRLINE DEMAND

American Airlines introduced the first frequent flyer (FF) program in 1981 as a marketing initiative to induce brand loyalty among its frequent business customers. Soon after, the vast majority of domestic airlines offered their own FF programs and the popularity of these programs skyrocketed. The basic idea of these programs is that passengers by traveling on a given airline accumulate mileage redeemable for free flights and/or upgrades at a later date. Today, tens of millions of people are members of at least one FF program.

Airlines introduced FF programs for several reasons including the desire to generate brand loyalty and repeat business from lucrative business travelers, to achieve product differentiation by taking advantage of their route networks and to obtain valuable demographic and travel-behavior data from their program members. However, these programs have gotten larger than the airlines ever anticipated, introducing a slew of financial issues for the airlines (*e.g.* revenue dilution, revenue passenger displacement, the liabilities of unused rewards), ethical issues[31] and legal issues such as whether FF programs are anti-competitive[32]. Discussing the above issues and related questions is beyond the scope of this research. However, several articles in the literature discuss the origins of these programs, their successes and failures, the profiles of FF members, the issues raised above and more (Toh and Hu 1988; Lefer 1988; Hu *et al.* 1988; Tretheway 1989; Kearney 1990; Ellis 1998).

The question of relevance for this dissertation is to determine what effect, if any, FF programs have on travelers' itinerary choice. It is expected that frequent flyer membership has a positive impact on airline choice since (all things being equal) a traveler will be inclined to choose an airline of which he/she is a frequent flyer member due to the promise of future benefits that can be earned by using that airline.

The literature reports mixed results concerning the effectiveness of FF programs. Toh and Hu (1988), Nako (1998) and Bruning and Rueda (2000) find that relative to other factors FF membership does not appear to be dominant, but rather has a marginal impact on carrier loyalty. The effectiveness of FF programs in inducing loyalty may be limited since many frequent flyers (especially those who fly most frequently) belong to

[31] For example, FF programs have created a principal-agent problem among frequent flyer employees and the employers who pay for their travel since the employees receive the benefits of FF programs. With these programs, there is a possible incentive for employees to book flights at higher prices than necessary or to take circuitous routings to receive benefits at the expense of their employer.

[32] The costs associated with setting up these programs and the fact that smaller airlines have to give more frequent and lucrative rewards through their programs (to compensate for their smaller route networks) may constitute a barrier to entry for potential new carriers. Additionally, it has been hypothesized that FF

more than one carriers' program (Toh and Hu 1988; Lefer 1988). Regardless, Bruning and Rueda (2000) find that FF programs are more important to frequent business travelers whose tickets are paid for by their employers, while Nako (1992) shows that the effectiveness of FF programs differs across airlines and depends on the presence the corresponding airlines have in the cities where their participating members reside.

In contrast to these studies, Proussaloglou and Koppelman (1995) find that FF membership and active participation in a FF program are the most important determinants of carrier choice relative to other carrier attributes. Proussaloglou and Koppelman (1999) confirm this result and estimate the premium that passengers are willing to pay to travel on a carrier of which they are frequent flyer members. Other authors have also recognized that air-travel passengers will pay a premium to travel on an airline of which they are frequent flyer members (Borenstein 1989; Berry 1990; Borenstein 1990; Dresner and Windle 1992).

Finally, it is hypothesized that since short distance air-travelers pay a higher per-mile fare than long distance travelers, they may not receive as much benefit from the typical FF program where benefits are accrued on the basis of mileage. Thus, from a customer preference and airline marketing standpoint, trip-based FF program schemes may be more attractive to travelers who typically travel short distances, while mileage-based FF programs may be more attractive to travelers who travel long distances (Suzuki

programs led the industry to consolidate in the mid 1980's since there was no economic or operational reasons for this consolidation.

2003). Finally, FF membership has also been shown to influence airport (and ultimately airline) choice (Suzuki *et al.* 2003).

2.3.9 THE IMPACT OF SERVICE QUALITY ON AIRLINE DEMAND

It is likely that air-carrier service quality has an impact on carrier choice. A carrier's service quality can refer to its on-time performance, baggage handling record, staff and crew friendliness, food quality, safety record, etc.

On-time performance (or perceived on-time performance) is the most common measure of service quality modeled in the literature (Nason 1981; Nako 1992; Proussaloglou and Koppelman 1995; Morash and Ozment 1996; Proussaloglou and Koppelman 1999; Bruning and Rueda 2000; Suzuki *et al.* 2001). All of these studies conclude that on-time performance improvements translate to increased carrier demand. Proussaloglou and Koppelman (1995) show that frequent travelers are more sensitive to on-time reliability than less-frequent travelers, and Suzuki (2000) develops a model showing that passengers who experienced flight delays on their last trip are more likely to switch airlines for their next trip than passengers who did not experience delays.

Other service quality attributes that have been measured with respect to their impact on demand include baggage handling record (Nason 1981; Morash and Ozment 1996; Suzuki *et al.* 2001), efficient check-in procedures (Nason 1981; Suzuki *et al.* 2001), staff friendliness (Nason 1981; Morash and Ozment 1996), food quality (Nason 1981; Suzuki *et al.* 2001), safety record (Nason 1981; Suzuki 2000; Suzuki *et al.* 2001), and flight cancellations or denied boardings (Morash and Ozment 1996).

Proussaloglou and Koppelman (1999) show that a general notion of perceived carrier quality of service positively influences carrier choice probabilities, while Suzuki *et al.* (2001) model system airline share as a function of service quality using a loss aversion framework.

2.4 AIRPORT-PAIR ITINERARY SHARE MODELING AND LOGIT SHARE TECHNIQUES FOR MODELING ITINERARY SHARE

2.4.1 MODELING AIRPORT-PAIR ITINERARY SHARE

None of the above-referenced aviation demand studies systematically model itinerary shares. As mentioned previously, understanding the relationship between some service attributes and airline demand is only possible by employing itinerary-level data. Until recently, this was a gap in the literature. However, in preparation for this dissertation, Coldren *et al.* (2003), Coldren and Koppelman (2005a) and Coldren and Koppelman (2005b) began the process of filling this gap by developing initial itinerary share models. These models link itinerary-level air-carrier attributes with airport-pair itinerary demand share. The research presented in this dissertation builds on that framework.

2.4.2 LOGIT SHARE TECHNIQUES FOR MODELING ITINERARY SHARE

This dissertation performs aggregate share analyses using advanced discrete choice model structures (in particular, GEV models (McFadden 1978)) to model the air-travel itinerary share problem. In addition to measuring the impacts of different service characteristics on itinerary share, the use of GEV models allows the underlying competitive relationships between itineraries to be captured.

The simplest and most widely used discrete choice model is the multinomial logit model derived by McFadden (1974). The MNL model has been used in many marketing and transportation applications. In particular, it is used in many of the above-mentioned air-carrier allocation studies (Nason 1981; Ghobrial and Soliman 1992; Nako 1992; Proussaloglou and Koppelman 1995; Yoo and Ashford 1996; Proussaloglou and Koppelman 1999; Algers and Beser 2001; Suzuki *et al.* 2001). However, the assumptions that underlie the MNL derivation produce choice-probability cross-elasticities that are equal across all other alternatives in response to a change in an attribute value of a single alternative. This limits its appropriateness for modeling itinerary shares[33].

McFadden (1978) introduced the generalized extreme value family of models allowing for much more modeling flexibility than the restrictive MNL (which itself is a member of the GEV family). The most widely used relaxation of the MNL model is the nested logit model. The NL model allows for the grouping of alternatives (with similar characteristics) into "nests". The estimation of this model quantifies the similarity of alternatives in a given nest, within which cross-elasticities are greater than between alternatives not in a common nest. For a given airport-pair, since it is likely that the itineraries of a given carrier (or departure time-period) share similar characteristics, the NL model may be relevant to itinerary share modeling. Koppelman and Wen (1998a, 1998b) detail the differences between the utility maximizing nested logit (UMNL) model developed at about the same time by McFadden (1978) and Williams (1977), and the non-normalized nested logit (NNNL) model widely attributed to Daly (1987). The latter model is not consistent with random utility maximization, produces unrealistic elasticity interpretations and may result in inconsistencies in the interpretation of parameters.

In addition to the NL model, GEV models which may be relevant to itinerary share modeling include the cross-nested logit (CNL) model (Vovsha 1997), the ordered generalized extreme value model (Small 1987), the principles of differentiation (PD) model (Bresnahan *et al.* 1997) and the generalized nested logit (GNL) model (Wen and Koppelman 2001) of which the MNL, NL, CNL, OGEV and PD models are special cases.

Properties of discrete choice models in general and specific GEV models in particular will be expanded upon in the remaining chapters of this dissertation. The use, description, and introduction of these models with respect to itinerary share modeling will also be detailed. However, for a very useful and comprehensive summary of GEV (and other closed-form) discrete choice models see Koppelman and Sethi (2000).

[33] This conclusion will be discussed extensively in Chapter 5.

CHAPTER 3: CONCEPTUAL AND MODELING FRAMEWORK

Discrete choice scenarios arise when decision-makers must choose from a set of mutually exclusive and collectively exhaustive alternatives. This chapter is divided into two sections: Section 3.1 briefly introduces discrete choice analysis and some of its techniques and Section 3.2 describes the discrete choice scenario presented in this dissertation, the data employed and aggregation issues arising from the use of this data.

3.1 INTRODUCTION TO DISCRETE CHOICE ANALYSIS AND TECHNIQUES

Following the framework established by Domencich and McFadden (1975), discrete choice scenarios are described by four elements: a decision-maker, the alternatives available to the decision-maker, attributes of these alternatives, and a decision rule.

The decision-maker is an individual person or group of persons (who make a common decision). The alternatives available to the decision-maker must be discrete, each of which has a vector of attributes. Finally, although there are numerous decision rules (*e.g.* dominance, satisfaction[34]), the decision rule commonly assumed is that decision-makers choose the alternative that maximizes their utility.

The analyst does not know the actual utility that decision-makers achieve from alternatives, however. Following convention, the utility of each alternative for a given decision maker is decomposed into two components: a deterministic component consisting of observable attributes of the alternative, and a random component representing unknown and/or unobservable components of the decision maker's utility.

[34] See Ben-Akiva and Lerman (1985) for definitions and examples of these and other decision rules.

Further following convention, the utility of alternative i, U_i, is expressed as the sum of the deterministic component, V_i, and the random component, ε_i:

$$U_i = V_i + \varepsilon_i \tag{3.1}$$

Since ε_i is a random variable the entire utility, U_i, is a random variable. V_i is generally assumed to be linear-in-parameters where:

$$V_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni} \tag{3.2}$$

That is, an alternative's deterministic value, V_i, is a linear function of explanatory variables (attributes, represented above by the $X_{ki}'s$) and their corresponding parameter estimates (represented above by the $\beta_k's$). The parameter estimates provide an understanding of the relative importance of different attributes on choice. In this formulation, decision-makers make tradeoffs among the attributes of the alternatives and are able to represent the attractiveness of each alternative as a scalar value. This entire formulation of alternative utility is referred to as linear in parameters with additive disturbance (LPAD) (Manski 1973).

Since it is assumed that a decision-maker selects the alternative from his/her choice set (denoted by C) that maximizes his/her utility, and the analyst treats utilities as random, it is necessary to calculate the probability that a decision-maker will choose a given alternative. The probability that a decision-maker chooses alternative i from the choice set C with J alternatives, $\Pr ob(i:C)$, is given by:

$$\begin{aligned} Prob(i:C) &= Prob(U_i \geq U_j \ \ \forall j) = \\ & Prob(V_i + \varepsilon_i \geq V_j + \varepsilon_j \ \ \forall j) = Prob(\varepsilon_j - \varepsilon_i \leq V_i - V_j \ \ \forall j) \end{aligned} \tag{3.3}$$

This is a multivariate cumulative distribution function for the differences between pairs of error terms.

Specific assumptions on the distribution of the ε's lead to different model structures. For example, the assumption that the error terms are independent and identically gumbel distributed across alternatives and decision-makers leads to the familiar multinomial logit model developed by McFadden (1974). However, to accurately model many choice situations (including the scenario in this research), it is necessary to relax the assumption of error independence across alternatives. The generalized extreme value family of models proposed by McFadden (1978) allow for this. These models are closed-form, are consistent with random utility maximization theory and assume that alternative error terms have a gumbel distribution with equal variance across alternatives (but not necessarily independently distributed across alternatives). GEV models can be generated from any function G, with arguments Y_i for each alternative i:

$$G(Y_1, Y_2, ..., Y_J) \quad s.t. \; Y_1, Y_2, ..., Y_J \geq 0 \tag{3.4}$$

The above G function must be non-negative, be homogeneous, approach infinity as each of its arguments approach infinity and have odd (even) order partial derivatives that are non-negative (non-positive) (McFadden 1978; Koppelman and Sethi 2000). With a GEV model it can be shown that the probability of an alternative (assuming the transformation $Y_i = \exp(V_i)$ to ensure positive Y_i), P_i, is given by:

$$P_i = \frac{e^{V_i} G_i\left(e^{V_1}, e^{V_2}, ..., e^{V_J}\right)}{G\left(e^{V_1}, e^{V_2}, ..., e^{V_J}\right)} \tag{3.5}$$

where $G_i(\;)$ is the first derivative of G with respect to Y_i and V_i is the deterministic component of the utility of alternative i.

The parameters for a given GEV discrete choice model can be estimated using full-information maximum likelihood on a random or choice-based data sample. This yields parameter estimates that maximize the likelihood (posterior probability) of the observed sample choices conditional on the model. The likelihood function, L, from a sample of discrete choice data is given by:

$$L = \prod_t \prod_j P_{jt}^{\delta_{jt}} \tag{3.6}$$

where the outer product is over all individuals in the sample and the inner product is over all alternatives in each individual's choice set. P_{jt} is the probability that individual t chose alternative j and $\delta_{jt} = 1$ if individual t chose alternative j ($\delta_{jt} = 0$ otherwise). For computational reasons (and because maximizing the log of a function achieves the same solution as maximizing the function itself), the log of the likelihood function is usually maximized in discrete choice applications. The log-likelihood function from a sample of discrete choice data is given by:

$$\ln(L) = \sum_t \sum_j \delta_{jt} \ln(P_{jt}) \tag{3.7}$$

The parameter estimates that maximize this log-likelihood function are obtained when the first derivative of the function with respect to each parameter is zero.

3.2 DISCRETE CHOICE SCENARIO OF RESEARCH PROBLEM, DATA USED AND DATA AGGREGATION ISSUES

The person or group who books an air-travel itinerary is the decision-maker in the modeling framework of this study. The alternative sets are modeled as all itineraries linking directional airport-pairs for each day of the week (since for a given day of the week and airport-pair the itineraries linking the airports are the mutually excusive and collectively exhaustive choices facing air-travelers). For example, all Monday itineraries from Boston to Los Angeles constitute an alternative set, as do all Friday itineraries from Allentown to Tucson. Each itinerary in each alternative set is described by a set of attributes such as level-of-service, fare and departure time. Finally, it is assumed that the air-travel passengers being modeled choose itineraries with the intent of maximizing their utility.

Passenger bookings data for this research was obtained from a compilation of computer reservation systems (CRS). These are data sources containing detailed records of booked itineraries. CRS data is commercially available and compiled from several computer reservation systems including Sabre (Sabre Travel Network 2000, 2001), Galileo (Galileo International 2000, 2001) and Worldspan (Worldspan, L.P. 2000, 2001) as well as Internet travel sites such as Orbitz (Orbitz, L.L.C. 2000, 2001), Travelocity (Travelocity.com, L.P. 2000, 2001), Expedia (Expedia, Inc. 2000, 2001) and Priceline (Priceline.com, Inc. 2000, 2001). The CRS data is believed to include 90% of all domestic air-travel bookings[35] during the study period. However, increasing use of direct-

---

[35] It is assumed that the small number of bookings not captured by the CRS source have the same characteristics as the bookings contained in CRS. This assumption may be questionable (and hence the

carrier bookings (via telephone and Internet) has since reduced the proportion of bookings reported by this source. This is likely to be an ongoing trend due to reduced travel agent commissions and reduced price differentials between "discount" Internet sites and the carriers' own websites. The CRS data contains variables describing each booked itinerary such as trip origin and destination, leg(s) origin and destination, leg number, party size, flight number(s), leg(s) departure and arrival time(s), departure and arrival date, and airline.

Since the itinerary building rules employed by the different computer reservation systems differ to a limited degree, a major carrier's itinerary building engine was used to generate the set of itineraries between all airport-pairs. This engine builds all feasible online itineraries (up to the double-connection level) using leg-based air-carrier schedule data obtained from the Official Airline Guide (OAG Worldwide Limited 2000, 2001). The OAG data contains the following variables: operating airline, codeshare airline (if a codeshare leg), origin, destination, flight number, departure and arrival time, equipment, days of operation, leg mileage and flight time. In building the itineraries, distance-based circuitry logic is used to eliminate unreasonable itineraries, and minimum and maximum connection times are incorporated to ensure that unrealistic connections are not allowed. Finally, itineraries are generated for each day of the week accounting for day-of-week differences in service offered.

---

presented model results may be slightly biased) since it is likely that direct-carrier bookings contain more leisure bookings than the entire population of bookings.

Because no disaggregate information is available from the individual bookings, individuals and groups who booked a given itinerary (for a given day-of-the-week) are treated as homogeneous and are summed over an entire month (to get a total passenger count) for each itinerary in the bookings data. The generated itineraries are merged with the booked itineraries to assemble the estimation datasets for this research. Because of the aggregate nature of the data, the log-likelihood function maximized in this research incorporates the frequency of the chance of each alternative. That is,

$$\ln(L) = \sum_{a}\sum_{i} Y_{ia} \ln(P_{ia}) \tag{3.8}$$

where the outer summation is over all alternative sets in the estimation data (i.e. all airport-pair-day-of-the-week combinations), the inner summation is over all itineraries in each alternative set, Y_{ia} is the total number of passengers who chose itinerary i in alternative set a and P_{ia} is the estimated probability of each passenger choosing itinerary i in alternative set a. Finally, using the estimation data, aggregate GEV models are estimated with full-information maximum likelihood techniques using the GAUSS modeling software (Aptech Systems, Inc. 2005). The specific forms of the different models tested are described in Chapters 4 – 6.

Even though the bookings data employed in this study is based on the choices of individual travelers, it does not include any information on the demographic characteristics of the individual that made the booking or any trip-related characteristics of the booking (such as income, business vs. leisure, number of days booked in advance of departure, duration of stay). Thus, since no individual data is available to identify differences among travelers, it may not be appropriate to count the full weight of the individual observations in calculating the statistics for the models. The most extreme adjustment can be accomplished by dividing the log-likelihood values for the models by the ratio of the number of booked passengers to the number of airport-pair, day-of-the-week combinations; and the parameter estimate t-statistics by the square root of this ratio. Intermediate adjustments that take account of the fact that individual choice behavior is observed can be justified as well. Statistics discussed in Chapters 5 – 6 and presented in Tables 5.4 – 5.5 and 6.1 – 6.2 refer to the unadjusted values as well as the values obtained after adjusting (according to the ratio described above) due to the aggregate nature of the data.

CHAPTER 4: MODELING THE IMPACT OF SERVICE CHARACTERISTICS ON AIR-TRAVEL ITINERARY SHARE USING THE MULTINOMIAL LOGIT MODEL

In this chapter, the quantitative relationship between air-travel service characteristics and itinerary share is determined via estimation of aggregate multinomial logit models.

4.1 INTRODUCTION

An important motivation for developing itinerary share models is to gain an understanding of the impact of different service factors on itinerary share. Researchers, policy-makers and air-carriers can use this information. As one example, information about passenger departure time preferences can be used by policy makers to better allocate slots at slot-controlled airports, while carriers can take advantage of this information to better match their schedule offerings with customer departure time preferences.

In this chapter, aggregate MNL models are developed to model itinerary shares for each domestic[36] airport-pair in the United States. The itinerary attributes modeled are level-of-service, connection quality, carrier attributes, aircraft size and type, and departure time. Factors not directly modeled – though discussed – include elapsed itinerary travel time, carrier frequent flyer program effectiveness and service quality.

4.2 MODELING FRAMEWORK

The United States was divided into five regions: each Continental time zone (East, Central, Mountain, West) and a region for Alaska and Hawaii. Using these regions, eighteen "entities" were defined: all sixteen combinations of the Continental time zones (East-East (E-E), East-Central (E-C), East-Mountain (E-M), East-West (E-W), . . . , West-West (W-W)), as well as an entity for Alaska and Hawaii to the Continental U.S. and an entity for the Continental U.S. to Alaska and Hawaii[37].

Using January 2000 data, aggregate multinomial logit models were estimated for each of these entities with a common specification using data from all airport-pairs in the respective entity. Under MNL model assumptions, it can be shown that the market share of passengers assigned to each itinerary between an airport-pair for a given day of the week is given by the following equation:

$$S_i = \frac{\exp(V_i)}{\sum_{j \in J} \exp(V_j)} \tag{4.1}$$

where S_i is the passenger share assigned to itinerary i,
$\exp(\ )$ is the exponential function,
V_i is the value (deterministic portion of utility) of itinerary i and
the summation is over all itineraries for the airport-pair-day-of-the week.

The independent variables used in these models are described in Table 4.1.

---

[36] Mexico-to-U.S. (and vice versa) and Canada-to-U.S. (and vice versa) airport-pairs are also included in the analysis.

[37] Airport-pairs within Hawaii or Alaska, and between Hawaii and Alaska were excluded from the analysis.

TABLE 4.1: Description of Explanatory Variables

Variable	Description
Level-of-Service	Dummy variable representing the level-of-service of the itinerary (nonstop, direct, single-connect or double-connect) with respect to the best level-of-service available in the airport-pair.
Second-Best Connection	For connection itineraries sharing a common leg, a dummy variable indicating that the itinerary is not the best connection (with respect to ground time) for the given incoming or outgoing leg at a transfer airport.
Second-Best Connection Time Difference	If the second-best connection indicator equals one, this variable measures the ground time difference between the itinerary and the best connection itinerary.
Best Connection Time Difference	Elapsed time difference between an itinerary involving a stop or connection and the fastest itinerary involving a stop or connection for each airport-pair independent of transfer airport.
Distance Ratio	Itinerary distance divided by the shortest itinerary distance for the airport-pair multiplied by 100.
Point of Sale Weighted Airport Presence	Carrier origin and destination presence (determined by percentage of departures) weighted by industry airport-pair point-of-sale percentages divided by 100 to give units of percent from 0 to 100.
Fare Ratio	Carrier average fare divided by the industry average fare for the airport-pair multiplied by 100.
Carrier	Dummy variable representing carriers having more than 0.5% of itineraries in the entity. All other carriers are combined together in a single category.
Code share	Dummy variable indicating whether any leg of the itinerary was booked as a code share.
Regional Jet	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a regional jet.
Propeller Aircraft	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a propeller aircraft.
Mainline Jet Seats	If an itinerary involves neither a regional jet nor a propeller aircraft leg, this variable measures the number of seats on the smallest aircraft for the itinerary.
Regional Jet Seats	If an itinerary includes a regional jet leg (but no propeller aircraft leg), this variable measures the number of seats on the smallest regional jet aircraft for the itinerary.
Propeller Aircraft Seats	If an itinerary includes a propeller aircraft leg, this variable measures the number of seats on the smallest propeller aircraft for the itinerary.
Departure Time	Dummy variable for each hour of the day (based on the local departure time of the first leg of the itinerary).

Estimation results for five entities (the entities originating in the Eastern time zone and the West-East entity) are reported in this chapter[38]. Table 4.2 reports some summary statistics for these entities and shows the scope of the data used in the estimations. The passenger number is the number of people who booked an itinerary to travel in the entity during the month. The especially large numbers for the East-East entity reflect the busy East Coast travel corridor.

TABLE 4.2: Summary Statistics for Five Entities

Entity	Airport-Pairs	Itineraries	Booked Passengers
East-West	2,294	559,073	1,128,916
East-Mountain	1,809	276,032	419,392
East-Central	6,755	1,034,861	2,002,554
East-East	11,147	1,669,354	5,861,742
West-East	2,293	476,664	1,143,671

A more intuitive description of the data is obtained by summarizing the itineraries in an alternative set along different service characteristic dimensions. Tables 4.3 - 4.5 show the distribution of itineraries and passengers for Monday itineraries linking Boston and Los Angeles for January 2000 according to level-of-service, carrier and departure time, respectively. Table 4.3 shows that, as expected, the nonstop itineraries in this market captured the majority of passengers. Direct itineraries attracted a small but not insignificant number of passengers. Single-connection itineraries captured a modest number of passengers in total but very few passengers per itinerary, and no passengers booked a double-connection itinerary. This is not surprising since both nonstop and direct service was available in this market. Table 4.4 reports that American Airlines and United Airlines captured the majority of passengers for this alternative set. Not coincidentally, these are the only carriers who offered nonstop service in this market (four nonstop flights each). Other airlines carried a much smaller number of passengers via connections through their respective hubs. Finally, Table 4.5 shows the distribution of passengers and itineraries by departure time. Note the large fraction of passengers who boarded during a one-hour morning and a one-hour afternoon peak period.

TABLE 4.3: Monday Boston – Los Angeles Itineraries by Level-of-Service

Level-of-Service	Itineraries	Booked Passengers	Average Number of Booked Passengers per Itinerary
Nonstop	8	2,009	251.1
Direct	5	104	20.8
Single-Connect	249	402	1.6
Double-Connect	61	0	0.0
Total	**323**	**2,515**	**7.8**

[38] Similar results were obtained for the other thirteen entities.

TABLE 4.4: Monday Boston – Los Angeles Itineraries by Carrier

Carrier	Itineraries	Booked Passengers
American Airlines	62	1,210
United Airlines	75	865
U.S. Airways	24	102
Trans World Airlines	14	70
Continental Airlines	30	69
America West Airlines	15	67
Delta Airlines	48	53
Northwest Airlines	22	39
Air Canada	4	0
Other	29	40
Total	**323**	**2,515**

TABLE 4.5: Monday Boston – Los Angeles Itineraries by Departure Time

Departure Time	Itineraries	Booked Passengers
Midnight-5 A.M.	0	0
5-6 A.M.	5	9
6-7 A.M.	30	18
7-8 A.M.	29	67
8-9 A.M.	30	744
9-10 A.M.	17	50
10-11 A.M.	24	13
11-12 Noon	29	410
12-1 P.M.	17	59
1-2 P.M.	30	237
2-3 P.M.	21	27
3-4 P.M.	31	292
4-5 P.M.	16	1
5-6 P.M.	26	544
6-7 P.M.	8	4
7-8 P.M.	8	20
8-9 P.M.	1	2
9-10 P.M.	1	18
10-12 Midnight	0	0
Total	**323**	**2,515**

4.3 THE IMPACT OF SERVICE CHARACTERISTICS ON ITINERARY SHARE

The parameter estimates for the five models are reported in Table 4.6[39]. These estimates can be interpreted in terms of the impact of a variable value change on the value of the alternative and its market share. Further, each model can be evaluated in terms of the implied importance of its variables and its statistical significance.

4.3.1 THE IMPACT OF LEVEL-OF-SERVICE ON ITINERARY SHARE

The estimation results presented here provide strong evidence of the importance of level-of-service on itinerary share. All parameter estimates for level-of-service, relative to the best level-of-service for the airport-pair[40], are large, negative and significant. The magnitude of preference among levels-of-service indicates that each reduction in level-of-service from the best available substantially reduces the value of the associated itinerary. These results provide evidence that passengers strongly prefer to avoid connections presumably due to their increased elapsed travel time, the inconvenience of switching planes, higher probability of delays, increased potential for lost baggage, etc.

The magnitude and relative differences among the level-of-service parameter estimates is very similar across these five entities. This result supports the idea that air passengers are similarly sensitive to differences in level-of-service across the entire domestic system.

---

[39] All parameter estimates in this table significant at the 0.05 level.

[40] Each airport-pair has a best level-of-service. For example, Boston – Los Angeles has nonstop service. Therefore, it is a "nonstop market". However, the best level-of-service linking many airport-pairs is a single or even double connection.

TABLE 4.6: Itinerary Share Models

Explanatory Variables	E-E	E-C	Entity E-M	E-W	W-E
Level-of-Service					
Nonstop Itinerary in Nonstop Market[41]	0.0000	0.0000	0.0000	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.6819	-1.5253	-1.7168	-1.6615	-1.5349
Single-Connect Itinerary in Nonstop Market	-3.1213	-2.8760	-2.7087	-2.9729	-2.8858
Double-Connect Itinerary in Nonstop Market	-7.7557	-7.2970	-7.5112	-7.2130	-6.5395
Direct Itinerary in Direct Market	0.0000	0.0000	0.0000	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.7901	-0.7768	-1.0989	-1.0615	-0.9847
Double-Connect Itinerary in Direct Market	-4.9442	-4.5565	-4.3308	-4.4475	-4.5869
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000	0.0000	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-3.0953	-3.1205	-2.5431	-2.8209	-2.7105
Connection Quality					
Second-Best Connection	-0.5560	-0.5058	-0.5322	-0.7269	-0.4377
Second-Best Connection Time Difference	-0.0157	-0.0190	-0.0178	-0.0162	-0.0180
Best Connection Time Difference	-0.0108	-0.0094	-0.0093	-0.0104	-0.0088
Distance Ratio	-0.0125	-0.0116	-0.0210	-0.0173	-0.0207
Carrier Attributes					
Point of Sale Weighted Airport Presence	0.0024	0.0100	0.0078	0.0071	0.0077
Fare Ratio	-0.0018	-0.0038	-0.0045	-0.0035	-0.0040
Carrier Constants	--------	--------	--------	--------	--------
Code share	-1.5911	-2.1255	-1.9220	-2.1658	-2.2082
Aircraft Size and Type					
Mainline Jet	0.0000	0.0000	0.0000	0.0000	0.0000
Regional Jet	-0.7046	-0.8317	-0.8844	-0.7079	-0.4110
Propeller Aircraft	-1.2420	-1.0599	-0.8081	-0.9988	-0.9269
Mainline Jet Seats	0.0041	0.0037	0.0047	0.0032	0.0036
Regional Jet Seats	0.0117	0.0101	0.0144	0.0080	0.0052
Propeller Aircraft Seats	0.0246	0.0146	0.0157	0.0184	0.0212
Departure Time					
Midnight - 5 A.M.	-1.1653	-1.1577	-1.2341	-1.2037	-0.7228
5 – 6 A.M.	-0.4653	-0.4577	-0.5341	-0.5037	0.1482
6 – 7 A.M.	0.0000	0.0000	0.0000	0.0000	0.0000
7 – 8 A.M.	0.2865	0.2693	0.2179	0.2297	0.0570
8 – 9 A.M.	0.2836	0.4130	0.3983	0.3168	0.2354
9 –10 A.M.	0.2046	0.3204	0.4156	0.3413	0.0311
10 - 11 A.M.	0.1219	0.2571	0.4539	0.3425	0.0592
11 - 12 noon	0.1022	0.2539	0.3842	0.2670	0.0196
12 – 1 P.M.	0.1452	0.3137	0.4252	0.2977	0.0570
1 – 2 P.M.	0.1732	0.3676	0.2298	0.2594	0.0219
2 – 3 P.M.	0.2622	0.4742	0.1964	0.2462	-0.0988
3 – 4 P.M.	0.3438	0.5254	0.2136	0.2557	0.1175
4 – 5 P.M.	0.4230	0.5853	0.2476	0.2097	0.0873
5 – 6 P.M.	0.4667	0.5959	0.2419	0.2324	0.1779
6 – 7 P.M.	0.4054	0.5516	0.2491	0.1889	-0.2437
7 – 8 P.M.	0.2047	0.4627	-0.0099	0.0472	0.2914
8 – 9 P.M.	-0.0362	0.0866	-0.0984	-0.2243	0.1417
9 – 10 P.M.	-0.3565	-0.2917	-0.6878	-0.3924	0.0541
10 – Midnight	-0.6468	-0.7454	-0.7637	-0.3170	-0.0573
Rho-square w.r.t. Zero	0.377	0.369	0.372	0.396	0.385

[41] "Nonstop Market" means the "best" level-of-service available in the airport-pair is a nonstop itinerary, "Direct Market" means the best level-of-service available in the airport-pair is a direct itinerary, etc.

4.3.2 THE IMPACT OF CONNECTION QUALITY ON ITINERARY SHARE

Several variables in the models measure the relative passenger preference for and quality of itineraries involving connections (in addition to the level-of-service variables). The results indicate that travelers making a connection strongly prefer the best connection (shortest ground time) among itineraries sharing a common leg at a transfer station. The models include two variables to measure the impact of second-best connection itineraries: the "second-best connection" and "second-best connection time difference" variables. Figures 4.1 and 4.2 show visual representations of two types of second-best connections. Both figures show Boston to Los Angeles single-connect itineraries with ORD (Chicago) as the connecting station and time represented along the vertical axis. The #123 – #789 connection is a "second-best connection" in both figures.

In addition, passengers compare all connections linking an airport-pair (possibly involving different connecting airports) with respect to elapsed time (flight plus ground time) differences. The "best connection time difference" variable captures passenger aversion to both excess ground and flight time. The magnitude of this parameter is approximately 1/2 to 2/3 of the "second-best connection time difference" parameter estimate suggesting that the comparison between connecting itineraries sharing a common leg carries more importance than the comparison across different routings.

A related comparison of itinerary connection quality is the ratio of each itinerary's distance to that of the shortest itinerary distance for the airport-pair ("distance ratio" variable). The parameter estimates for this variable were negative in all of the models indicating that circuitous itineraries have lower value than more direct itineraries.

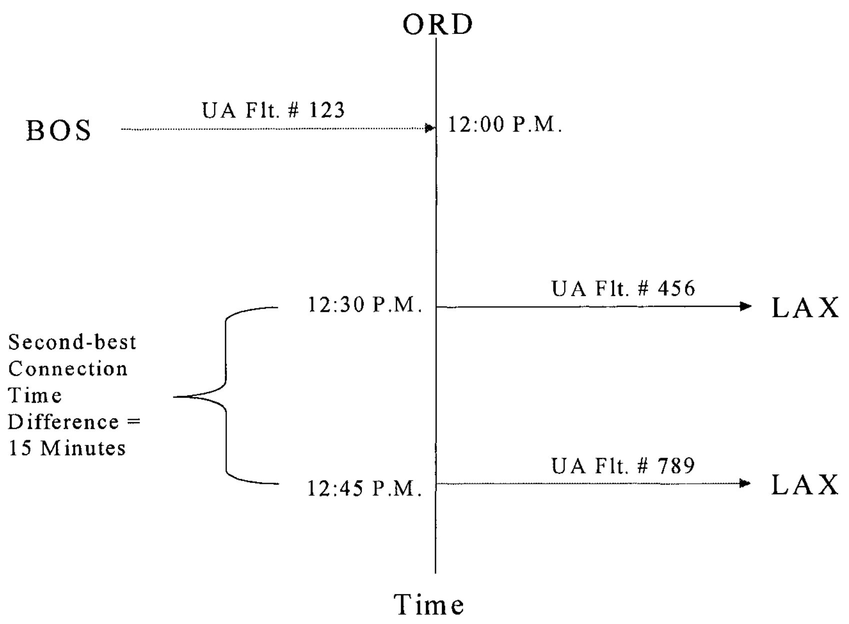

FIGURE 4.1: Visual Representation of Second-best Connection and Second-best Connection Time Difference

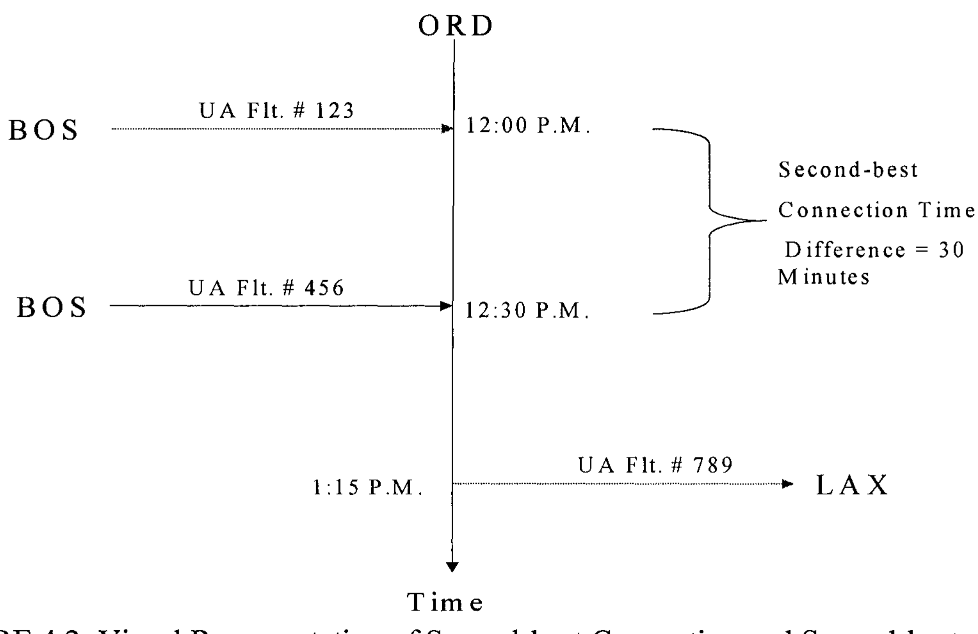

FIGURE 4.2: Visual Representation of Second-best Connection and Second-best Connection Time Difference

4.3.3 IMPACT OF CARRIER PRESENCE, FARES AND CARRIER ON ITINERARY SHARE

Carrier market presence can be represented as the amount of service provided in the airport-pair (nonstop frequency for example) or as a function of the carrier's overall presence in the origin and/or destination airports (percentage of departures for example). Several functional forms of carrier presence were experimented with in the models. The specification yielding the most reasonable results was the "point of sale weighted airport presence" variable included in the present models. The parameter estimates for this variable are positive matching *a priori* expectations since increased presence at an airport allows a carrier to offer passenger's greater frequencies and more destinations. Additionally, increased carrier presence provides marketing opportunities and enables carriers to attract more frequent flyer customers.

Itineraries with higher fares have a lower value than lower-cost itineraries (other things being equal). Fare data for models estimated in this dissertation was obtained from the "Superset" data source (Data Base Products, Inc. 2000, 2001). The parameter estimates for the "fare ratio" variable in the models are negative as expected. Though the fare data employed is superior to other revealed preference aviation demand studies, it is important to recognize that the data is based on averages for each carrier across all itineraries for each airport-pair. Detailed fare data for specific itineraries and ticket classes was not available for this research.

Carrier (represented by dummy variables in the models) is an important component of itinerary choice. These variables capture the many influences that a carrier's image has on itinerary choice such as marketing effectiveness, frequent flyer program quality, food quality, safety record, on-time performance, propensity to overbook, baggage handling record, flight crew friendliness, etc. These parameter estimates are all highly significant and (for a given carrier) varied greatly across the entities reflecting differences in a carrier's quality of service and image in different entities. The carrier-specific constants have been suppressed from Table 4.6 for proprietary reasons. However, exclusion of these parameter estimates does not impact the behavioral interpretation of the models.

4.3.4 IMPACT OF AIRCRAFT SIZE AND TYPE ON ITINERARY SHARE

Clearly, the type(s) of aircraft used on an itinerary and the number of seats offered on these aircraft types affects the value of the itinerary. The estimation results confirm the belief that passengers prefer large (mainline) jets to regional jets to propeller aircraft due to their increased speed, more comfortable cabins and higher perceived levels of safety. Also, passengers prefer larger aircraft to smaller aircraft (within an aircraft type). Figure 4.3 shows the contribution of aircraft type and number of seats to itinerary value for the East-East entity displaying typical aircraft sizes of 15-40 for propeller aircraft, 50-80 for regional jet aircraft and 100+ for mainline aircraft. Value is determined by applying the appropriate equipment variable and the corresponding seats variable. Passenger sensitivity to an increase in aircraft size is particularly strong for smaller aircraft.

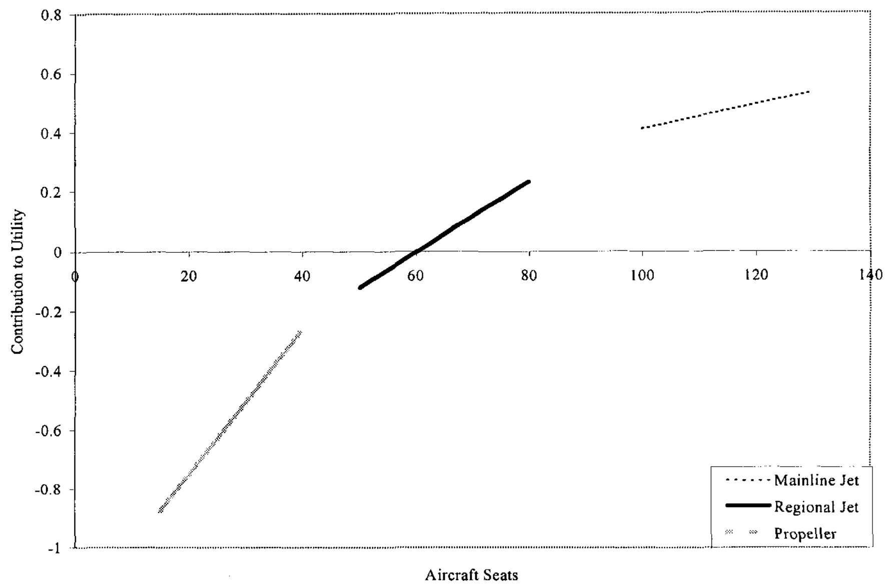

FIGURE 4.3: Contribution of Aircraft Size and Type to Itinerary Value for the East-East Entity

4.3.5 IMPACT OF DEPARTURE TIME ON ITINERARY SHARE

The five entities displayed in Table 4.6 show distinctly different preference patterns for departure time. However, in all cases early morning (before 7 A.M.) and late evening (after 8 P.M.) itineraries are generally not preferred. Complete interpretation of these parameter values depends on the flight schedules supplied in each entity. A negative parameter indicates that the time period is not preferred after taking account of less frequent flight offerings during the period. These departure time-of-day preferences reflect different origin and destination constraints across entities. For example, itineraries departing after 8 P.M. have a higher preference in the West-East entity than the westbound entities. This is a result of these travelers taking advantage of “red eye” itineraries allowing them to arrive on the East Coast early the following morning.

4.3.6 IMPACT OF ELAPSED TIME, FREQUENT FLYER PROGRAM EFFECTIVENESS AND SERVICE QUALITY ON ITINERARY SHARE

Elapsed itinerary trip time was not included in the models presented in this dissertation because it did not yield significant parameter estimates. This is due, primarily, to the other variables in the model specification that indirectly measure elapsed time. These include the level-of-service indicators, the connection quality variables (second-best connection variables, best connection time difference, distance ratio) and the aircraft size and type variables.

Due to data unavailability, the impact of frequent flyer programs and service quality (*e.g.* on-time performance, baggage handling record, overbooking rates) on itinerary share were not explicitly modeled in this research. However, it is believed that the carrier constants included in the presented models capture many of the service quality effects as well as the attractiveness of a given carriers frequent flyer program.

4.4 VALIDATION

The full set of eighteen logit-based models was implemented as a component of a major U.S. carrier's itinerary share model. For a given month, validation was undertaken by allocating demand to itineraries between each airport-pair and assigning the itinerary-level demand to the flight-segments that define each itinerary. For each of the carrier's segments the total number of forecasted passengers was compared to onboard passenger count data. Absolute errors were averaged across segments and compared to predictions using the carrier's previous QSI-based itinerary share model for the twelve months of 1999, August 2001, March 2002 and May 2002.

The results, reported in Table 4.7, are consistently superior for the logit-based models for every month based on model estimation from a single month (January 2000). This suggests very good model stability from month to month. Note that similar results were obtained when validation was applied to the selected months in 2002; that is, even including months after the September 11th, 2001 terrorist attacks.

TABLE 4.7: Onboard Segment-Level Model Validation Analysis (Mean Absolute Percentage Deviation)

Month	QSI-Based Model	Logit-Based Model	Difference
January 1999	19.52	17.91	1.61
February 1999	18.98	17.21	1.77
March 1999	18.40	16.48	1.92
April 1999	17.31	15.94	1.37
May 1999	17.22	16.23	0.99
June 1999	18.33	16.87	1.46
July 1999	16.70	14.91	1.79
August 1999	16.41	14.57	1.84
September 1999	17.77	15.93	1.84
October 1999	18.13	15.66	2.47
November 1999	16.14	14.42	1.72
December 1999	19.28	17.73	1.55
August 2001	17.47	15.79	1.68
March 2002	17.91	16.14	1.77
May 2002	19.08	17.73	1.35
Average	**17.91**	**16.23**	**1.68**

4.5 SUMMARY AND CONCLUSIONS

This chapter studies the influence of various service attributes on itinerary share. Eighteen aggregate multinomial logit models are developed, each covering a major region-to-region (or intra-region) of the United States. The development of these models provides a better understanding of the relative importance of different service factors on itinerary share compared to previous studies.

Detailed analysis is performed on the relative importance of itinerary level-of-service, connection quality, aircraft type and size, departure time, carrier presence, fares and carrier on itinerary share. The parameter estimates for all of these itinerary characteristics have an intuitive interpretation and are consistent over all eighteen entities.

Validation results using a major U.S. carrier's onboard segment-level data consistently favor the logit-based models relative to previously used QSI-based models.

Additionally, itinerary service attributes controllable by a carrier to improve market share are clearly identified. Among these, the most important are the provision of higher levels of service (more nonstop and direct itineraries) suggesting the potential market value of moving away from the dominant hub-and-spoke network structure of the major legacy carriers, flying mainline jets, matching service offerings with customer departure time preferences, optimizing connection times and maintaining a substantial presence in primary markets.

CHAPTER 5: MODELING THE UNDERLYING COMPETITIVE DYNAMIC AMONG AIR-TRAVEL ITINERARIES WITH NESTED LOGIT MODELS

5.1 INTRODUCTION

Chapter 4 modeled itinerary shares using aggregate multinomial logit functions of the itineraries' attributes. MNL models are adequate for describing the impact of service attributes on airport-pair itinerary share. However, the across-itinerary independence of the itinerary error terms (inherent in the derivation of the MNL model structure) implies that all itineraries "compete" equally with each other for a given airport-pair-day-of-the-week. That is, the underlying competition among air-travel itineraries is assumed to be "uniform" when modeled with a MNL function. This is further demonstrated by a well-known property of MNL models, the independence of irrelevant alternatives (IIA) property, which states that the relative probabilities of any pair of itineraries are independent of the attributes (or even the presence) of any other itineraries. This can be seen mathematically by looking at the ratio of MNL probabilities for two itineraries, i and j:

$$\frac{P_i}{P_j} = \frac{\dfrac{\exp(V_i)}{\sum_{k \in K} \exp(V_k)}}{\dfrac{\exp(V_j)}{\sum_{k \in K} \exp(V_k)}} = \frac{\exp(V_i)}{\exp(V_j)} \tag{5.1}$$

which is only a function of the itineraries i and j. The IIA property of the MNL model is also apparent by examining the cross-elasticity equation for the change in the probability of itinerary j due to changes in an attribute of itinerary i:

$$\eta_{X_{ik}}^{P_j} = \frac{\partial P_j}{\partial X_{ik}} \frac{X_{ik}}{P_j} = -P_i X_{ik} \beta_k \tag{5.2}$$

where X_{ik} is the value of itinerary $i's$ k^{th} attribute and β_k is attribute $k's$ parameter estimate. Note that the expression on the right side is not a function of j. That is, changing an attribute of itinerary i affects all other itineraries in the same proportion.

The central hypothesis of this chapter is that the underlying competition among air-travel itineraries for a given airport-pair and day-of-the-week is not uniform. Rather, it is asserted that groups of itineraries sharing one or more common attributes will exhibit more competition (as measured by cross-elasticities) amongst themselves than with itineraries not sharing these attributes.

Generalized extreme value models allow for the possibility of correlation between error terms for groups of alternatives (McFadden 1978; Koppelman and Sethi 2000) and permit the estimation of differential alternative competition measurements simultaneously with the value function parameters. This allows for flexible and complicated inter-alternative competition dynamics to be modeled. Using aggregate GEV (in particular, variations of the NL model) itinerary share models, this chapter measures the differential inter-itinerary competition dynamics that are hypothesized to exist among groups of air-travel itineraries.

In this chapter and Chapter 6 it is hypothesized that the competition among air-travel itineraries is differentiated by proximity in departure time, carrier, level-of-service or a combination of these dimensions[42]. Clearly, itineraries with similar departure times share unobserved characteristics that air-travelers consider in their itinerary selection process. For example, it is likely that air-travelers have desired departure times and consider groups of itineraries with departure times that are close to their desired departure times. That is, if an air-traveler desires to depart in the early morning, he/she will compare/contrast the attributes of different early-morning itineraries more closely than afternoon or evening itineraries. Thus, itineraries within a given time period are likely to compete with each other more than with itineraries in a different time period. Similarly, it is likely that itineraries of a given carrier compete more with each other than with itineraries of different carriers due to loyalty factors including frequent flyer program affiliations. Finally, it is believed that itineraries with the same level-of-service share many unobserved characteristics that differentiate them from itineraries with different levels-of-service. For example, (by definition) single-connect itineraries involve one stopover with a change of airplane[43]. Therefore, for a given airport-pair, the error terms for itineraries of a given level-of-service are likely to be correlated (resulting in greater substitutability between the itineraries than with itineraries of different levels-of-service).

Section 5.2 outlines the modeling framework for this chapter and Chapter 6. Section 5.3 presents the estimation of a "base" MNL model. Due to the belief that the

[42] Other itinerary competition dynamics likely exist (*e.g.* the competition among fare-classes within and across itineraries). However, the data constraints of this dissertation do not permit these relationships to be explored.

[43] From the travelers' perspective this involves a layover, the inconvenience of switching planes, the possibility of missing his/her connection, etc.

competition among itineraries is differentiated by proximity in departure time, carrier or level-of-service, the constraints of the MNL model are then relaxed in Section 5.4 by estimating two-level nested logit models. These models permit itineraries to be grouped (nested) by departure time, carrier or level-of-service. In addition to the value function parameter estimates, these models also contain an inverse logsum parameter[44] representing the level of itinerary competition within nests. For a given two-level NL model, the estimated value of the inverse logsum parameter (along with the overall model fit) indicates whether it is valid to group the itineraries according to the selected nesting structure (i.e. whether increased competition exists among itineraries in the nests, therefore rejecting the MNL model).

It is likely that the competition among air-travel itineraries is differentiated by proximity along more than one of the above-mentioned dimensions. To address this, a two-level weighted nested logit (WNL) model is formulated and estimated in Section 5.5 combining the competitive results of different two-level NL models. This model allows the simultaneous consideration of parallel two-level nesting structures (each structure yields an inverse logsum parameter estimate indicating the amount of itinerary competition within the nests of that structure) with a weight parameter indicating the relative importance of each structure. The formulation of this model is similar to the principles of differentiation models developed by Bresnahan *et al.* (1997).

Clearly, the two-level WNL model has advantages over the more restrictive two-level NL model structure. However, this model cannot capture the inter-itinerary competition dynamic that may exist among itineraries sharing a common attribute within another attribute. For example, (as mentioned above) it is believed that itineraries sharing a common time period compete more closely with each other than with itineraries of different time periods. However, within these time periods it is likely that itineraries of the same carrier exhibit even more competition amongst themselves. As a result, Section 5.6 presents the estimations of three-level nested logit models. These models group (nest) itineraries at an upper-level for a given dimension, and within each upper-level nest group itineraries according to a second dimension (lower-level nest). In addition to the value function parameter estimates, these models yield upper and lower-level inverse logsum parameter estimates indicating the differential amount of itinerary competition within the upper and lower-level nests. The estimated values of these inverse logsum parameters (along with the overall model fit) indicate whether a three-level nesting specification is supported (i.e. the hypothesis that a two-level NL model is adequate to describe the underlying itinerary competition dynamic should be rejected).

Finally, a three-level WNL model and a nested weighted nested logit (NWNL) model are estimated in Sections 5.7 and 5.8, respectively. The three-level WNL model is a direct extension of the two-level WNL model (it simultaneously estimates parallel three-level structures with a weight parameter indicating the relative importance of each structure). The NWNL model combines properties of the three-level nested logit and weighted nested logit models. Both of these models are new to the literature and

---

[44] The inverse logsum parameter is defined in Section 5.4.

incorporate all three attribute dimensions (departure time, carrier, level-of-service) of interest.

All of the above mentioned models are members of the generalized extreme value family of models (McFadden 1978) and are shown to outperform the base MNL model with respect to statistical tests and behavioral interpretations, leading to a clearer understanding of the air-travel inter-itinerary competition dynamic.

5.2 MODELING FRAMEWORK

As mentioned above, the motivation in developing the itinerary share models of this chapter and Chapter 6 is to understand the underlying competitive structure of air-travel itineraries. Using May 2001 data, models are estimated using all airport-pairs from the East to the West (as determined by time zone) regions of the United States and Canada.

Tables 5.1 – 5.3 report the distribution of itineraries and booked passengers by level-of-service (Table 5.1), carrier (Table 5.2) and departure time (Table 5.3) for the estimation dataset. All models estimated in this chapter and Chapter 6 have a common specification with respect to the independent variables representing itinerary service characteristics[45].

As discussed in Section 3.2, due to the aggregate nature of the data, the log-likelihood values for the models estimated in this chapter and Chapter 6 are adjusted by dividing by the ratio of the number of booked passengers to the number of airport-pair,

[45] Some value function variables (e.g. point of sale weighted airport presence, aircraft size) had insignificant or incorrect (sign) parameter estimates in the advanced models. Therefore, the value function specification for models contained in this chapter and Chapter 6 is simpler than that found in Chapter 4.

day-of-the-week combinations; and the parameter estimate t-statistics by the square root of this ratio. For the estimation data, the log-likelihood adjustment factor is $469{,}078/14{,}893 = 31.50$; and the t-statistic adjustment factor is $Sqrt(31.50) = 5.61$.

TABLE 5.1: Number of Itineraries and Booked Passengers by Level-of-Service for the East-West Region Dataset

Level-of-Service	Itineraries	Booked Passengers
Nonstop	3,382	144,814
Direct	3,675	17,180
Single-Connect	252,561	294,421
Double-Connect	369,504	12,663
Total	**629,122**	**469,078**

TABLE 5.2: Number of Itineraries and Booked Passengers by Carrier for the East-West Region Dataset

Carrier	Itineraries	Booked Passengers
United Airlines	133,111	93,704
American Airlines	83,497	72,642
Continental Airlines	89,805	43,070
Delta Airlines	92,968	72,124
Northwest Airlines	88,759	40,487
U.S. Airways	36,073	41,246
Other	104,909	105,805
Total	**629,122**	**469,078**

5.3 MULTINOMIAL LOGIT MODEL

The base or reference model for this chapter is the MNL model reported in Table 5.4. The parameter estimates are reported in groups corresponding to level-of-service, connection quality, carrier attributes, aircraft type, and departure time variables. The parameter estimates for this model are very similar to the estimates from Chapter 4, where the interpretation of these estimates is detailed. For this chapter and Chapter 6 it is sufficient to state that the value function parameter estimates have the correct sign, are of reasonable magnitude and all are significant at the 0.05 level after adjustment.

TABLE 5.3: Number of Itineraries and Booked Passengers by Departure Time for the East-West Region Dataset

Departure Time	Itineraries	Booked Passengers
5-6 A.M.	8,099	3,327
6-7 A.M.	92,019	44,212
7-8 A.M.	64,561	53,929
8-9 A.M.	37,265	37,718
9-10 A.M.	49,586	36,629
10-11 A.M.	37,833	20,427
11-12 Noon	39,321	21,604
12-1 P.M.	48,808	26,138
1-2 P.M.	55,426	23,576
2-3 P.M.	41,574	21,535
3-4 P.M.	46,849	30,311
4-5 P.M.	44,189	32,899
5-6 P.M.	36,713	45,257
6-7 P.M.	15,687	31,565
7-8 P.M.	8,085	21,414
8-9 P.M.	2,644	11,160
9-10 P.M.	401	5,355
10-12 Midnight	62	2,022
Total	**629,122**	**469,078**

The value function parameter estimates for the advanced models (presented in Tables 5.4 – 5.5) are very similar across the different model specifications[46]. However, in general, the estimates for the advanced models are smaller in magnitude than the corresponding estimates for the MNL model. This represents the lower sensitivity (substitutability) between groups of alternatives not in a common nest. The increased sensitivity between groups of alternatives in common nests is captured by the inverse logsum variables.

5.4 TWO-LEVEL NESTED LOGIT MODEL

Initial two-level nested logit estimations assumed nesting based on each of the three dimensions described above. In these models, itineraries are grouped into nests according to departure time (morning, 5:00 – 9:59 A.M.; midday, 10:00 A.M. – 3:59 P.M; evening, 4:00 P.M. – Midnight), carrier (six major U.S. carriers and a group of "other" carriers) or level-of-service (nonstop, direct, single-connect, double-connect). Visual representations of the two-level time NL model and the two-level carrier NL model are shown in Figures 5.1 and 5.2, respectively.

[46] In Table 5.4, all parameter estimates significant at the 0.05 level after the adjustment procedure. In Table 5.5, parameter estimates in bold not significant at the 0.05 level after the adjustment procedure. In both tables, the significance of value function parameter estimates is with respect to their hypothesized value of zero; the significance of inverse logsum parameter estimates is with respect to their hypothesized value of one.

TABLE 5.4: Itinerary Share Models: MNL, Two-Level NL's and Two-Level WNL

Explanatory Variables	Model			
	MNL	2-Level NL Time	2-Level NL Carrier	2-Level WNL: Time | Carrier
Level-of-Service				
Nonstop Itinerary in Nonstop Market	0.0000	0.0000	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.9595	-1.6271	-1.7798	-1.4253
Single-Connect Itinerary in Nonstop Market	-2.8371	-2.3540	-2.5695	-2.0624
Double-Connect Itinerary in Nonstop Market	-6.6264	-5.4663	-5.7872	-4.6364
Direct Itinerary in Direct Market	0.0000	0.0000	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.7370	-0.6207	-0.6579	-0.5323
Double-Connect Itinerary in Direct Market	-3.9250	-3.2331	-3.4217	-2.7375
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.6015	-2.1915	-2.3043	-1.8997
Connection Quality				
Second-Best Connection	-0.4208	-0.3331	-0.3216	-0.2396
Second-Best Connection Time Difference	-0.0087	-0.0071	-0.0074	-0.0058
Distance Ratio	-0.0135	-0.0109	-0.0131	-0.0103
Best Connection Time Difference	-0.0056	-0.0047	-0.0051	-0.0041
Carrier Attributes				
Fare Ratio	-0.0060	-0.0052	-0.0039	-0.0033
Carrier Constants (Proprietary)	--------	--------	--------	--------
Code share	-1.8601	-1.5241	-1.6861	-1.3383
Aircraft Type				
Mainline Jet	0.0000	0.0000	0.0000	0.0000
Regional Jet	-0.4560	-0.3856	-0.4225	-0.3464
Propeller Aircraft	-0.4201	-0.3496	-0.3658	-0.2919
Departure Time				
5 – 6 A.M.	-0.2184	-0.1931	-0.2084	-0.1814
6 – 7 A.M.	0.0000	0.0000	0.0000	0.0000
7 – 8 A.M.	0.1385	0.1118	0.1235	0.0964
8 – 9 A.M.	0.2381	0.1907	0.2150	0.1663
9 –10 A.M.	0.2646	0.2135	0.2365	0.1848
10 - 11 A.M.	0.2672	0.1873	0.2412	0.1619
11 - 12 noon	0.2290	0.1643	0.2168	0.1507
12 – 1 P.M.	0.2476	0.1761	0.2293	0.1593
1 – 2 P.M.	0.1614	0.1043	0.1507	0.0956
2 – 3 P.M.	0.1686	0.1058	0.1599	0.0982
3 – 4 P.M.	0.1856	0.1219	0.1709	0.1100
4 – 5 P.M.	0.0960	0.0486	0.0934	0.0523
5 – 6 P.M.	0.0972	0.0490	0.0840	0.0429
6 – 7 P.M.	0.1760	0.1179	0.1535	0.1007
7 – 8 P.M.	0.0833	0.0443	0.0857	0.0502
8 – 9 P.M.	-0.0803	-0.0807	-0.0563	-0.0541
9 – 10 P.M.	-0.2587	-0.2243	-0.2131	-0.1778
10 – Midnight	-0.3407	-0.3179	-0.2847	-0.2546
Inverse Logsum Parameter (Time)	--------	1.2244	--------	1.5435
Inverse Logsum Parameter (Carrier)	--------	--------	1.1768	1.4519
WNL Weight Parameter (Time Structure)	--------	--------	--------	0.5364
Log Likelihood at Zero	-2,173,197	-2,173,197	-2,173,197	-2,173,197
Log Likelihood at Convergence	-1,558,186	-1,557,443	-1,556,663	-1,555,632
Adjusted Log Likelihood at Convergence	-49,466	-49,443	-49,418	-49,385
Rho-square w.r.t. Zero	0.2830	0.2833	0.2837	0.2842

TABLE 5.5: Itinerary Share Models: Three-Level NL's, Three-Level WNL and NWNL

Explanatory Variables	Model			
	3-Level NL: Time, LOS	3-Level NL: Time, Carrier	3-Level WNL: T|C, T|L	NWNL
Level-of-Service				
Nonstop Itinerary in Nonstop Market	0.0000	0.0000	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.6479	-1.6570	-1.6754	-1.6833
Single-Connect Itinerary in Nonstop Market	-2.3401	-2.3802	-2.3703	-2.3655
Double-Connect Itinerary in Nonstop Market	-5.5099	-5.2215	-5.2649	-5.2761
Direct Itinerary in Direct Market	0.0000	0.0000	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.5935	-0.6362	-0.6096	-0.5982
Double-Connect Itinerary in Direct Market	-3.2467	-3.1347	-3.1467	-3.1465
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.2118	-2.1679	-2.1869	-2.1922
Connection Quality				
Second-Best Connection	-0.3290	-0.2504	-0.2473	-0.2449
Second-Best Connection Time Difference	-0.0071	-0.0063	-0.0062	-0.0062
Distance Ratio	-0.0108	-0.0112	-0.0111	-0.0110
Best Connection Time Difference	-0.0047	-0.0049	-0.0048	-0.0048
Carrier Attributes				
Fare Ratio	-0.0051	-0.0036	-0.0036	-0.0035
Carrier Constants (Proprietary)	--------	--------	--------	--------
Code share	-1.5082	-1.5408	-1.5257	-1.5179
Aircraft Type				
Mainline Jet	0.0000	0.0000	0.0000	0.0000
Regional Jet	-0.3827	-0.4019	-0.3985	-0.3961
Propeller Aircraft	-0.3459	-0.3294	-0.3258	-0.3233
Departure Time				
5 - 6 A.M.	-0.1925	-0.2152	-0.2139	-0.2127
6 - 7 A.M.	0.0000	0.0000	0.0000	0.0000
7 - 8 A.M.	0.1099	0.1163	0.1146	0.1136
8 - 9 A.M.	0.1880	0.1944	0.1918	0.1902
9 -10 A.M.	0.2108	0.2148	0.2125	0.2109
10 - 11 A.M.	0.1850	0.1971	0.1950	0.1933
11 - 12 noon	0.1625	0.1808	0.1791	0.1777
12 - 1 P.M.	0.1740	0.1845	0.1824	0.1810
1 - 2 P.M.	0.1029	0.1152	0.1140	0.1132
2 - 3 P.M.	0.1045	0.1175	0.1163	0.1155
3 - 4 P.M.	0.1199	0.1328	0.1309	0.1299
4 - 5 P.M.	0.0455	0.0577	0.0550	0.0535
5 - 6 P.M.	0.0457	0.0383	0.0355	0.0340
6 - 7 P.M.	0.1139	0.1068	0.1032	0.1013
7 - 8 P.M.	0.0416	0.0546	0.0521	0.0507
8 - 9 P.M.	-0.0818	-0.0586	-0.0599	-0.0607
9 - 10 P.M.	-0.2234	-0.1982	-0.1973	-0.1965
10 – Midnight	-0.3161	-0.2821	-0.2807	-0.2802
Upper-Level Inverse Logsum Parameter (Time)	1.2124	1.0667	1.0594	1.0584
Lower-Level Inverse Logsum Parameter (Carrier)	--------	1.3568	1.3642	1.3637
Lower-Level Inverse Logsum Parameter (LOS)	1.2376	--------	**1.5492**	**1.7650**
Weight Parameter (Time|Carrier Structure)	--------	--------	**0.9529**	**0.9490**
Log Likelihood at Zero	-2,173,197	-2,173,197	-2,173,197	-2,173,197
Log Likelihood at Convergence	-1,557,435	-1,554,227	-1,554,219	-1.554,210
Adjusted Log Likelihood at Convergence	-49,442	-49,341	-49,340	-49,340
Rho-square w.r.t. Zero	0.2833	0.2848	0.2848	0.2848

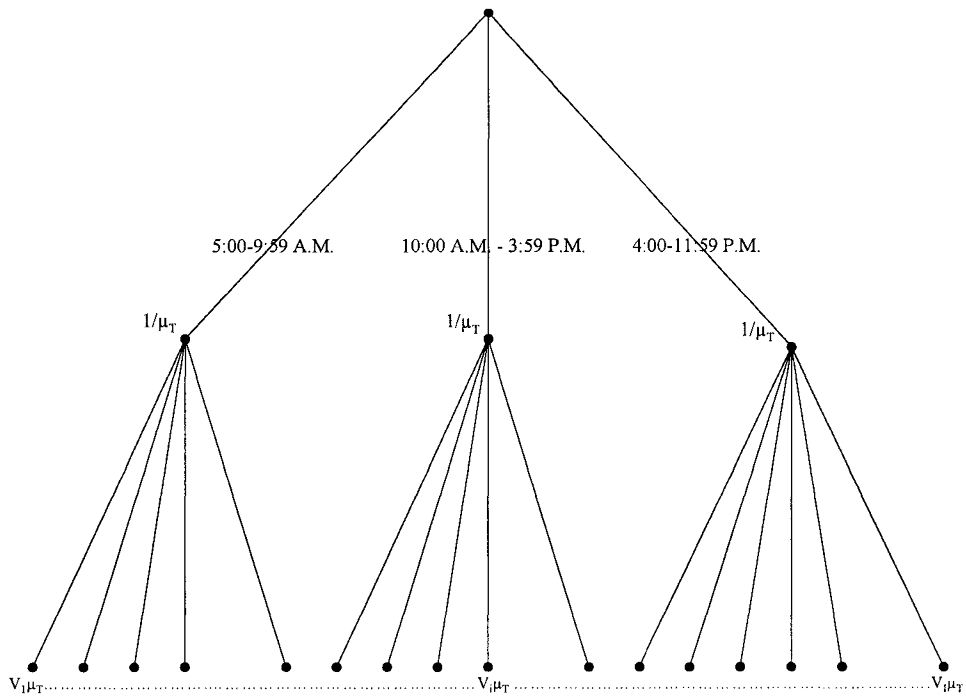

FIGURE 5.1: Two-Level NL Time Model Structure

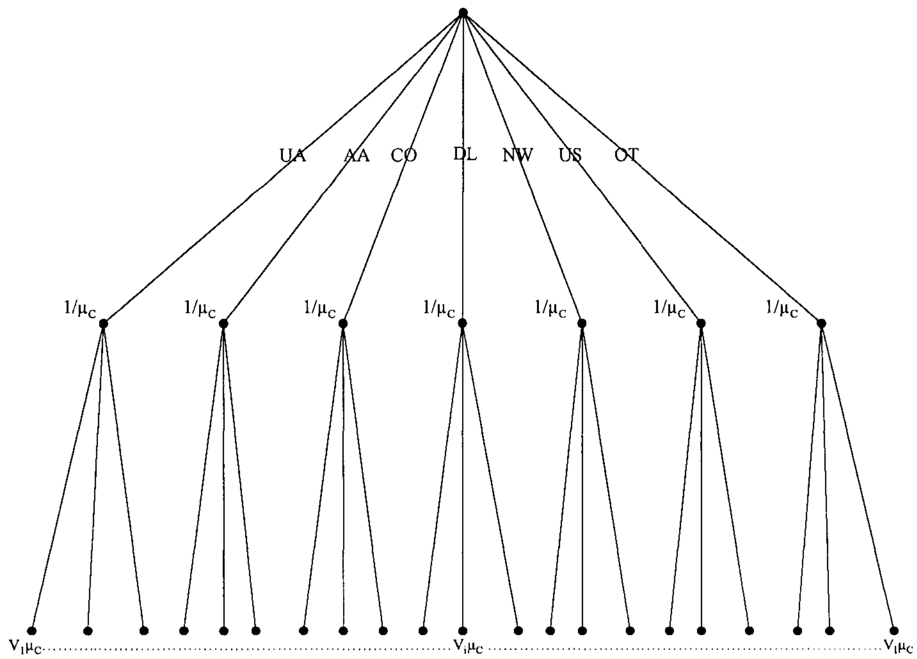

FIGURE 5.2: Two-Level NL Carrier Model Structure

For the two-level NL models estimated in this chapter, the total variance of the itinerary error terms is set equal to $\frac{\pi^2}{6}$ as is commonly done in GEV models. The error term for each itinerary is decomposed into two components: an independent component specific to the itinerary and a component common to all itineraries in its nest. That is,

$$U_i = V_i + \varepsilon_i + \varepsilon_c \tag{5.3}$$

where the error variance for the independent component is given by $\frac{\pi^2}{6\mu^2}$ (μ is the inverse of the logsum parameter[47]) and the error variance for the common component is given by $\frac{\pi^2}{6}\left(1-\frac{1}{\mu^2}\right)$. The inverse logsum parameter must be greater than one to ensure that the independent component variance is smaller than the total variance. Inverse logsum parameter estimates can be interpreted as indicating the amount of "competition" among itineraries sharing a common nest with larger values indicating a higher level of substitution within nests. This is because a larger inverse logsum implies that more of the total error variance is associated with the common error term (that is, itineraries sharing a nest with a large inverse logsum parameter have a high level of correlation). Assuming this alternative error term structure and random utility maximization, it can be shown that with a two-level nested logit specification the share of passengers assigned to each itinerary between an airport-pair for a given day of the week is given by:

$$S_i = S_{n'} \times S_{i|n'} = \frac{\exp\left(\frac{1}{\mu}\Gamma_n\right)}{\sum_{n' \in N} \exp\left(\frac{1}{\mu}\Gamma_{n'}\right)} \times \frac{\exp(\mu V_i)}{\sum_{i' \in n} \exp(\mu V_{i'})} \tag{5.4}$$

where S_i is the passenger share assigned to itinerary i,

$S_{n'}$ is the passenger share assigned to nest n',

$S_{i|n'}$ is the passenger share assigned to itinerary i given nest n',

μ is the inverse logsum parameter associated with the nests[48],

$\Gamma_n = \ln\left(\sum_{i' \in N_n} \exp(\mu V_{i'})\right)$ and

V_i is the value of itinerary i.

The increased competition among itineraries sharing a common nest can be seen from cross-elasticity formulas gotten from the two-level NL model. The cross-elasticity equation for the change in the probability of alternative j due to changes in the k^{th} attribute of alternative i (where i and j belong to the same nest) is given by:

---

[47] In many other studies the error variance for the independent component is given by $\frac{\pi^2\theta^2}{6}$, where the logsum parameter is represented by θ and must be less than one. In the current derivation, μ, the inverse logsum, is equal to $\frac{1}{\theta}$ leading to identical results.

$$\eta^{P_j}_{X_{ik}} = \frac{\partial P_j}{\partial X_{ik}} \frac{X_{ik}}{P_j} = -P(i \mid n')\beta_k X_{ik}\left[P(n') + (\mu - 1)\right] \tag{5.5}$$

where $P(i \mid n')$ is the probability of itinerary i given nest n', $P(n')$ is the probability of nest n' and μ is the inverse logsum parameter. Alternatives not sharing a common nest have cross-elasticities of the same form as the MNL model (however, there may be differences in the magnitude of the value function parameters between the MNL and NL model). The cross-elasticities for alternatives sharing a common nest in the two-level NL model are larger in magnitude than the cross-elasticities for alternatives that do not share a common nest. This is shown below:

$$\begin{aligned} \left|-P_i\beta_k X_{ik}\right| &\le \left|-P(i \mid n')\beta_k X_{ik}\left[P(n') + (\mu - 1)\right]\right| \\ P_i &\le P(i \mid n')\left[P(n') + (\mu - 1)\right] \\ P_i &\le P_i + P(i \mid n')(\mu - 1) \end{aligned} \tag{5.6}$$

The above inequalities hold since μ must be greater than or equal to one. The higher the value of μ, the more sensitive itineraries are to attribute changes in other itineraries sharing a common nest.

The estimation results for two-level NL models with itineraries nested by departure time and carrier are reported in Table 5.4 showing that itineraries within a common time period and itineraries flown by the same carrier have common attributes that passengers consider in their itinerary selection process. These models reject the hypothesis that the MNL model is the true model at the 0.001 level after adjustment. On

---

[48] Early experimental estimations yielded similar parameter estimates for the inverse logsum variables

the other hand, grouping itineraries by level-of-service did not yield theoretically acceptable results (the inverse logsum parameter was estimated to be less than one, which is inconsistent with utility theory). This was surprising since it seems likely that an itinerary within a given level-of-service nest would share many characteristics with the other itineraries within the same nest and thus should have higher cross-elasticities among themselves than with itineraries of different levels-of-service.

5.5 TWO-LEVEL WEIGHTED NESTED LOGIT MODEL

The two-level weighted nested logit model simultaneously estimates parallel two-level nesting structures (each structure is equivalent to a two-level NL model) with a weight parameter indicating the relative importance of each structure. Each itinerary in each alternative set appears twice in the model, once in each of the parallel structures. The WNL model can be shown to be a special case of the generalized nested logit model (Wen and Koppelman 2001).

Due to the strong empirical results from the two-level time and carrier NL models, a two-level WNL model was estimated with a time structure and a carrier structure (see Figure 5.3). The share of passengers assigned to each itinerary between an airport-pair for a given day of the week is:

---

across nests in each nesting dimension. Given this result and a desire for consistency across nests of the same type, the inverse logsum parameter is constrained in each case to be equal across all common nests.

$$
\begin{aligned}
S_i &= w_t \times S_t \times S_{i|t} + w_c \times S_c \times S_{i|c} \\
&= w_t \times \frac{\exp\left(\dfrac{1}{\mu_t}\Gamma_t\right)}{\sum_{t' \in N} \exp\left(\dfrac{1}{\mu_t}\Gamma_{t'}\right)} \times \frac{\exp(\mu_t V_i)}{\sum_{i' \in t} \exp(\mu_t V_{i'})} \\
&\quad + w_c \times \frac{\exp\left(\dfrac{1}{\mu_c}\Gamma_c\right)}{\sum_{c' \in N} \exp\left(\dfrac{1}{\mu_c}\Gamma_{c'}\right)} \times \frac{\exp(\mu_c V_i)}{\sum_{i' \in c} \exp(\mu_c V_{i'})}
\end{aligned} \tag{5.7}
$$

where c represents the carrier nests,
t represents the time nests,
w_c is the weight given to the carrier structure and
$w_t = 1 - w_c$ is the weight given to the time structure.

Estimation results for this model are reported in Table 5.4. The inverse logsum parameters for both the time and carrier nests are significantly greater than one (at all levels of significance after adjustment) indicating increased itinerary competition among itineraries sharing a common time period or carrier. The weight parameter is close to ½ and significantly different than zero or one (at all levels of significance after adjustment) indicating that each portion of the structure is important. Finally, this model outperforms both the two-level time and carrier NL models at the 0.001 level after adjustment.

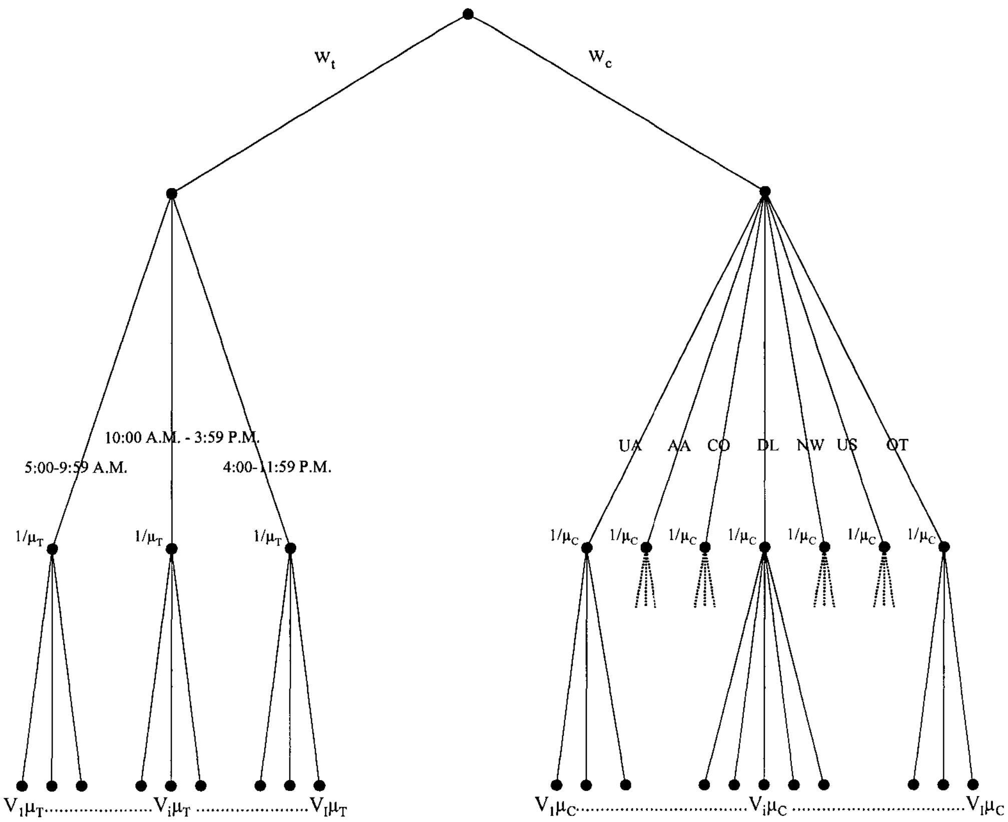

FIGURE 5.3: Two-Level WNL Time | Carrier Model Structure

5.6 THREE-LEVEL NESTED LOGIT MODEL

Six three-level nested logit model specifications were estimated representing all possible three-level combinations for the three itinerary dimensions under study (upper-level time and lower-level carrier (time, carrier); carrier, time; time, level-of-service; level-of-service, time; carrier, level-of-service; level-of-service, carrier).

For the three-level NL models, the total variance of the itinerary error terms is decomposed into three components: an independent component specific to the itinerary, a component common to all itineraries in its lower-level nest (nest n') and a component common to all itineraries in its upper-level nest (nest m'). That is,

$$U_i = V_i + \varepsilon_i + \varepsilon_{n'} + \varepsilon_{m'} \tag{5.8}$$

where the error variance for the independent component, ε_i, is given by $\frac{\pi^2}{6\mu_n^2}$ (μ_n is the inverse logsum parameter for the lower-level nest), the error variance for the random component distinct to the lower-level nest but not including the error component of the independent elemental alternative is given by $\frac{\pi^2}{6\mu_m^2}-\frac{\pi^2}{6\mu_n^2}$ (μ_m is the inverse logsum parameter for the upper-level nest), and finally the error variance for the random component associated with the upper-level nest but excluding the error component of the lower-level nest and the elemental alternative is $\frac{\pi^2}{6}-\frac{\pi^2}{6\mu_m^2}$. Thus, both upper and lower-level inverse logsum parameters must be greater than one and the lower-level inverse logsum parameter must be greater than the upper-level inverse logsum parameter. The share of passengers assigned to each itinerary between an airport-pair for a given day of the week is given by:

$$S_i = S_{m'} \times S_{n'|m'} \times S_{i|n'}$$
$$= \frac{\exp\left(\frac{1}{\mu_m}\Gamma_m\right)}{\sum_{m' \in M}\exp\left(\frac{1}{\mu_m}\Gamma_{m'}\right)} \times \frac{\exp\left(\frac{\mu_m}{\mu_n}\Gamma_n\right)}{\sum_{n' \in N}\exp\left(\frac{\mu_m}{\mu_n}\Gamma_{n'}\right)} \times \frac{\exp(\mu_n V_i)}{\sum_{i' \in n}\exp(\mu_n V_{i'})} \tag{5.9}$$

where $S_{m'}$ is the passenger share assigned to upper-level nest m',

$S_{n'|m'}$ is the passenger share assigned to lower-level nest n' given upper-level nest m',

μ_m is the inverse logsum parameter associated with the upper-level nests,

μ_n is the inverse logsum parameter associated with the lower-level nests,

$$\Gamma_n = \ln\left(\sum_{i' \in N_j} \exp\left(\mu_n V_{i'}\right)\right) \text{ and}$$

$$\Gamma_m = \ln\left(\sum_{n' \in N_m} \exp\left(\frac{\mu_m}{\mu_n}\Gamma_{n'}\right)\right).$$

The requirement that the lower-level inverse logsum parameter be greater than the upper-level inverse logsum parameter implies that itineraries within the same lower-level nest (and hence within the same upper-level nest) share the most unobserved attributes and compete more closely with each other than with other itineraries. Itineraries sharing a common upper-level nest (but not a lower-level nest) have less competition among themselves than with itineraries that share the same lower-level nest, but a greater level of competition than with itineraries in a different upper-level nest.

Of the six three-level NL models estimated, only two satisfied the inverse logsum conditions described above. These models, reported in Table 5.5, are for time, level-of-service and time, carrier. A visual representation of the three-level time, level-of-service NL model is shown in Figure 5.4 and the three-level time, carrier NL model is shown in Figure 5.5. The time, carrier model rejects both the time and carrier two-level NL models at the 0.001 level after adjustment.

The time, level-of-service model does not reject the two-level time NL model after adjustment. However, it does improve upon the two-level NL time model before adjustment and both its inverse logsum parameters are significant after adjustment (they are significantly different from each other after adjustment as well). Regardless, the marginal significance of this model implies that this three-level nesting specification may not be valid. However, variations of this model are revisited in the next two sections.

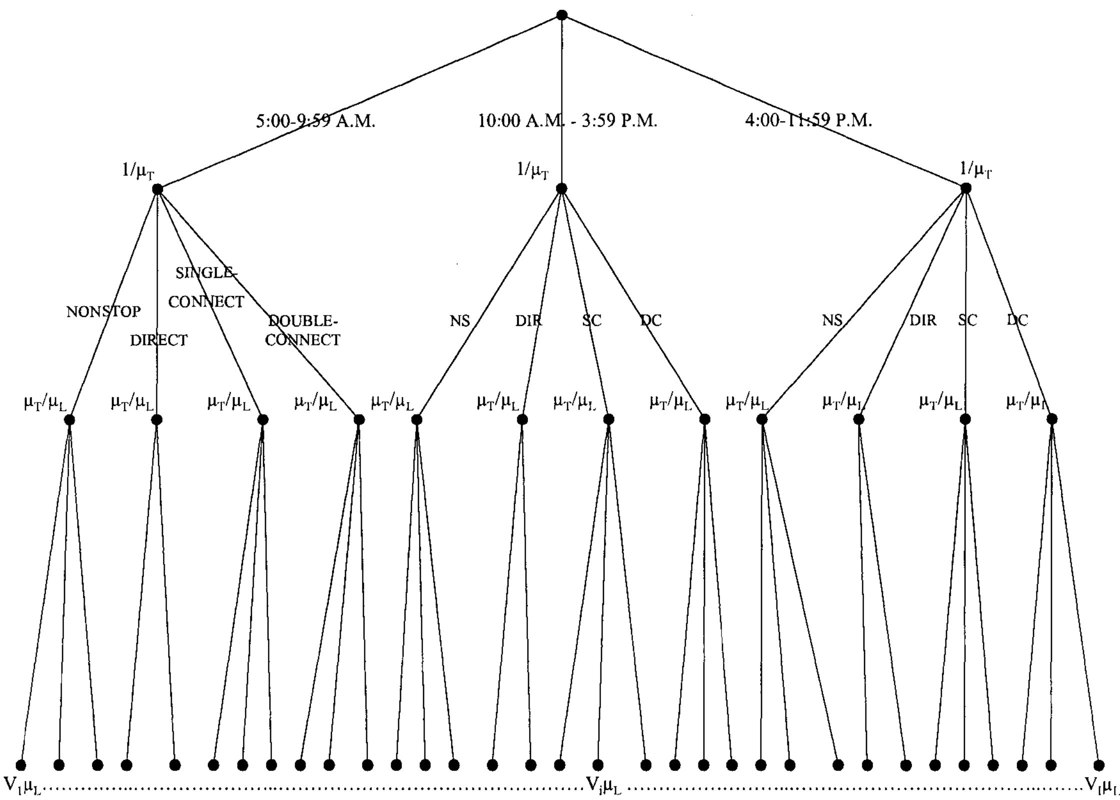

FIGURE 5.4: Three-Level NL Time, Level-of-Service Model Structure

These three-level NL results indicate that there is moderate itinerary competition among itineraries sharing a common time period and greater competition among itineraries sharing both time period and carrier or (to a lesser extent) time period and level-of-service. This demonstrates the importance of conditioning the within carrier (level-of-service) competition dynamic by time period. However, note that the three-level time, carrier NL model implies that itineraries of the same carrier, but of different time periods, do not exhibit much competition amongst themselves.

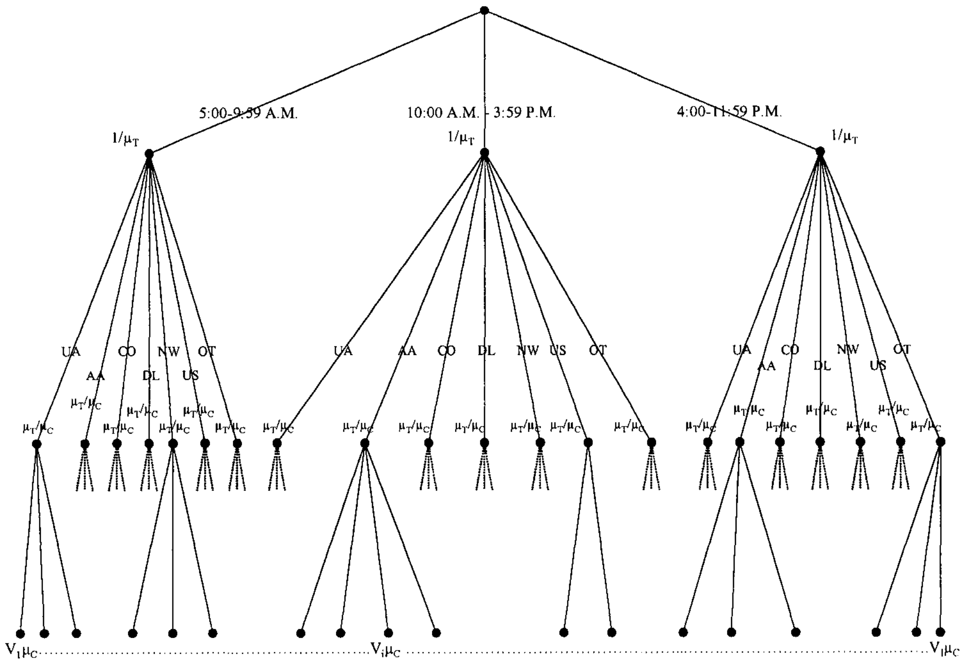

FIGURE 5.5: Three-Level NL Time, Carrier Model Structure

5.7 THREE-LEVEL WEIGHTED NESTED LOGIT MODEL

The three-level weighted nested logit model is a direct extension of the two-level weighted nested logit model. The mathematical structure of the model is:

$$
\begin{aligned}
S_i &= w_{ct} \times S_{ict} + w_{st} \times S_{ist} \\
&= w_{ct} \times S_t \times S_{c|t} \times S_{i|ct} + w_{st} \times S_t \times S_{s|t} \times S_{i|st} \\
&= w_{ct} \times \frac{\exp\left(\frac{1}{\mu_T}\Gamma_t\right)}{\sum_{t' \in T}\exp\left(\frac{1}{\mu_T}\Gamma_{t'}\right)} \times \frac{\exp\left(\frac{\mu_T}{\mu_C}\Gamma_c\right)}{\sum_{c' \in C}\exp\left(\frac{\mu_T}{\mu_C}\Gamma_{c'}\right)} \times \frac{\exp(\mu_C V_i)}{\sum_{i' \in c}\exp(\mu_C V_{i'})} \\
&\quad + w_{st} \times \frac{\exp\left(\frac{1}{\mu_T}\Gamma_t\right)}{\sum_{t' \in T}\exp\left(\frac{1}{\mu_T}\Gamma_{t'}\right)} \times \frac{\exp\left(\frac{\mu_T}{\mu_S}\Gamma_s\right)}{\sum_{s' \in S}\exp\left(\frac{\mu_T}{\mu_S}\Gamma_{s'}\right)} \times \frac{\exp(\mu_S V_i)}{\sum_{i' \in s}\exp(\mu_S V_{i'})}
\end{aligned}
\tag{5.10}
$$

Due to the marginal significance of the three-level time, level-of-service NL model and the significance of the three-level time, carrier NL model, a three-level WNL model was estimated with parallel three-level structures for time, carrier and time, level-of-service. Figure 5.6 gives a visual representation of this model and the estimation results are reported in Table 5.5. The upper-level inverse logsum parameter for time is significant at the 0.001 level (after adjustment) and the lower-level inverse logsum parameter for carrier is significant (at the 0.001 level) after adjustment indicating a high level of competition among itineraries flown by a carrier within a time period. However, the lower-level inverse logsum parameter for level-of-service is not significant after adjustment. Additionally, after adjustment, the weight on the time, carrier structure is not significantly different from one. These results indicate that the time, carrier side of the

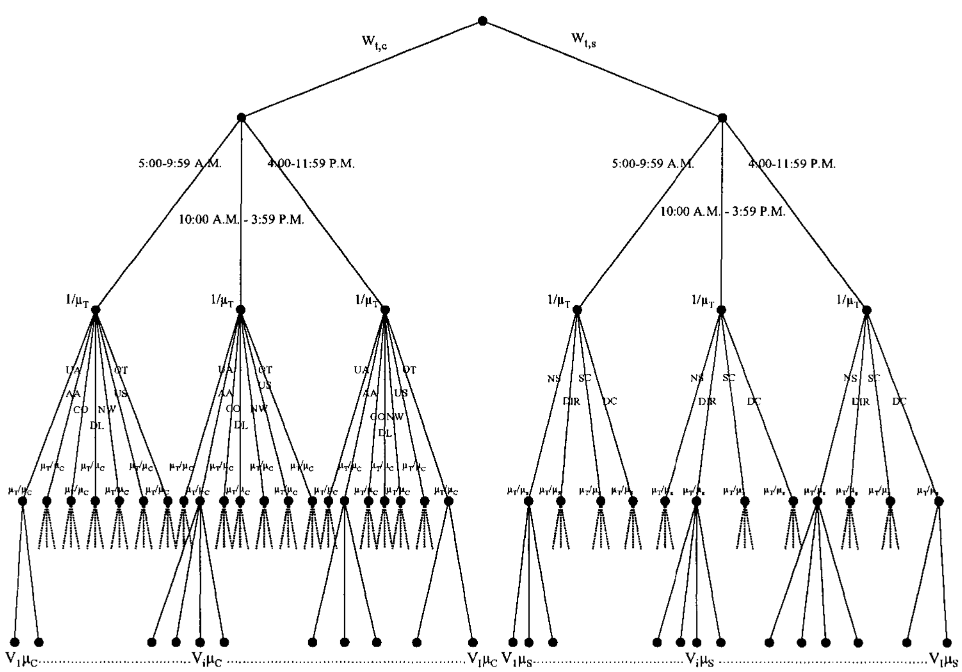

FIGURE 5.6: Three-Level WNL Time | Carrier, Time | Level-of-Service Model Structure

model “dominates”. Finally, after adjustment, the model is only marginally better than the three-level time, carrier NL model.

5.8 NESTED WEIGHTED NESTED LOGIT MODEL

The nested weighted nested logit model, a “hybrid” version of the NL and WNL models, is in the GEV class of models and is new to the literature. The motivation for estimating this model came from the estimation results of the three-level NL and WNL models discussed in Sections 5.6 and 5.7, respectively. Both three-level NL models that yielded significant (or marginally significant) results (the time, level-of-service and time, carrier models) had itineraries nested at the upper level by departure time. However, when these models were “combined” into a three-level WNL model, the weight

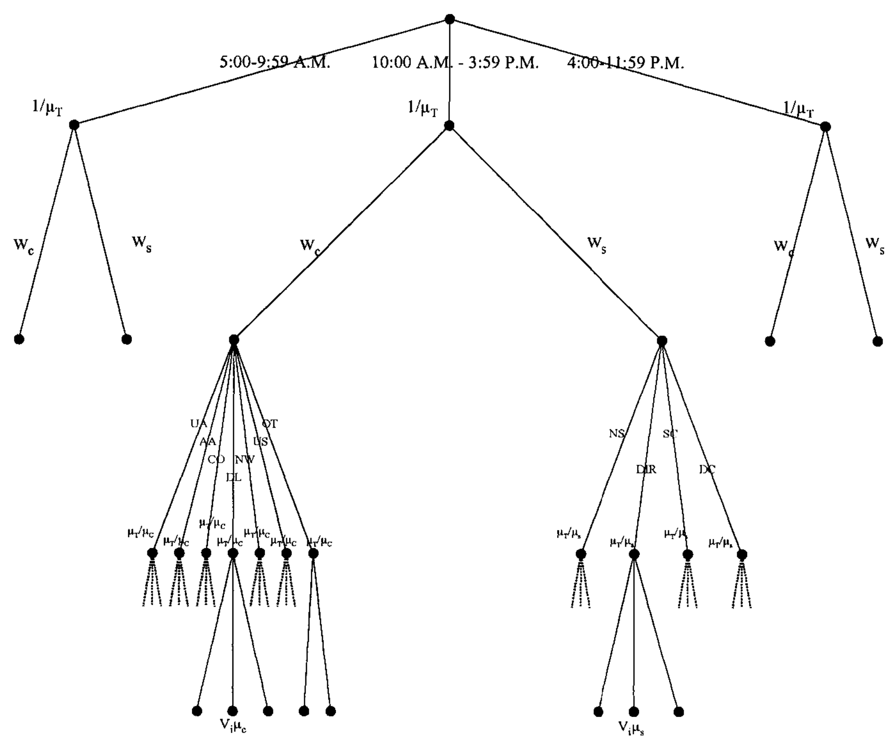

FIGURE 5.7: Nested Weighted Nested Logit Model Structure

parameter and the lower-level inverse logsum parameter for level-of-service in the time, level-of-service structure were not significant. To address this, the NWNL was developed. Figure 5.7 gives a visual representation of this model. In the structure shown in the figure, itineraries are grouped at the upper level by departure time. On the lower level (within a time period) each itinerary is placed into two parallel nesting structures: a carrier structure and a level-of-service structure. A weight parameter indicates the relative importance of each of these structures within the upper-level departure time nest. The share of passengers assigned to each itinerary between an airport-pair for a given day of the week is:

$$
\begin{aligned}
S_i &= S_t \times \left[\left(w_c \times S_{c|t} \times S_{i|ct} \right) + \left(w_s \times S_{s|t} \times S_{i|st} \right) \right] \\
&= \frac{\exp\left(\frac{w_c}{\mu_t} \Gamma_{tc} + \frac{w_s}{\mu_t} \Gamma_{ts} \right)}{\sum_{t' \in T} \exp\left(\frac{w_c}{\mu_{t'}} \Gamma_{t'c} + \frac{w_s}{\mu_{t'}} \Gamma_{t's} \right)} \times \\
&\quad \left(\left(w_c \times \frac{\exp\left(\frac{\mu_t}{\mu_c} \Gamma_c \right)}{\sum_{c' \in C} \exp\left(\frac{\mu_t}{\mu_{c'}} \Gamma_{c'} \right)} \times \frac{\exp(\mu_c V_i)}{\sum_{i' \in c} \exp(\mu_c V_{i'})} \right) + \left(w_s \times \frac{\exp\left(\frac{\mu_t}{\mu_s} \Gamma_s \right)}{\sum_{s' \in S} \exp\left(\frac{\mu_t}{\mu_{s'}} \Gamma_{s'} \right)} \times \frac{\exp(\mu_s V_i)}{\sum_{i' \in s} \exp(\mu_s V_{i'})} \right) \right)
\end{aligned} \tag{5.11}
$$

The estimation results for this model are presented in Table 5.5. The results indicate that this model is almost identical to the three-level WNL model discussed in Section 5.7.

That is, its value function parameter estimates are very similar to the three-level WNL, and its lower-level inverse logsum parameter for level-of-service and weight parameter are not significant after adjustment[49]. Finally, the NWNL model only marginally improves upon the three-level time, carrier NL model and the three-level WNL model. Thus, it appears that "weighting" carrier and level-of-service structures within a time period does not improve upon our understanding of the true itinerary competition dynamic.

Although the three-level WNL and NWNL models yield insignificant (after adjustment) parameter estimates (indicating their nesting structures may not be valid for the current application), the improvement in these models (over previously estimated models with respect to log-likelihood values) demonstrates the potential of improving model goodness of fit through more complex nesting. This possibility is explored in Chapter 6.

5.9 SUMMARY AND CONCLUSIONS

This chapter shows that the competition among air-travel itineraries is not "uniform". Thus, itinerary share models employing multinomial logit methodology are not adequate. Two-level nested logit models are estimated showing that itineraries sharing a common time period or carrier (but not level-of-service) exhibit a strong amount of competition among themselves. Using these results, a two-level weighted nested logit model (a new variation in the GEV family of models) with parallel time and

[49] The weight parameter is not significantly different than one indicating that (like the three-level WNL model) the carrier structure "dominates" the level-of-service structure.

carrier nesting structures is estimated. It significantly rejects the standard two-level NL models and has advantages over the more restrictive NL model structure.

Three-level nested logit models are estimated. The results of these models show that itineraries sharing a common time period have a moderate amount of competition amongst themselves, while itineraries sharing both time period and carrier (and to a lesser extent time period and level-of-service) exhibit a strong amount of competition amongst themselves.

This chapter introduces the three-level WNL and NWNL models. These models are new to the literature and are in the GEV family of models. However, even though these models simultaneously incorporate the three dimensions of departure time, carrier and level-of-service, they only marginally improve upon the three-level NL model with itineraries nested at the upper level by departure time and at the lower level by carrier.

Finally, while the estimations for models with itineraries nested by level-of-service were generally not significant, it is still reasonable to expect that increased competition exists within level-of-service nests (especially when the level-of-service nests are within upper-level time period nests as demonstrated by the marginal significance of the time, level-of-service three-level NL model). Regardless, it appears that the underlying competition among air-travel itineraries can almost fully be described by nesting itineraries by the departure time and carrier dimensions.

CHAPTER 6: MODELING THE UNDERLYING COMPETITIVE DYNAMIC AMONG AIR-TRAVEL ITINERARIES WITH ORDERED GENERALIZED EXTREME VALUE MODELS

6.1 INTRODUCTION

The models estimated in Chapter 5 demonstrate the importance of considering the differential competition or substitution among air-travel itineraries connecting airport-pairs. In particular, various model structures were developed showing that inter-itinerary competition is differentiated by departure time, carrier and (to a lesser extent) level-of-service. This demonstrates that itinerary share models employing multinomial logit methodology are not adequate for describing the underlying competition among air-travel itineraries.

The model specifications estimated in Chapter 5 included several variations of the nested logit model structure. However, these models (those that nest itineraries by departure time) group itineraries by arbitrary discrete time periods and recognize differences in competition only between and within the pre-defined time periods (nests). This imposes unrealistic constraints on the departure time-of-day competition dynamic; for example, it implies that an itinerary within a given nest will compete more closely with an itinerary sharing the nest than with an itinerary in an adjacent nest that is closer in departure time.

The models formulated and presented in this chapter capture a more complicated and realistic itinerary competition structure (for the departure time dimension) than the nested logit models. The motivation for developing these models was the belief that – within an airport-pair – the amount of competition between itineraries is differentiated by the relative degree of proximity in their departure times. This property, named "proximate covariance" by Small (1987), implies that itineraries that are "closer" to each other (by departure time in this case) exhibit a higher amount of error correlation (substitution and competition) with each other than with itineraries that are more separated in time. The level of correlation between itineraries increases the closer they are to each other. Models estimated in this chapter capture this property by grouping (nesting) itineraries (according to their departure times) into narrow time periods and ordering these time periods by departure time (early morning to late evening). These models hypothesize that an itinerary will compete most closely with itineraries in the same narrow time period and less closely as the difference in time periods increases.

This chapter begins by estimating several ordered generalized extreme value models (Small 1987). Small's development of the OGEV model was for the case of distinctly ordered alternatives[50]. Bhat (1998) used a combined MNL-OGEV structure to model mode and departure time choice for distinct trips. In the current application, the OGEV structure is used to model the underlying competition among air-travel itineraries (for a given airport-pair-day-of-the-week) along the departure time dimension. As will be shown below, the nesting structure of these models consists of overlapping time periods where each itinerary is allocated to contiguous nests according to allocation parameters. The values and significance of these allocation parameters indicate whether the assumption underlying the OGEV model (i.e. the proximate covariance property) is valid

[50] For example, a household decision scenario of how many automobiles to own.

(that is, whether the hypothesis that the nested logit model is adequate to describe the departure time competition dynamic can be rejected).

As demonstrated (using variations of the nested logit model) in Chapter 5, it is desirable to model the intra-carrier (and potentially intra-level-of-service) competition dynamic within an upper-level departure time structure. The OGEV models described in the preceding paragraph cannot accomplish this (since they only model the inter-itinerary competition dynamic along the departure time dimension). To rectify this, "hybrid" OGEV models are developed in this chapter. These models are new to the literature and incorporate the traditional OGEV model structure (described above) at the upper level with a GEV component such as the NL model at the lower level. Modeling air-travel itinerary shares using hybrid OGEV specifications (representing realistic and complicated competition dynamics) is an important step forward for aviation demand modeling.

6.2 ORDERED GENERALIZED EXTREME VALUE MODEL

Figure 6.1 gives a visual representation (for a generic airport-pair-day-of-the-week) of an OGEV model with six time periods (TP's) (5:00-6:59 A.M., 7:00-9:59 A.M., 10:00 A.M.-12:59 P.M., 1:00-3:59 P.M., 4:00-6:59 P.M. and 7:00-10:59 P.M.) in which each itinerary is allocated to two nests. Figure 6.2 gives a visual representation of an OGEV model with eight time periods (5:00-6:59 A.M., 7:00-8:59 A.M., 9:00-10:59 A.M., 11:00 A.M.-12:59 P.M., 1:00-2:59 P.M., 3:00-4:59 P.M., 5:00-6:59 P.M. and 7:00-10:59 P.M.) in which each itinerary is allocated to three nests. With the OGEV model, the share of passengers assigned to each itinerary between an airport-pair for a given day of the week is as follows:

$$S_{i \subset k} = \sum_{j=k}^{k+M} P(i \subset k \mid N_j) P(N_j) \tag{6.1}$$

where $i \subset k$ indicates that itinerary i departs during time period k,

$M+1$ is the number of nests to which each itinerary is allocated,

N_j is nest j that includes alternative i (where $j = 1,2,\ldots,K+M$),

K is the total number of time periods,

$P(i \subset k \mid N_j)$ is the probability of choosing alternative i from nest j and

$P(N_j)$ is the (unobserved) probability of choosing nest j.

The components of equation (6.1) can be expanded in terms of the probability of choosing a specific itinerary, i, from nest j to which it is allocated as follows:

$$P(i \subset k \mid N_j) = \frac{\alpha_{j-k} \exp(\mu V_i)}{\sum_{i' \subset k' \in N_j} \alpha_{j-k'} \exp(\mu V_{i'})} \tag{6.1a}$$

where $\sum_{i' \subset k' \in N_j}$ is the summation over all itineraries, i', belonging to nest j,

α_{j-k} is the allocation parameter for an itinerary belonging to time period k assigned, in part, to nests $j = k, k+1, \ldots, k+M$ subject to $\alpha_i \geq 0$ and $\sum_{i=0}^{M} \alpha_i = 1$,

μ is the inverse logsum parameter associated with the nests and

$V_{i'}$ is the deterministic portion of the utility for alternative i'.

and the probability of choosing nest j is as follows:

$$P(N_j)=\frac{\exp\left(\frac{1}{\mu}\Gamma_{N_j}\right)}{\sum_{\forall N_m}\exp\left(\frac{1}{\mu}\Gamma_{N_m}\right)} \tag{6.1b}$$

where $\Gamma_{N_j}=\ln\left(\sum_{i'\subset k'\in N_j}\alpha_{j-k'}\exp(\mu V_{i'})\right)$.

The estimation results for the models represented in Figures 6.1 and 6.2 are reported in Table 6.1[51]. In both models, the inverse logsum parameter estimate is significantly larger than one (after adjustment) indicating increased itinerary competition within nests and the allocation parameters are significantly different from zero and one (after adjustment) indicating increased itinerary competition across time periods. Statistically, both models reject the two-level NL time model (Table 5.4) at the 0.001 level after adjustment. Additionally, these OGEV models are behaviorally superior to the nested logit model since they allow for differential itinerary competition across time period boundaries.

The three-allocation OGEV model significantly rejects the two-allocation OGEV model at all levels of significance after adjustment. Additionally, it is behaviorally superior since (for a given itinerary) it yields four differential "levels" of inter-itinerary

[51] In Table 6.1, all parameter estimates significant at the 0.05 level after the adjustment procedure. The significance of value function parameter estimates is with respect to their hypothesized value of zero; the significance of inverse logsum parameter estimates is with respect to their hypothesized value of one; the significance of allocation parameter estimates is with respect to their hypothesized values of zero and one.

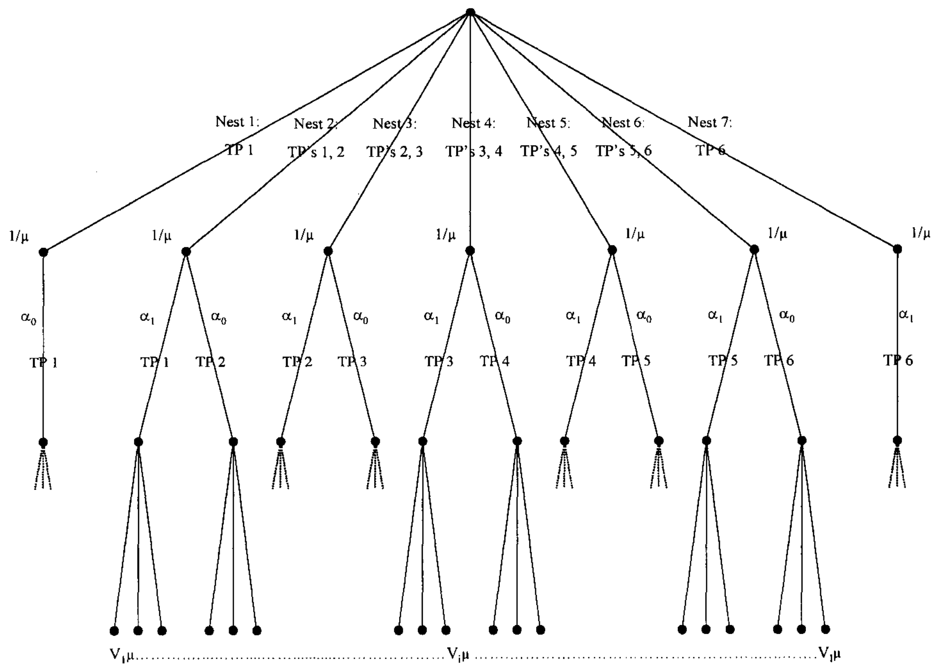

FIGURE 6.1: Two-Allocation OGEV Model Structure

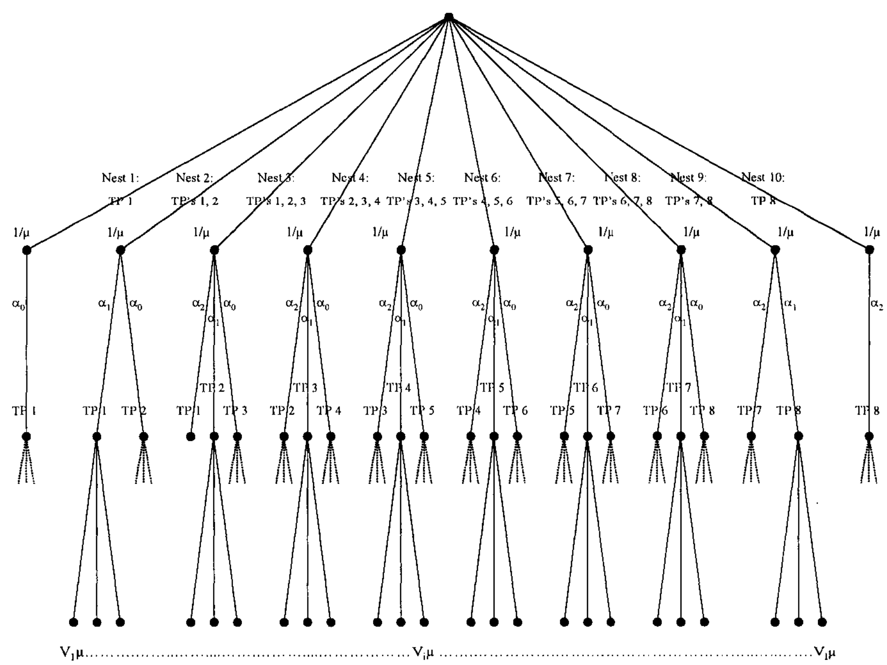

FIGURE 6.2: Three-Allocation OGEV Model Structure

TABLE 6.1: Itinerary Share Models: Two and Three-Allocation OGEV's

Explanatory Variables	Model 2-Allocation OGEV	Model 3-Allocation OGEV
Level-of-Service		
Nonstop Itinerary in Nonstop Market	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.6049	-1.5549
Single-Connect Itinerary in Nonstop Market	-2.3157	-2.2380
Double-Connect Itinerary in Nonstop Market	-5.3363	-5.1295
Direct Itinerary in Direct Market	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.6106	-0.5905
Double-Connect Itinerary in Direct Market	-3.1604	-3.0386
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.1618	-2.0988
Connection Quality		
Second-Best Connection	-0.3161	-0.2966
Second-Best Connection Time Difference	-0.0070	-0.0067
Distance Ratio	-0.0107	-0.0102
Best Connection Time Difference	-0.0046	-0.0044
Carrier Attributes		
Fare Ratio	-0.0051	-0.0050
Carrier Constants (Proprietary)	--------	--------
Code share	-1.4842	-1.4229
Aircraft Type		
Mainline Jet	0.0000	0.0000
Regional Jet	-0.3764	-0.3631
Propeller Aircraft	-0.3435	-0.3329
Departure Time		
5 - 6 A.M.	-0.1825	-0.1749
6 - 7 A.M.	0.0000	0.0000
7 - 8 A.M.	0.2335	0.2050
8 - 9 A.M.	0.3132	0.2808
9 -10 A.M.	0.3360	0.2993
10 - 11 A.M.	0.3069	0.2991
11 - 12 noon	0.2812	0.2532
12 - 1 P.M.	0.2978	0.2622
1 - 2 P.M.	0.2321	0.2052
2 - 3 P.M.	0.2285	0.1957
3 - 4 P.M.	0.2425	0.2348
4 - 5 P.M.	0.1883	0.1677
5 - 6 P.M.	0.1888	0.1686
6 - 7 P.M.	0.2507	0.2293
7 - 8 P.M.	0.0638	0.0756
8 - 9 P.M.	-0.0521	-0.0371
9 - 10 P.M.	-0.1814	-0.1612
10 – Midnight	-0.2607	-0.2412
Inverse Logsum Parameter	1.2607	1.3182
Alpha 1 (Allocation Parameter)	0.2215	0.0728
Alpha 2 (Allocation Parameter)	--------	0.2520
Log Likelihood at Zero	-2,173,197	-2,173,197
Log Likelihood at Convergence	-1,557,214	-1,556,869
Adjusted Log Likelihood at Convergence	-49,435	-49,424
Rho-square w.r.t. Zero	0.2834	0.2836

competition: itineraries sharing the same time period, itineraries in adjacent time periods, itineraries that are separated by two time periods, and itineraries that are separated by three or more time periods[52]. This confirms the belief that itinerary competition is differentiated by proximity in departure time (higher competition with close proximity). Examining the cross-elasticity equations of this three-allocation OGEV model for the change in the probability of itinerary j due to changes in an attribute of itinerary i illustrates these relationships. If itinerary i is three or more time periods away from itinerary j, the elasticity is given by:

$$\eta_{X_{im}}^{P_j} = \frac{\partial P_j}{\partial X_{im}} \frac{X_{im}}{P_j} = -X_{im}\beta_m P_i \tag{6.2}$$

where X_{im} is the value of itinerary i's m^{th} attribute and β_m is the parameter corresponding to attribute m. This is the same elasticity formula as that obtained for the MNL model. However, if itinerary j belongs to time period k and itinerary i belongs to time period $(k-2)$[53], the elasticity is given by:

[52] The two-allocation OGEV model allows for three differential levels of itinerary competition: itineraries sharing the same time period, itineraries in adjacent time periods, and itineraries that are separated by two or more time periods. The two-level NL model (with itineraries nested by departure time) allows for only two differential levels of itinerary competition: itineraries within the same nest and itineraries not in the same nest.

[53] An analogous formula applies if itinerary i belongs to time period $(k+2)$.

$$\eta^{P_j}_{X_{im}} = -X_{im}\beta_m\left[P_i + \frac{(\mu-1)P(i\,|\,N_k)P(j\,|\,N_k)P(N_k)}{P_j}\right] \tag{6.3}$$

This elasticity is larger in magnitude than the elasticity in equation (6.2) since μ must be larger than one. Note that the conditional probability of each itinerary given the nest (and hence the elasticity) is a function of its allocation parameter with respect to the nest[54]. Next, if itinerary j belongs to time period k and itinerary i belongs to time period $(k-1)$[55], the elasticity is given by:

$$\eta^{P_j}_{X_{im}} = -X_{im}\beta_m \times\left[P_i + \frac{(\mu-1)\left[P(j\,|\,N_k)P(i\,|\,N_k)P(N_k)+P(j\,|\,N_{k+1})P(i\,|\,N_{k+1})P(N_{k+1})\right]}{P_j}\right] \tag{6.4}$$

This elasticity is larger in magnitude than the elasticities in equations (6.2) and (6.3). Again, the magnitude of the elasticity is a function of the values of the allocation parameters. Finally, if itineraries j and i both belong to time period k, the elasticity is given by:

$$\eta^{P_j}_{X_{im}} = -X_{im}\beta_m \times\left[P_i + \frac{(\mu-1)\begin{bmatrix}P(j\,|\,N_k)P(i\,|\,N_k)P(N_k)+P(j\,|\,N_{k+1})P(i\,|\,N_{k+1})P(N_{k+1})\\ +P(j\,|\,N_{k+2})P(i\,|\,N_{k+2})P(N_{k+2})\end{bmatrix}}{P_j}\right] \tag{6.5}$$

This elasticity is larger in magnitude than the elasticities in equations (6.2 – 6.4).

---

[54] That is, $P(i\,|\,N_k)$ and $P(j\,|\,N_k)$ are functions of the underlying allocation parameters.

Hybrid OGEV specifications incorporating inter-itinerary competition along the carrier and/or level-of-service dimensions, under the time dimension, are presented in the following sections.

6.3 THREE-LEVEL NESTED LOGIT ORDERED GENERALIZED EXTREME VALUE MODEL

Following the results obtained from the three-level NL models presented in Chapter 5, three-level nested logit ordered generalized extreme value (NL-OGEV) models are developed where the OGEV model structure (described in Section 6.2) is incorporated in the upper level of the three-level NL model structure (described in Chapter 5). These models, which are new to the literature, have itineraries allocated to nests at the upper level according to an OGEV structure and nested at the lower level by carrier or level-of-service. Visual representations of a three-level time, level-of-service NL-OGEV model and a three-level time, carrier NL-OGEV model (with itineraries allocated to two OGEV nests) are shown in Figures 6.3 and 6.4, respectively. A visual representation of a three-level time, carrier NL-OGEV model (with itineraries allocated to three OGEV nests) is shown in Figure 6.5.

[55] An analogous formula applies if itinerary i belongs to time period $(k+1)$.

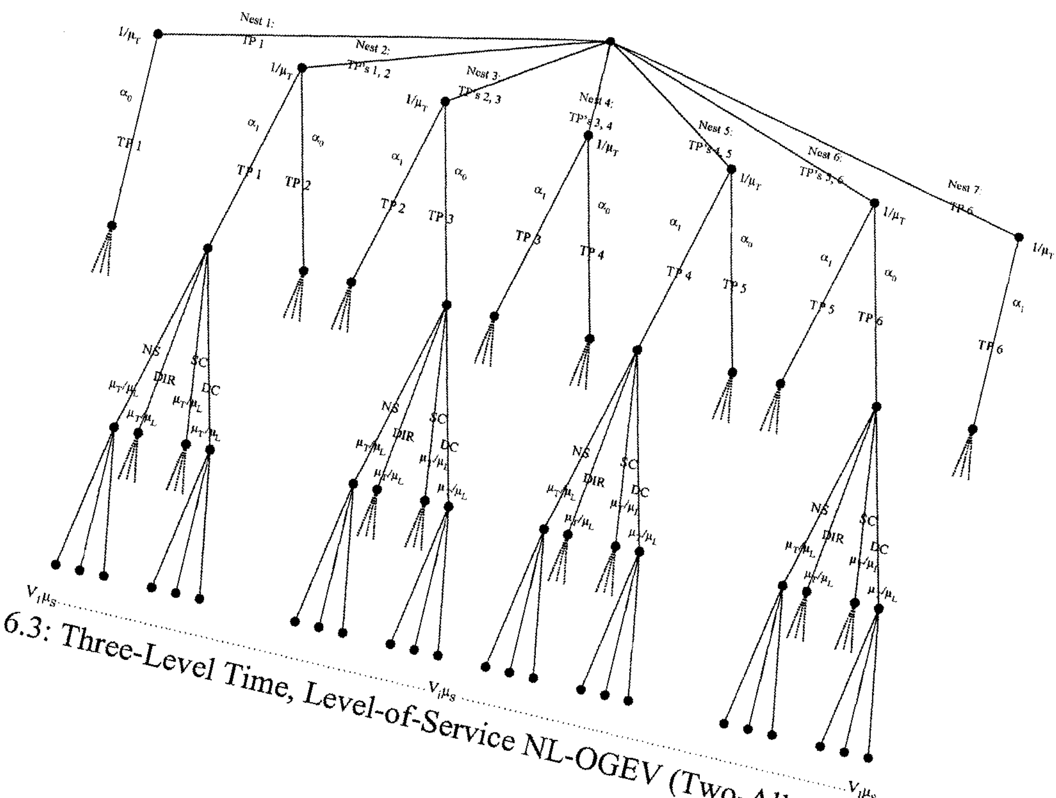

FIGURE 6.3: Three-Level Time, Level-of-Service NL-OGEV (Two-Allocation) Model Structure

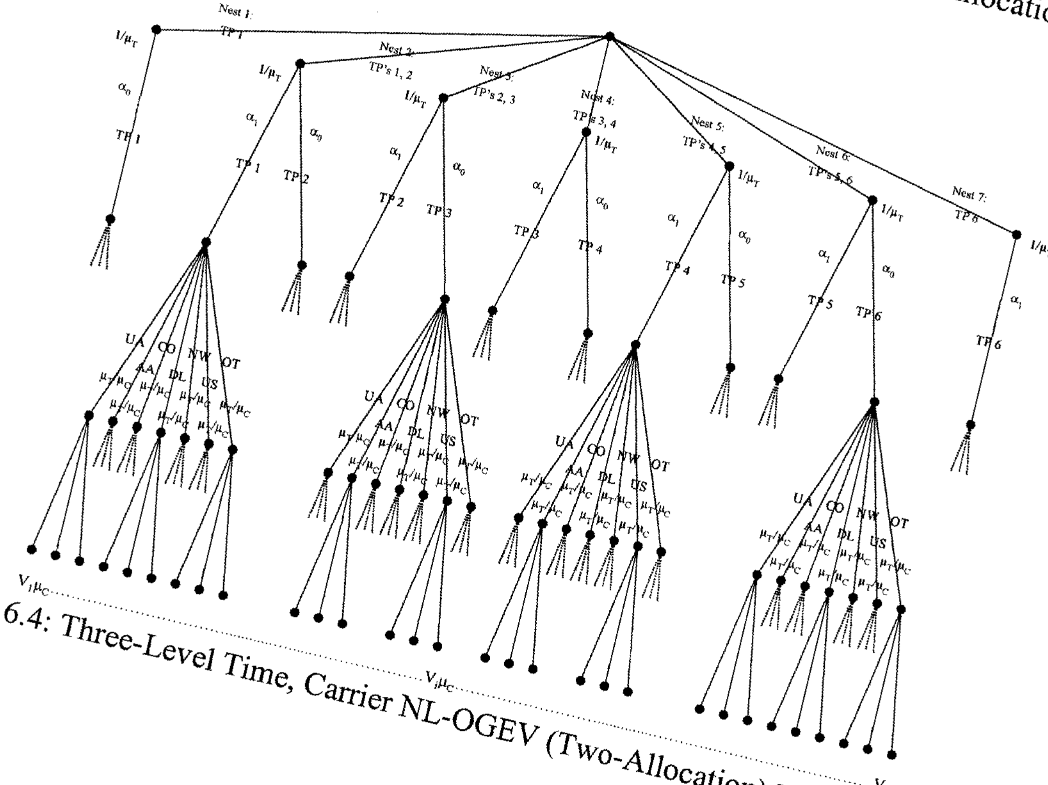

FIGURE 6.4: Three-Level Time, Carrier NL-OGEV (Two-Allocation) Model Structure

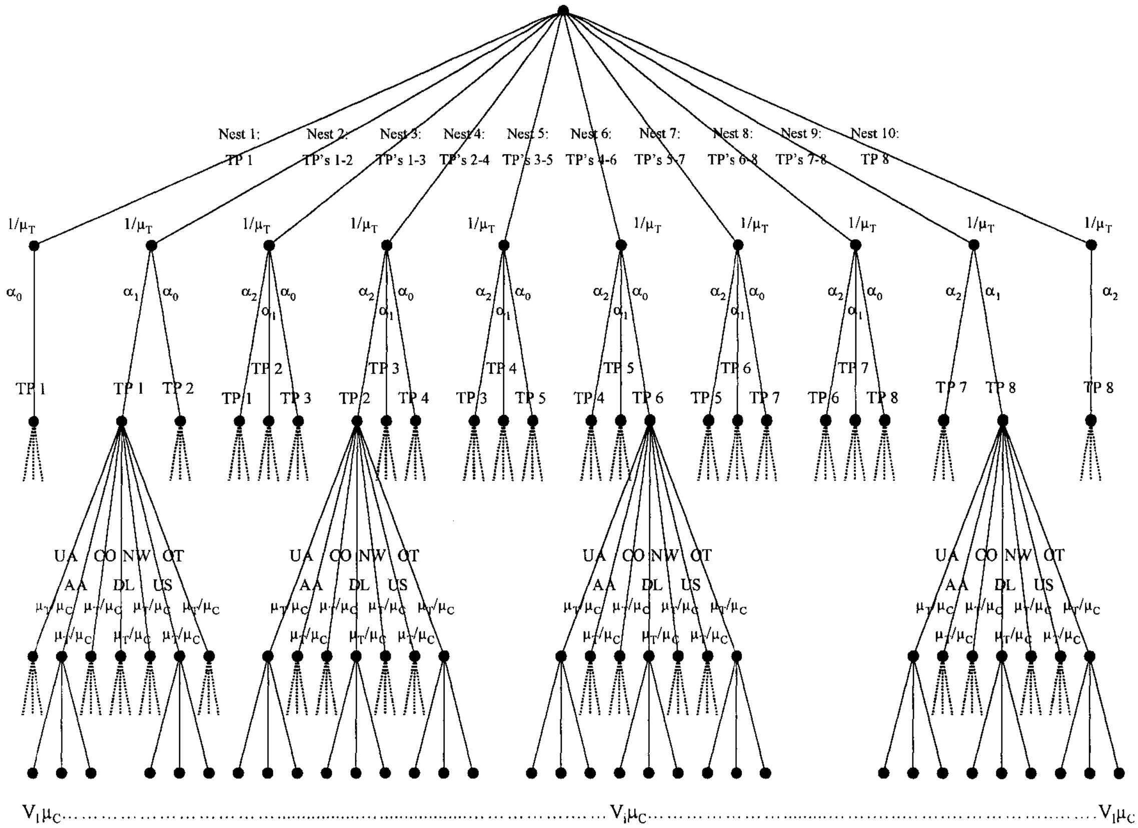

FIGURE 6.5: Three-Level Time, Carrier NL-OGEV (Three-Allocation) Model Structure

For the three-level time, carrier NL-OGEV models, the share of passengers assigned to each itinerary between an airport-pair for a given day of the week is given by:

$$S_{i\subset k,c} = \sum_{j=k}^{k+M} P(OGEV_j)P(c \mid OGEV_j)P(i \mid c, OGEV_j)$$

$$= \sum_{j=k}^{k+M} \frac{\exp\left(\dfrac{1}{\mu_{OGEV}}\Gamma_j\right)}{\sum_{j'\in J}\exp\left(\dfrac{1}{\mu_{OGEV}}\Gamma_{j'}\right)} \times \frac{\exp\left(\dfrac{\mu_{OGEV}}{\mu_{NL}}\Gamma_c\right)}{\sum_{c'\in C}\exp\left(\dfrac{\mu_{OGEV}}{\mu_{NL}}\Gamma_{c'}\right)} \times \frac{\alpha_{j-k}\exp(\mu_{NL}V_i)}{\sum_{i'\subset k',c}\alpha_{j-k'}\exp(\mu_{NL}V_{i'})} \quad (6.6)$$

where $P(OGEV_j)$ is the passenger share assigned to the j^{th} upper-level OGEV nest,

$P(c \mid OGEV_j)$ is the passenger share assigned to carrier $c's$ lower-level NL nest given the j^{th} upper-level OGEV nest,

$P(i \mid c, OGEV_j)$ is the passenger share assigned to itinerary i given lower-level carrier NL nest c and upper-level OGEV nest j,

μ_{OGEV} is the inverse logsum parameter associated with the upper-level OGEV nests,

μ_{NL} is the inverse logsum parameter associated with the lower-level carrier NL nests,

$$\Gamma_c = \ln\left(\sum_{i' \subset k', c} \alpha_{j-k'} \exp\left(\mu_{NL} V_{i'}\right)\right) \text{ and}$$

$$\Gamma_j = \ln\left(\sum_{c' \in C} \exp\left(\frac{\mu_{OGEV}}{\mu_{NL}} \Gamma_{c'}\right)\right).$$

and similarly for the three-level time, level-of-service NL-OGEV model. Consistent with the three-level NL model, the OGEV and nested logit inverse logsum parameter estimates must be greater than one and the NL inverse logsum parameter must be larger than the OGEV inverse logsum parameter.

The estimation results for the three-level NL-OGEV models are reported in Table 6.2[56]. The three-level time, level-of-service (two-allocation) NL-OGEV model does not reject the two-allocation OGEV model (Table 6.1) after adjustment. However, it does improve upon the two-allocation OGEV model before adjustment, both its inverse logsum parameters are significant after adjustment (they are significantly different from each other after adjustment as well), and the allocation parameter is significant after adjustment. These results are similar to the results of Section 5.6 where the three-level time, level-of-service NL model was shown to only marginally improve upon the two-level time NL model.

TABLE 6.2: Itinerary Share Models: Three-Level NL-OGEV's and Three-Level WNL-OGEV

Explanatory Variables	Model			
	3-Level NL-OGEV (2-Allocation): Time|LOS	3-Level NL-OGEV (2-Allocation): Time|Carrier	3-Level NL-OGEV (3-Allocation): Time|Carrier	3-Level WNL-OGEV (2-Allocation): T|C, T|L
Level-of-Service				
Nonstop Itinerary in Nonstop Market	0.0000	0.0000	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.6459	-1.6060	-1.5840	-1.6469
Single-Connect Itinerary in Nonstop Market	-2.3144	-2.3175	-2.2837	-2.3053
Double-Connect Itinerary in Nonstop Market	-5.4438	-4.9888	-4.8732	-5.1164
Direct Itinerary in Direct Market	0.0000	0.0000	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.5766	-0.6203	-0.6148	-0.5710
Double-Connect Itinerary in Direct Market	-3.2063	-3.0155	-2.9596	-3.0597
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.2079	-2.1151	-2.0968	-2.1635
Connection Quality				
Second-Best Connection	-0.3123	-0.2144	-0.1936	-0.2084
Second-Best Connection Time Difference	-0.0069	-0.0059	-0.0057	-0.0058
Distance Ratio	-0.0107	-0.0107	-0.0101	-0.0105
Best Connection Time Difference	-0.0046	-0.0048	-0.0048	-0.0047
Carrier Attributes				
Fare Ratio	-0.0051	-0.0034	-0.0034	-0.0033
Carrier Constants (Proprietary)	--------	--------	--------	--------
Code share	-1.4720	-1.4865	-1.4594	-1.4593
Aircraft Type				
Mainline Jet	0.0000	0.0000	0.0000	0.0000
Regional Jet	-0.3749	-0.3947	-0.3894	-0.3892
Propeller Aircraft	-0.3409	-0.3176	-0.3138	-0.3118
Departure Time				
5 – 6 A.M.	-0.1829	-0.1911	-0.1922	-0.1894
6 – 7 A.M.	0.0000	0.0000	0.0000	0.0000
7 – 8 A.M.	0.2356	0.2638	0.2308	0.2675
8 – 9 A.M.	0.3151	0.3407	0.3092	0.3430
9 –10 A.M.	0.3380	0.3622	0.3193	0.3651
10 - 11 A.M.	0.3047	0.3424	0.3363	0.3437
11 - 12 noon	0.2790	0.3205	0.2747	0.3221
12 – 1 P.M.	0.2958	0.3370	0.2833	0.3385
1 – 2 P.M.	0.2327	0.2645	0.2196	0.2678
2 – 3 P.M.	0.2288	0.2572	0.2067	0.2602
3 – 4 P.M.	0.2415	0.2718	0.2534	0.2735
4 – 5 P.M.	0.1911	0.2201	0.1878	0.2246
5 – 6 P.M.	0.1912	0.2060	0.1840	0.2103
6 – 7 P.M.	0.2523	0.2678	0.2471	0.2703
7 – 8 P.M.	0.0600	0.0948	0.1058	0.0910
8 – 9 P.M.	-0.0531	**-0.0014**	**0.0122**	**-0.0014**
9 – 10 P.M.	-0.1794	-0.1141	-0.0944	-0.1089
10 – Midnight	-0.2585	-0.2018	-0.1837	-0.1986
Upper-Level OGEV Inverse Logsum Parameter (Time)	1.2325	1.0896	1.1087	1.0698

[56] The three-level time, level-of-service (three-allocation) NL-OGEV model did not yield reasonable results. Additionally, in Table 6.2, parameter estimates in bold not significant at the 0.05 level after the adjustment procedure.

Lower-Level NL Inverse Logsum Parameter (Carrier)	--------	1.4539	1.5196	1.4778
Lower-Level NL Inverse Logsum Parameter (LOS)	1.2718	--------	--------	**1.5765**
Alpha 1 (Allocation Parameter)	0.1903	0.1787	0.0205	0.1648
Alpha 2 (Allocation Parameter)	--------	--------	0.2425	--------
Weight Parameter (Time|Carrier Structure)	--------	--------	--------	**0.9012**
Log Likelihood at Zero	-2,173,197	-2,173,197	-2,173,197	-2,173,197
Log Likelihood at Convergence	-1,557,199	-1,553,430	-1,552,661	-1,553,397
Adjusted Log Likelihood at Convergence	-49,435	-49,315	-49,291	-49,314
Rho-square w.r.t. Zero	0.2835	0.2852	0.2855	0.2852

The three-level time, carrier (two-allocation) NL-OGEV model rejects the three-level time, carrier NL model (Table 5.5) and the two-allocation OGEV model (Table 6.1) at the 0.001 level (before and after adjustment).

The time, carrier (three-allocation) NL-OGEV model rejects the three-level time, carrier NL model (Table 5.5), the three-allocation OGEV model (Table 6.1) and the time, carrier (two-allocation) NL-OGEV model (Table 6.2) at the 0.001 level (before and after adjustment). Of the models discussed up to this point in the dissertation, the three-level time, carrier (three-allocation) NL-OGEV model has the best overall model statistics (by far). Additionally, its inverse logsum and allocation parameter estimates are all highly significant after adjustment. This indicates a high level of competition among itineraries flown by the same carrier within the same, adjacent or plus/minus two time periods.

These strong three-level NL-OGEV results indicate that imposing an upper-level OGEV structure and a lower-level NL structure on the itinerary competition dynamic dramatically improves upon the more rigid NL model structure. In addition to reinforcing the finding from Chapter 5 that the within carrier and (to a lesser extent) within level-of-service competition dynamic should be conditioned by time period, the significance of the OGEV allocation parameters in these models indicate that itineraries do indeed have several differential levels of competition (with respect to departure time) with other itineraries. That is, the closer itineraries are to each other (with respect to departure time) the more they will compete with each other.

In the following section, a three-level weighted nested logit OGEV model is estimated. This model combines the results of three-level NL-OGEV models.

6.4 THREE-LEVEL WEIGHTED NESTED LOGIT ORDERED GENERALIZED EXTREME VALUE MODEL

The three-level weighted nested logit ordered generalized extreme value (WNL-OGEV) model combines the OGEV model with the three-level WNL model (described in Chapter 5). This allows for the simultaneous estimation of parallel three-level NL-OGEV nesting structures with a weight parameter indicating the relative importance of each structure.

Due to the marginal significance of the three-level time, level-of-service (two-allocation) NL-OGEV model and the significance of the three-level time, carrier (two-allocation) NL-OGEV model, a three-level WNL-OGEV (two-allocation) model is estimated with parallel three-level NL-OGEV structures for time, carrier and time, level-of-service. That is, within each structure, itineraries are nested at the upper level according to an OGEV specification and are nested at the lower level by level-of-service or carrier. Figure 6.6 gives a visual representation of this model and its mathematical structure is given by:

$$S_{i\subset k,c,l} = w_{t,c}\sum_{j=k}^{k+M} P(OGEV_j)P(c\mid OGEV_j)P(i\mid c,OGEV_j) + w_{t,l}\sum_{j=k}^{k+M} P(OGEV_j)P(l\mid OGEV_j)P(i\mid l,OGEV_j) \tag{6.7}$$

where $w_{t,c}$ is the weight given to the time | carrier structure,

$w_{t,l} = 1 - w_{t,c}$ is the weight given to the time | level-of-service structure,
the first summation is identical to equation (6.6) and
the second summation is identical to equation (6.6) except for the substitution of level-of-service, l, for carrier, c.

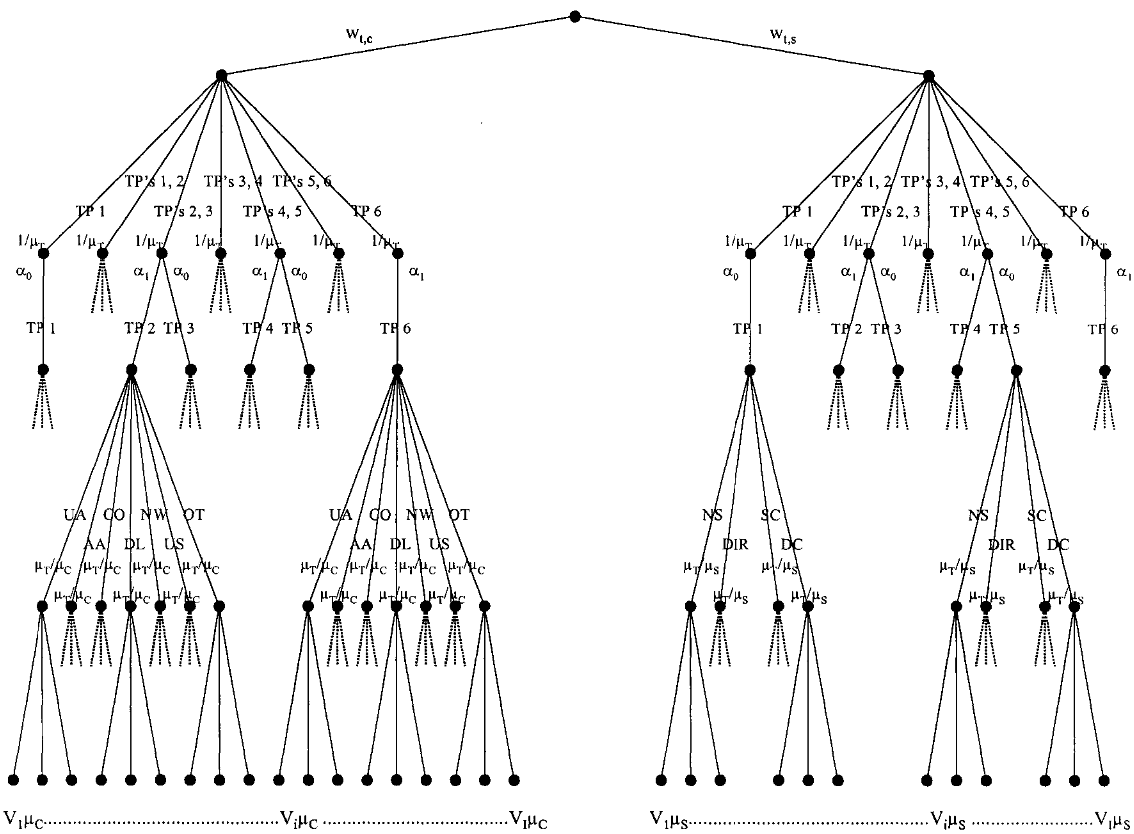

FIGURE 6.6: Three-Level Time | Carrier, Time | Level-of-Service WNL-OGEV (Two-Allocation) Model Structure

The estimation results for this model are reported in Table 6.2 The upper-level OGEV inverse logsum parameter is significant at the 0.001 level after adjustment and the lower-level NL inverse logsum parameter for carrier is significant (at the 0.001 level) after adjustment indicating a high level of competition among itineraries flown by a carrier within the same or adjacent time periods. However, the lower-level inverse logsum parameter for level-of-service is only significant at the 0.10 level after adjustment. Additionally, after adjustment, the weight on the time, carrier structure is only significantly different from one at the 0.10 level. Similar to the findings from Chapter 5 (with respect to the three-level WNL model), these results indicate that the time, carrier side of the model "dominates". Finally, after adjustment, the model is only marginally better than the three-level time, carrier (two-allocation) NL-OGEV model.

6.5 SUMMARY AND CONCLUSIONS

In this chapter, models are estimated capturing a realistic inter-itinerary competition dynamic along the departure time dimension. Each of the models tests the hypothesis that air-travel itineraries (for a given airport-pair-day-of-the-week) exhibit proximate covariance; that is, the amount of competition (substitution) between itineraries increases as the difference in their departure times decreases. The nested logit models estimated in Chapter 5 are not capable of capturing the phenomenon.

Two and three-allocation OGEV models are estimated. Both of these models show that air-travel itineraries do indeed exhibit the proximate covariance property. Thus, the hypothesis that nested logit models (with itineraries nested by departure time) are adequate to describe the inter-itinerary competition dynamic along the departure time dimension is rejected. The three-allocation OGEV model captures, for each itinerary, four differential levels of competition with respect to other itineraries in its airport-pair along the departure time dimension (depending on the proximity of the itineraries' departure times). These OGEV models are the first in the aviation demand literature to capture the proximate covariance property.

Advanced hybrid OGEV models are estimated (these models are new to the literature) incorporating an OGEV structure at the upper level with a GEV (in particular, NL and WNL) structure on the lower level. In addition to capturing the proximate covariance property of air-travel itineraries, these models also measure differential inter-itinerary competition dynamics along the carrier and/or level-of-service dimensions. Of these models, the three-level time, carrier (three-allocation) NL-OGEV model yielded superior model statistics and behavioral interpretations. This model is the preferred specification of the dissertation.

Finally, the models estimated in this chapter are shown to have advantages over the more restrictive model structures estimated in the previous chapters and are shown to outperform these models with respect to statistical tests and behavioral interpretations. It is also interesting to note that the results of this chapter "parallel" those from Chapter 5. First, the "superior" model in both chapters has a specification where itineraries are nested at the lower level by carrier within an upper level departure time structure. Second, the hybrid OGEV models in this chapter with level-of-service in their specification do not yield significant results (similar to the nested logit models of Chapter 5 with level-of-service in their specification not yielding significant results).

CHAPTER 7: SUMMARY AND CONCLUSIONS

7.1 SUMMARY AND RESULTS

The research detailed in this dissertation describes the development of aggregate air-travel itinerary share models. These models forecast the number of passengers expected to travel on each itinerary between any airport-pair (conditional on the forecasted airport-pair volume) and are essential to strategic planning in the aviation industry.

Using state-of-the-art logit share techniques, this is the first study to model aviation demand at the itinerary level. Modeling itinerary-level demand is critical because itineraries are the products that are ultimately purchased by passengers and itinerary-level attributes directly influence demand. As a result, this study provides the most comprehensive and thorough explanation of air-carrier demand to-date and fills a gap in the airline forecasting literature.

The motivation for this research is twofold given that the overarching goal is to model itinerary shares: 1) to understand the impact of different air-carrier service attributes on itinerary share and 2) to understand the underlying competitive dynamic between itineraries.

To address the first goal, eighteen aggregate multinomial logit models are developed, each covering a major region-pair of the United States. In specifying these models, independent variables are used relating each itinerary with its airport-pair demand share. Due to the fact that these models employ itinerary-level data (which has previously not been reported in the literature), the development of these models provides a better understanding of the relative importance of different service factors on aviation demand compared to previous studies. Detailed analysis is performed on the relative importance of itinerary level-of-service, connection quality, aircraft type and size, departure time, carrier presence, fares and carrier on itinerary share.

Level-of-service is shown to have a strong impact on itinerary share with each reduction in level-of-service from the best available (in an airport-pair) substantially reducing the value (and hence market share) of the associated itinerary. This confirms the well-known belief that, if possible, people prefer to avoid connections in their air-travel (all things being equal). For itineraries that do involve a connection however, itineraries with shorter ground times and less circuitous routings are shown to be preferred. Increased carrier presence in an airport-pair is shown to positively influence a carrier's itinerary shares in that airport-pair. This is undoubtedly a result of the many marketing advantages (most notably, the advantage of offering customers an attractive frequent flyer program) afforded carriers with significant presence in a market. As expected, lower-cost itineraries are shown to be preferred to higher-cost itineraries (all things being equal). The results also indicate that passengers prefer larger aircraft to smaller aircraft types and (within an aircraft type) prefer larger planes to smaller planes. This confirms the belief that passengers prefer larger airplanes due to their higher levels of comfort and perceived levels of safety. Finally, the results show that passengers have varying preferences for departure times (depending on the region-pair of the different airport-pairs). However, in general, early morning and late evening itineraries are not preferred.

These multinomial logit models were implemented as a component of a major U.S. carrier's itinerary share model and were validated using onboard segment-level data. The logit-based models consistently and significantly improved upon the carrier's previous QSI-based model with respect to passenger accuracy. This demonstrates that logit-based models are not only conceptually superior to QSI-based models (as described in Chapter 1), but are superior in practice as well.

The second goal of this dissertation is to model the underlying competitive dynamic among air-travel itineraries. It is hypothesized that inter-itinerary competition among itineraries is not "uniform", but is rather differentiated by proximity in departure time, carrier, level-of-service or a combination of these dimensions. MNL models cannot capture this phenomenon and, as a result, this hypothesis is tested via the development and estimation of variations on the nested logit model. These models group (nest) itineraries according to combinations of the above-mentioned dimensions and measure differential levels of inter-itinerary competition as a function of the value of their estimated inverse logsum parameter(s). Some of these nested logit model structures have never been used in the aviation demand literature, and some are new to the logit share literature in general.

Two-level nested logit models are estimated showing that itineraries sharing a common departure time period or carrier (but not level-of-service) exhibit a strong amount of competition amongst themselves. These models reject the hypothesis that the MNL model is adequate for modeling air-travel itinerary shares. Using these results, a two-level weighted nested logit model with parallel time and carrier nesting structures is estimated. It significantly rejects the standard two-level NL models, and has advantages over the more restrictive NL model structure since it allows for the inter-itinerary competition dynamic to be measured across two dimensions. The two-level WNL model cannot capture the inter-itinerary competition dynamic that may exist among itineraries sharing a common attribute within another attribute, however. To address this, three-level nested logit models are estimated. The results of these models show that itineraries sharing a common departure time period have a moderate amount of competition amongst themselves, while itineraries sharing both time period and carrier (and to a lesser extent time period and level-of-service) exhibit a higher level of competition amongst themselves. Finally, three-level WNL and NWNL models are introduced. These GEV models, new to the literature, simultaneously incorporate the three dimensions of departure time, carrier and level-of-service. Their estimation results only marginally improve upon the three-level NL model with itineraries nested at the upper level by departure time and at the lower level by carrier, however.

These nested logit specifications represent an important step forward for air-travel itinerary share modeling since they demonstrate that inter-itinerary competition is not "uniform". These models (those that employ itinerary departure time in their nesting structure) are hypothesized to inadequately describe the inter-itinerary competition dynamic along the departure time dimension, however. This is because they group itineraries by arbitrary and rigid time periods.

It is hypothesized that the competition (substitution) among air-travel itineraries (for a given airport-pair-day-of-the-week) along the departure time dimension increases (with several breakpoints) the closer itineraries are to each other with respect to departure time. To capture this more realistic approach to itinerary share modeling, ordered generalized extreme value (and variations thereof) models (Small 1987) are estimated. OGEV models have the desired property of proximate covariance (Small 1987) where itineraries that are closer to each other by departure time exhibit greater correlation with each other than with itineraries that are separated in time. The level of correlation between itineraries increases the closer they are to each other.

Two and three-allocation OGEV models are estimated. Both of these models show that air-travel itineraries do indeed exhibit the proximate covariance property. Thus, the hypothesis is rejected that nested logit models (with itineraries nested by departure time) are adequate to describe the inter-itinerary competition dynamic along the departure time dimension. The three-allocation OGEV model captures (for each itinerary) four differential levels of competition with other itineraries in its airport-pair along the departure time dimension (depending on the proximity of the itineraries' departure times). These OGEV models are the first in the literature to capture the proximate covariance property of airline demand.

Because it is shown (via three-level nested logit models) that the within carrier or within level-of-service competition dynamic is especially strong within narrow periods of time, advanced hybrid OGEV models are developed. These models are new to the literature and incorporate the traditional OGEV model structure (at the upper level) with other GEV components such as the NL model (at the lower level). These models capture complicated – yet realistic – competitive structures since (in addition to capturing the proximate covariance property) they measure differential inter-itinerary competition dynamics along the carrier and/or level-of-service dimensions within narrow departure time periods. Several of these models are estimated. Of these, the three-level time, carrier (three-allocation) NL-OGEV model yields superior model statistics and behavioral interpretations. This model is the preferred specification of the dissertation.

Finally, it is interesting to note that none of the variations of the nested logit or OGEV models that have itineraries nested by level-of-service in their specification yielded significant results. Thus, it appears that the underlying competition among air-travel itineraries can almost fully be described by nesting itineraries by the departure time and carrier dimensions. This assertion is discussed in the following section.

7.2 LIMITATIONS OF RESEARCH AND FUTURE DIRECTIONS

The research contained in this dissertation has several limitations. First, the models developed are itinerary "share" rather than itinerary "choice" models. This is because even though the bookings data (obtained from CRS data sources) employed in this study is based on the choices of individual travelers, it does not include any information on the socio-economic or demographic characteristics of the individual that made the booking or any trip-related characteristics of the booking (such as business vs. leisure, number of days booked in advance of departure, duration of stay). This is compensated for by the fact that comprehensive individual itinerary bookings data is employed. This breadth of data allows for itinerary-level demand models to be estimated for the first time as well as the use of specifications that have never been employed before.

Regardless, this absence of data about individual travelers (and trip characteristics) introduces measurement error that has the potential to produce bias in the model parameters. Disaggregate bookings data that includes personal and trip characteristics should be obtained so that the effect of these variables can be estimated via modification of the value function or market segmentation. Such models would yield behavioral insights into the influence of different itinerary service characteristics on different groups of the population.

Similarly, it is believed that the CRS data employed in this dissertation captured around 90% of the U.S. Domestic air-travel bookings during the study period. It is assumed that the remaining 10% of bookings not captured by this data source have the same underlying characteristics as the CRS bookings. This assumption may not be warranted since the non-CRS bookings are more likely to be leisure bookings than the bookings contained in CRS. Further, this problem will worsen since industry-wide bookings data (CRS data sources) are capturing less and less bookings.

Second, fare is obviously an important determinant of itinerary choice (especially for leisure travelers). Detailed fare-class (or even itinerary-level) fare data was not available for this study, however. Rather, average fare by carrier (across itineraries) for each airport-pair is employed as an independent variable in the presented itinerary share models. This is the best fare data currently available for a revealed preference air-carrier demand allocation study.

Third, all models are estimated on data that precede the September 11, 2001 terrorist attacks. Obviously, much has changed in the industry since this event occurred.

However, use of this data for forecasting is likely to be valid since the underlying behavioral relationships between itinerary characteristics (*e.g.* departure time, level-of-service, connection quality, equipment used) and itinerary shares is believed to be robust. Chapter 4 presents validation results that confirm this belief.

Fourth, as demonstrated in Chapters 5 and 6, models employing specifications with itineraries nested by level-of-service generally were not significant. However, because itineraries of a given level-of-service exhibit many common characteristics, it is reasonable to expect that they should have increased competition amongst themselves. It is possible that the insignificance of these models is related to the itinerary generation engine that was employed in creating the itineraries for these models. In particular, in some nonstop markets this engine creates hundreds of connecting (single and double-connect) itineraries. Surely, passengers do not systematically consider each of these alternatives in their itinerary selection process. To address this in the future, different choice set generation logic should be tested and/or models employing probabilistic choice set consideration should be employed.

Fifth, each of the models estimated in this dissertation group airport-pairs by time-zone combination (*e.g.* East-East, East-Central, East-Mountain, East-West). However, other airport-pair groupings are also possible such as grouping by stage length, high yield vs. low yield, the size of the airports or a combination of these measures.

Finally, though the purpose of this dissertation is to present the benefits of logit-based air-travel itinerary share models, it should be noted that there are certain advantages to QSI-based models. The advantages of QSI models lie in their simplicity, intuitive interpretation of model parameters and adequate forecasting ability. Logit model estimations on the other hand (in particular, the models presented in this dissertation) require large amounts of data, specialized software and a trained analyst to estimate the model and interpret the model results. Additionally, by definition, when using a logit-based itinerary share model the analyst cannot "turn off" variables as is possible with QSI-based models. However, due to their simultaneous and optimal estimation of parameters (allowing for variable interactions to be captured), forecasting superiority and (most importantly) ability to model complex itinerary substitution patterns (allowing for rapidly changing market conditions to be accurately modeled), the advantages of logit-based models far outweigh the disadvantages.

7.3 CONCLUSIONS

This dissertation details the development of aggregate air-travel itinerary share models. These models forecast the number of passengers expected to travel on each itinerary between any airport-pair (conditional on the forecasted airport-pair volume). As the first study to model aviation demand at the itinerary level, this work is an important step forward in aviation demand modeling.

The first goal of the dissertation was to understand the impact of different air-carrier service attributes on itinerary share. The models incorporated in the presented research contain independent variable specifications linking itinerary and airline characteristics to itinerary share. Explanatory variables used in this study such as level-of-service indicators differing by market type, connection quality variables, departure time variables, and equipment size and type variables have previously not been reported in the literature. Additionally, the presented logit-based models are shown to outperform a QSI-based model.

The second goal of the dissertation was to model the underlying competitive dynamic among air-travel itineraries. To accomplish this, increasingly complex generalized extreme value specifications (in particular, variations of the nested logit and ordered generalized extreme value models) are used. Some of these structures have never been used in the aviation demand literature and some are new to the logit share literature in general. These new models may be applicable to areas outside of aviation demand as well. The advanced models estimated in this dissertation have advantages over the more restrictive model structures and outperform these models with respect to forecasting accuracy, statistical tests and behavioral interpretations, leading to a clearer understanding of the air-travel itinerary competition dynamic.

REFERENCES

Abrahams, M. 1983. A Service Quality Model of Air Travel Demand: An Empirical Study. *Transportation Research – Part A* 17, no. 5:385-393.

Algers, S. and M. Beser. 2001. Modelling Choice of Flight and Booking Class – A Study using Stated Preference and Reveled Preference Data. *International Journal of Services Technology and Management* 2, no.1/2:28-45.

Anderson, J. E. and M. Kraus. 1981. Quality of Service and the Demand for Air Travel. *Review of Economics and Statistics* 63, no. 4:533-540.

Aptech Systems, Inc. 2005. *GAUSS*. Maple Valley, WA, USA.

Ashford, N. and M. Benchemam. 1987. Passengers' Choice of Airport: An Application of the Multinomial Logit Model. *Transportation Research Record* 1147:1-5.

Augustinus, J. G. and S. A. Demakopoulos. 1978. Air Passenger Distribution Model for a Multiterminal Airport System. *Transportation Research Record* 673:176-180.

Basar, G. and C. Bhat. 2004. A Parameterized Consideration Set Model for Airport Choice: An Application to the San Francisco Bay Area. *Transportation Research – Part B* 38, no. 10:889-904.

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

Ben-Akiva, M., T. Morikawa and F. Shiroishi. 1989. Analysis of the Reliability of Stated Preference Data in Estimating Mode Choice Models. In *Selected Proceedings of the Fifth World Conference on Transport Research* 4:263-277. Ventura, CA: Western Periodicals Co.

Berry, S. T. 1990. Airport Presence as Product Differentiation. *American Economic Review* 80, no. 2:394-399.

Bhat, C. R. 1998. Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips. *Transportation Research – Part B* 32, no. 6:361-371.

Borenstein, S. 1989. Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry. *RAND Journal of Economics* 20, no. 3:344-365.

Borenstein, S. 1990. Airline Mergers, Airport Dominance, and Market Power. *American Economic Review* 80, no. 2:400-404.

Bresnahan, T. F., S. Stern and M. Trajtenberg. 1997. Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s. *RAND Journal of Economics* 28, no. 0, Special Issue:S17-S44.

Brown, S. L. and W. S. Watkins. 1968. The Demand for Air Travel: A Regression Study of Time-Series and Cross-Sectional Data in the U.S. Domestic Market. *Highway Research Record* 213:21-34.

Brown, S. L. and W. S. Watkins. 1971. Measuring Elasticities of Air Travel from New Cross-sectional Data. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 310-315. Washington, D.C.: American Statistical Association.

Bruning, E. and V. Rueda. 2000. The Importance of Frequent Flyer Programs as a Barrier to Entry into Air Travel Markets. *Journal of Transportation Law, Logistics and Policy* 67, no. 4:367-380.

Coldren, G. M. and F. S. Koppelman. 2005. Modeling the Competition among Air-travel Itinerary Shares: GEV Model Development. *Transportation Research – Part A* 39, no. 4:345-365.

Coldren, G. M. and F. S. Koppelman. 2005. Modeling the Proximate Covariance Property of Air-travel Itineraries along the Time of Day Dimension. *Transportation Research Record*, In Press.

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

Corsi, T., M. Dresner and R. Windle. 1997. Air Passenger Forecasts: Principles and Practices. *Journal of the Transportation Research Forum* 36, no. 2:42-62.

Daly, A. 1987. Estimating "Tree" Logit Models. *Transportation Research – Part B* 21, no. 4:251-267.

Data Base Products, Inc. 2000. 2001. *Superset*. Dallas, TX, USA.

De Vany, A. 1974. The Revealed Value of Time in Air Travel. *Review of Economics and Statistics* 56, no. 1:77-82.

De Vany, A. S. 1975. The Effect of Price and Entry Regulation on Airline Output, Capacity and Efficiency. *The Bell Journal of Economics* 6, no. 1:327-345.

De Vany, A. S. and E. H. Garges. 1972. A Forecast of Air Travel and Airport and Airway Use in 1980. *Transportation Research* 6, no. 1:1-18.

Domencich, T. A. and D. McFadden. 1975. *Urban Travel Demand – A Behavioral Analysis*. New York: American Elsevier Publishing Company, Inc.

Douglas, G. W. 1971. Excess Capacity, Service Quality and the Structure of Airline Fares. In *Proceedings of the Transportation Research Forum*, 109-122.

Douglas, G. W. and J. C. Miller III. 1974. *Economic Regulation of Domestic Air Transport: Theory and Policy*. Washington, D.C.: The Brookings Institution.

Douglas, G. W. and J. C. Miller III. 1974. Quality Competition, Industry Equilibrium, and Efficiency in the Price-constrained Airline Market. *American Economic Review* 64, no. 4:657-669.

Dresner, M. and R. Windle. 1992. Airport Dominance and Yields in the U.S. Airline Industry. *The Logistics and Transportation Review* 28, no. 4:319-339.

Dresner, M., J-S C. Lin and R. Windle. 1996. The Impact of Low-cost Carriers on Airport and Route Competition. *Journal of Transport Economics and Policy* 30, no. 3:309-328.

Ellis, J. M. 1998. The New Role of Frequent Flyer Programs. In *Handbook of Airline Marketing*, ed. G. F. Butler and M. R. Keller, 577-585. Aviation Week Group.

English, J. M. and G. L. Kernan. 1976. The Prediction of Air Travel and Aircraft Technology to the Year 2000 Using the Delphi Method. *Transportation Research* 10, no. 1:1-8.

Expedia, Inc. 2000. 2001. *Expedia.com*. Bellevue, WA, USA.

Furuichi, M. and F. S. Koppelman. 1994. An Analysis of Air Travelers' Departure Airport and Destination Choice Behavior. *Transportation Research – Part A* 28, no. 3:187-195.

Galileo International 2000. 2001. *Galileo CRS*. Parsippany, NJ, USA.

Ghobrial, A. and S. Y. Soliman. 1992. An Assessment of Some Factors Influencing the Competitive Strategies of Airlines in Domestic Markets. *International Journal of Transport Economics* 19, no. 3:247-258.

Gordon, S. and R. De Neufville. 1973. Design of Air Transportation Networks. *Transportation Research* 7, no. 3:207-222.

Gosling, G. D. 1994. *Development of a Recommended Forecasting Methodology for Aviation System Planning in California*. University of California, Berkeley – Institute of Transportation Studies Research Report UCB-ITS-RR-94-7. Berkeley: Institute of Transportation Studies, University of California at Berkeley.

Harvey, G. 1987. Airport Choice in a Multiple Airport Region. *Transportation Research – Part A* 21, no. 6:439-449.

Hess, S. and J. W. Polak. 2005. Mixed Logit Modelling of Airport Choice in Multi-airport Regions. *Journal of Air Transport Management* 11, no. 2:59-68.

Hu, M. Y., R. S. Toh and S. Strand. 1988. Frequent-flier Programs: Problems and Pitfalls. *Business Horizons* 31, no. 4:52-57.

Huston, J. H. and R. V. Butler. 1988. The Effects of Fortress Hubs on Airline Fares and Service: The Early Returns. *Logistics and Transportation Review* 24, no. 3:203-215.

Ippolito, R. A. 1981. Estimating Airline Demand with Quality of Service Variables. *Journal of Transport Economics and Policy* 15, no. 1:7-15.

Jung, J. M. and E. T. Fujii. 1976. The Price Elasticity of Demand for Air Travel: Some New Evidence. *Journal of Transport Economics and Policy* 10, no. 3:257-262.

Kanafani, A. 1980. Price Elasticities of Nonbusiness Air Travel Demand. *Transportation Engineering Journal* 106, no. TE2:217-225.

Kanafani, A. 1981. Aircraft Technology and Network Structure in Short-haul Air Transportation. *Transportation Research – Part A* 15, no. 4:305-314.

Kanafani, A. and P. Chang. 1979. *A Systems Analysis of Short Haul Air Transportation*. University of California, Berkeley – Institute of Transportation Studies Research Report UCB-ITS-RR-79-9. Berkeley: Institute of Transportation Studies, University of California at Berkeley.

Kanafani, A. K. and S-L Fan. 1974. Estimating the Demand for Short-haul Air Transport Systems. *Transportation Research Record* 526, 1-15.

Kanafani, A. and A. Ghobrial. 1982. Aircraft Evaluation in Air Network Planning. *Transportation Engineering Journal* 108, no. TE3:282-300.

Kanafani, A. and A. A. Ghobrial. 1985. Airline Hubbing – Some Implications for Airport Economics. *Transportation Research – Part A* 19, no. 1:15-27.

Kearney, T. J. 1990. Frequent Flyer Programs: A Failure in Competitive Strategy, with Lessons for Management. *Journal of Consumer Marketing* 7, no. 1:31-40.

Koppelman, F. S. and V. Sethi. 2000. Closed Form Discrete Choice Models. In *Handbook of Transport Modeling*, ed. D. A. Hensher, K. J. Button, 211-227. Oxford: Pergamon Press.

Koppelman, F. S. and C-H Wen. 1998. Alternative Nested Logit Models: Structure, Properties and Estimation. *Transportation Research – Part B* 32, no. 5:289-298.

Koppelman, F. S. and C-H Wen. 1998. Nested Logit Models – Which are You Using? *Transportation Research Record* 1645, 1-7.

Lefer, H. 1988. Frequent Flyer Programs: Boon or Bane? *Air Transport World* 25, no. 7:16-19.

Manski, C. F. 1973. The Analysis of Qualitative Choice. Ph.D. diss., The Massachusetts Institute of Technology.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka, 105-142. New York: Academic Press.

McFadden, D. 1978. Modeling the Choice of Residential Location. *Transportation Research Record* 673, 72-77.

Morash, E. A. and J. Ozment. 1996. The Strategic Use of Transportation Time and Reliability for Competitive Advantage. *Transportation Journal* 36, no. 2:35-46.

Morrison, M. 2000. Aggregation Biases in Stated Preference Studies. *Australian Economic Papers* 39, no. 2:215-230.

Mumayiz, S. A. and R. W. Pulling. 1992. Forecasting Air Passenger Demand in Multi-airport Regions. In *Proceedings of the Transportation Research Forum*. Arlington: Transportation Research Forum.

Murphy, J. J., P. G. Allen, T. H. Stevens and D. Weatherhead. 2005. A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics*, In Press.

Nako, S. M. 1992. Frequent Flyer Programs and Business Travellers: An Empirical Investigation. *The Logistics and Transportation Review* 28, no. 4:395-414.

Nason, S. D. 1981. The Airline Preference Problem: An Application of Disaggregate Logit. Presented at the AGIFORS Symposium, Santa Barbara, California.

Ndoh, N. N., D. E. Pitfield and R. E. Caves. 1990. Air Transportation Passenger Route Choice: A Nested Multinomial Logit Analysis. In *Spatial Choices and Processes*, ed. M. M. Fischer, P. Nijkamp and Y. Y. Papageorgiou, 349-365. North-Holland: Elsevier Science Publishers B.V.

OAG Worldwide Limited 2000. 2001. *Official Airline Guide.* Bedfordshire, LU5 4HB, United Kingdom.

Orbitz, L.L.C. 2000. 2001. *Orbitz.com.* Chicago, IL, USA.

Priceline.com, Inc. 2000, 2001. *Priceline.com.* Norwalk, CT, USA.

Proussaloglou, K. and F. Koppelman. 1995. Air Carrier Demand: An Analysis of Market Share Determinants. *Transportation* 22, no. 4:371-388.

Proussaloglou, K. and F. S. Koppelman. 1999. The Choice of Air Carrier, Flight, and Fare Class. *Journal of Air Transport Management* 5, no. 4:193-201.

Reiss, P. C. and P. T. Spiller. 1989. Competition and Entry in Small Airline Markets. *Journal of Law and Economics* 32, no. 2:S179-S202.

Sabre Travel Network 2000. 2001. *Sabre GDS.* Southlake, TX, USA.

Skinner, R. E. Jr. 1976. Airport Choice: An Empirical Study. *Transportation Engineering Journal* 102, no. TE4:871-882.

Small, K. A. 1987. A Discrete Choice Model for Ordered Alternatives. *Econometrica* 55, no. 2:409-424.

Suzuki, Y. 2000. The Relationship between On-time Performance and Airline Market Share: A New Approach. *Transportation Research – Part E* 36, no. 2:139-154.

Suzuki, Y. 2003. Airline Frequent Flyer Programs: Equity and Attractiveness. *Transportation Research – Part E* 39, no. 4:289-304.

Suzuki, Y., M. R. Crum and M. J. Audino. 2003. Airport Choice, Leakage, and Experience in Single-airport Regions. *Journal of Transportation Engineering* 129, no. 2:212-218.

Suzuki, Y., J. E. Tyworth and R. A. Novack. 2001. Airline Market Share and Customer Service Quality: A Reference-dependent Model. *Transportation Research – Part A* 35, no. 9:773-788.

Toh, R. S. and M. Y. Hu. 1988. Frequent-flier Programs: Passenger Attributes and Attitudes. *Transportation Journal* 28, no. 2:11-22.

Transportation Research Circular 348. 1989. *Aviation Forecasting Methodology: A Special Workshop*. Washington, D.C.: Transportation Research Board.

Transportation Research E-Circular E-C040. 2002. *Aviation Demand Forecasting: A Survey of Methodologies*. Washington, D.C.: Transportation Research Board.

Travelocity.com, L.P. 2000. 2001. *Travelocity.com*. Dallas, TX, USA.

Tretheway, M. W. 1989. Frequent Flyer Programs: Marketing Bonanza or Anti-competitive Tool? *Journal of the Transportation Research Forum* 30, no. 1:195-201.

Verleger, P. K. Jr. 1972. Models of the Demand for Air Transportation. *Bell Journal of Economics and Management Science* 3, no. 2:437-457.

Vovsha, P. 1997. Application of Cross-nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record* 1607, 6-15.

Wen, C-H and F. S. Koppelman. 2001. The Generalized Nested Logit Model. *Transportation Research – Part B* 35, no. 7:627-641.

Williams, H. C. W. L. 1977. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A* 9, no. 3:285-344.

Windle, R. J. and M. E. Dresner. 1995. The Short and Long Run Effects of Entry on U.S. Domestic Air Routes. *Transportation Journal* 35, no. 2:14-25.

Windle, R. and M. Dresner. 1995. Airport Choice in Multiple-airport Regions. *Journal of Transportation Engineering* 121, no. 4:332-337.

Worldspan, L.P. 2000. 2001. *Worldspan CRS*. Atlanta, GA, USA.

Yoo, K. E. and N. Ashford. 1996. Carrier Choices of Air Passengers in Pacific Rim: Using Comparative Analysis and Complementary Interpretation of Revealed Preference and Stated Preference Data. *Transportation Research Record* 1562, 1-7.