The Evolution of Airline Distribution Channels and Their Effects on Revenue Management Performance

by
Diana M. Dorinson

S.B., Civil Engineering (1996)
Massachusetts Institute of Technology

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2004

Signature of Author ____________________________
Department of Civil and Environmental Engineering
May 14, 2004

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

Accepted by ____________________________
Cynthia Barnhart
Professor, Department of Civil and Environmental Engineering

Accepted by ____________________________
Nigel H. M. Wilson
Chairman, Transportation Education Committee

Accepted by ____________________________
Heidi Nepf
Chairman, Departmental Committee on Graduate Students

The Evolution of Airline Distribution Channels and Their Effects on Revenue Management Performance

by

Diana M. Dorinson

Submitted to the Department of Civil and Environmental Engineering
on May 14, 2004
in partial fulfillment of the requirements for the degree of
Master of Science in Transportation

Abstract

Over the past ten years, the development of more advanced computer systems and the growth in the use of the Internet have led to numerous changes in airline ticket distribution strategies. For example, the use of websites for booking and ticketing air travel continues to increase, and the Internet is often cited as the preferred model for a low-cost distribution channel. At the same time, Network Revenue Management methods are now viewed as a key tool for airlines to maximize revenue in an increasingly competitive marketplace.

These new systems and tools have helped the airlines achieve record profits in the strong economy of the late 1990s, but these profits may have masked hidden costs of using the new technology. Examples of hidden costs include the added computational burden of increased search engine requests to the computer reservations system as well as the increased opportunity for automated systems to bypass the booking limits set by the revenue management system. Such costs have yet to be examined and quantified in an academic research effort. The purpose of this thesis research is to understand a variety of issues related to how the technologies of more advanced distribution channels and more sophisticated revenue management systems interact with each other and impact air travel providers.

First, an empirical analysis of ticketing data is used to demonstrate that there are significant differences in ticket purchasing behavior among customers who use different distribution channels. Second, a review of previous experiments showing the negative revenue impacts of Inventory Control Bypass are presented, together with a discussion of some of the more promising solutions to Bypass. Next, these prior results are compared to a new set of experiments covering both path-based and leg-based Caching techniques. The new experiments show that the negative revenue impacts of Caching are at least as serious as those of Bypass, and may be more serious, depending on an airline's choice of how to interface with distributors who cache.

Thesis Supervisor: Peter P. Belobaba
Title: Principal Research Scientist, Department of Aeronautics and Astronautics

Acknowledgements

First and foremost, I would like to thank my advisor, Peter Belobaba. He has always been extremely supportive of my work, and I am grateful for his guidance and advice over the past two years. In addition, I am deeply indebted to him for providing me with the opportunity to work on two fascinating projects.

Thank you to my research sponsors for their support, both financial and academic. Each of the airline members of the PODS Consortium has made key contributions to the simulation experiments, and I appreciate the years of work that they put into the PODS model before I ever arrived. Craig Hopperstad was always eager to program my latest experiments into PODS—even the major overhaul required to implement leg-based Caching. His talents made my job infinitely easier. I would also like to thank Glenn Colville for encouraging me to pursue this area of research and providing me with a fabulous crash course to start me on my way.

My sponsors at Amadeus helped a great deal with the "real-world" side of this thesis, including offering me a fantastic summer internship experience. Olivier Muller, Robert Berrez, and Kedrin Wurmser worked very hard to provide the data and help me understand it. Natalie Evan and Wei Chen helped to keep me smiling all summer. François-Marc Levointurier-Vajda and Guglielmo Guastalla provided both help and friendship throughout the process.

Two other individuals deserve recognition for their valuable assistance on my thesis research. Karl Swartz generously supplied reference data on flight distances that would have taken me days to calculate on my own. Rick Zeni helped me to understand the technical background on several topics and provided the context that linked together many disparate ideas into one research agenda.

I would also like to thank the many professors at MIT who have always shown considerable enthusiasm for my endeavors, both academic and personal, including Nigel Wilson, Joe Sussman, Cynthia Barnhart and Amedeo Odoni.

◆ ◆ ◆ ◆ ◆

Thanks to all my friends in ICAT, CEE, and CTL. There are so many ways in which each of you has contributed to my experience here, and I am glad we got to share two years and lots of laughs together. Thanks also to the women (and coaches) of MIT Club Ice Hockey and to my dedicated crew of yoga students. I would not have made it through without you.

Several fellow students contributed more than just their friendship to this thesis. Thank you to the PODS crew past and present, including Tom Gorin, Emmanuel Carrier, Kendell Timmers, Andy Cusano, Adeem Usman, and Alex Lee. Your help and your efforts on previous experiments made a big difference to my work. Thanks especially to Kendell for getting me off the ground with SAS—the data analysis would not have been possible without your help. Thank you to Shiro Yamanaka for all your help when we were cleaning the data. Thank you to Ryan Tam for the airport location data. Thank you to Demian Raspall for sharing lots of crepes and waffles at Arrow Street. Thank you to Alex Modzanowska for always checking on me even when I thought I was too busy to chat. Thank you to Jeff Busby, both for your help with the statistical analysis and for always finding new and creative ways for us to procrastinate.

Last, but not least, I thank my family for their wonderful support. Mark, I could not have survived without your patience, encouragement and love. Dad, those early flights in our Piper set in motion a love of aviation that grows stronger each day. Hillary, you have always provided a burst of sunshine, happiness, and warmth just when I needed it most. Mom, thanks for setting an example for me—both in graduate school and in life—you gave me the confidence to believe this was possible.

I would like to dedicate this thesis to my grandparents, who have provided me with inspiration, a love of learning, and the passion to pursue my dreams. I love you all!

Contents

List of Figures

List of Tables

1 INTRODUCTION

Over the past ten years, the development of more advanced computer systems and the growth in the use of the Internet have led to numerous changes in airline ticket distribution strategies. For example, the use of websites for booking and ticketing continues to increase, and the Internet is often cited as the preferred model for a low-cost distribution channel. At the same time, Network Revenue Management methods are now viewed as a key tool for airlines to maximize revenue in an increasingly competitive marketplace.

These new systems and tools have helped the airlines achieve record profits in the strong economy of the late 1990s, but these profits may have masked hidden costs of using the new technology. Examples of hidden costs include the added computational burden of increased search engine requests to the computer reservations system as well as the increased opportunity for automated systems to bypass the booking limits set by the revenue management system. Such costs have yet to be examined and quantified in an academic research effort. The purpose of this thesis research is to understand a variety of issues related to how the technologies of more advanced distribution channels and more sophisticated revenue management systems interact with each other and impact air travel providers.

1.1 Motivation

Today, more than 22% of U.S. airline tickets (by revenue) are sold on the Internet [Carpenter, 2004]. The airlines cite lowered costs and better customer databases as benefits of using web-based distribution channels. At the same time, the success of the Internet has made it possible to search price and schedule information much more quickly than ever before. Also, the increased use of automated technologies has made it more difficult to track and control the way that the various computer systems communicate with each other. There have been numerous anecdotes among industry insiders about how robotic search engines have been able to bypass ticket purchasing restrictions, or how more transparent pricing structures are hurting industry performance. However, little has been done to quantify exactly how these new technologies are impacting airlines. And because the true scale of revenue impacts is not well known, airlines have been reluctant to devote significant resources to investigating these issues, for fear of distracting themselves from core problems such as over-capacity and competition from low-cost carriers.

There are so many factors contributing to the current financial difficulties in the airline industry that it would be incorrect to claim that any or all of the solutions or ideas presented here could magically restore the airlines to profitability. But if the revenue impacts of new distribution technologies are as disruptive and lasting as some in the industry have suggested, then it is imperative that airlines understand the trends early, and chart a course of action quickly. This thesis will begin to explore these issues more systematically, so that as website usage continues to grow, airlines are better positioned to react and respond, and hopefully increase their long-term viability as key transportation providers.

1.2 Research Objectives

The specific goals of this research are both qualitative and quantitative. The first goal is to provide a description of some of the most significant distribution challenges facing the airline industry today. This includes explaining the essence of the technical issues and cataloging the most prominent potential solutions to these challenges. The second goal of this research is to develop an analytical understanding of just how these distribution challenges impact the various industry players. In particular, a numerical scale with which to measure the true revenue impacts can help provide the context and motivation for future work on this essential developing topic.

1.3 Literature Review

The foundation of much of this thesis rests on the large body of work already developed in the area of revenue management. This topic has a rich history that is well explained in several key theses and survey papers. Williamson's Ph.D. dissertation (1992) covers the mathematical models that are fundamental to the practice of network revenue management. Talluri and van Ryzin (1999) provide an excellent history of the origins of revenue management together with a detailed glossary and an extensive bibliography. More recently, Belobaba (2002a) has provided an update on state of the art revenue management methods, and Barnhart, Belobaba, and Odoni (2003) place revenue management in context with other operations research problems in air transportation.

In the area of distribution channels, the topics covered in this thesis have primarily been covered in less formal settings such as presentations and panel discussions at industry conferences and working group sessions of professional organizations such as INFORMS. Several graduate theses have been identified addressing more qualitative issues of distribution. One of the earliest references is Wattanakamolchai (1996) who evaluated the convenience of airline reservation systems on the Internet. Zhang (2001) also discusses airline websites, focusing on issues such as customer service, marketing, and network resources. Lane (2003) covers the effect of the internet as part of an investigation into elite flyers. To the author's knowledge, this is the first academic effort to catalog, describe, and quantify these topics.

1.4 Thesis Organization

This thesis is organized into six chapters, including this introduction. Chapter 2 gives much of the historical context that explains how the airline industry evolved to its current state and why key players are concerned about the future. It also outlines the research methodology used in later chapters. Chapter 3 contains an empirical analysis of a sample of actual ticketing data with a particular emphasis on significant differences between the types of purchasing behavior observed for different distribution channels. In Chapter 4, the subject of Inventory Control Bypass is covered from both a qualitative and quantitative perspective. The Chapter contains a description of how Bypass happens, a summary of the results of previous simulation experiments that measure the impacts of Bypass, and an explanation of some ideas for how to mitigate airlines' revenue losses due to Bypass. Chapter 5 focuses primarily on the issue of Caching, but also includes some details about its major alternative, Selective Polling. In addition to the descriptive elements, two detailed sets of simulation experiments are presented that demonstrate the revenue impacts of Caching. Finally, Chapter 6 concludes with both a summary of the key findings, and an extensive listing of potential future research tasks which could reinforce the conclusions presented here.

2 BACKGROUND AND INDUSTRY TRENDS

While many of the topics discussed in this thesis have been a part of the airline industry for years—in some cases decades—the current convergence of these issues has resulted in complex and sometimes confusing revenue results for airlines. This chapter provides important background information on the developments and challenges in each of five key areas: (1) historical methods of ticket distribution, (2) the rise of the Internet as a major distribution channel, (3) the development of Network Revenue Management, (4) the evolution of ticketing bypass mechanisms, and (5) the increasing volume of real-time availability messages. Much of the information presented in this chapter is based on the way the industry has developed in the United States, but it is clear that many of these factors are already present in Europe as well. Following this contextual material, the chapter concludes with a description of the particular problems analyzed in this thesis and the research methodology employed in the study.

2.1 Historical Distribution Methods

There are two primary ways to book and purchase airline tickets today. The first is to contact the airline directly, either by speaking with their call-center telephone agents, by using their branded internet site, or in person at their Airport Ticket Office ("ATO") or City Ticket Office ("CTO"). In this case, an employee of the airline or the website host computer directly communicates with the airline's computer reservations system (CRS) to determine seat availability, prices, and fare rules for the purchase. One potential disadvantage of this method for travelers is that they will only receive information about the carrier they select and its code-share partners. They may not obtain the lowest fare for their travel, because another carrier may offer the same journey for less.

The other alternative for prospective passengers is to use a travel agent, either a human or an internet site. Although agents are usually experts on the travel industry, it would be quite cumbersome for human agents to contact each potential airline themselves, so over time, an intermediary service has developed, known as a Global Distribution System (or "GDS"). A GDS is a central communications provider who provides CRS services that facilitate the booking of reservations for airline tickets, as well as a variety of other travel-related services such as hotels, rental cars, tour packages, etc. Multiple airlines agree to share their seat inventory and pricing information with the GDS, who in turn organizes it and makes it easily available to travel agents and computerized websites. There are five major GDSs today: Abacus, Amadeus, Galileo, Sabre, and Worldspan. Some of these GDSs were initially owned by one or more airlines, but later sold or spun off as independent entities so that services could be cross-marketed. Some are still owned by their airline sponsors. Today, each airline typically participates in several, or all of these systems to maximize coverage in the marketplace.[1]

There is a highly symbiotic relationship between the three parties (airlines, agents, and GDSs). By accessing many airlines at once, agents feel more comfortable that they are finding the lowest price travel option for their customers by using a GDS. And, because most agents use a GDS, airlines receive better exposure of their product when they participate, usually leading to higher ticket sales. The installed base of travel agents using the GDS is a key factor in convincing airlines to participate, and at the same time the GDS depends on having many participating airlines as one of its major selling points to agents. Market share is a critical factor in the success of GDSs, because agents usually choose to affiliate with only one GDS. This is partly due to the complexity of learning the commands and procedures of multiple systems, but also because agents pay to use the GDS through subscription fees, and as airline participation has become more comprehensive, there is less and less incremental benefit from subscribing to multiple GDSs. One of the reasons that airlines participate in so many GDSs is that, unlike agents, they pay the GDS for their distribution services with transaction-based fees, which can be passed on to the customer in fares for each ticket. In fact, many low-fare carriers do not use GDSs, because the per-transaction fees cut too deeply into their profit-margins on low-fare tickets.

As low fare carriers have gained prominence, price pressure on the legacy carriers has increased, driving airlines to investigate ways to reduce the costs associated with

[1] As a part of its responsibilities to review competitiveness in the airline industry, the General Accounting Office has prepared reports and congressional testimony which contain informative descriptions of the historical development of CRSs (1986) and GDSs (2003), the 1984 Civil Aeronautics Board regulations governing conduct of CRS providers, also known as the "CRS rules" (1988), and the structural relationships between airlines, travel agencies, and CRSs (1992).

using GDSs and travel agents for the distribution of airline tickets. The two components of distribution costs are the GDS transactions fees, mentioned above, and commissions paid by the airlines to travel agents to encourage bookings. Historically, these commissions were based on a percentage of the ticket price, and certain "override" commissions were added when agents booked large dollar volumes with one carrier. This practice created the last link in the symbiotic relationship. Agents could earn a larger commission by booking higher priced tickets for their customers, so it was to their advantage to find a slightly more expensive ticket that could satisfy the customer's travel needs. Put differently, once a reasonably priced ticket was located, agents had no incentive to continue searching for lower fares.

At their peak in 1994, U.S. airlines paid commissions of more than 10% on domestic tickets and more than 16% on international tickets [GAO, 1999]. Then, during the economic recession of the early 1990s, airlines began to feel that these high rates cut too deeply into their profit margins. Since 1994, the airlines have gradually reduced the base commission percentages as a cost-saving measure, and more recently base commissions have been eliminated entirely, while increasing the focus on override commissions in an attempt to generate more agent loyalty. An unintended consequence of this gradual shift in commission structure is that the payments that travel agents receive from airlines now have little to do with the price of individual tickets. While travel agents may feel more loyal to a single airline, they have much less incentive to try to convince their clients to buy higher priced tickets, which is contributing to the erosion of airlines' pricing power. This revenue-side effect was greatly masked by the economic boom in the late 1990s, and airlines were able to make these moves unilaterally, because they are much less dependent on the travel agent community now that the Internet has become a more viable distribution method for their products and services.

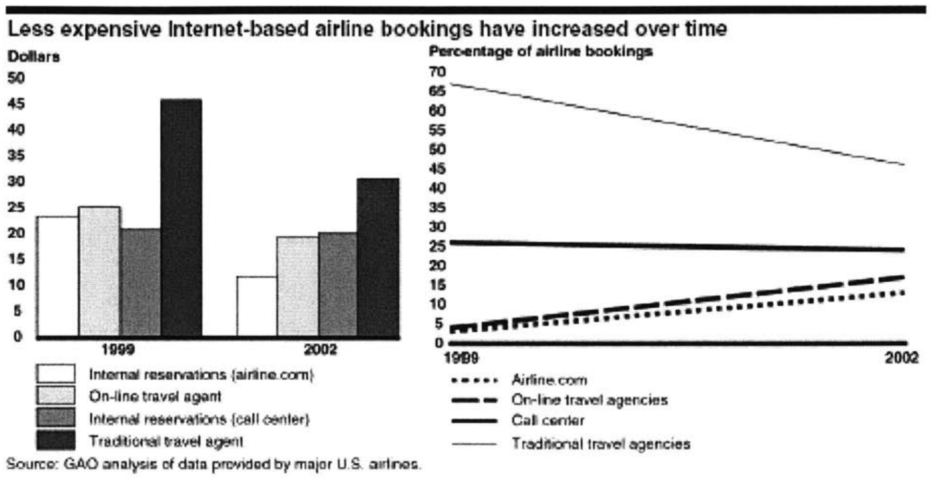
2.2 Rise of the Internet

The airline industry was one of the first to capitalize on the Internet as a distribution method, and travel remains one of the largest sources of online commerce of any business, accounting for more than 40% of online sales volume in 2002 [Transportation Group International, L.C., 2002]. Initially, airline websites offered simple database services such as looking up flight schedules, quoting fares, and helping frequent fliers track their mileage accounts. Some airline websites also promoted special offers and fare sales to try to sell highly distressed inventory, but to purchase travel, passengers still had to go to an ATO/CTO or a travel agent in order to issue the ticket.

With the advent of electronic tickets in the mid-1990s, it became possible to issue airline tickets without requiring paper documentation. At the same time, some entrepreneurs observed that booking airline travel through a GDS had become a highly automated process, one that was well-suited to serving up over the Internet. In particular, they realized that the cryptic codes used in the command-line GDS displays could be translated into an easily understandable graphical user interface making it possible to sell tickets from a GDS directly to the end customer. A number of "travel & shopping" websites were created such as BizTravel, Expedia, and Travelocity. Some were internet extensions of existing businesses, such CheapTickets.com which had operated a network of call centers since 1987. Also, some websites were directly affiliated with and/or sponsored by one GDS (e.g., Travelocity uses Sabre and CheapTickets.com uses Galileo), while others built their own search engine technology from scratch. Together, this set of travel websites pioneered the idea of passengers booking their own travel, replacing human travel agents.

To remain competitive, the airlines quickly moved to create interactive websites for selling tickets directly to the customer, and the airlines found a number of cost-saving side benefits of taking the "middle-men" out of the process. First, selling tickets over a website frees up airline customer service personnel for more complicated tasks. More importantly, online bookings do not pay a travel agency commission, and for many airlines who were able to connect their online website directly to their own CRS without going through a GDS, there are no GDS bookings fees to pay either. Airlines began to promote website booking, and once freed from their dependence on travel agents and GDSs, they were free to reduce commissions without significant backlash from the agent community. The move to website distribution helped airlines reduce their commission expenses by more than 50% between 1993 and 2000 [Lavere, 2000]. With the onset of the economic recession in the year 2000, airlines continued to encourage their customers to use website distribution channels to help manage costs. Although the dollar value of tickets is lower in recent years due to general decreases in air travel demand, the share of tickets booked on-line has continued to grow, as shown in Figure 2.1.[2]

Figure 2.1:
U.S. Revenue and Bookings Share of Major Distribution Channels



At the same time, there has also been some consolidation and realignment in the online travel industry. In some cases, these industry changes emerged from new business models for independent travel websites, such as Orbitz, which has developed a technology for booking tickets from a wide variety of airlines without utilizing a GDS. According to Nielsen estimates for website visibility among the top ten independent travel sites for airline bookings, the top three—Expedia, Orbitz, and Travelocity—have 55% of the market between them, with each of the three having at least 17% of the total [Smith, 2004]. A white paper produced by McGee (2003) for the Consumers WebWatch organization contains an excellent framework for understanding the differences among travel website offerings today. We will adopt this framework for our discussion, and highlight its major elements below.

2.2.1 Branded Sites

According to Consumer WebWatch, "these sites are owned by one or more travel suppliers and are basically dealerships selling a single line of products (e.g., American Airlines) or consortia of partner products (e.g., Northwest Airlines-KLM Royal Dutch

[2] General Accounting Office, 2003.

Airlines)." These sites have many of the characteristics of Integrated Transparent Sites (below), but are explicitly promoting one or more airlines.

2.2.2 Integrated Transparent Sites

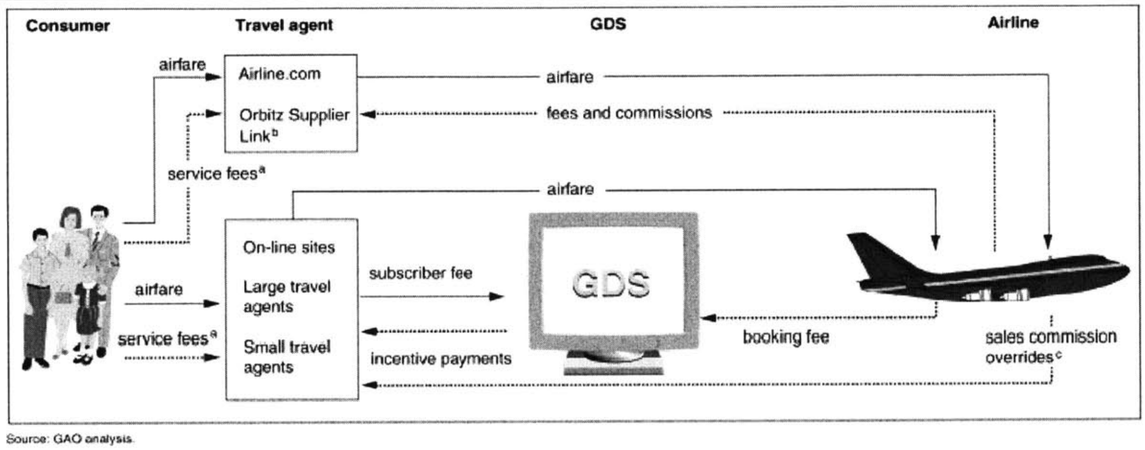
In this category, Consumer WebWatch explains that "sites may or may not be owned by travel suppliers but they act as online travel agencies offering multiple products from competing companies at varying fares and rates (e.g., Expedia, Orbitz, Travelocity); the identities of the travel suppliers as well as the fares and rates are provided to the consumer prior to booking." Note that the ownership of Integrated Transparent Sites may include airlines, for example Orbitz in the United States and Opodo in Europe both have substantial airline ownership. However, they market the website to consumers as a third party alternative, without relying on the branding of the airlines involved.

2.2.3 Integrated Opaque Sites

Opaque sites differ from transparent sites by offering some, but not all, of the travel information to the customer at the time of purchase. Consumer WebWatch writes, "these sites may or may not be owned by travel suppliers but they offer multiple products from competing companies at varying fares and rates and the identities of the travel suppliers are not provided to the consumer prior to booking (e.g., Hotwire, Priceline). In addition, the consumer may be required to bid for a fare or rate prior to booking (e.g., Priceline)." These websites form contractual agreements with the major carriers to help liquidate some of their most distressed inventory, often at lower fares than can be found on the transparent sites. The customer agrees to some level of flexibility in their travel arrangements, as well as forfeiting their right to claim frequent flier miles, in exchange for the lower fares. As a result, the airline is able to sell seats that would otherwise go empty without severely diluting their brand name, because the name of the airline is one of the elements which are hidden from the traveler during the booking process. Hotwire and Priceline are the two best known opaque sites in the U.S. today.

In the end, the rise in the use of the Internet has created another path for the purchase of airline tickets that does not involve a travel agent and that may or may not involve a GDS. As a result, there are now more possibilities for booking tickets, and a number of new financial flows among the parties involved in ticket distribution. Figure 2.2[3] depicts the roles and relationships of the major elements today.

Figure 2.2:
Relationship of Ticket Distribution Entities



[a]Consumers pay services fees.
[b]Airlines that subscribe to Orbitz Supplier Link pay less fees (including the commission per transaction) than GDS booking fee.
[c]Airline commission and override payments vary and are based on travel agencies meeting certain sales goals.

2.3 Development of Network Revenue Management

Over the 35 years since airline routes and fares in the domestic United States were deregulated, airlines have experimented with a variety of pricing policies and decision tools. One of the most important practices developed to date is that of differential pricing, in which airlines charge different prices for a seat on the aircraft based on a number of associated fare restrictions. The restrictions make the ticket more or less flexible, and thus more or less valuable to the consumer. If properly managed, this practice helps airlines to increase their total revenue, and it also allows the airline to provide air travel options to a wider variety of passengers.

One of the most challenging parts of implementing differential pricing for air travel is that the passengers who are willing to pay for added flexibility typically do not arrive until very close to departure, while more price-sensitive passengers often plan well in advance. Also, both types of passengers want to fly at the same times of the day and of the week, and so airlines need a mechanism to make sure that there are enough seats

[3] General Accounting Office, 2003.

remaining on the flight close to departure time to accommodate the late-booking, high-revenue passengers. Computerized revenue management systems ("RM systems") have been developed to calculate how many seats should be made available for each fare product on each flight. The objective of every RM system is to fill each available seat with highest possible revenue. Typically, this means that on high demand flights, RM systems limit the availability of discount fares while on low demand flights, empty seats are offered at very low fares. Through the use of customized software tools, RM systems can collect and maintain historical booking data by flight and fare class for each past departure date; forecast future booking demand and no-show rates by flight departure date and fare class; calculate booking limits to maximize total flight revenues; and provide interactive decision support for RM analysts.

The most common RM algorithms in use today make use of a concept called "nested" booking limits, in which the fares available are ranked and grouped together into fare-classes. Then, "the total seats made available to the highest [fare-]class include all seats available to all lower classes, so that a high fare request cannot be refused as long as there remain seats to be sold."[4] This nesting is accomplished by calculating the number of high-fare passengers expected on each flight and working backwards from the aircraft capacity, setting aside the appropriate number of seats for each class in descending order of revenue value to the airline. By limiting the total number of discount tickets sold early in the booking process, enough seats are "protected" for the high-fare passengers who will arrive later on.

The simplest of these RM systems perform the necessary calculations exclusively on a flight-by-flight basis by calculating the optimal mix of fare products on each flight leg in isolation. This is known as Fare Class Yield Management ("FCYM"), leg-based revenue management, or simply Leg Revenue Management. A widely used method for setting booking limits in a leg-based system is a probabilistic approach that focuses on the expected marginal seat revenue ("EMSR") anticipated for each fare class. It has been estimated that leg RM methods can lead to revenue gains of two to five percent compared to when the airline does not attempt to control seat inventory. Together with "overbooking" practices, which help airlines compensate for passengers who fail to show

[4] Belobaba, 2002b.

up for high demand flights at departure time, leg-based revenue management can increase airline revenues by a total of four to six percent with effectively no increase in airline operating costs [Belobaba, 2002a].

While these gains are impressive, in a network operation, further gains may be possible. This is due to the fact that a single flight can be used by many different itineraries, so that "high demand" flights may carry passengers traveling on vastly different routes, each with very different revenue contributions. As a result, there may be further revenue gains possible by distinguishing between seats available to single-leg ("local") vs. multi-leg ("connecting") passengers as well as between different types of connecting itineraries. Whereas the early RM systems focused exclusively on the fare class mix of the individual flights, the current state of the art is to define the set of Origin-Destination-Fare Classes ("ODFs") which use the flight leg, and optimize the chosen mix over the entire network using this much larger set. This more advanced method is known as Origin-Destination Revenue Management, O-D Control, or Network Revenue Management.

There are a variety of mathematical algorithms in development and/or current use which perform the computations required for network RM. Some methods map all ODFs by revenue value into a notional set of virtual classes. Each of these virtual classes may include local and connecting itineraries from a variety of the actual fare classes that will be sold. Then, the nested booking limits are set for the virtual classes. Methods which forecast expected passenger demands at the leg level are known as Greedy Virtual Nesting ("GVN"), while a more sophisticated network optimization approach is called Displacement Adjusted Virtual Nesting ("DAVN"). Another set of network RM methods uses a concept called bid-price control, in which calculations are performed on equations which are arithmetically equivalent to virtual nesting, but expressed in terms of the minimum acceptable fare value (the bid-price) for each unique itinerary/fare combination. In this way, the decision rule to accept any one fare is much simpler than the nested allocation approach, but unless the bid values are re-calculated very often, bid-price control cannot perform as well as other RM methods. Bid-price methods that store their data and forecast demand on a leg level are known as Heuristic Bid Price ("HBP") while the full-up network optimization is known as Network Bid Price, or sometimes more specifically, Probabilistic Bid Price ("ProBP"). Regardless of the method used, network RM can increase revenues a further one to two percent above the gains from using a leg RM approach. For a moderately sized U.S. airline, this can translate to millions of dollars a year in additional revenues [Belobaba, 2002b].

It is important to highlight the underlying sources of these revenue gains. In either the leg-based or the network-based strategies, the benefit of revenue management comes from matching supply and demand. Booking limits on high-demand flights help channel low-fare demand to empty flights, while protecting seats for the highest fare passengers on flights that are expected to depart full. In addition, an airline can match or initiate almost any low fare because of its ability to limit the availability of these low-fare products to a small subset of total seats. Thus, the airline can maintain a competitive pricing posture without the risk of revenue dilution. This is particularly important now that website search engines allow customers to compare prices across travel alternatives much more rapidly than ever before. Airlines count on being instantaneously competitive on price in order to protect their market share. The booking limits set by a revenue management system help the airline to match a competitor's prices directly and immediately, but with limited revenue exposure.

2.4 Evolution of Ticketing Bypass

For many years, passengers and travel agents have attempted to purchase tickets at prices below what the airline would otherwise wish to charge for the chosen itinerary by exploiting loopholes in fare rules and global distribution systems. While such activities have always been considered inappropriate by airlines, they were initially limited by the need for a human travel agent interface. These loopholes were often dealt with through the personal relationships between airlines and the travel agent sales force, or else they were not a large enough problem to merit much attention from airline management. More recently, the proliferation of websites and other automated ticketing options has led to less reliance on the travel agent community, leaving the airline with less ability to influence purchasing behavior. There are now many more opportunities for passengers to undertake "alternative ticketing," which undermines the performance of RM systems. As a result, understanding the dynamics of consumer purchasing behavior and the mechanics of ticket distribution are a high priority for airlines as they attempt to shore up their revenue position.

There are a variety of practices which might fall into the general category of alternative ticketing. We will distinguish between two major types, specifically "Fare Rule Evasion" and "Inventory Control Bypass."

2.4.1 Fare Rule Evasion

The first type of alternative ticketing includes a variety of tactics in which the agent or traveler would deliberately construct their itinerary by using only selected pieces of one or more tickets which did not reflect the true itinerary of the traveler. This was a common practice if sum of the fares of the various tickets was lower than the current lowest fare on the actual route of travel. A report by the General Accounting Office (2001) describes several of the most common forms of fare rule evasion including:

- **Back-to-Back Ticketing**—In order to avoid the Saturday-Night-Stay requirement, a passenger would buy two round-trip tickets to and from their destination, each of which involves a Saturday night, thus qualifying for the discount fare. The dates of the tickets would overlap such that the outbound portion of one ticket and the return portion of the other would encompass the desired mid-week journey. Many airlines have attempted to eliminate back-to-back ticketing by instituting a requirement that the first segment of an airline ticket must be used or the entire ticket would be cancelled. Some frequent travelers are able to buy two round-trip tickets where the four segments can be combined to form two complete mid-week trips. This practice is highly discouraged by airline threats to terminate the frequent flyer rights of any passengers caught using back-to-back tickets. Alternatively, if prices are low enough, a traveler may choose to purchase two round-trip tickets, intending to use only the outbound portion of each ticket for the actual travel, and forfeiting the return portion of each ticket. This issue requires a great deal of vigilance on the part of the pricing department, but it is not specifically a revenue management challenge.

- **Hidden-City Ticketing**—Sometimes a passenger wants to travel to an airline's hub city, but the fares from their origin to the hub are either prohibitively expensive, or completely unavailable. Quite often, at the same time, there are many fares beyond the hub market are available at lower prices. The passenger would purchase a round trip connecting ticket, intending to leave the airport at the hub without using the continuing portion of the journey. Upon their return, they would board only the last fight in the itinerary to return from the hub back home. The name derived from the fact that the traveler's true destination, the hub, was "hidden" from the airline at the time of purchase. As with back-to-back ticketing, airlines now cancel tickets when intermediate flight coupons are not used, so this practice is not as widely seen today.
- **Married Segment**—This type of loophole also occurs when a passenger cannot obtain their desired local itinerary. In this case, the traveler needs the assistance of a travel agent, who manually connects to a GDS to secure a booking for a connecting itinerary that uses the desired flight leg and is ultimately bound for a spurious destination. Then, only the second leg is cancelled, allowing the passenger to retain a low-fare, local seat on the first leg and the customer proceeds to purchase only this low-priced segment. Most airlines with network RM are well aware of this practice. It has come to be known as "Married Segment Abuse," because it breaks apart two segments which would otherwise be linked. In order to eliminate this practice, many airlines have successfully implemented "Married Segment Logic" in their CRSs, in which agents cannot cancel a portion of a connecting reservation, but must cancel the entire booking. This prevents agents from acquiring the local seats at low prices when the airline wants to protect those seats for higher-fare connecting passengers.

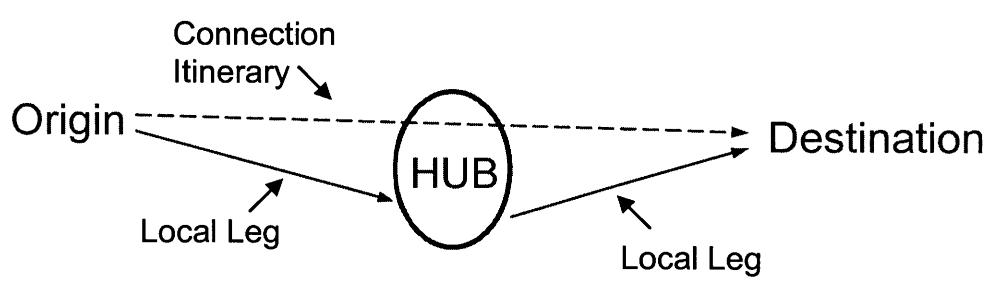
None of these practices is technically illegal, but they certainly have a negative impact on airline revenues. As a result, the airlines have worked to eliminate these practices as much as possible, chiefly through the redefinition of fare rules and CRS availability logic.

2.4.2 Inventory Control Bypass

The second type of alternative ticketing is the subject of Chapter 4 in this thesis. In contrast with Fare Rule Evasion, where a passenger or agent is deliberately manipulating the situation, Inventory Control Bypass can occur without the end-user's knowledge. The term "bypass" refers to the fact that whether by accident or by design, the passenger is able to get around the inventory controls set up by a revenue management system and obtain tickets at prices that are not actually available for sale. Revenue management systems base their calculations on the expected future revenue for each seat remaining for sale. Through Bypass, a customer contributes less revenue than the system expected, resulting not only in immediate financial losses on that ticket, but also in a loss of accuracy in future forecasts that limit the ability of the RM system to effectively maximize revenue.

At the present time, two types of Inventory Control Bypass have been identified, and they occur only in the case of passengers traveling on connecting itineraries. As shown in Figure 2.3, connecting passengers have multiple path choices from origin to destination, either on a pre-defined "through" or "connection" itinerary or on an itinerary constructed from multiple local legs. These itinerary types are referred to as connecting paths and local paths, respectively. Either path may involve a change of planes and even carriers at the hub airport. Also, it is possible for both types of paths to use the same physical flight legs. We distinguish between the two path types because airlines who use network RM allocate different seats to each path type, and the seats are typically *only* available to passengers flying on the corresponding itinerary.

Figure 2.3:
Multiple Ways to Assemble Connecting Itinerary



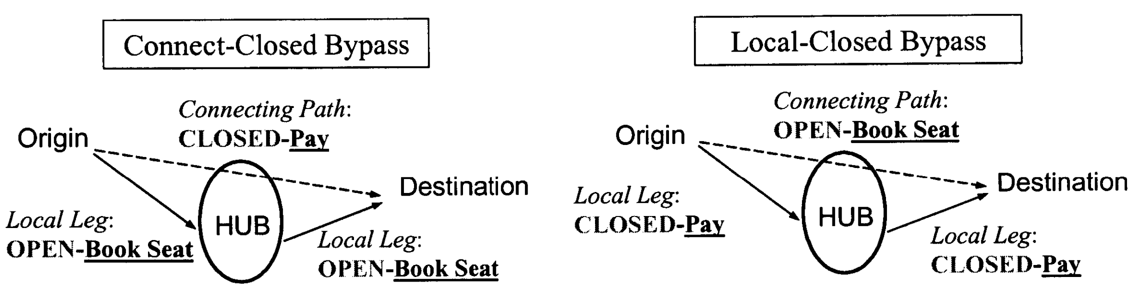
In the past, manually searching over all available local legs via a CRS to find a connecting itinerary was a fairly cumbersome process; most human agents relied on the pre-defined connecting itineraries to make reservations for their connecting passengers. With the rise of the Internet, powerful and efficient computer search engines have become much more prevalent in the industry today. From the customer's perspective, there is no difference between the two itinerary types, so adding the local-leg itineraries to the search is a logical way to improve the chances of finding a satisfactory travel option for the customer.

Once an acceptable travel option has been found and reserved, the passenger proceeds to the next step, which is to purchase the actual ticket. Because the availability search is performed independently from the pricing/ticketing action, passengers can sometimes obtain lower fares by securing seats out of inventory on an open path, but pricing according to fares that correspond to closed paths. This phenomenon is known as Closed Path Bypass, and it appears to be occurring more frequently, as website and GDS pricing search engines look at more and more itinerary combinations in order to find the lowest fares.

We delineate between two types Closed Path Bypass, because there are two different scenarios under which it can occur. They are briefly described here and depicted below in Figure 2.4.

- **Connect-Closed Bypass**—Booking seats on two local flights when the connecting inventory is not available, and then pricing the ticket at the connecting fare in the same booking class as the local seats
- **Local-Closed Bypass**—Booking seats from connecting inventory, but computing the total fare paid by joining together two local fares of the same booking class as the connecting seats, even when the local inventory is no longer available for sale

Figure 2.4:
Comparison of Connect-Closed Bypass and Local-Closed Bypass



More information about the different bypass mechanisms and the revenue impacts on the airlines are explained with a detailed example in Chapter 4.

2.4.3 Segment Pricing Inversion

It is easy to confuse Connect-Closed Bypass with another, separate phenomenon called Segment Pricing Inversion. If the sum of two local fares on the desired route is less than the connecting fare, the search engine or the travel agent may simply book the passenger on two local paths, and charge this connecting passenger two local fares. However, this is not considered a form of Inventory Control Bypass, because the fare paid corresponds exactly to the seats obtained. We restrict the term "bypass" to those cases when the selected inventory and the price paid do not match, because only these scenarios have material impacts on the performance of revenue management algorithms. For comparison purposes, segment pricing inversion is also discussed more thoroughly in Chapter 4.

2.5 Increases in Real-Time Availability Requests

The spread of internet distribution channels has resulted in a large increase in the volume of communications messages passed back and forth between the airline CRSs, GDSs, and the websites. For example, customers are becoming more and more comfortable with purchasing tickets on the Internet, and individual customers tend to comparison shop much more online than an experienced travel agent would. It has been estimated that passengers will make the same availability request at an average of three websites before making their purchase [Zeni, 2003]. This is only one of many reasons that the total number of availability requests to airline CRSs has mushroomed in the last few years. The volume of messaging traffic between computer systems—particularly real-time or "seamless" availability requests among GDSs and end-users—is now becoming a computational burden for airlines.

Different solutions to reduce the number of messages exchanged have been proposed, and some of these have already been developed and implemented. Each solution relies on determining seat availability by some mechanism other than a real-time request to the CRS. The idea is that because agents and consumers often browse through many travel options before making their selection, it is not necessary to respond to every availability request with real-time information. Instead, some sort of stored data or decision rule could be used to provide the customer with a reasonable approximation of their options, and a real-time availability request would be used once the customer had committed to actually purchasing a ticket. The total number of ticket purchases is much, much less than the number of requests for availability information, and so reductions in the number of "browsing" requests can have a significant impact on the resources required to respond to potential customers. The following sections briefly describe the major proposals that have been offered to date to attempt to reduce availability requests. Further information is provided in the detailed discussion of these issues in Chapter 5.

2.5.1 Selective Polling

One proposal is to use the stored availability from Availability Status Messages or ("AVS") messages on certain flights where the revenue impact of using leg-level (and potentially stale) data is expected to be low. The airline would use some decision criteria for deciding which flights could be handled this way and pass this information on to the GDS. If the flight in question is "AVS-ok", the GDS would use this stored data instead of sending an availability request back to ("polling") the CRS. At least one GDS already has the capability to use Selective Polling, but information is not available on how widely it is being used by the airlines, or if it is used at all.

2.5.2 Caching

In this method, the results of any real-time availability requests made by the search engine are stored in a memory cache. When a consumer asks for details on the same flight itinerary, the system consults the data stored in the memory cache instead of making a real-time poll. The data in a cache are likely to be fresher than the information received through AVS, but this would depend on the time between similar requests, and the mechanism used to keep the data elements in the cache accurate and up to date. In addition, a cache system requires much more memory than Selective Polling. Caching is already in use today by at least two integrated transparent websites and one GDS. In fact, Caching is particularly notable because of the fact that there is no need to develop industry standards in order to implement it. Websites often strive to develop a unique and powerful type of caching system as a source of competitive advantage.

2.5.3 Two-Pass Systems

Because of the chance of inaccuracies when using Selective Polling or Caching, some systems rely on a combination of methods. Essentially, they use the AVS message or the memory cache to filter the list of all possible flights down to the set which is most likely to have seats available. Then, they perform a real-time request for availability only on the subset of flights chosen, and these real-time requests are shown to the customer. As with other systems, if there is any delay in the AVS messages or in updating the cache, there could be inconsistencies. While the real-time results received from the final search would be accurate, by relying on AVS or a cache to select flights to poll, the system may miss flights which could have become available since the last update.

2.5.4 Proxy Systems

In order to make sure their cache is up to date, some website search engines continuously poll for availability on popular flights. These additional requests further exacerbate the problems, and as the airlines' computational resources become more and more constrained, GDSs have begun to emphasize entirely new distribution models. One GDS has a patent pending on a proposal to decouple the browsing and purchasing functions by using a proxy server. The idea is that a central repository, built to the communications standards of the GDS, is more readily scalable and able to handle the massive numbers of messages coming in and out of the system. The airline would propagate all of its inventory information in real time to such a proxy server, and search engines would poll the proxy server instead of the CRS. Because the airline actively sends its availability out to the proxy server, it will always be up to date. This structure would completely eliminate polls of the CRS for simple browsing requests, and sharply reduce or perhaps even eliminate the need to rely on cache data that might be erroneous. Actual ticket purchases would still be passed all of the way back to the CRS, so the airline would still control the inventory as before, but without the clutter caused by excessive communications messages.

2.5.5 Inventory Hosting

Taking the proxy idea a step further, at least one GDS has moved forward on developing complete inventory hosting capabilities. This would involve not only handling the messaging traffic involved in a proxy server, but also the operation of the physical CRS on behalf of the airline. Airlines still make their own decisions about prices, fare rules, and seat availability, but this information is regularly passed to the GDS, who then manages the communications with all of the various booking channels on the airline's behalf, completes ticketing transactions, and sends the final passenger data back to the airline. Virtually all of the messaging traffic remains with the GDS, and the airline focuses on the key issues of defining the product offering and handling the customers.

Both proxy servers and inventory hosting could be very attractive options to those airlines that are facing severe resource constraints. At the same time, these options are potentially very expensive and complicated to implement, particularly because they require a transition to new information technology architecture. Also, some of the options discussed above such as Selective Polling and proxy servers require the airlines' cooperation; some others, like Caching, do not. It is very important to understand the tradeoff between the cost to implement these solutions and costs of errors from not using real-time availability data. Airlines would like to have more information about the revenue-side impacts of some of the issues raised above before making a commitment to these new paradigms. Chapter 5 presents more detailed information about these topics as well as the results of a number of experiments regarding investigating the revenue impacts of Caching.

2.6 Research Methodology

As described within this chapter, airlines have moved to reduce their distribution costs by encouraging development of internet sites, but this has had the unintended consequence of increasing demands on their reservations systems and potentially allowing travelers to bypass their inventory controls. While there is a great deal of anecdotal information about the effects of these changes, the analytical portion of thesis seeks to model and quantify some of the effects more specifically as a first step to developing appropriate and practical solutions. The research methodology employed for this task combines both empirical data analysis and computer simulation.

2.6.1 Empirical Data Analysis

Ticketing transaction data was obtained from Amadeus Global Travel Distribution, S.A., a major international GDS. This data contains itinerary, fare, and transaction history information about a variety of tickets issued over two days in January, 2004. Although it is by no means comprehensive of Amadeus or of the industry in general, we use this snapshot to try to determine whether or not there are meaningful statistical variations that might justify future study in this area. In particular, we discuss the level of internet usage and develop a better understanding of how the web is used to book travel today. This analysis is presented in Chapter 3.

2.6.2 Simulation Study

While the empirical data was used to measure the characteristics of internet usage in general, an *ex-post* analysis cannot tell us whether or not the passengers or their travel agents (human or web-based) actually used the kinds of bypass or caching mechanisms described above, or even if they had the opportunity to. Simulation is one of the best tools to measure the behavior of passengers in a controlled environment where all inputs are known. The Passenger Origin-Destination Simulator, ("PODS") combines a passenger choice module and a revenue management module to evaluate their interaction in just such an environment. The use of simulation is an excellent way to shed light on the extent of the problems in the industry today. Specifically, Chapter 4 contains a review of several recent simulation studies related to Inventory Control Bypass and Chapter 5 presents two sets of new experiments on the problem of Caching.

3 EMPIRICAL DATA ANALYSIS

As described in Chapter 2, GDSs provide the critical link between the hundreds of different travel providers and the hundreds of thousands of agents who sell travel to the general public. In the case of air travel, the GDSs provide a centralized source of information about what flights are available and the fares charged for each ticket. But GDSs are not simply repositories of data; they act as a sophisticated communications network in order to organize and coordinate the millions of data transfers that need to occur between all of the parties in order to efficiently process and sell airline tickets. As such, they are uniquely positioned at the heart of the ticketing transaction. As a member of the Global Airline Industry Program at MIT, Amadeus Global Travel Distribution, S.A. ("Amadeus") offered to provide actual ticketing data to MIT for research purposes. This chapter discusses the preparation and analysis of that data.

3.1 Developing a Data Set

3.1.1 EDIFACT Messages

Today, buying and selling of travel services is conducted through a highly complex communications system. Many commercial transactions—including the purchase of airline tickets—are accomplished by the exchange of different types of electronic messages collectively known as "EDIFACT" messages, which stands for "Electronic Data Interchange for Administration, Commerce and Transport." EDIFACT messages contain important communications protocols designed to help the various computer systems talk to each other in the same "language" as well as the specific relevant details about the transaction, in this case the travel itinerary being planned. Messages are transmitted every time a computer system requests, sends, receives, or acknowledges information from another computer system. As the transaction progresses, different types of messages are used to share different types of information.

One type of message in particular, called an "AIRRQT" message, is sent from the entity selling the ticket to the GDS every time an airline ticket is purchased by an end consumer. The message contains many different data elements, including, but not limited to, the following:

- Flight itinerary details such as air carrier, origin, destination, flight number, departure & arrival time, equipment type, etc.
- Booking class, fare basis, and valid dates of travel
- Fare paid, including taxes
- Transaction information such as the identity of the ticketing office, the type of office, and the status of the ticket in the airline computer reservations system

When an AIRRQT message is received and successfully processed, it is acknowledged by the central system at Amadeus with an "AIRACK" message. This particular message exchange is one of the few times during the ticketing transaction that all of these data elements are synchronized and recorded simultaneously. As such, AIRRQT messages offer great potential for comparing the travel purchases of different characteristic groups within the Amadeus system. In support of several research efforts ongoing at MIT, Amadeus staff developed a new data filter that captured and processed a sample of incoming AIRRQT messages in order to obtain a data set of ticketing records for further analysis.

3.1.2 Obtaining the Data Sample

After working with the staff at Amadeus to determine the format and content that would be most appropriate for analysis, the constraints on collecting ticketing data were assessed. The Amadeus GDS operates twenty-four hours a day around the world, with major customers in all geographical areas. Thus, each data pull encompassed full twenty-four hour blocks in order to minimize geographical bias. Mid-week days were preferred to a Monday, Friday, or weekend, again to minimize bias from choosing a time period that would have only non-business hours in certain regions of the world. Finally, potential sampling dates were compared to international calendars to ensure that major holidays did not distort ticketing activity on a regional basis.

Because the data filter was a new procedure being added to the normal workings of the Amadeus system, data sampling was limited to a very small number of days in order to reduce the impact on labor and computational resources. A two-day period was selected in the middle of January, and the filter was run for 48 hours on all ticketing transactions processed by Amadeus. The processed data was stored in multiple raw-text files, which were later consolidated into master files for each day in the sample.

3.1.3 Cleaning and Preparing the Raw Data

Before the data could be analyzed, a number of actions were performed to clean, organize, and extend the data set. Steps were taken to identify and eliminate those data records which would not be meaningful in our analysis. In particular, there are often omissions in important data fields such as ticket number or ticket type that would make it difficult to identify unique travel records. The software program SAS was used to merge the individual text files and to clean out those records with missing or problematic data. Several checks were performed to eliminate records with any of the following conditions:

- Missing information required to identify a unique ticket from the sample, such as ticket number, Passenger Name Record number ("PNR"), and certain itinerary information
- Inappropriate transaction type, such as refunds, voided tickets, and cancellations.
- Duplicate records

Several data elements required for the analysis are not directly included in the message records, but can be derived based on the values contained in one or more fields. SAS was used to compare the appropriate field values, compute the additional variables, and add them to each record. In some cases the values can be determined directly from the contents of the AIRRQT records. In other cases, reference tables provided by Amadeus were used to determine appropriate values for cross-sectional analysis variables such as the cabin of service (First, Business, or Economy) and the city and region where ticketing took place.

3.1.4 Ticket Consolidation

Once the data had been cleaned of erroneous data records and the additional reference variables added, they were grouped and organized for analysis. As part of the normal ticketing procedures, a separate AIRRQT message is generated for each flight leg processed, resulting in a separate line of data in the file. These individual legs must be grouped together to form up the entire travel itinerary. Because this analysis examines ticketing characteristics, a consolidation step was necessary to convert the multiple leg-based rows into single rows that represent only the important details for each ticket purchased.

In processing the data from legs to tickets, there were several important considerations and assumptions. The most important of these assumptions relates to the actual itinerary of travel. In the same way that multiple flight legs can be grouped together and purchased as one ticket, it is possible for groups of tickets to be collected together under a higher identifier called a Passenger Name Record (PNR). This might happen if, for example, a passenger needs to travel on two different airlines that do not use the same reservations system. The travel agent simply acquires the various tickets separately and groups them together in one PNR in the GDS for the convenience of the passenger. In some cases, the different tickets may be purchased over a period of several days as the agent works on finding the right travel service(s) requested by the client. A further complication is that passengers may add to or change their travel plans in such a way that an additional set of tickets and/or PNRs become necessary. This flexibility is built into the system to allow agents to meet the needs of their customers, but as a result of these practices, there is no way to be sure that any one ticket examined from the sample represents the entire "journey" for that passenger. Other tickets purchased on different days or on other airlines—and in fact other modes of transportation entirely—may comprise one or more links of a passenger's total journey. As a result, no effort was made to determine the overall routing or destination of the traveler, and the records were not consolidated any further beyond the level of individual tickets.

While the itinerary itself was difficult to determine systematically, one piece of information that can easily be obtained is the total distance traveled on each ticket. Using another reference table [Swartz, 2004], the flight distance between the boarding point and off point of each leg was added up to find a total trip distance for each ticket. Trip distance is important, because it is used to normalize fares. Specifically, fares were converted from their original currency into U.S. dollars using currency exchange rates on the date of ticketing, and two key fare values were isolated. The first is the total purchase amount paid by the passenger, i.e. the final "price" of traveling. When divided by total miles traveled on the purchased itinerary, the pro-rated price per mile is useful for comparing fares among different customer groups. The second fare value is the base fare, or the price net of taxes, which is equivalent to the revenue received by the airline for providing the travel service.[5] The resulting base fare figure represents the revenue paid per revenue passenger mile (RPM) traveled, which is also known as yield.

Two other variables added to each ticket record were calculated directly from the date of the transaction: "advance purchase" and "advance planning." Advance purchase represents the amount of time between the day the ticket was purchased and the departure date of the first flight in the itinerary. Many airline tickets sold around the world include restrictions on their use that is directly related to the amount of time in advance that the ticket is purchased, so there could be correlations within the data which could be revealed by examining this attribute. Advance planning is a more subtle variable representing the amount of time between when the PNR was first created and the day that the ticket was purchased. A variety of ticketing practices influence how much time a passenger or their agent spends researching and arranging travel. While it is likely that there is less structured correlation related to advance planning than to advance purchase, some trends may be observed which could be partially explained by the amount of time devoted to planning the trip before the actual transaction was completed. Both advance purchase and advance planning were measured in number of days compared to the date of ticketing.

3.1.5 Ticketed Cabin

It is expected that this analysis will reveal important differences in ticketing behavior between passenger types, such as economy passengers compared to first class passengers, and so this analysis attempts to identify them and treat them separately where

[5] Taxes vary greatly by region, country, and even particular airports. In addition, they are retained by the government entity imposing the taxes, and so they do not contribute to the revenue results of the air carrier.

possible. In reality, there may be some overlap between the groups, for example, some passengers who are willing to pay for a first class ticket may end up in a lower class due to seat availability, and some passengers who pay the economy fare are hoping to be upgraded to higher classes because of their membership in loyalty programs. Although we cannot measure the passengers' expectations or willingness to pay, we do know that passengers who purchased a ticketed cabin above Economy paid for and received that higher level of service. In order to identify these passengers, the highest cabin purchased on any flight was used to screen for those passengers who received either First or Business class. In about 18% of ticket records, it was impossible to compute the cabin because of one or more elements not being available in the cross reference table. This happens for specific airlines that were not listed in the Amadeus reference table, and so the records were retained in order to prevent accidental bias against one or more air carriers or regions in the global observations.

3.1.6 Distribution Channel

The Amadeus computer systems need to know who they are communicating with so that the right EDIFACT messages are generated for each transaction. As a result, there are several different message fields which help identify what kind of user is executing the ticketing transaction request. Based on the values contained in four different data fields in the AIRRQT message, we can group the records according to the distribution channel that was used to make the purchase.

The idea of a "Distribution Channel" has different meanings in different contexts. For example, some people use the phrase to refer to the physical means of communicating with the GDS, such as using a command-line terminal connection or a web-based graphical user interface. In this thesis, "distribution channel" is used describe the way that the end customer—the airline passenger—relates to the GDS, which could be through a human agent, such as a professional travel agent or call center representative, or through the use of a travel website. In this context, there are more than a dozen unique distribution channels defined within the Amadeus system, including various types of office locations, call centers, and websites. For the purposes of this analysis, the distribution channels are grouped together into five functional categories, each of which is described below.

Airline

The first group represents the airline suppliers who make their seat inventory available through the Amadeus GDS. This group includes System Users, Non-System Users and Airline Alliances. System User Airlines are air carriers who have their inventory information "hosted", i.e., made available for purchase, primarily through the Amadeus GDS. The information technology links between System User Airlines and Amadeus are highly integrated. Non-System Users are the other airline suppliers that can be booked using the Amadeus system but their inventory is not hosted on the Amadeus computer network. Most likely they are using Amadeus to help increase their global presence, but they do not invest a large portion of their resources in promoting their travel products through Amadeus. This could be due to the fact that they do not have many customers in countries where Amadeus is a leading GDS, or several other factors. Finally, Airline Alliances are the multi-airline cooperative marketing and/or code-sharing partnerships such as oneworld™, Star Alliance™, and SkyTeam®.

Passengers who ticket directly with these airlines or airline alliances—at the Airport Ticket Office (ATO), City Ticket Office (CTO), or through a call center—are using the "Airline" channel. In this case, a trained staff member helps the traveler arrange their flights. Because the passenger is interacting with an airline or alliance representative, their choice set is likely to be limited to flights operated by the airline or the alliance partners. Also, the passenger has the benefit of being assisted by a human agent who is highly familiar with the inventory available and the way the system functions, and who may be able to quickly respond to the customer's concerns.

Airline Websites

Many airlines have developed their own branded websites in an attempt to reduce labor costs associated with serving passengers in person. A passenger who purchases a ticket through one of the Airline Website channel will have similar limitations on their choice set as the regular Airline group, but they will not have the aid of a human to guide them through the process. In addition, some airlines offer special "web-only" fares as an enticement to their passengers to use the website, and so certain fare products may be available to a website passenger that cannot be obtained through other channels. This group also includes the branded websites of airline alliances which market themselves based on the brand identity of the alliance members.

Travel Agent

The third category of distribution is the Travel Agent. This is another channel in which the passenger is served by a trained professional who is very familiar with industry practices and product offerings. The Travel Agent channel includes all agencies that are independent from direct airline control, from the major national and multi-national chains such as American Express, Carlson-Wagonlit, or Rosenbluth International, to the private "mom-and-pop" agents in smaller communities, to the specialty agencies serving backpackers or tour operators. Travel agency transactions are conducted in person at a travel agent office, or they are carried out over the phone either with the physical office or through a dedicated call center. The major difference between the Airline channel and the Travel Agent channel is that the travel agent who is assisting the passenger does not necessarily have an affiliation to one airline, and is free to offer a much wider variety of choices to the customer. However, it should be recognized that the airline commission structure described in Chapter 2 provides a strong incentive to the travel agents to concentrate their ticketing activity with just a few airlines. To the extent that the agent has a goal of reaching a certain breakpoint in override commissions, they may promote the travel services of one airline over another, so long as it meets the needs of their customer.

Travel Agent Websites

Similar to the airline website channel discussed above, there is a parallel Travel Agent Website channel, because a number of the larger companies now have website interfaces where the user may purchase their travel directly online. The agency receives credit for this purchase in their monthly volume of transactions, but is able to reduce the costs of serving customers by letting them book and purchase tickets for themselves through the branded website. Again, the passenger will have a very similar choice set as the Travel Agent channel, but will not have the benefit of a knowledgeable person to walk them through the process.

Integrated Websites

The last major functional group includes all other types of websites. As opposed to the various branded websites that would fall into the categories above, these websites are not directly affiliated with a single airline or with a traditional "bricks-and-mortar" travel agency. These integrated websites can be further categorized, as shown in Table 3.1 below. Although these groups may have distinct characteristics, the overall volume of tickets purchased on Amadeus in any of these categories is relatively small, and so they are grouped together in this study as one distribution channel.

Table 3.1:
Examples of Integrated Website Types

<table>
<tr><th>Website Type</th><th>Description</th><th>Website Category</th><th>Examples</th></tr>
<tr><td>Provider Web</td><td>Website of other travel providers such as hotel or rental car companies</td><td>Transparent</td><td>Hertz, Avis, Marriott, Princess Cruises</td></tr>
<tr><td rowspan="2">Travel & Shopping Web</td><td rowspan="2">Website offering travel products from many different providers</td><td>Transparent</td><td>Expedia, Travelocity</td></tr>
<tr><td>Opaque</td><td>Hotwire, Priceline</td></tr>
<tr><td>Joint Venture Web</td><td>Partnership website of multiple airlines & a GDS, marketing itself without reference to its owners</td><td>Transparent</td><td>Opodo, Orbitz</td></tr>
</table>

Passengers purchasing through one of these un-affiliated websites should have a fairly broad choice-set, but again will not have any assistance from human agents. While override commissions will not be a factor, contractual arrangements between the website and the airlines may influence which travel options are promoted most heavily by the website. There are currently no rules about how websites display their inventory for sale, and various display structures, pop-up ads, and front-page fare sales are often used to try to sway the customer's purchasing decision. This is in marked contrast to the direct use of a GDS display, which is subject to a number of rules and regulations that attempt to minimize bias, as mentioned in Chapter 2.

3.1.7 Final Data Set

After ticket consolidation, 430,885 complete ticket records were identified for the first day sample. Based on the results of the distribution categorization described above, there was very little web representation in the data sample. This is not surprising because Amadeus focuses primarily on servicing travel agencies and major supplier partners. In order to have enough observations in each of the cross-sectional categories to support statistical significance tests, it was decided to combine together all web transactions, including Airline Websites, Travel Agent Websites, and all types of Integrated Websites. Thus, records were grouped for analysis into three categories instead of five: Airline, Travel Agent, and Website.

3.2 Descriptive Statistics

One of the first tasks under this research effort was to group the data along several categories in order to determine the overall makeup of the ticket sample. There are several primary dimensions of our classification: Geography, Cabin, and Distribution Channel. The following tables present data for each of these categories individually, as well as some compound data tables showing deeper cross-sections.

Table 3.2:
Number of Tickets by Geographical Region

Region	Number of Observations	Share of Total
Africa & Middle East	12,298	3%
Asia & Pacific	75,909	18%
Europe	282,128	65%
Latin America	29,295	7%
North America	31,255	7%
Total	430,885	100%

Table 3.3:
Number of Tickets by Aircraft Cabin

Cabin	Number of Observations	Share of Total	Share of Known
First	2,946	1%	1%
Business	29,919	7%	8%
Economy	319,154	74%	91%
Subtotal: Known	352,019	82%	100%
Unknown	78,866	18%	
Total	430,885	100%	

Table 3.4:
Number of Tickets by Distribution Channel

Distribution Channel	Number of Observations	Share of Total
Airline	113,773	26%
Travel Agent	302,421	70%
Website	14,691	3%
Total	430,885	100%

A review of these high level tables leads to the following observations:

- European tickets represent the majority of tickets in the sample with more than 65% of the total. Asia represents 18%, followed by North America and Latin America with 7% each, and Africa/Middle East with 3%.
- Of the 82% of tickets in the sample where a cabin could be determined, 91% were purchased in Economy class, 8% in Business class and 1% in First class.
- There are very few tickets purchased on the web in the sample, amounting to 3% over the sample as a whole. Travel agency tickets dominate the sample at 70%. System User tickets comprise 26%.

Several cross-tabulations were also performed in order to evaluate where there might be concentrations of data. Tables 3.5 and 3.6 below show the number of tickets in the sample purchased on the web in each region and cabin, as well as the share of the total that these web tickets represent.

Table 3.5:
Number of Tickets Purchased on the Web by Region

Region	Number of Tickets	Web Share of Region Total
Africa & Middle East	-	0%
Asia & Pacific	605	1%
Europe	13,049	5%
Latin America	435	1%
North America	602	2%
Total	14,691	3%

Table 3.6:
Number of Tickets Purchased on the Web by Cabin

Cabin	Number of Tickets	Web Share of Cabin Total
First	35	0%
Business	654	0%
Economy	13,664	3%
Subtotal: Known	14,353	3%
Unknown	338	0%
Total	14,691	3%

Some more interesting observations become clear at this level:

- There are no tickets in the sample purchased through websites from the Africa/Middle East region.
- Europe has more tickets purchased on the web than any other region, with 89% of all web tickets in the sample. Other regions with web-based ticket purchases have between 3 and 4% of the total number of web tickets.
- Within each region, the share of tickets purchased on the web varies in different regions. The share of web tickets in the Amadeus sample is much higher in Europe, at 5%, than in other regions, which have only 1-2% of tickets purchased on the web.
- The Amadeus sample contains virtually no tickets purchased on the web in any cabin except Economy. Although not shown in the tables above, the few tickets in Business or First cabin that were purchased on the web were purchased through airline websites, not travel agency or integrated websites.

There are several key factors which could explain the distribution of data that has been described above. The primary driver of the results is of course the fact that this is a one-day sample of tickets from one GDS. A Tuesday in January at Amadeus may not be representative of all days, all GDSs, or the industry as a whole. For example, the large share of tickets in the European region is most likely attributable to the fact that Amadeus is a GDS based in Europe and owned by three of Europe's largest airlines. Naturally there will be a large number of tickets purchased using the Amadeus system in Europe, and somewhat lower numbers in other regions where other GDSs might be equally or more popular than Amadeus. The high concentration of web tickets in Europe is also influenced by the strong local presence of Amadeus. In addition, Amadeus has a relatively new division called eTravel which develops online booking tools for suppliers and agencies. As these services mature and proliferate, they may enhance the ability of customers in different market segments to make web-based purchases using the Amadeus GDS.

Even within the relatively established European market, the percentage of tickets purchased on the web is somewhat lower than expected based on anecdotal evidence. However, these figured should not be interpreted as an overall "web penetration rate," because the web tickets in this sample represent only those tickets purchased on the web through the Amadeus GDS. There are other GDSs which may have higher or lower rates of web ticketing. More importantly, many airlines such as easyJet and Ryanair choose the Internet for distribution precisely because they can set up reservations systems that do not have to go through a GDS, thereby reducing their total distribution costs. As a result, many web purchases in the airline industry today will not be reflected in a sample of GDS activity.

Turning to the distribution of tickets by cabin, the results are probably driven by industry practices more than other factors. The proportions of tickets that are purchased in each cabin are clearly a direct reflection the number of seats for sale in each cabin on aircraft worldwide. Intuitively, it is not possible to buy more tickets for certain cabin than there are seats available to sell. Even so, the percentage of tickets in the Economy cabin appears to be very high, possibly higher than the share of Economy seats available in the worldwide fleet. This is most likely due to the practice mentioned in Section 3.1.5, in which frequent travelers buy tickets in lower classes of service hoping to be upgraded at departure time. Upgrades are not reflected in these ticketing records. In addition, airline overbooking is not used as often in the premium cabins.

When specifically considering the cabin distribution of tickets purchased on the web, we see that most web purchases are for economy cabin tickets. This is not surprising because many travelers in the higher cabins are traveling on business, where a travel agent is more likely to handle their transaction. Even if the purpose of travel is non-business, passengers at higher fare levels typically focus on service more than price, and might prefer to use the services provided by a human agent rather than arrange travel on their own.

3.3 Mean Values

After establishing the distribution of data among the different cross sections, we turn to the question of whether differences between cross-section populations are statistically significant. We will address two types of comparisons. The first is a set of hypothesis tests in which we estimate whether there are significant differences between the distribution channels, in particular web versus non-web channels and airline versus travel agent channels. The second set of hypothesis tests examines whether or not sub-sets of the population along the regional or cabin cross-sections show significant differences from the overall averages

3.3.1 Advance Planning

Differences Between Channels

Based on the mean value of advance planning in each cross-sectional category, the time between initiating a booking and completion of ticketing for web bookings is virtually zero for all regions and all cabins, implying that the PNR is created at the time of purchase in most cases. The mean value of advance planning for non-web channels for the entire sample is 11 days. This difference in advance planning for web customers versus non-web customers is highly significant ($p<0.01$), both in the aggregate and for all combinations of region and cabin for which web data was available.

Considering only the non-web customers, we can also demonstrate differences between the airline and travel agent channels. The average length of advance planning for the airline channel is 10 days, while the travel agent channel is slightly longer at 11 days ($p<0.01$). There are nearly three times as many travel agent observations as airline observations in the sample, so it is no surprise that their characteristics dominate the aggregate results. The longer time for advance planning among travel agent passengers as compared to airline passengers is consistent regardless of cabin and significant at the $p<0.01$ level for all cabins. The mean values for each cabin group in each channel are given in Table 3.7, which shows that First cabin passengers have the greatest disparity. Although the results are not as uniform for the different regional groupings, all of the differences in mean value of advance planning by channel and region are statistically significant at $p<0.01$. Mean values by region are given in Table 3.8.

Table 3.7:
Average Value of Advance Planning by Cabin for Non-Web Channels

Cabin	Airline Channel	Travel Agent Channel	Difference
First	9	20	+11
Business	8	10	+2
Economy	10	12	+2
Unknown	7	9	+2
Total	10	11	+1

Table 3.8:
Average Value of Advance Planning by Region for Non-Web Channels

Region	Airline Channel	Travel Agent Channel	Difference
Africa & Middle East	14	9	-5
Asia & Pacific	15	17	+2
Europe	6	10	+4
Latin America	14	6	-8
North America	11	13	+2
Total	10	11	+1

Differences Between Passenger Groups

Due to the fact that advanced planning through web channels is generally at or very close to zero, it is difficult to make conclusive statements about the differences in advanced planning by region or cabin for web tickets. Although the differences in mean value represent only fractions of days, in some cases, it is possible to obtain statistically significant results:

- Business cabin passengers purchasing on the web complete their transaction slightly sooner than the average passenger in the sample. ($p<0.01$)
- Passengers from Asia, Latin American and North America who purchase tickets on the web also make their purchase in slightly less time than the overall average for the sample. ($p<0.01$)

Turning specifically to the non-web customers, we observe that mean value of advance planning differs by region. Advance planning is higher than the 11 day overall average in Asia (16 days) and North America (13 days) and lower than the average in Africa/Middle East, Europe, and Latin America (9 days in all three regions). These differences from the entire sample mean are all significant at the $p<0.01$ level. Similarly, there are highly significant differences ($p<0.01$) in advance planning outcomes by cabin type, with the First cabin tickets being purchased 16 days from the start of booking, compared to 9 days from initiating booking for Business cabin tickets, 11 days from booking for Economy cabin tickets, and 9 days from booking for those tickets where a cabin could not be determined. It should be noted that while the mean value for Economy cabin tickets in given as 11 days, or the same value as the overall average for non-web tickets, the use of additional precision shows that the mean for the Economy cabin is actually 11.3 days, while the mean value for all non-web tickets in the sample is 10.7 days, and the difference was also significant at the $p<0.01$ level.

In addition, the non-web channels exhibit highly significant differences in advance planning between cabin and region sub-populations. In the data shown in Tables 3.7 and 3.8, the mean value of advance planning for each of the cabin or region groups differed from the average value for the channel as a whole in most of the categories. These differences were tested for statistical significance with the following results:

- All differences between mean value of advance planning by region and overall mean value for the channel (Table 3.8) were significant at $p<0.01$ in both channels.

- All differences in mean value of advance planning by cabin in the travel agent channel (second column, Table 3.7) were significant at $p<0.01$.
- The difference in mean value of advance planning between business class passengers and total passengers in the airline channel is significant at $p<0.01$, but results for the other cabins are inconclusive.

Implications

As mentioned earlier in this section, the length of advance planning for tickets from the sample that were purchased through web channels is nearly zero. Those web-tickets where the advance planning time is greater than zero may be for trips that are "paid for" by redeeming loyalty program miles. These special tickets can usually be reserved several days or more in advance of the final issuance of tickets. However, the large majority of tickets from this sample that were purchased on the web have zero advance planning, which is consistent with the fact that while a traveler may spend a lot of time browsing before committing to a final purchase, most websites require the traveler to complete all elements of the final ticketing transaction simultaneously.

A travel agent, on the other hand, often requests and holds various seat reservations that might meet the traveler's needs inside an active PNR, later releasing unneeded seats when the preferred itinerary has been identified. For very complex trips, it could take several weeks to establish which travel options are most appropriate, so the average value of 11 days for the non-web tickets in the sample seems reasonable. It should also be noted that advance planning does not necessarily refer to the amount of time that a reservation for air travel is held by a passenger. Advance planning it is measured from the point when the PNR is first created, and PNR creation can be triggered by a number of different events. In general, any reservation of travel products, from car rentals to hotel reservations and even cruise or tour packages requires the creation of a PNR in which to store the information about the trip. These non-air elements can, and oftentimes must, be reserved well before a passenger begins contemplating the appropriate air travel for their needs.

Passengers arranging trips with multiple travel partners often rely on travel agents to process the various documents and payments required. This could partially explain the reason why advance planning is slightly longer in the travel agent channel than in the airline channel. Although some airlines do have partnerships with car and hotel providers, the overall complexity of the trip is likely to be lower for airlines than for travel agents. Also, there is a strong incentive for the airline employee to encourage the customer to purchase an airline ticket, which may shorten the time horizon to ticketing.

When considering the differences between airline and travel agent by cabin and region, it is harder to draw specific conclusions. In terms of cabin, one explanation for the smaller average values in First and Business cabins is that these tickets are often fully refundable, so the passenger does not risk very much by committing to their purchase right away, while Economy passengers might proceed with more caution. Unfortunately, this hypothesis cannot explain the considerably larger value of advance planning for First cabin customers using travel agents.

In the regions, the values of advance planning for the airline channel are several days higher than the channel average in Africa/Middle East, Asia/Pacific, and Latin America—all regions where there is a strong flag carrier system still in place. It is possible that due to limited competition for the provision of air travel services makes the passenger less concerned about securing a low priced ticket promptly. This might also explain the very low value of advance purchase in Europe, where there has been a recent increase in the number of extremely low-fare carriers in the market. Although the low-fare carrier bookings themselves are unlikely to be included in this sample, their influence on the product offerings of the legacy carriers market might still be a relevant influence on this data.

3.3.2 Advance Purchase

As mentioned previously in this chapter, advance purchase—the number of days between ticket purchase and departure—is often strongly influenced by fare restrictions. While the observations below are valid in the statistical sense, they may not be completely independent of other key factors not addressed in this analysis.

Differences Between Channels

In the aggregate sample, passengers using non-web channels purchased their tickets closer to departure than those passengers who used web channels. The mean values of advance purchase are 24 days for non-web channels and 59 days for web channels, a difference of 35 days which is highly significant ($p<0.01$). These results are dominated by the large number of Economy cabin observations in the sample, for which the mean values are 24 days advance purchase for non-web channels and 60 days advance purchase for web channels. However, Business and First cabins have somewhat different characteristics. In the Business cabin observations, non-web customers purchased an average of 11 days in advance using non-web channels and 28 days in advance using web channels, a difference of 16 days, which is significant at $p<0.01$. The First cabin passengers showed the largest disparity between channels, with mean values of advance purchase of 32 days for non-web channels and 113 days for web channels. This difference of 81 days is also significant at the $p<0.01$ level. These results are contained in Table 3.9 below.

Turning to the regions, first recall that there are no web bookings in Africa/Middle East. The average value of advance purchase for the non-web channels in that region was 8 days. Secondly, in Asia and Latin America, there is fairly limited web data, so that conclusions can only be drawn about Economy cabin customers. In both regions, the mean value of advance purchase was shorter for web channels than for non-web channels. In Asia, the values are 4 days and 23 days, respectively, while in Latin America they are 9 days and 12 days, respectively. European and North American customers in non-web channels purchase tickets closer to departure than those in web channels, both for the regions as a whole, and when the observations in each region are grouped by cabin. Table 3.10 contains these regional results.

In considering only the non-web channels, we see a consistent pattern of longer advance purchase times with airline channels than with travel agent channels across all cabin and region subgroups except for the Unknown cabin designation. The mean values for each sub-population and each channel are given in Tables 3.9 and 3.10, and all differences greater than zero are highly significant ($p<0.01$). These differences are even significant a $p<0.01$ for all combinations of region and cabin, except for First cabin customers in Africa/Middle East and Unknown cabin customers in Africa/Middle East, Europe, Latin America, and North America.

Table 3.9:
Average Value of Advance Purchase by Cabin and by Channel

Cabin	Airline Channel	Travel Agent Channel	Airline vs. T.A. Difference	All Non-Web Channels	Web Channels	Web vs. Non-Web Difference
First	46	23	-23	32	113	+91
Business	25	6	-19	11	28	+16
Economy	35	22	-13	26	60	+34
Unknown	19	19	-	19	58	+39
Total	34	20	-14	24	59	+35

Table 3.10:
Average Value of Advance Purchase by Region and by Channel

Region	Airline Channel	Travel Agent Channel	Airline vs. T.A. Difference	All Non-Web Channels	Web Channels	Web vs. Non-Web Difference
Africa & Middle East	15	7	-8	8	-	n/a
Asia & Pacific	32	10	-22	22	4	-17
Europe	38	22	-16	26	62	+37
Latin America	12	11	-1	11	9	-2
North America	50	28	-22	31	77	+46
Total	34	20	-14	24	59	+35

Differences Between Passenger Groups

While the differences tests between the advance purchase behaviors of different channel types showed very high levels of significance, the tests for differences between the cabin groups were not as consistent. In particular, for the web channel observations, only the mean value of advance purchase in the Business cabin (28 days) can be shown to be statistically different than the mean value for all web observations (59 days) at the $p<0.01$ level. The difference between the First cabin (113 days) and all web observations (59 days) is significant at $p<0.05$. It is not possible to draw conclusions for the remaining cabin groups in the sample of web observations. Among the non-web observations in Table 3.9, all comparisons between the mean values by cabin and the mean value for the entire channel are statistically significant at $p<0.01$ for both airline channels and travel agent channels, as well as the non-web channels taken together.

The differences between mean values of advance purchase for a region as compared to the overall channel average are highly significant. Specifically, all regional values shown in Table 3.10 are statistically different from the mean value given for the channel at $p<0.01$, regardless of channel grouping.

Implications

In the overall sample and even in the cabin sub-groups, we can see that web customers purchase their tickets furthest from departure, followed by airline customer and then travel agent customers. This is probably driven largely by the fact that business travel—which is often arranged only a few days or a week before departure—is more likely to be arranged through a corporate travel agent than directly with the airline. At the other end of the spectrum, passengers looking for a low priced ticket may have learned that the best time to search is well in advance of their travel dates in order to take advantage of discount fares which have advance purchase restrictions. These customers probably use the web to facilitate comparison shopping, and once they find a reasonably priced ticket, they snatch it up quickly to lock in their price. Although this sample does not include an excessively large share of tickets from the United States, it is interesting to note that the overall average value of advance purchase is 3 days longer than the familiar "21-day advance purchase excursion fare" ticket popularized by American Airlines and later copied by many others.

As was seen with advance planning, there is a fair amount of consistency in advance purchase results when comparing results across cabins. Specifically, the higher cabins tend to book earlier than the Economy cabin. This makes sense because passengers in Business and First are competing for a much smaller number of seats available—they probably want to make sure they get the higher class product before it sells out. Also, as mentioned in the previous section, there is less risk in committing to a fully refundable ticket, so there is no benefit to waiting once the decision to travel has been made. One further observation regarding the cabin-level data is that the extremely high value of advance purchase for First cabin tickets purchased on the web comes from a relatively small number of observations. Though statistically significant at the $p<0.01$ level, it is surprising to find that the average value is 113 days, or almost four months from the date of travel. These 35 observations may have had some underlying characteristics, for example being part of a special event such as a conference or the Olympics, which would not necessarily be replicated in every sample.

Much of the pattern behavior that can be used to describe the results among the cabin groups is completely absent when the data are grouped by region. One straightforward observation is that there is very little difference in the advance purchase behavior for the different channels in the Latin America region, and the mean values of advance purchase in Latin America are also some of the lowest of any region. Pricing policies or long-standing conventions in the industry could influence people to purchase in a much more uniform manner than in other regions. The largest discrepancies between web and non-web advance purchase are in Europe and North America, regions known to have both relatively high internet usage and extensive low-fare carrier markets. As mentioned above, the pressure to comparison shop using the Internet may be driving the very high values of advance purchase seen on the web in these regions.

3.4 Summary

The data presented in this analysis represent one day of operations of one GDS, so the conclusions drawn here may not be generalizable to other seasons of the year, other GDSs, or the industry as a whole. However, in many cases, this sample showed very high levels of significance in the differences between regions, between cabins, and between distribution channels. Further research may be warranted to see just how pervasive these differences are. If they are found to be general industry characteristics, airlines may want to revisit the assumptions underlying their revenue management algorithms, such as booking arrival curves. In particular, as market share shifts among the regions and channels, the differences in purchasing behavior could have a material impact on revenue results.

4 INVENTORY CONTROL BYPASS

As mentioned in Chapter 2, one of the concerns held by airlines related to the development of internet distribution methods is the potential for increased Inventory Control Bypass. This chapter is devoted to describing the problem of Bypass with more detailed examples and summarizing recent research on how Bypass affects revenue results. For completeness, we will also describe the phenomenon of Segment Pricing Inversion and compare it to Bypass. Finally, this chapter details some of the key proposals for addressing Bypass, including Journey Control and Price-As-Booked.

4.1 Detailed Examples

Recall from Chapter 2 that there are two major types of Closed Path Bypass:

- **Connect-Closed Bypass**—Booking two local flights when connecting flights are not available, and then pricing the ticket at the connecting fare in the same booking class
- **Local-Closed Bypass**—Booking seats from connecting availability, but pricing at the sum of two local fares of the same booking class

To illustrate the difference between the two scenarios for Closed Path Bypass as well as Segment Pricing Inversion, we will use the following example. A passenger wishes to travel from Seattle (SEA) to Boston (BOS), and is willing to travel via an intermediate hub airport (HUB). When the request for travel is made, the reservations system checks the seat availability and returns results for the four example fare classes, Y, B, M, and Q, which are ranked in descending order of their fares. Because a customer will choose the least expensive travel option which meets their needs, they first consider the lowest fare class. If that fare class is shown as available, then a seat can be reserved from inventory at the corresponding price. Otherwise, the customer must choose another (higher) fare, or choose not to travel.

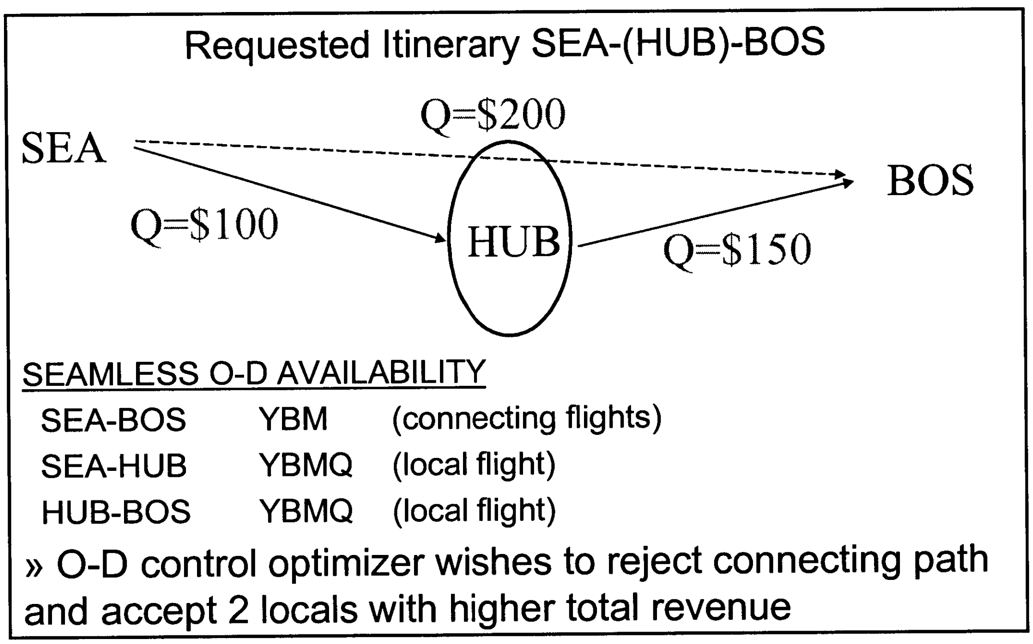
4.1.1 Connect-Closed Bypass Example

In the first example, depicted below in Figure 4.1, the reservations system shows that SEA-BOS connecting inventory in the lowest class (Q=$200) is unavailable. Normally, the connecting passenger must choose the next higher fare (M=$275).

However, in some cases, booking search engines will offer the passenger two seats in Q-class that are obtained separately from the local inventory for each leg. At the time of ticket purchase, the fare calculation is performed on the total itinerary, which is a connecting itinerary, and so the customer is charged the $200 Q-class connecting fare.

The airline's network RM system has determined that in order to maximize network revenues, the $200 Q fare on the SEA-BOS connection should be rejected. At the same time, the Q fare remains open on the two local SEA-HUB and HUB-BOS legs, with expectation of $100+$150 = $250 in total revenue. Unfortunately, when the complete SEA-BOS itinerary is priced at the connect Q fare, seats that are worth $250 have been sold for only $200, leading to a $50 network revenue loss for the airline.

Figure 4.1:
Example of Connect-Closed Bypass

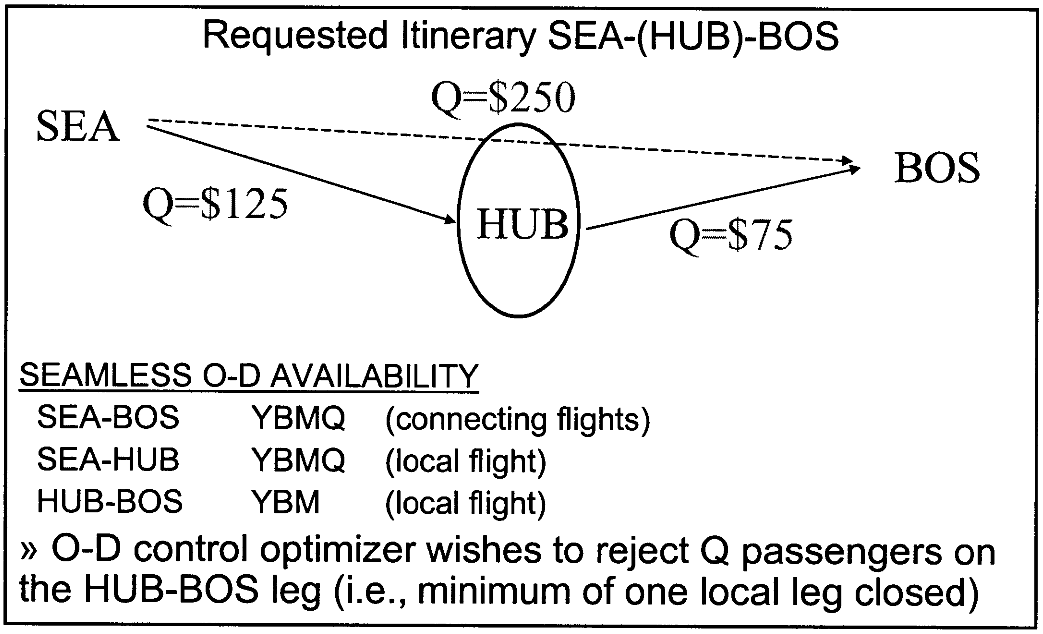


Source: Belobaba, 2002b.

4.1.2 Local-Closed Bypass Example

Figure 4.2 shows an example of Local-Closed Bypass. In this case, connecting inventory is still available in Q-class at a price of $250, and the sum of the two local fares in Q-class is $200, although the second Q-class local leg ($75) is no longer available because of revenue management controls. A seat is booked from connecting Q-class inventory, and the travel agent or fare search engine then attempts to obtain the lowest price available for the selected itinerary. Because the seats in the PNR are in Q-class for both legs, the ticket is constructed by assembling the fares on two local legs in the same class, which are priced at the Q fare levels for the local legs, or \$125 + \$75 = \$200. In this case, the connecting inventory that has been selected is priced at \$250, but the system provides a total fare of \$200, leading to \$50 network revenue loss for airline. The airline's network RM system has determined that \$75 Q fare for HUB-BOS is unavailable, but the separation of availability and pricing functions means that the lower priced ticketing can be obtained anyway.

Figure 4.2:
Example of Local-Closed Bypass



Source: Belobaba, 2002b.

4.1.3 Sources of Revenue Loss Under Bypass

Revenue losses in the presence of Closed Path Bypass come from two sources. First, there is a direct impact from an individual ticket sold through Bypass. The passenger obtains a ticket at a price below the value of the remaining inventory for sale, and the difference between the amount paid and what the airline could otherwise expect to receive for the same seat is permanently lost to the airline. In terms of opportunity cost, the airline has also lost the chance to sell that seat to a customer who might arrive later, and who could be willing to pay the true fare. When external systems are able to circumvent inventory controls, expected revenues do not materialize.

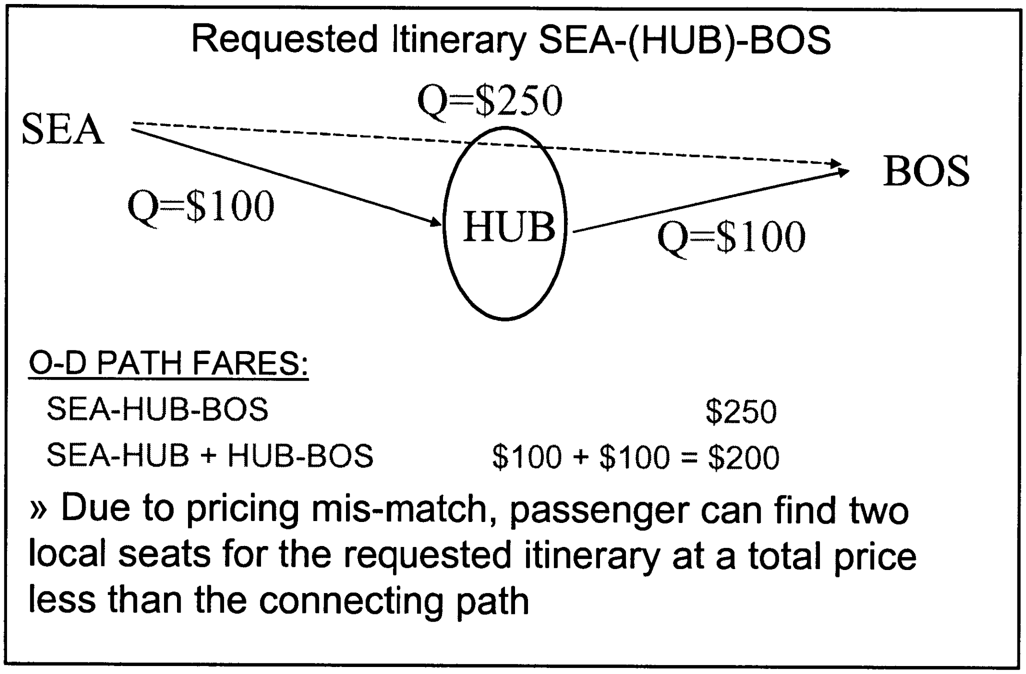
The second source of revenue loss is forecast error. The demand forecasters inside revenue management systems make projections about future demand based primarily on historical booking data. A customer who buys a ticket from a fare class that has already been closed will not be properly accounted for in the historical database used to develop future forecasts. This distortion of future demand forecasts leads to errors in the estimation of network displacement costs and bid prices within the revenue management algorithm, and ultimately to incorrect booking limits. Typically this means that too few seats will be protected for late arriving, high-paying passengers. This, in turn, limits the total revenue potential that the airline could achieve on that flight, even if no other bypass bookings were to be made.

4.1.4 Segment Pricing Inversion Example

There is a subtle difference between Closed Path Bypass and Segment Pricing Inversion. In Closed Path Bypass, passengers obtain seats from an open path, but pay prices that correspond to closed paths. In Segment Pricing Inversion, two or more paths may be open simultaneously, but due to the pricing structure in the two hub-based markets which make up the trip, the passenger can find lower fares using the local legs instead of the connecting path. See Figure 4.3 for an example of Segment Pricing Inversion.

This type of problem is more systematic, as CRS and website automated pricing engines routinely search for this possibility, and in fact, any passenger wishing to take advantage of cheap local fares could easily book the separate tickets themselves. Although there is some revenue loss in these situations, inventory controls are properly set, seats are purchased from open inventory, and the correct amount of expected revenue is achieved for each occupied seat. In addition, it is easy to maintain forecast accuracy in the presence of segment pricing inversion as long as the passengers are correctly accounted for upon ticketing. Thus we define this situation as a pricing problem, not a revenue management problem.

Figure 4.3:
Example of Segment Pricing Inversion



Source: Belobaba, 2002b.

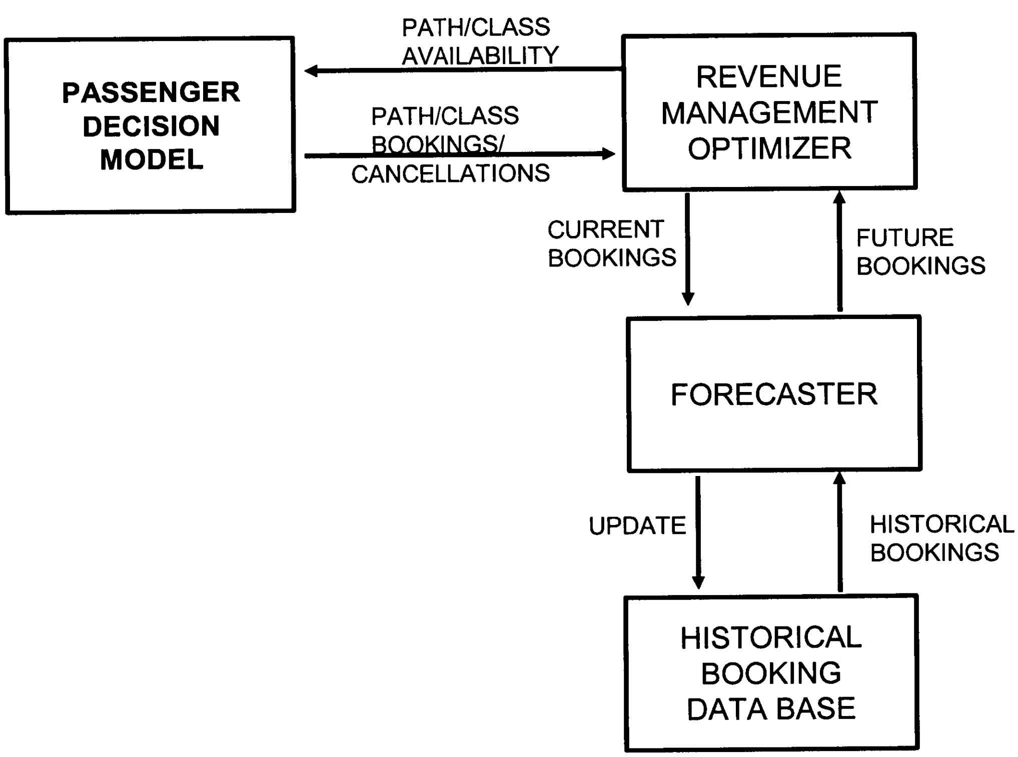
4.2 Summary of Recent Bypass Research

In conjunction with a consortium of several major international airlines, researchers at MIT have used the Passenger Origin-Destination Simulator, ("PODS") to perform a wide variety of experiments pertaining to issues of revenue management, forecasting methods, and passenger choice behavior. In the past few years, there have been several experiments looking specifically at the area of bypass of inventory controls. These experiments have quantified the revenue impacts of different types of Bypass, estimated how these impacts change depending on the proportion of passengers attempting Bypass, and tested the effectiveness of methods that account for Bypass during forecasting. The following sections describe some basics of the PODS system, highlight the key results of recent Bypass research, and discuss an experiment on Segment Pricing Inversion.

4.2.1 Description of PODS

The Passenger Origin Destination Simulator is a unique analysis tool in that it combines several different components in a real time simulation involving two or more airlines, so that the effects of competitive interactions can be observed. The first major component of the simulator is each airline's demand forecaster, which takes historical booking data developed prior to and during the simulation run as inputs, and then estimates anticipated future demand on each flight. The second component is a revenue management optimizer for each airline which takes the anticipated future demands as an input, as well as bookings on hand to date, and computes the appropriate booking limits for the different fare classes. Finally, these are combined with a passenger decision model which simulates the arrival of passengers and their choice process when purchasing a ticket, in particular their real-time selection from a set of available seats on different airlines, schedules, and prices. The general relationship between these elements is shown in Figure 4.4. These components interact repeatedly over a simulated booking period of 63 days for each flight, in order to estimate the total booking process for a given set of flights.

Figure 4.4:
Major Elements of PODS Model



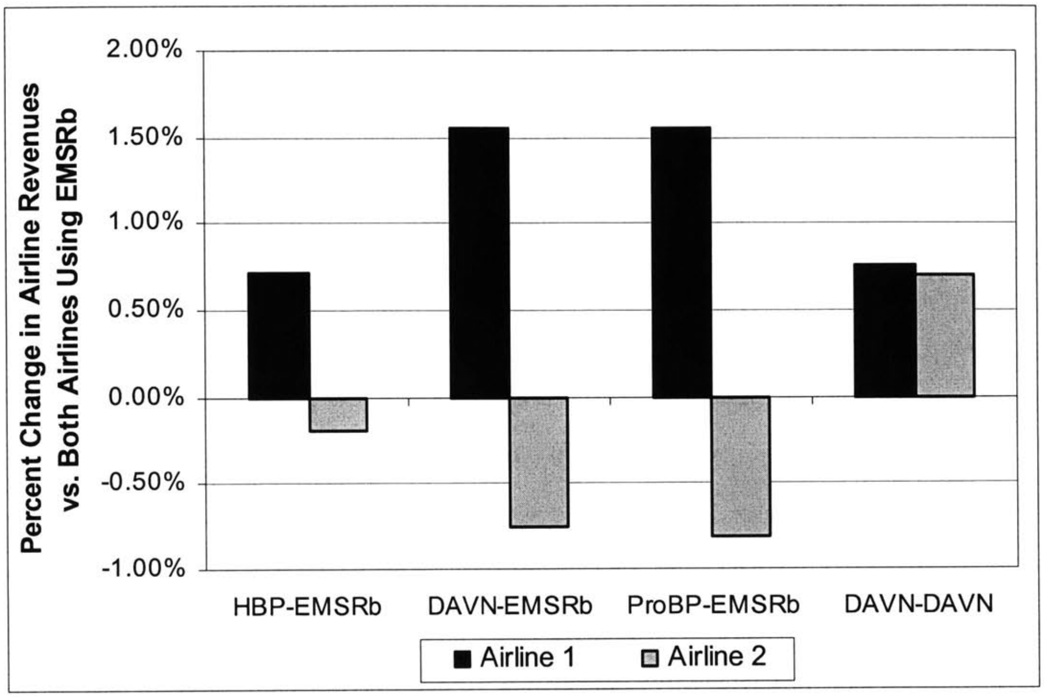
Source: Belobaba, 2002b.

Although there are many revenue management algorithms which can be simulated using PODS, the work described in this thesis focuses on four primary methods. The first is a leg RM method which makes use of Expected Marginal Seat Revenue calculations (EMSR, or sometimes EMSRb or simply "Eb"). There are also three network RM methods: Displacement Adjusted Virtual Nesting (DAVN), Heuristic Bid Price (HBP) and Probabilistic Bid Price (ProBP). For simulation experiments, leg RM is typically used as a base case, and then the benefit of using one of the network RM methods is computed as the incremental revenue gain over the use of leg RM. Also, various competitive combinations are modeled. In some cases, only one airline is given the network RM method to test its performance against a competitor with a less sophisticated RM system. In other cases both airlines are given the network RM methods, in order to understand whether revenue gains are sustainable if all carriers were to migrate to network revenue management.

All of the PODS experiments presented in this thesis used PODS Network D, which contains two simulated airlines each operating a U.S.-domestic network based around their own mid-continent hub. Each airline has three banks of flights per day serving 40 spoke cities, for a total of 252 flight legs and 482 markets. The revenue results presented in this chapter were achieved in most cases with average network load factors between 80 and 85%, depending on RM methods used and particulars of the Bypass experiments.

Figure 4.5 shows typical revenue results for the most common pairings of RM methods studied in PODS. It can be seen that two of the network RM methods, DAVN and ProBP, both produce revenue gains of more than 1.5% compared to the case where both carriers use leg RM methods, provided that their competitor is still using a leg method. However, these gains are not entirely lost if the competitor upgrades to a network RM method. For example, the far right pairing in Figure 4.5 shows that the carriers can each gain 0.70% or more under these conditions.

Figure 4.5:
Base Case Revenue Results for PODS Simulations



4.2.2 Results of Experiments on Bypass

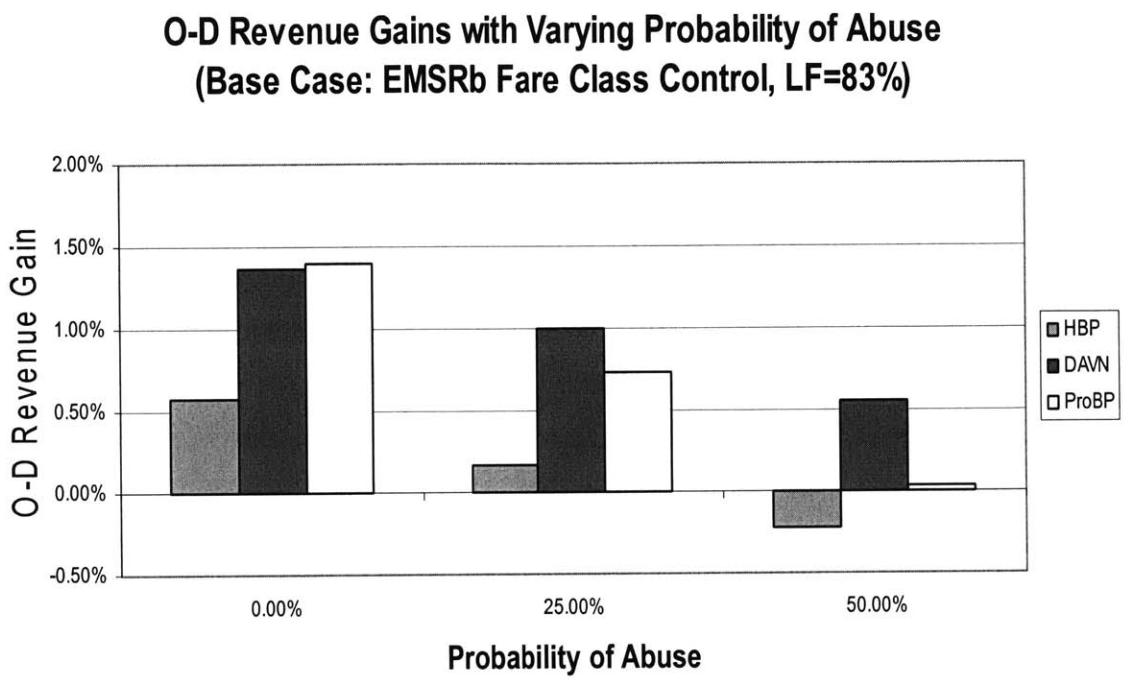
Currently, the only PODS experiments dealing with Bypass have looked at Connect-Closed Bypass, which was previously referred to as "OD-Control Abuse." To simulate this scenario, the PODS model included two sets of paths for each connecting itinerary: the normal connecting path and a new path made up of two local-legs. Then, a passenger who encountered closed connecting inventory would consider the local-leg path—priced at the corresponding connect fare—as another option in their choice set. This passenger behavior was controlled by a variable parameter between zero and 1.0 indicating the probability that a given passenger would attempt to bypass inventory controls. The value of this variable corresponds to whether or not a passenger would attempt bypass, given the opportunity, but bypass may not be necessary depending on the availability of their preferred path choice.

The Connect-Closed Bypass experiments only modified the passenger choice process, so the airlines in the simulation still performed revenue management assuming that the sale of two local seats would generate revenue equal to the sum of two local fares. As a result, the simulated airlines experienced both types of revenue losses described in Section 4.1.3 above. There are direct losses from taking a lower (connecting) fare than the expected sum of two local fares and there is a distortion of future demand forecasts due to accepting a booking after a fare class was closed.

The first experiments, performed by Lee (2001), tested a scenario in which all passengers attempted Bypass. In these cases, virtually all of the revenue gains attributable to the use of the network RM system were wiped out by the occurrence of Connect-Closed Bypass. In later simulations conducted by Cusano (2002a), variable levels of bypass were tested to measure the sensitivity of the revenue impacts. Results from the variable Bypass experiments are shown in Figure 4.6. In this case, one carrier always used a leg RM method (EMSRb), while the other carrier used one of the three network RM methods: DAVN, HBP, or ProBP. The bars on the graph show the increase in revenues by the network RM carrier compared to when both carriers were using leg RM. The three sets of results show how the ability of network RM to generate incremental revenue gains is significantly diminished by Connect-Closed Bypass. It can be seen that, for example, the DAVN method, which normally produces revenue gains of 1.4% to 1.6% over leg RM when no Bypass is present, produces gains of 1% at only 25% probability of bypass, and approximately 0.55% gain at 50% probability of bypass.

From the figure, we see that in addition to the fraction of passengers attempting Bypass, the revenue impact also depends the particular RM method used. Leg RM methods such as EMSRb do not distinguish between the different itineraries in the same fare class. As a result, there is no "local" or "closed" inventory to be controlled, and thus Bypass does not have any impact. Of the network RM methods, DAVN revenue gains are least affected. Revenue gains from ProBP and HBP drop to almost zero and may even generate small negative revenue impacts compared to the use of leg RM. The two bid price methods (HBP and ProBP) are more seriously affected than DAVN, because forecasting distortions influence probabilistic bid prices more than deterministic LP shadow prices. Although not shown here, the revenues that are lost by the network RM carriers due to Bypass are greater than the revenues gained by the EMSRb competitor. And, even if both carriers are using network RM, the presence of Bypass makes them both worse off than the base case where they both used leg RM.

Figure 4.6:
Connect-Closed Bypass Simulation Results—
Incremental Revenue Gains for Airline 1 when Airline 2 Always Uses EMSRb



Source: Cusano, 2002b.

The results shown above were achieved for a network load factor of approximately 83%. At higher load factors, more passengers are expected to arrive during the booking process, and thus more of the lower path/fare class combinations are closed by the revenue management system. With more paths closed, it is more likely that a given passenger will be confronted with closed inventory on their desired path/class, which might then lead them to bypass. Moreover, when Bypass is not present the revenue gains from using network RM are driven higher by higher load factors. Thus, at higher load factors, there is more to lose if bypass passengers are successful in their attempt.

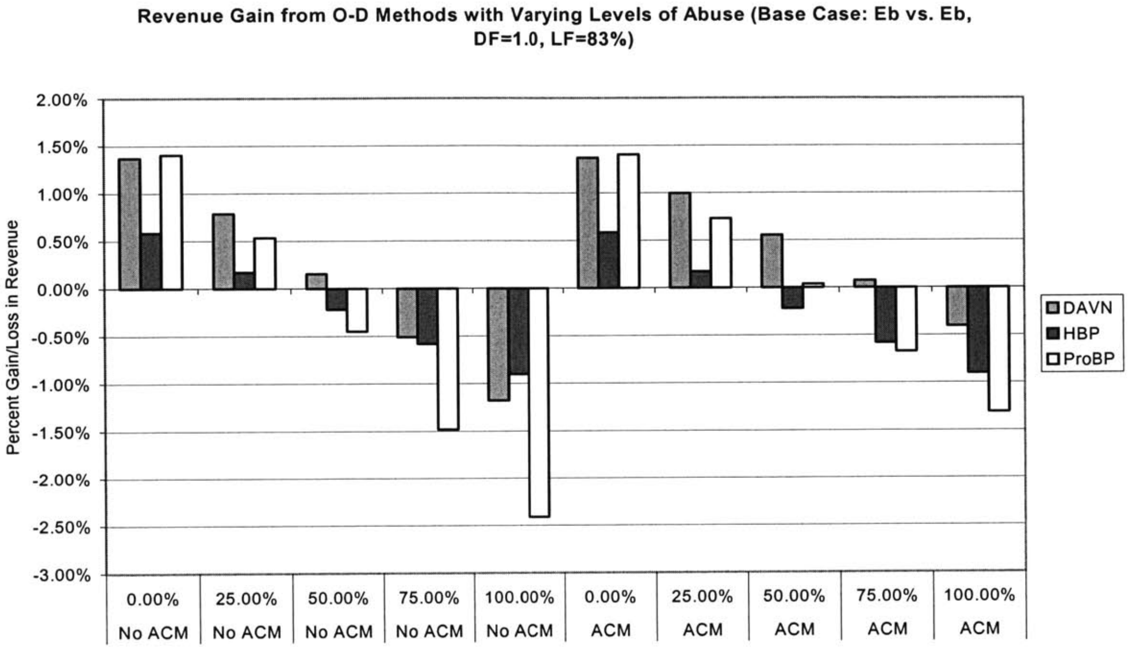
Clearly, bypass bookings can reduce the incremental revenue gains of network RM methods over leg fare class methods. This is of particular concern to those airlines that have chosen to upgrade their RM systems to network RM. What was formerly a revenue advantage may cause them to fall behind their leg-RM competitors unless appropriate solutions can be found.

4.2.3 Bypass Compensation Methods

Some airlines attempt to counteract forecasting errors with "Bypass Compensation Methods" ("BCM", but previously known as "ACM") within the RM system. The idea is to keep track of the bypass bookings and use the values to adjust the forecast, hopefully stemming losses from under-protections. This concept is relatively easy to implement in network RM systems, although not all airlines using network RM record bypass bookings in the same way. Some airlines maintain a separate tally of bypass bookings and add an average value to the historical data for path bookings. Other airlines add bypass connecting bookings to the local paths that are associated with the bypassed paths, because the seats themselves were purchased from local inventory.

As an example of BCM, Figure 4.7 shows simulation results from experiments by Cusano (2002b) that incorporate the local-path method. In this experiment, the first carrier was given one of the network RM methods while the other carrier always used the leg RM method, EMSRb. Simulations were run at different probabilities of bypass, both with and without the use of BCM by the carrier using network RM. The results show that BCM is critical to preserving revenue gains in a competitive environment. Although increasing the probability of bypass still results in decreasing revenues, these losses are not as dramatic when the airline uses BCM, and thus revenue levels are somewhat protected. The airline using network RM uses BCM to protect against Bypass, but the carrier using EMSRb does not experience Bypass and therefore does not use BCM. When both carriers are using network RM in the presence of Bypass, each carrier can choose whether or not to use BCM. Clearly if only one of the carriers is using BCM, it will be better off, but later experiments in this area have also shown that if both airlines use BCM, then they are both better off than if neither of them were using BCM. We see that properly accounting for bypass bookings can help to counteract the revenue losses. Even so, BCM only helps to adjust future forecasting; financial losses due to reduced fare revenue will persist unless appropriate inventory controls can be developed.

Figure 4.7:
Simulation Results for Bypass Compensation Methods



Source: Cusano, 2002b.

4.2.4 Segment Pricing Inversion Experiments

For completeness, we also describe a recent experiment related to Segment Pricing Inversion conducted by Lee (2001). This simulation set up is similar but not identical to the set-up for Connect Closed Bypass. For every connecting origin-destination pair in the network, two path alternatives are generated. The first is the connecting path priced at the published connecting fare, and the second is a connecting path comprised of the two local legs, *priced as the sum of the two local fares*. When the connecting path is closed by the network RM system, passengers look for the "local" alternative. Again only the passenger choice process is affected. When the airlines' RM systems set the booking limits, they do not anticipate that connecting passengers will attempt to purchase local seats.

The results of the simulations showed that unlike Inventory Control Bypass, Segment Pricing Inversion affects all RM methods equally. Regardless of RM method, Airline 1 always had a small revenue loss, while Airline 2 always had a small gain. This is because the revenue impact of segment pricing inversion depends primarily on the pricing structure in effect at the time of sale, in particular the proportion of path/classes that have two local fares that sum to less than the connecting fare. This will determine how often a passenger would prefer an itinerary composed of local fares to an inventory at the connecting fare.

In addition, it should be noted that the revenue impact from Segment Pricing Inversion in these simulations is much, much smaller than the impact from Inventory Control Bypass. This is due to the fare structure within the PODS model, so different airlines may have different experiences in this regard. However, because Segment Pricing Inversion can be corrected with vigilance on the part of the pricing department at an airline, this is not considered a revenue management problem. In contrast, Inventory Control Bypass continues to present a challenge because the current transaction structure does not give the airlines a mechanism to prevent passengers from acquiring seat inventory at the wrong fare.

4.3 Other Proposed Mitigation Methods

As previous PODS research has shown, there is a legitimate concern over revenue losses associated with Connect-Closed Inventory Control Bypass. Realizing that correcting for forecast errors is not enough to restore these losses, airlines have begun to think about ways that they can enforce network inventory controls without negative impacts on their revenue. Two of the most prominent ideas to date are Journey Control and Price-from-Availability, each of which is described below.

4.3.1 Journey Control

Recall that connecting passengers who are unable to secure their desired travel may attempt to bypass inventory controls on the O-D connecting path by purchasing seats on each of the local leg paths that make up their total journey. In order to obtain seats out of local leg inventory, each leg must be sent to the CRS for a separate check of availability. For passenger convenience, each of the local legs is added to the same PNR as the trip is assembled by the agent, be it human or web. As a result, when subsequent legs are requested from the CRS, it is possible to identify the request as belonging to a PNR which already contains a local leg for another part of a connecting itinerary.

This has now been made explicit by passing "journey data" to the CRS, which indicates that the flight legs are part of a larger itinerary. An EDIFACT message standard exists for embedding journey data in availability requests, although not all GDS systems currently transmit journey data to the CRS when they are passing their transaction requests.

Some airlines have considered—and a few airlines have implemented—a system known as "Journey Control" for PNR bookings. When an availability request is received for a PNR, computer systems use the journey data to test whether or not a previous local leg has been requested on the same itinerary. If so, then the CRS evaluates that request by using the connecting path availability for the entire journey (instead of the local path availabilities for the current leg being requested). It is this connecting availability which determines whether or not to accept the passenger on the local leg, and the CRS will refuse to allow the second local leg to be booked if the connecting path is not available. Presumably a passenger or agent who is unable to obtain the second local leg cancels their entire itinerary and re-books some other way.

Although only practiced by a few major airlines at the present time, there are already several different existing or proposed implementations of Journey Control. For example, some CRSs can monitor up the three legs as a connecting path, while others can only review two legs at a time. Some systems use this check of connecting inventory for all requests (both availability requests and sell requests), while others may use local availability right up until the actual ticketing request is made, at which point the check of connecting paths is carried out, and the entire ticket may be rejected.

Journey control prevents connecting passengers from accessing local inventory when the corresponding connecting inventory is sold out. It is an effective way to counteract Connect-Closed Bypass, but it has no effect on Local-Closed Bypass. There are also three significant drawbacks to Journey Control, each of which is discussed below.

Rejecting "sum of locals"

The first drawback of Journey Control is that the carrier may potentially lose some revenue, because the passenger may have been willing to pay the sum of the local leg fares, but instead they are rejected out of hand simply because connecting availability in that fare class has been closed down. In many cases, the sum of the two local legs turns out to be higher than the connect fare that has been closed down, although its value would not exceed the next highest open connecting fare (otherwise the passenger would not be interested in purchasing the sum of locals fare). The existence of the higher connecting fare is what creates the revenue dilemma. Because the airline had previously optimized their local availability, refusing a passenger who is willing to buy open seats on the aircraft at the correct price may result in increased levels of spill. On the other hand, the connecting passenger might have been willing to purchase that higher fare, and by leaving open the option to purchase the two local legs, the airline may be giving up the incremental revenue between these two fare values. Ultimately the airline must decide whether the revenue lost from rejecting the local-leg passenger is likely to be recovered by a later arrival of either two locals or a connecting passenger at the higher fare.

Marriage Control Issues

The second concern about Journey Control is a potential conflict with marriage control. When a connecting itinerary is directly requested by an agent, the segments are married together at booking time so that they cannot be individually cancelled to access low fare local inventory. Journey control does not disallow all bookings of two separate local legs, but only those in which there is no equivalent connecting inventory available; in other words, in some cases, the separate booking of local legs may be allowed. However, when the subsequent legs are successfully added to the PNR, they are added one at a time, and thus the CRS cannot marry the leg to the previous segment(s), because there is no other "live" segment under consideration when the response is passed back from the CRS. As a result, the use of Journey Control potentially exposes an airline to married segment problems. At the moment, there is one GDS which has developed a solution, and several others have announced plans to develop their own solutions, but at the present time, this issue still potentially poses a threat to airline revenues.

Code-Sharing Issues

The third drawback to Journey Control relates to code-sharing between carriers who have different types of revenue management control systems. Whenever availability or sell requests passed between the carriers are not consistently handled, there is a chance of revenue leakage from one carrier to the other. For example, when one carrier receives a request for a seat on one of their code-share partner's flights, the two carriers exchange information about the requested itinerary during the booking process in order to secure the reservation. If the two carriers' RM systems show different levels of availability for the same itinerary request, then the carrier that has tighter controls on their inventory may lose passengers while the partner is able to accept the passenger and enjoy increased revenue as a result. When that happens, one carrier has gained revenue while the other missed an opportunity to sell a ticket, thus we say revenue has leaked out of one carrier's network into the other.

Revenue leakage can have varying impacts depending on the financial structure of the code-sharing agreement between the carriers, but at a minimum, it leads to a distortion of traffic patterns and purchasing behavior that may have long-term consequences. The most obvious distinction between code-sharing carriers would be when one carrier is using a leg RM system while the other carrier is using a network RM system. This problem—without the presence of Inventory Control Bypass—has been researched already in several MIT theses, including Fernandez de la Torre (1999) and Darot (2001). However, even two carriers who both use the same basis for their RM logic may experience problems if one carrier has Journey Control and their code-share partner does not. Although both carriers would offer inventory on each other's flights, it is likely that some passengers would be spilled from the carrier who does use Journey Control and possibly recaptured by the carrier who does not use Journey Control.

4.3.2 Price-as-Booked

Given the potential risks of Journey Control, as well as its inability to combat both types of Closed Path Bypass, an alternative has been proposed that is a "price-as-booked" or "price-from-availability" policy. This idea is simply a basic enforcement of inventory controls throughout the entire ticketing transaction. If the seats requested come from local inventory, then they would have to be priced as two locals. If the seats requested come from connecting inventory, then they would have to be priced as a connection, not two locals. Such a policy would mitigate both types of Closed Path Bypass without risking lost passengers who are willing to purchase the sum of the local fares.

This concept, while simple to understand, would be extremely difficult to implement. Price-as-booked requires significant changes to GDS processing and communications architectures in order to integrate the availability and pricing functions. Gaining consensus from the wide variety of industry parties on the technical standards and protocols may not come easily. From the airlines' perspective, it is much easier to change the decision rules within one's own CRS to evaluate journey data on local leg requests than it is to convince the entire industry to overhaul the system. In addition, the industry players would like to know what they stand to gain or lose before choosing to make a transition of this magnitude.

4.4 Summary

As shown in Section 4.2.2, revenue losses from Inventory Control Bypass in the PODS simulations depend on the probability that customers will attempt Bypass. Translated to the real world, this probability represents the proportion of passengers who are using advanced search engines and internet sites to book their travel. Although estimates indicate that the number of U.S. domestic airline tickets booked on the web is currently at about 30%, the popularity of the Internet among consumers continues to grow [GAO, 2003]. In addition, further promotion of the web as the preferred distribution channel—whether by new low-cost carriers or the struggling legacy carriers—will only increase web usage even further. This means that the benefits of implementing Journey Control or Price-As-Booked will be a moving target in the future, continuing to increase as web usage spreads. Airlines will need to decide at what point the continuing costs of Inventory Control Bypass are outweighed by the costs and benefits of the solutions.

5 CACHING

The numerous electronic interactions between CRSs, GDSs, and end-users (travel agents, websites, and their individual customers) are placing a burden on the computational abilities of many airline CRSs. Chapter 2 presented a brief introduction to several of the more popular proposals trying to reduce the volume of availability requests. Two of these—Selective Polling and Caching—have already been implemented by a number of industry entities with mixed success. This chapter is devoted to further explaining the two concepts and analyzing their strengths and weaknesses. In particular, we develop the PODS architecture necessary to perform simulation experiments to quantify the revenue impacts of implementing either concept. Results of two sets of experiments are also presented.

It should be noted that other, more straightforward alternatives do exist for dealing with the ever increasing volume of messages coming into the CRS. For example, instead of focusing on reducing the number of messages received, the airline could purchase hardware to provide more communications capacity or invest in a customized software solution that allowed the existing machines to work more efficiently. It is relatively easy to evaluate the costs and benefits of such "off-the-shelf" solutions. But because the costs and benefits of Selective Polling and Caching are still little understood, airlines would like to know more about the trade-offs of implementing them before deciding which option to pursue.

5.1 Sources of Increasing CRS Message Activity

As described in Section 2.5, the rise of the Internet as a travel distribution channel has led to a dramatic spike in the number of communications messages passed between GDSs, CRSs, and end-users. There are several factors driving this increase. First, the faster speed of today's computer search engines means that many more itinerary combinations can be included in each availability search without creating any noticeable inconvenience for the end user. Second, some travel sites use sophisticated "robotic" algorithms to continuously extract availability from first-line data sources, such as airline websites and GDS feeds, in order to compile a broader database for their customers.

Finally, consumers themselves are also taking advantage of their newfound access to detailed availability information and comparison shopping across many websites before making their purchase. Together, these forces are contributing to a dramatic increase in the "look-to-book" ratio, or the total number of availability requests divided by the total number of actual reservations booked.[6]

It has been estimated by one GDS that experienced travel agents have a look-to-book ratio of approximately 12-to-1 [Ratliff, 2003]. These agents are familiar with the most common travel options, and they are paid based on total sales volume, so it is to their advantage to process transactions as quickly as possible. On-line users, on the other hand, are only searching for their own trip. They are willing to spend more time and energy searching if they believe it will help them obtain the lowest possible fare, or if the search will at least reassure them that they can do no better than what they have found on their first search. Look-to-book ratios for website users are less uniform, and somewhat more difficult to estimate, but 200-to-1 is considered a very conservative value [Ratliff, 2003]. The on-line user has a look-to-book ratio nearly ten times that of the traditional travel agent. Although bookings from website users comprise less than half of the total bookings today, the share is likely to continue growing as more and more airlines promote online booking, driving up the overall look-to-book ratio along with it. While GDSs have been structured for high volume message transmission from travel agents, airline CRSs are not as well equipped to deal with this kind of escalation.

5.2 Proposals for Reducing Message Volume

Various solutions have been proposed to ensure that both availability and sell requests are handled efficiently, and each solution has associated costs, either in direct implementation, or in potential lost revenue. Several of the ideas focus on reducing the number of requests for real-time availability. The two that are addressed in this thesis are Selective Polling and Caching. These are different practices relating to how third party distribution channels (such as websites) store and use seat inventory information for presentation to their customers. For Selective Polling, the airline and the third party

[6] Note that a very large number of reserved bookings are never ticketed, but airlines must pay the GDS a fee for each booking. To the extent that speculative bookings are also increasing with website usage, distribution costs may actually be rising for airlines with no appreciable increase in ultimate revenue.

agree in advance that certain real-time requests should use broadcast AVS messages for determining availability, while the remaining requests continue to poll seamless availability. In the case of Caching, the results of availability requests are stored in an independent memory cache, so that the third party or website does not have to take the time to call out to the airlines to obtain availability for each flight individually. In either case, the goal is to maintain a high level of accuracy and customer visibility at the third party site, while reducing message traffic into the CRS and freeing up computational resources.

5.2.1 Selective Polling

The idea of Selective Polling relies on an existing communications process known as Availability Status Messages or "AVS." Historically, airlines have communicated major changes in their availability with GDSs and other airlines by exchanging AVS messages. These messages are typically sent out whenever booking classes on a flight are closed, or if certain booking restrictions have been lifted. AVS messages are essentially an alert to the other parties to update their records, leftover from the days when there was no internet backbone to streamline communications. With the speed of real-time requests today, AVS messages are not as important, but some industry experts advocate giving them a new purpose.

Their idea for eliminating some portion of real-time requests is to use stored availability from AVS messages on selected flights, while the remaining flights continue to poll seamless availability. For example, an airline may know in advance that certain flights do not fill up very quickly. They could allow GDSs to sell tickets on this flight without a real-time check of the status of the flight in their CRS, because they know that there are always plenty of seats. They would broadcast an AVS message letting the GDSs know that the flight is open at some level of availability, and then as the flight began to fill up close to capacity, a second AVS message would be sent telling the GDSs that for the remainder of the time until departure, a real-time request would be required.

The booking capacity threshold approach is only one of a variety of possible decision rules that might be used to determine which requests could rely on AVS and which would have to actively poll the CRS. Other possible decision rules include:

- **Time to departure**—At a pre-determined point prior to the flight departure, stored AVS availability would no longer be used, regardless of the capacity or current bookings on the flight.
- **Key Revenue flights**—Use historical data to identify the set of flights which provide key contributions to the overall revenue results, for example the "top 100" flight legs or all flights which represent the top 25% of total revenue. These flights would always use seamless availability, while the rest would rely on AVS messages.
- **High Connect flights**—Use historical data to identify flights which contain a very high number of connecting passengers as opposed to local passengers. Because the errors from using stored data on these flights will be high, they would always poll actual availability, while other flights would rely on AVS.

One of the major challenges to implementing Selective Polling is that currently, AVS messages are used for individual flight legs, not for complete itineraries. If an airline is using a leg RM system, the AVS messages are based directly on the current fare-class availability for each leg, but for an airline using a network RM method, the situation is much more complex. For example, on a single flight, a fare class might be closed to local traffic but open for one or more of the connecting itineraries that use that leg. A leg-based message cannot distinguish between the alternatives, and it is unclear what type of AVS message would be appropriate—in the example just described, sending an "open" AVS message would frustrate local passengers trying to book, but sending a "closed" AVS message would turn away too many connecting customers. If the messages cannot provide consistent availability information to all passengers, the overall effectiveness of Selective Polling under leg-based AVS is called into question.

Some industry leaders have proposed a migration to O-D based AVS, but this idea could be extremely difficult to implement due to the sheer volume of AVS message traffic that would be generated. Currently, large airlines operate a few thousand flight legs each day, and each ticket sold would generate, at most, one AVS update message for each leg traversed in the itinerary. Indeed, for flight legs that are very open, the sale may not trigger a new AVS update at all. In the case of path-based inventory, a single airline may sell many tens of thousands of possible paths in the network, and each flight leg now carries passengers on hundreds of paths. Thus, when a single ticket is sold, availability updates might need to be issued for every one of the paths which uses that flight leg somewhere in the itinerary, potentially on the order of 5,000 messages per ticket sold. Multiplying this problem out to every carrier in the system and a full year of flights available for booking, the transmission and storage of this amount of information is not reasonably manageable. Again, as a way to help reduce the demands on an already over-burdened system, a combination of Selective Polling with O-D AVS is not likely to be helpful.

5.2.2 Caching

The second major proposal is to use a memory cache to maintain a private, external database of availability information that contains a "best guess" as to actual availability. Specifically, any third party, such as a GDS or website, can store the results of their recent availability requests for some predetermined amount of time. When a customer searches for flight options, if stored data for a particular flight in the request happens to be on-hand, the third party would use the stored data instead of a direct request back to the CRS, thereby eliminating the need for one or more seamless availability queries.

There are two types of Caching. The first is Active Caching, which uses the robotic tools mentioned in Section 5.1 to continuously update the cache. This is sometimes implemented with a rolling update, whereby some fraction—perhaps 20%—of the matrix is updated every so often in a continuous matrix "sweep." While the entire matrix is updated at regular intervals, there is always a chance of receiving a request in between updates that could cause errors. To keep this from happening too often, some third parties use more advanced algorithms which calculate the probability of receiving a request for a certain itinerary in the near future. Those flights which are most likely to be requested are kept up to date by re-checking the availability more often.

The second type of Caching is Passive Caching. In this implementation, each time a ticket is sold, a real-time availability request is made for the flights involved, providing information about the current status of the market in question. This data would be stored in the cache and considered reliable for a certain amount of time, after which it is discarded. Then, when new passengers arrive, if fresh data exists in the cache, it is used; otherwise the system is forced to revert to a real-time request. These real-time requests can also contribute to the data in the passive cache, so that after a time, a relatively broad database of the most requested flights is available in the cache.

In both types of Caching, old results are periodically purged to ensure that the GDS or website does not rely on information that is likely to be outdated. As a result, the stored data is likely to be somewhat fresher and more accurate than the AVS data, but there are still several drawbacks to Caching. The first drawback is that Caching requires a great deal more memory than AVS to implement effectively, typically involving large networks of computers dedicated to the task. The server farms are necessary because instead of waiting for airlines to distribute status messages, the parties who use Caching are hoping to be ahead of the game by keeping track of partial knowledge from previous results. Also, many websites have developed extremely complex methods to search for flight options for their customers, which would not be possible if each flight option had to be independently verified in the CRS.

The second and much more serious drawback of Caching is that caches today are typically stored on a leg-level, resulting in errors when websites or GDSs attempt to guess at path availability for an airline that uses network RM. Recall that the use of O-D AVS is considered far too memory intensive for current technology, and O-D Caching would be even more complex. In addition, even if an ambitious distributor did try to maintain all of the data required, the closure of one virtual class can affect multiple path classes. It may take some time for the third party to make enough real-time requests to find out about closures and update all of the affected path-classes, and so there still may be substantial errors in a path-based caching system.

Both airlines and third parties wish to avoid errors, because a reputation for not providing accurate information could drive away customers. One integrated transparent travel website claims that their cache error rate is currently 1%, but even this is considered too high by the company's management [Zeni, 2003]. From the airline perspective, errors in the cache might result in losing the chance to sell a ticket. With the current economic downturn, airlines are concerned that these kinds of mistakes are eating away at their already shrinking revenue.

5.2.3 Risks and Costs of Selective Polling and/or Caching

The revenue losses from Selective Polling and Caching stem from the same fundamental risks. First, airlines are concerned about the opportunity costs of using these alternative availability methods. Bookings and the associated revenue may be lost if the latest AVS message or cached availability shows there are no more seats remaining in the requested fare-class when there is actually a seat available to sell, perhaps due to a cancellation or a change in booking limits. Thus, frequent updates are critical to ensuring accuracy. But, the fact that Selective Polling and Caching are stored on a leg-level is a more significant handicap to performance. If a real-time request shows that a flight leg is closed to local passengers, the third party might eliminate the connecting itineraries that use that leg from its future searches. If that happens, then an airline that uses network revenue management may lose the chance to sell a seat.

The second concern over Selective Polling and Caching is that of customer disservice. Potential customers may become dissatisfied if an AVS- or cached-based display shows that a certain fare is available, but the booking is rejected when the customer attempts to make the purchase because the class is actually closed. Sometimes the customer may attribute this problem to the website, GDS, or travel agent, but there is a substantial risk that they blame the airline for what appears to be deceptive selling practices and ultimately become a lost customer.

The third concern relates to the problem of Inventory Control Bypass described in Chapter 4. One of the main reasons that third parties use a cache or AVS system is that it allows them to search over many more inventory combinations than a human agent. And because airlines have not yet found effective means to combat Bypass, they are hesitant to support the concepts of Selective Polling and Caching at this time. Airlines are worried that any increase in the websites' ability to aggressively search the CRS may result in additional revenue losses if the search engines are able to bypass their inventory controls.

Finally, the combinations of these three major concerns (revenue opportunity, customer service, and Inventory Control Bypass) might lead an airline to try to restrict the use of these methods to a small portion of their flight network, typically the least active flights in terms of bookings so that there would be less chance of errors. For the very active, high demand flights, airlines do not want to risk errors from a stored-availability system, and yet these are precisely the flights that generate a disproportionate share of CRS message traffic. As a result of these concerns, airlines have not yet embraced either of these two alternatives.

5.2.4 Other Alternatives for Reducing Real-Time Requests

As mentioned in Section 2.5, there are three other ideas for reducing availability requests to the CRS besides Selective Polling and Caching. Two of the methods, proxy systems and inventory hosting, require a substantial coordination with a third party. Because neither method is currently modeled in PODS, these two concepts are beyond the scope of this thesis. In the third method—the two-pass system—either stored AVS messages or cached data is used to select a subset of flights which will be shown to the end-user and then real-time requests are made for only this subset of flights. Two-pass systems are already in use today by some websites and GDSs who are concerned about customer service and want to make sure that their displays are accurate. Although the overall risk of revenue losses is likely to be less in a two-pass system, airlines are still concerned, because of the chance that errors in the stored data will keep their available inventory from qualifying for the real time request. This topic certainly deserves further study, but a clear understanding of the Selective Polling or Caching that are the basis of the first pass is a pre-requisite to developing a useful model for a complete two-pass system. Thus, the remainder of this chapter is devoted to a better understanding of Selective Polling and Caching.

5.3 PODS Modeling Issues

Both Caching and Selective Polling concepts can be tested with the use of a "shadow matrix" in PODS. The PODS simulation already contains a table of actual path-class availabilities that is used to determine which paths are in the passenger's choice set. In the experiments presented in this chapter, we also maintain a separate table of availabilities that represents the inventory information as it is stored by third party distribution channels. Each simulated passenger will have some probability of choosing one of these third-party channels for their transaction, and consequently of relying on this stored information instead of actual, real-time CRS data. Different methods of storing and updating the shadow matrix can be tested for their revenue impacts, along with different passenger and airline responses to error conditions in the shadow matrix.

5.3.1 Populating the Shadow Matrix

The dimensions of the shadow matrix will be based on the fare classes that are available on the network of flights in the schedule for each airline in the simulation. The airline's choice of RM method will determine what kind of information will be stored in each element of the shadow matrix. For example, EMSRb and DAVN, which both calculate discrete numbers of seats to be made available to each fare-class, and so the matrix would simply have a binary variable reflecting whether or not each fare-class has availability greater than zero. However, the virtual fare-classes used by DAVN to control seat availability are not replicated in the shadow matrix, because third party sites would never have direct access to this information. Closing a virtual class may affect multiple path-classes, but this is consistent with the maximum information that third parties could obtain.

The choice to model either Selective Polling or Caching determines the way in which the stored shadow matrix will be populated with data. For Selective Polling, there are a number of potential decision rules that could determine which flights would rely on the shadow matrix and which would check the real-time matrix, as described in Section 5.2.1. The appropriate thresholds and cutoff factors would have to be calculated based on the airline's network, fare structure, and risk tolerance. Caching, on the other hand, is a much more straightforward procedure in which the entire matrix uses a single update procedure. For this reason, Caching is the primary focus of the research presented here, but many of the modeling concepts described in this section are valid for both methods.

As discussed in Section 5.2.2, third parties who use Caching have a variety of different algorithms for deciding the interval on which to update elements of their stored data. We have used a relatively simple implementation for initial experiments, in which we refresh the entire matrix at certain fixed intervals. Specifically, there are 16 checkpoints called "timeframes" during each run of the PODS simulation, and at the beginning of each timeframe, we re-populate the whole cache with fresh data before processing the next passenger arrival. All airlines in the simulation would always be subject to the same update rates because a third-party distributor makes the decisions that control these factors.

5.3.2 Passenger Choice Process

The key to the revenue impacts of a system like Caching is that between updates of the shadow matrix, some ticketing transactions will be accepted directly by the CRS. These will not be reflected in the shadow matrix, so some portion of passengers would make their choice of paths based on erroneous data, and then might not be able to purchase the desired itinerary. Several new additions to the passenger choice process in PODS will help make our model more realistic.

First and foremost, only a fraction of passengers today purchase tickets through distribution channels where Caching or Selective Polling are in use. As a result, we use a variable parameter set between zero and 1.0 to control the probability that a simulated passenger will choose a booking channel that relies on the shadow matrix. Increasing this parameter corresponds to an increase in website usage, so we can see how revenue results vary with more and more internet bookings.

Some passengers will use the third party booking channel, but have no problems booking their travel, because they happen to select a combination of origin, destination, and fare-class (an "ODF") that is open in the CRS as well as the shadow matrix. Other passengers who rely on the shadow matrix will choose an ODF where the stored information in the shadow matrix is wrong. When they attempt to make their booking, the airline's records will show that no seat is available, so they will be rejected by the CRS. Three possibilities have been developed for the outcome of an error in the shadow matrix.

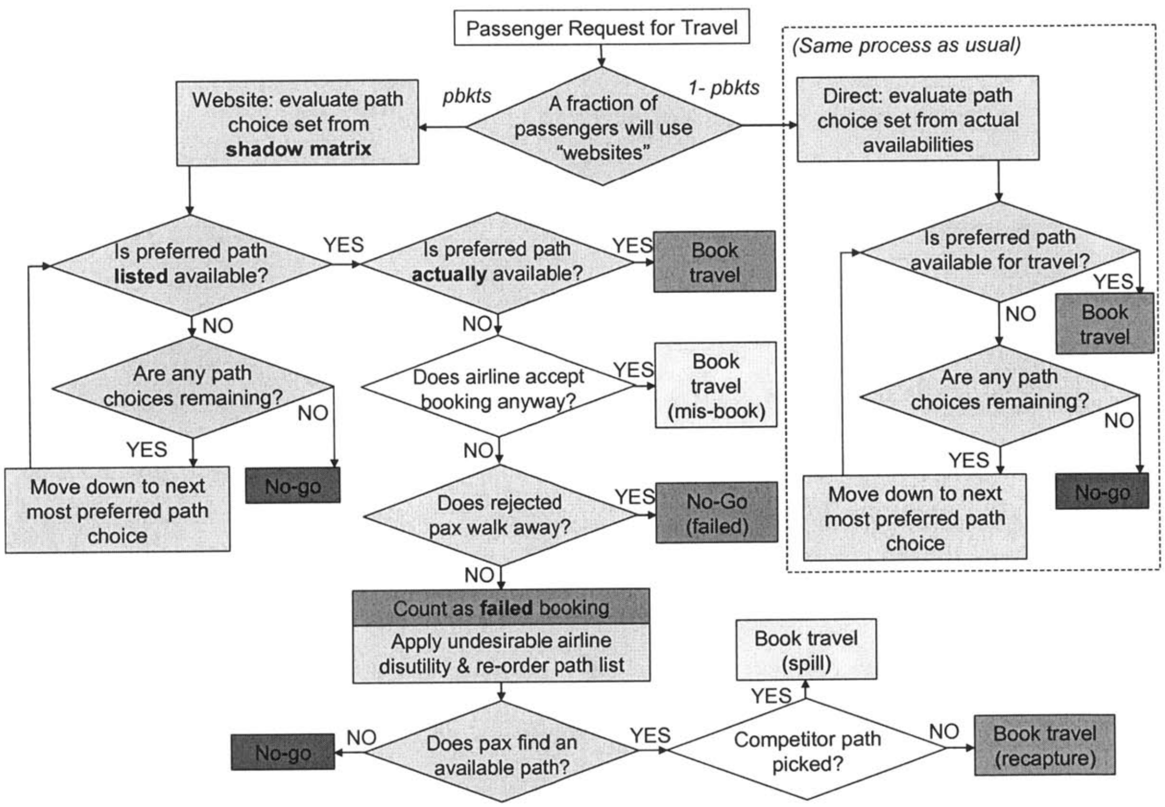
The first possibility is that some airlines may decide to honor the transaction, even though the cache was wrong. This would depend on the contractual arrangements between the airline and the distributor, and possibly the level of sophistication in the technological systems. We use a binary flag variable to indicate whether each carrier would reject or accept bookings for closed classes that come from a caching website (aircraft capacity permitting). This option is somewhat of a worst case scenario, because it is highly unlikely that airlines would choose to override the authorizations of the RM system in blanket fashion. We simply want to take advantage of the PODS simulations to quantify how the revenue impacts of accepting the incorrect bookings change as the use of third party channels increases. If there are noticeable trends or inflection points, this might point out new strategies for airlines to pursue.

In practice, it is unlikely that airlines would accept booking requests based on caching errors. Instead, the RM system would reject any attempted booking that involves a closed fare-class, and this may lead to customer service problems. For example, we model the second possible outcome of a shadow matrix error where the passenger who cannot obtain their desired ticketing decides not to travel, or becomes a "No-Go" passenger. This is also a rather extreme case, but could be realistic for certain circumstances, such as passengers who were only planning to travel because of a special advertised discount fare. Again, we can use the simulations to understand how this scenario impacts revenue results, hoping to gain insight about the passenger choice process.

The third possibility when a passenger is rejected by the CRS is that they have a more negative perception of the airline that rejected them. Many people who have used internet sites to book travel have had an experience where they select flights and enter all of the information, only to find out at the end of the transaction that the seats they had selected are "no longer available," and they have to start the entire booking process over again. Some passengers would assign blame for this poor service to the particular website, while others would blame the airline. As a first approximation, we have developed a variable parameter where the passenger attributes a disutility to the rejecting airline before re-planning their trip. A very high value of this parameter means that the passenger does not select any path from the airline that rejected them, unless they have no other available alternative. A more moderate value simply influences the choice of paths from among the passenger's options, potentially leading to spill to the competing carrier.

Figure 5.1 shows a composite diagram of the modifications to the passenger choice process in PODS. The label "pbkts" is the variable parameter determining website usage. On the right side, the box contains a representation of the normal process, which will be followed by those passengers who go directly to the airline CRS instead of using a website. That same process is also replicated on the far left side, because passengers will use a shadow matrix to make their first determination of availability. In the center, a decision tree is presented giving the sequential order in which the potential scenarios described above would be evaluated.

Figure 5.1:
Modified PODS Passenger Choice Process



This model of the passenger choice process assumes that most revenue losses will come either from accepting passengers in the wrong booking class (mis-bookings in Figure 5.1) or from excessive failed bookings and spill. In this real world, revenues are also negatively affected by the fact that booking cancellations can return seats to inventory. A shadow matrix may not reflect these updates in a timely fashion, which could mean lost revenue if a passenger would have traveled, but they do not know the seat is available. At this time, we do not model cancellations within PODS, so this effect is not incorporated here. But, due to the complex revenue interactions that would arise if cancellations were included, these experiments are a reasonable first step towards understanding the problem as a whole.

5.3.3 Path-Based vs. Leg-Based Caching

Because these Caching experiments are introducing several new parameters and data elements simultaneously, the first set of experiments is focused on a simplified scenario in which the cache is stored on a path-class basis for the whole network, exactly parallel to the actual table of availabilities. The shadow matrix mirrors the way that the simulated airlines store and control their own inventory, so there would not be any revenue losses from the discrepancy between leg-based caches and network-based RM. However, revenues can be lost due to the fact that availability data in the matrix may be out of date for any individual booking. As discussed in Section 5.2.2, an O-D based cache is extremely memory intensive and unlikely to reflect actual industry practice, but using a more detailed cache will actually simplify the analysis of the revenue interactions, helping to provide the foundation for later work. The simulation experiments examine the effects under different RM methods and at different levels of website usage. Each of the three passenger choice scenarios described above ("Airline Accepts", "No-Go" and "Disutility") is studied separately. Details of the simulation set-up and experimental results for path-based Caching are presented in Section 5.4.

In the second set of experiments, we turn our attention to the more realistic scenario in which the cache typically only contains leg-based availability and must somehow "compute" availability estimates for multi-leg paths. While data freshness will still be a problem in this scenario, we expect the discrepancies between a leg-based cache and path-based availabilities will also be an important factor in revenue losses.

In the leg-based Caching experiments, the shadow matrix is stored on a leg-basis, regardless of the RM method of the airlines. This means that for an airline using a network RM method, the local availability value will be passed to the shadow matrix, irrespective of any connecting path availabilities over that leg. Now there will be differences in how passengers assemble their choice set depending on the type of itinerary desired. Local passengers who only have single-leg availability requests can easily be handled by directly polling the corresponding element of the shadow matrix. If a leg is chosen by the passenger based on shadow matrix availability, they proceed to the booking process as before. The only errors in local leg availability would be due to stale data, and these errors will be handled using the same passenger choice structure we have already developed for path-based Caching, potentially including Airline Accepts, No-Go, or Disutility behaviors.

The handling of connecting passengers with multi-leg itineraries will be somewhat more complicated. The third party website will evaluate multi-leg requests by using leg-by-leg local availability data to decide whether or not there is an available seat to sell on any given path. Thus, if the local inventory has been closed down by an airline, then that leg will show as closed to all connecting paths using that leg. If a connecting passenger can find a path where all of the desired legs are locally available, then they would proceed to the booking process, at which point there are two approaches that airlines may take:

- **Sell connecting inventory**—In this case, the airline evaluates the booking request based on the inventory availability of the actual connecting path, which may already be closed. Those passengers who are rejected will be subject to the usual No-Go/Disutility passenger choice structure as before, which may mean they return to the choice process to try to find an alternative option to book. Note that this alternative has many of the characteristics of Journey Control, which was described in Section 4.3.1. The passenger is hoping to book an itinerary made up of local legs, but the airline sells connecting path inventory if and only if it is still available.
- **Sell local inventory at the connect fare**—For this approach, the airline chooses to sell seats out of local inventory to travel-site connecting passengers in an attempt to match what the passengers are expecting from a leg-based cache. This reduces the chance of rejecting a passenger due to a local/connect discrepancy, but it gives rise to potential forecast errors unless the passengers are properly accounted for. This alternative most closely resembles Connect-Closed Bypass, due to the fact that a connecting passenger is obtaining local seats at the connect fare. However, the connecting path may be open in this case, and connecting availability is not considered when the airline evaluates this request.

Because of the experiments presented in Chapter 4 showing the effectiveness of Bypass Compensation Methods, several alternatives were considered for addressing the potential forecast errors created by selling local inventory. First, the airline must choose how to record a connecting booking that uses local inventory in their historical database. The ticket purchase can be recorded as either a connecting passenger (even though the connecting path-class might already be closed) or as a local passenger on each of the two legs (because that is the inventory that is the source of the seats purchased). If the connecting path-class is closed, then the passengers would not be included in the forecast, so we can also turn on a compensation method in which a separate count of passengers is maintained and the average value is added to the forecast before re-optimization. This compensation method is not relevant for the "record as local" alternative because the passenger will always be successfully counted with the other local leg bookings.

Details of the simulation set-up and experimental results for leg-based Caching—including both concepts for selling inventory to connecting passengers presented above—are presented in Section 5.5.

5.4 Path-Based Caching Experiments

As described in the previous section, the first set of Caching experiments was conducted with a path-based shadow matrix as a means to refine our understanding of the basic mechanisms of revenue loss, error rates, and key performance indicators such as load factor. All experiments in this section were conducted using the EMSRb leg RM method and the DAVN network RM method in the following pairings: EMSRb vs. EMSRb, DAVN vs. EMSRb, and DAVN vs. DAVN.

5.4.1 Model Parameters

The base case for these experiments was both airlines using leg RM with no website usage and thus no Caching. The idea is to quantify the negative impacts of current practice, so that airlines understand how much they have to gain by eliminating the practices. In later work, other researchers can also test potential remedies that could restore the airline to the expected levels of revenues gains.

Once the baseline case where both airlines use leg RM was completed, we compared it to a similar set of runs for the case where one airline uses network RM while the other continues to use leg RM and to the case where both airlines use network RM. For the Caching scenarios, five levels of website usage were tested (specifically 0.2, 0.4, 0.6, 0.8, and 1.0) in order to measure the impact of Caching as website usage grows.

In addition to the level of website usage, there were three other variable parameters which had to be set for each simulation run. These were the flag indicating whether or not each airline accepts website passengers who attempt to book an ODF that has already been closed, the flag indicating whether or not website passengers who are rejected become No-Gos, and the value of the disutility that is assigned to a rejecting airline for those passengers who do remain in the booking process. Each of these variables was tested in a separate set of simulation runs, so that their effects could be independently measured. Thus, there were three "scenarios" to be tested: Airline Accepts, No-Go, and Disutility. The Baseline required no additional parameters—all of the features were turned off except Caching. The parameters for Airline Accepts and No-go were simply 1/0 flag values to turn each feature on or off as appropriate. The initial value selected for the Disutility experiments was 1000. This value is quite large compared to the other disutilities in the PODS model, which are on the order of hundreds not thousands. However, this value helps to measure an extreme case, and more typical values will be the subject of the second phase of experiments. All of the parameters described above are identical for both simulated airlines except for the revenue management method they are assigned.

5.4.2 Simulation Results

No-Go Scenario

We present the No-Go scenario first, because it is actually the most straightforward, both in terms of the passenger choice process, and in terms of revenue results. In this scenario, if a passenger finds a fare-class available in the shadow matrix that turns out to be closed in the airline CRS, then they leave the booking process and their revenue is lost.

The chart in Figure 5.2 shows the revenue results for all three combinations of RM methods tested. The vertical axis shows total revenue collected by each airline ("AL1" or "AL2"), and the horizontal axis has the different levels of website usage tested in these experiments, from the base case of no Caching (probability=0.0) to the case where all passengers use websites that cache (probability=1.0). As expected, revenues decline for both carriers, regardless of RM method. Also, the declines are relatively uniform as the probability of using a caching site increases, showing a nearly linear relationship between the level of path-based Caching and the level of revenue loss under these scenarios.

Figure 5.2 essentially presents the impact of "turning on" Caching. Notice that even in the base case with no Caching (shown at the far left), revenue results differ widely because of the different RM methods employed in each pairing. To evaluate the effect on the different RM pairings more readily, Figure 5.3 shows the revenue losses on a percentage basis compared to revenues earned under the same RM pairing with no Caching.

Figure 5.2:
Revenue Results for Three RM Method Pairings
(No-Go Scenario)

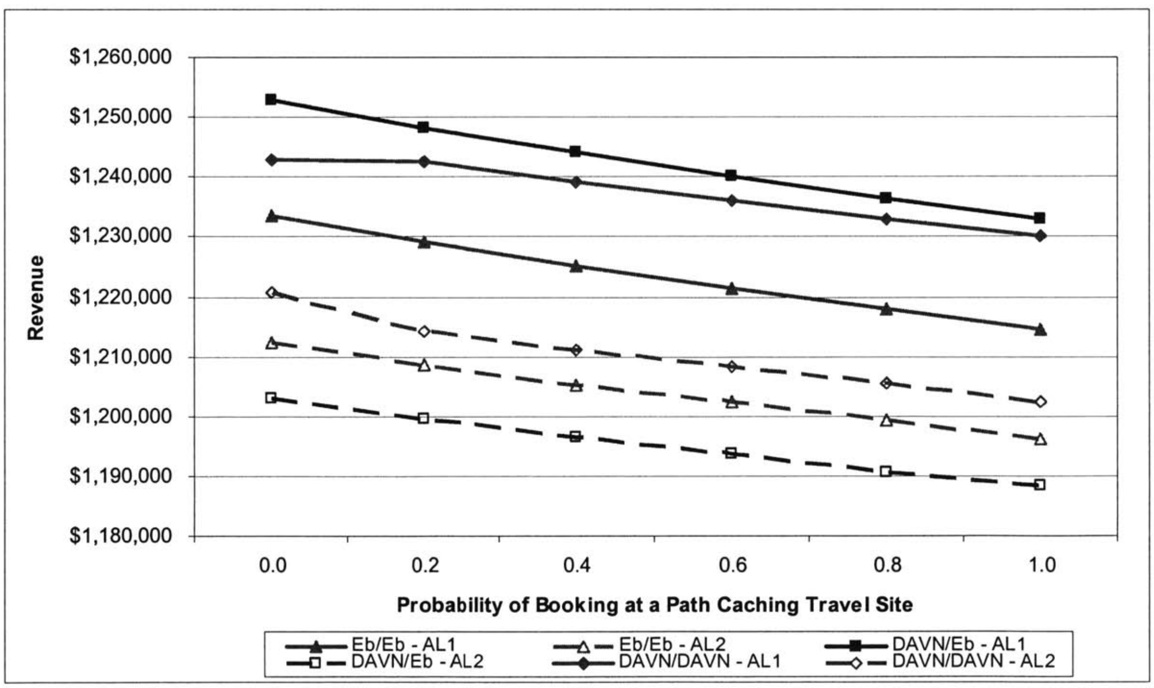
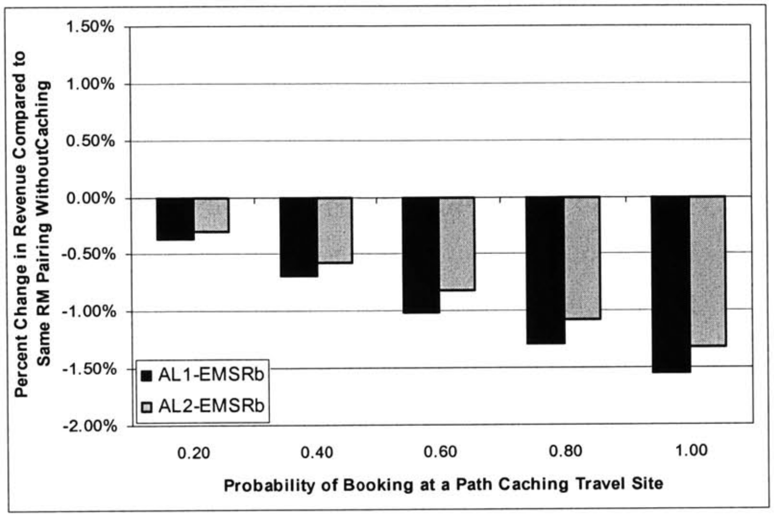
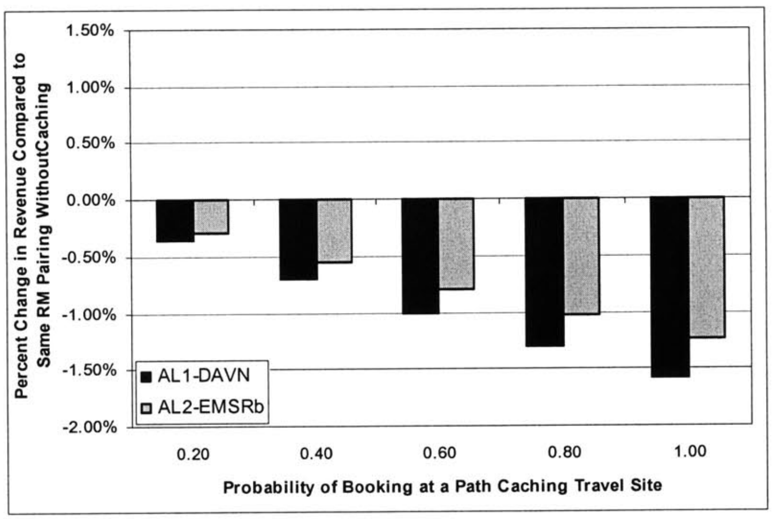


Figure 5.3:
Revenue Gains/Losses Due to Caching for Three RM Method Pairings
(No-Go Scenario)





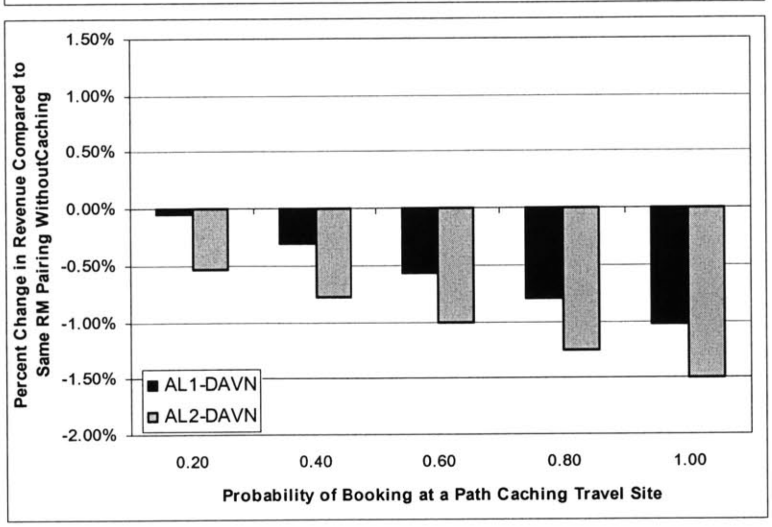


Figure 5.3 shows that the order of magnitude of revenue loss in the various RM pairings is very similar, particularly in the first two cases, ranging from -0.3% at low levels of website usage to -1.3% or more at 100% website usage. In the last RM pairing (DAVN/DAVN), the losses are slightly smaller for Airline 1 and slightly larger for Airline 2, although losses for both airlines exceed 1% of revenues at 100% website usage.

The main reason that there are revenue losses from the No-Go scenario is that a passenger who is rejected in the booking process can never be recovered through any sort of re-planning or sell-up, even if the passenger had been willing to pay a higher fare than their first choice. At each re-optimization, the RM optimizer can make small adjustments, but in each time period, there is always a fraction of demand that does not materialize. Some of the revenue decline is explicitly seen in lower load factors, but also, because lower fare-classes stay open longer than in the base case, the traffic mix shifts down towards the lower classes, resulting in a decrease in overall revenues.

Although the carriers lose revenue as the use of caching websites increases regardless of RM method, the use of the network RM method helps Airline 1 to improve its position relative to when both carriers were using leg RM, as shown in Figure 5.4. With no Caching, network RM produced gains of 1.55% over the use of leg RM, and this value declines only slightly as path-based Caching increases, to 1.52% at 100% caching website usage. Thus, while path-based Caching does impact revenues for both carriers, network RM maintains its ability to achieve incremental revenue gains over the use of leg RM methods.

Figure 5.5 shows the same graph for the case where both airlines are using network RM. Recall that when Caching is not present, two carriers using network RM can each gain approximately 0.7% in revenue compared to when both carriers use leg RM. Here we see that because of Caching and the No-Go behavior, Airline 1 gains as much as 1.3% in revenue while Airline 2 gains only about 0.5%. This is largely due to the fact that Airline 1 has a stronger schedule than Airline 2 in the simulation. The high number of No-Gos due to Caching leaves empty seats on both carriers, and when high-fare passengers arrive late in the booking process, there are plenty of seats available. The high-fare passengers are also typically more schedule sensitive, leading them to prefer Airline 1 at the expense of Airline 2.

Figure 5.4:
Incremental Revenue Gains of Network RM When One Carrier Uses Network RM (No-Go Scenario)

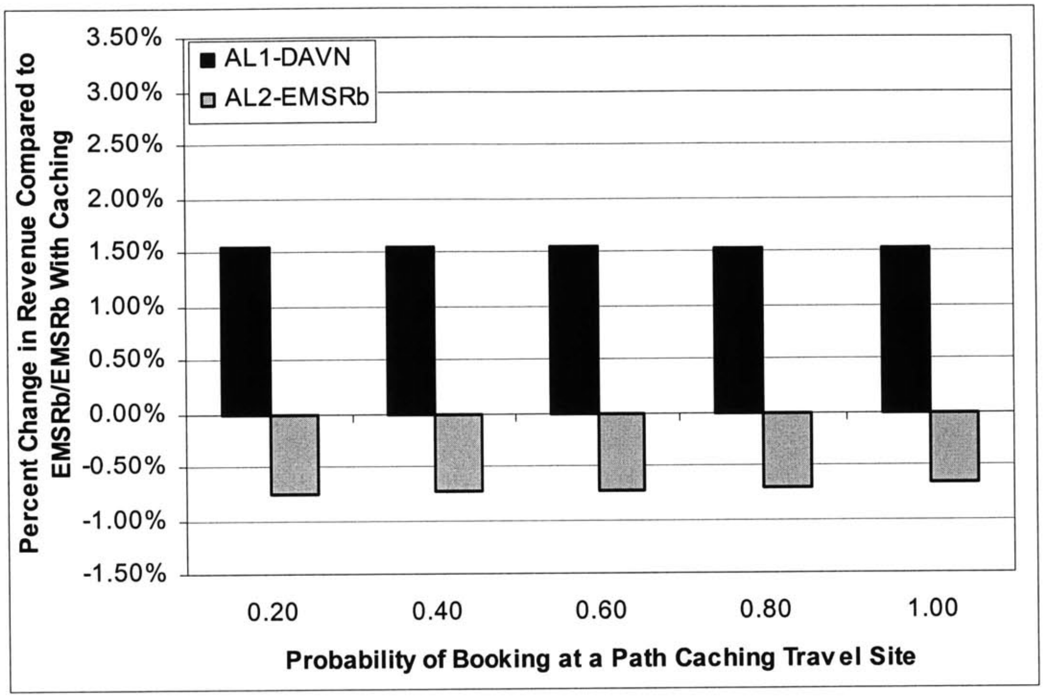
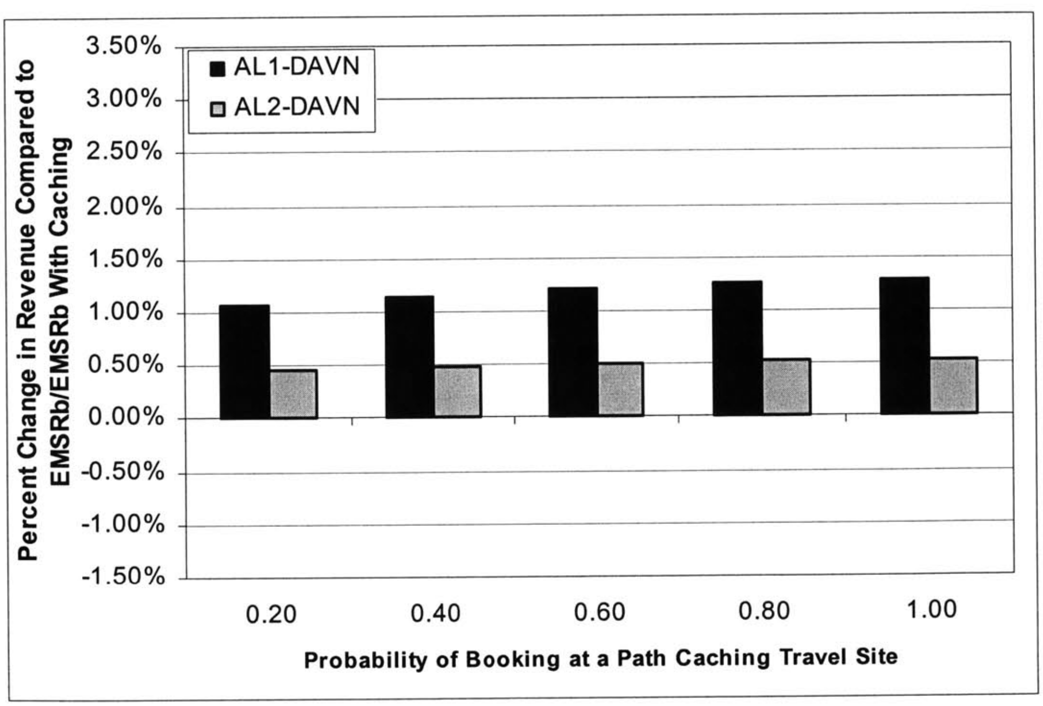


Figure 5.5:
Incremental Revenue Gains of Network RM When Both Carriers Use Network RM (No-Go Scenario)



Airline Accepts Scenario

The second scenario measures the revenue impact of accepting passengers into closed fare-classes if they happen to come from a caching website. The idea behind this strategy is to minimize problems attributable to poor customer service as seen above. Clearly the airlines will experience fairly significant revenue losses from accepting passengers in contradiction of the RM system booking limits, but these will need to be measured against airlines' other strategic options for dealing with Caching. Figure 5.6 shows that the revenue results for the different combinations of RM methods are somewhat mixed.

In the case where both carriers use leg RM, revenues are slightly higher than in the No-Go scenario, but both carriers still gradually lose revenue from accepting passengers in the presence of path-based Caching. When only one carrier uses network RM, their losses are much less than the No-Go scenario, but their leg RM competition continues to feel serious losses. If both carriers use network RM, Airline 1 gains slightly while Airline 2 loses revenues at nearly the same rate as the other two cases. Figure 5.7 shows these revenue impacts on a percentage basis for each of the three RM pairings.

Figure 5.6:
Revenue Results for Three RM Method Pairings
(Airline Accepts Scenario)

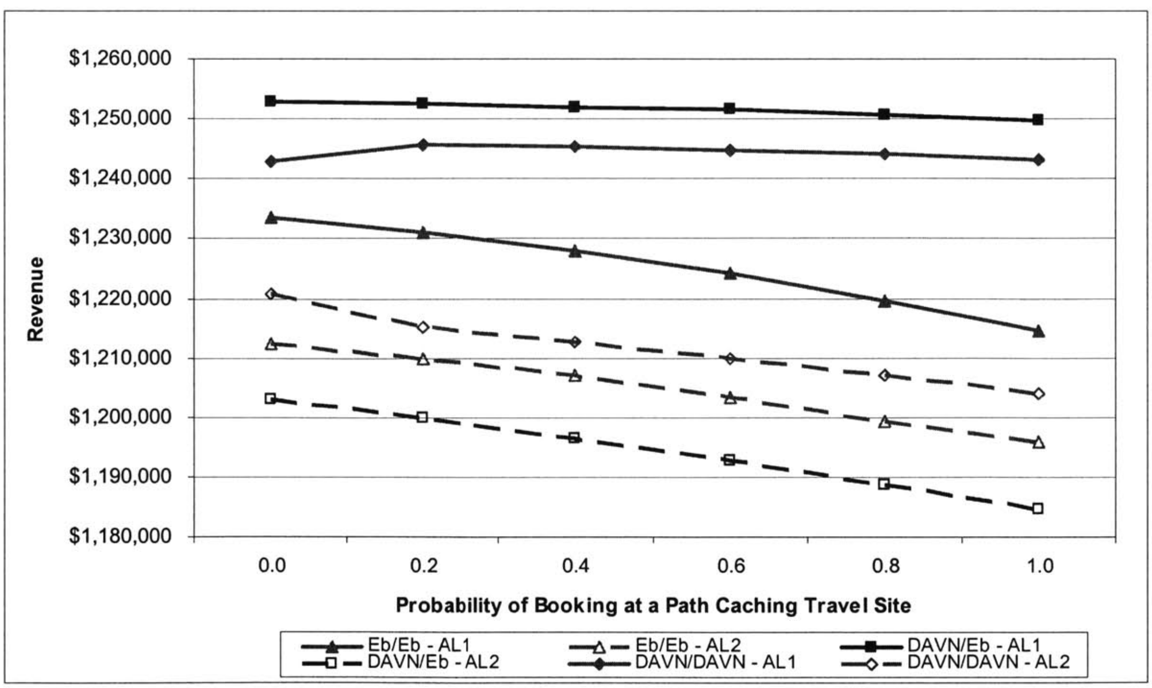
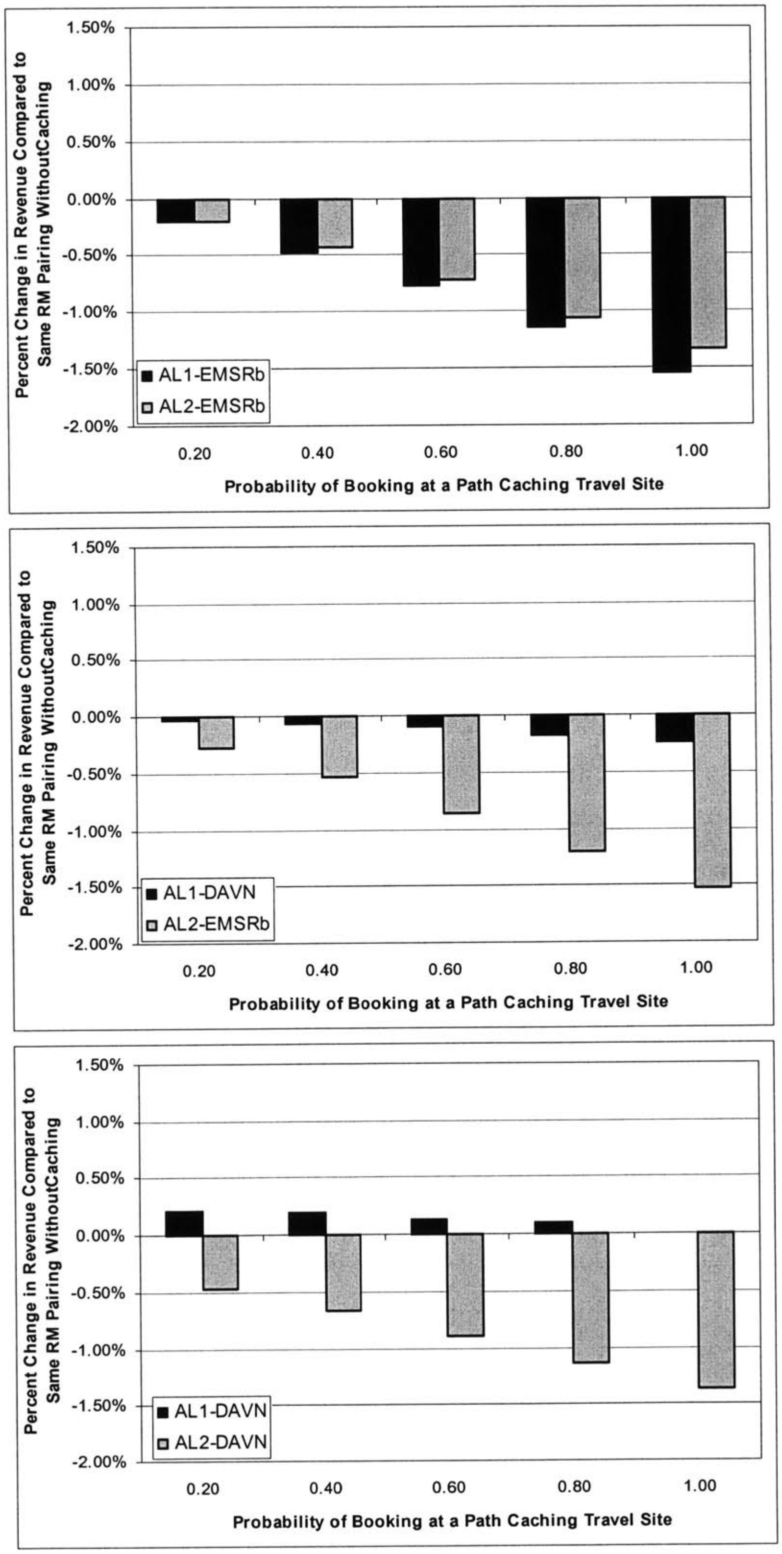


Figure 5.7:
Revenue Gains/Losses Due to Caching for Three RM Method Pairings
(Airline Accepts Scenario)



Load factors increase for both carriers in the Airline Accepts scenario, sometimes by as much as 1%, so it may seem somewhat surprising that revenues decline. However, the revenue losses come from taking too many low fare passengers early, which means that some late arriving, high fare passengers are unable to find their desired seat. By comparison, in the No-Go scenario, losses come mainly from the fact that some later-arriving, high fare passengers do not travel at all when they encounter a shadow matrix error while trying to book their first choice.

As with the No-Go scenario, we also evaluate the performance of network RM in the presence of Caching. The revenue results for one carrier using network RM are presented in Figure 5.8. When the carriers accept erroneous bookings, network RM produces extremely high revenue gains over leg RM methods, ranging from the baseline level of 1.55% at 20% caching website use up to 2.89% at 100% use. The reason for these gains is that the use of virtual buckets in the DAVN method allows the airline to separate the fares in each class into high-value and low value groupings and to control availability differently for each group. Although the airline might accept a few low value bookings initially, later re-optimizations by the network RM carrier can adapt to these accepted bookings and continue to map the most valuable fares into open classes to be sure that enough seats are available. On the other hand, the EMSRb logic employed by the leg RM carrier cannot distinguish between high-value and low-value fares. Accepting a low-value fare blocks a seat from all of the potential customers who might use that fare class, even the highest-value customers.

When both airlines use network RM and the Airline Accepts strategy, Airline 1 continues to achieve strong revenue gains and Airline 2 is able to benefit as well, as shown in Figure 5.9. In this case, both carriers are re-optimizing and re-bucketing, so they can make adjustments which allow them to gain revenues even in the presence of path-based Caching. As before, the stronger schedule of Airline 1 leads to a larger share of the gains. Airline 1 achieves approximately 2.5% gains, while Airline 2 achieves nearly 0.7%. Thus, both carriers reach or exceed the same level of incremental revenue gains from the use of network RM as they do without the presence of Caching.

Figure 5.8:
Incremental Revenue Gains of Network RM When One Carrier Uses Network RM (Airline Accepts Scenario)

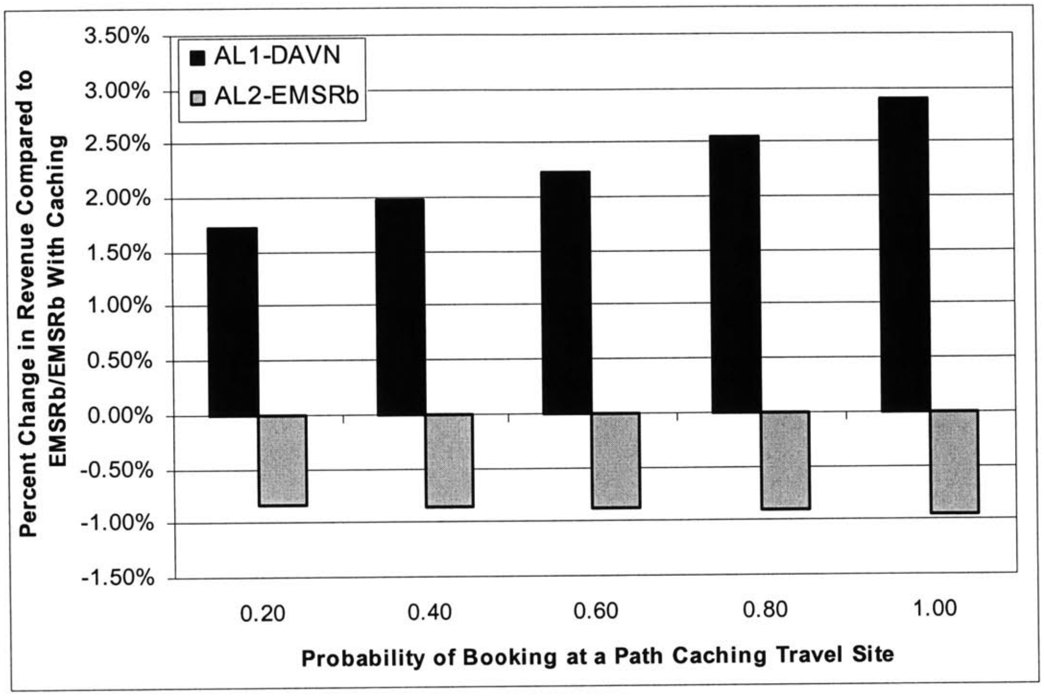
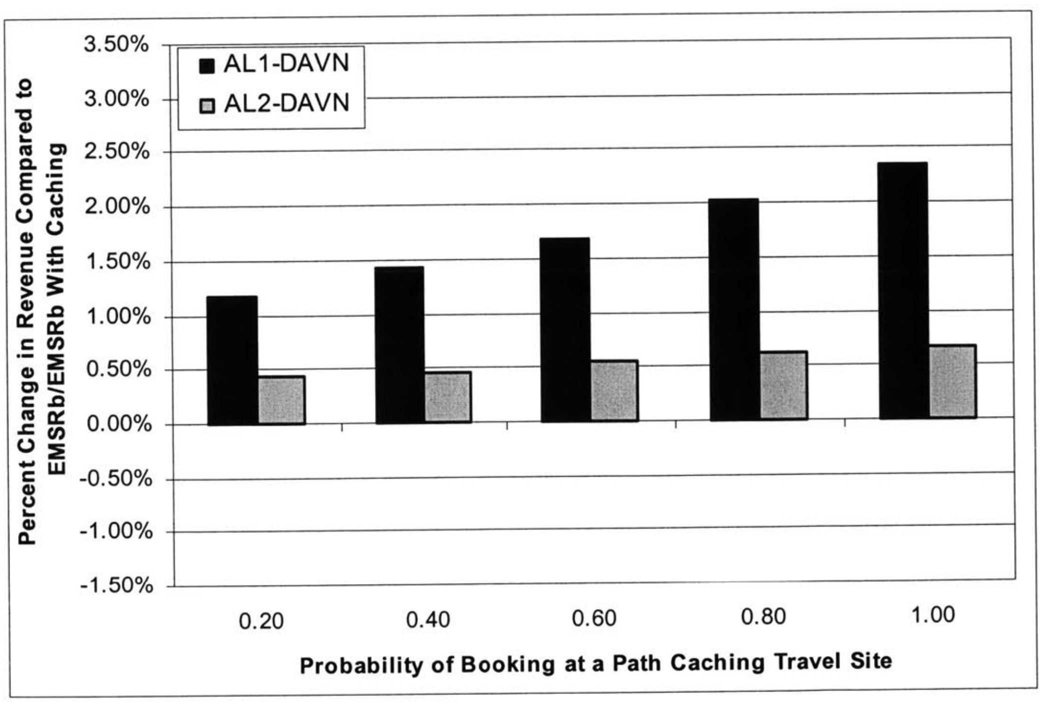


Figure 5.9:
Incremental Revenue Gains of Network RM When Both Carriers Use Network RM (Airline Accepts Scenario)



Disutility Scenario

The Disutility scenario is the most moderate of the three, because passengers are not accepted or rejected automatically due to shadow matrix errors. Instead, they have a chance to re-plan their trip by trying to book in another fare-class further down in their re-ordered choice set. As a result, the revenue losses in this scenario are not as extreme as in either of the first two cases of path-based Caching presented. Passengers in this scenario have a chance to re-plan their trip, and in fact will still purchase on the rejecting carrier if no other travel options are available. Figures 5.10 and 5.11 contain the comparison of revenues for all three RM method pairings. As shown in both figures, revenue gains and losses show similar patterns to the Airline Accepts scenario, but the magnitude of gains and losses is smaller.

Figure 5.10:
Revenue Results for Three RM Method Pairings
(Disutility Scenario)

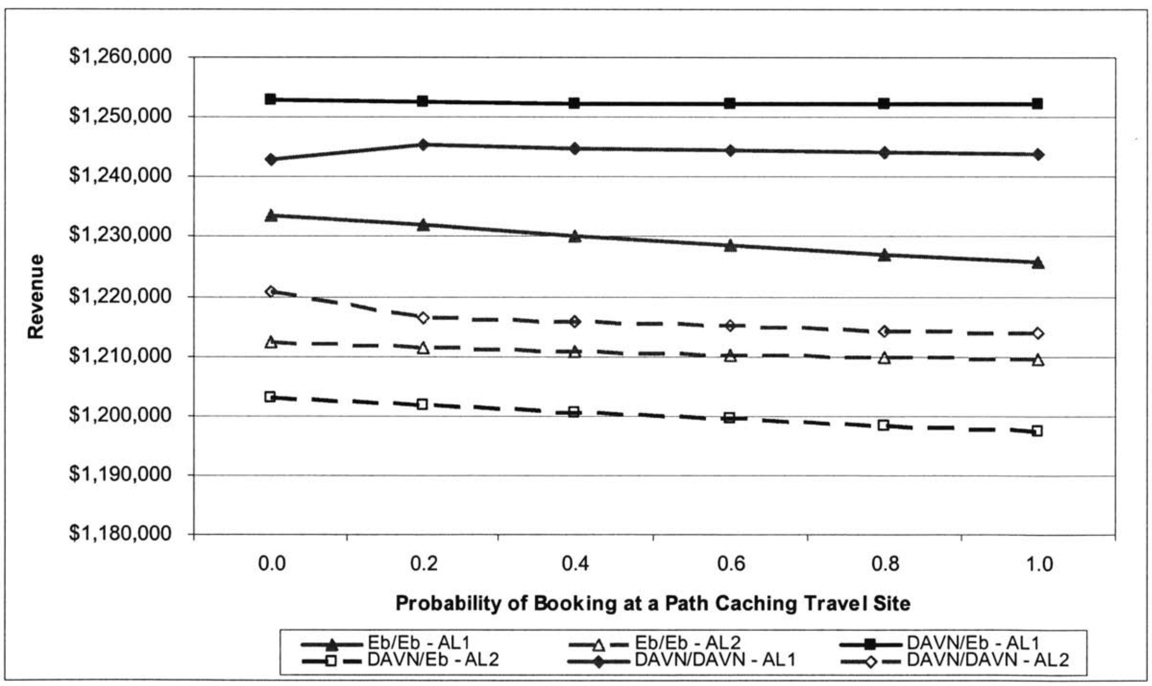
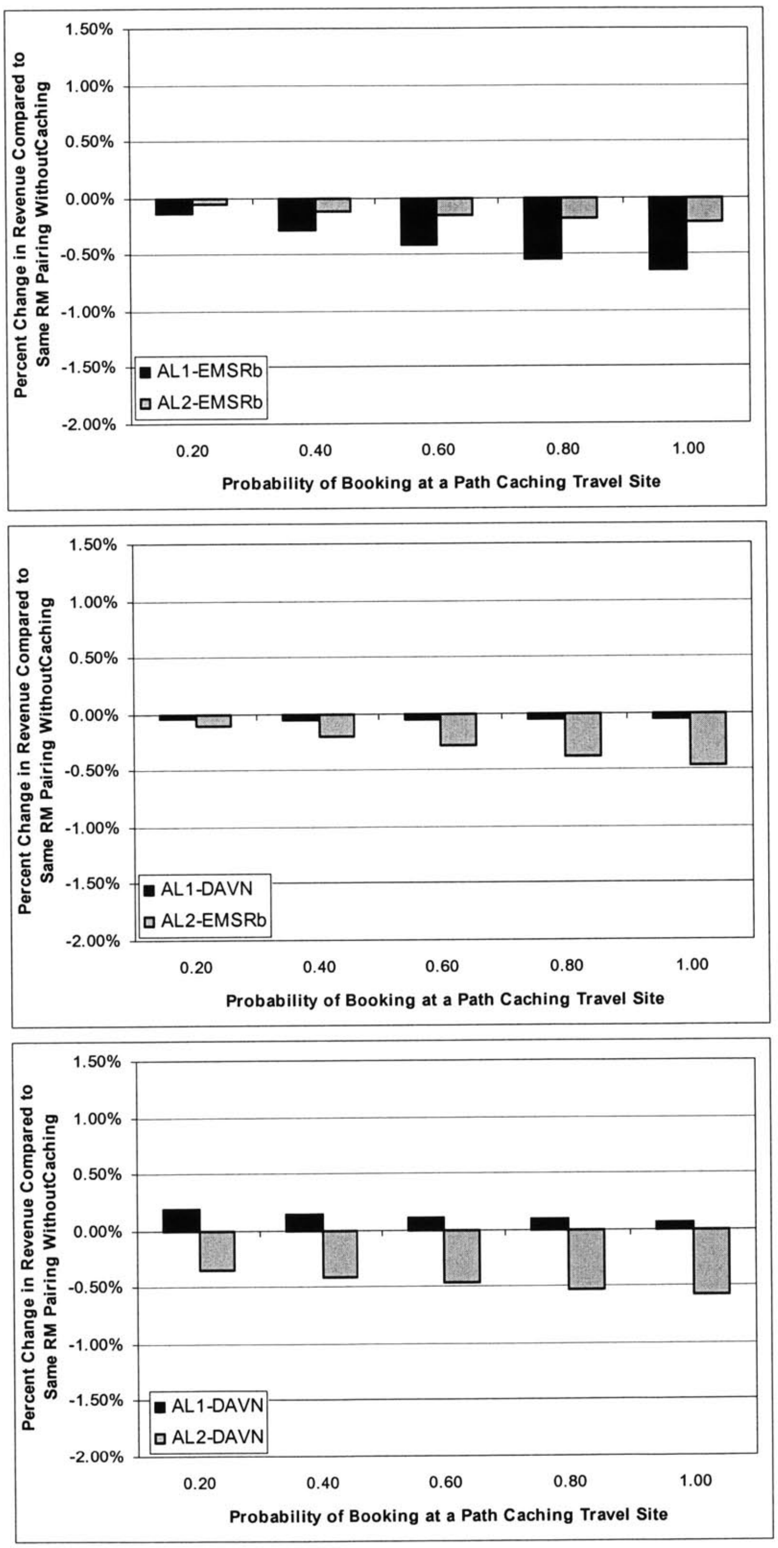


Figure 5.11:
Revenue Gains/Losses Due to Caching for Three RM Method Pairings
(Disutility Scenario)



This is another scenario in which the disparities between the two carriers that are inherent in Network D become apparent. For example, in the case with two leg RM carriers, with both carriers subject to the exact same parameters and disutilities, the revenue declines are steeper for Airline 1 than for Airline 2. Because Airline 1 is initially the stronger of the two, it has more to lose when Caching interrupts the normal booking processes. When only Airline 1 moves to network RM, then its revenues stay nearly flat, because the DAVN method can respond to the changes induced by passenger disutility behavior. The EMSRb carrier ends up leaving lower classes open too long and losing revenues at an even faster rate than it did when both carriers used leg RM.

The results presented in Figure 5.11 may create the sense that Caching, even with a strong disutility, is not a particularly bad thing if you are using a network RM method. However, the levels of gains and losses shown have most likely been mitigated by the fact that this is a closed, 2-carrier experiment, and we have set the parameters so that both carriers are equally affected by the disutility. Passengers who spill from one carrier end up on the other carrier, and vice versa, so that the two carriers simply end up "trading" passengers back and forth during the simulation. Further experiments should be conducted to measure the revenue impacts when carriers do not necessarily benefit from capturing the entire amount of spilled revenue coming from their competitors.

Just as the overall revenue losses are milder in this scenario, the revenue gains for from the use of network RM are also somewhere between the two extreme cases of No-Go and Airline Accepts. For Airline 1, these gains range from 1.66% with only 20% website usage up to 2.15 % at 100% website usage. These values are depicted in Figure 5.12. Figure 5.13 shows the revenue gains when both carriers are using network RM. In this case, Airline 1 does gain slightly more than in the No-Go scenario. However, Airline 2 gains less than either the No-Go or Airline Accepts cases.

Figure 5.12:
Incremental Revenue Gains of Network RM When One Carrier Uses Network RM (Disutility Scenario)

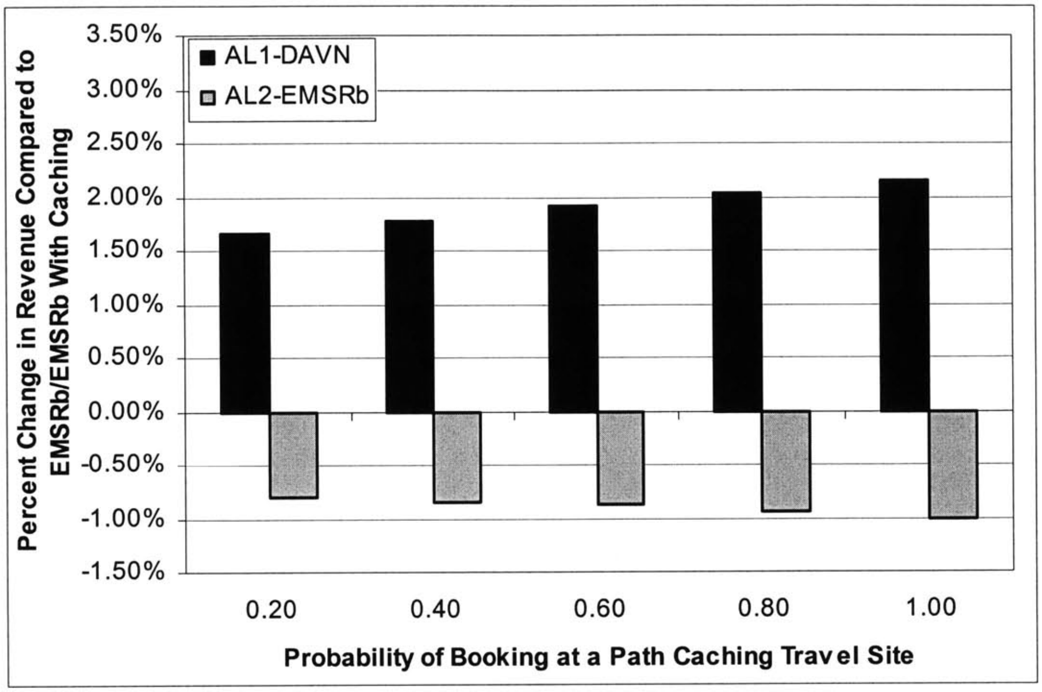
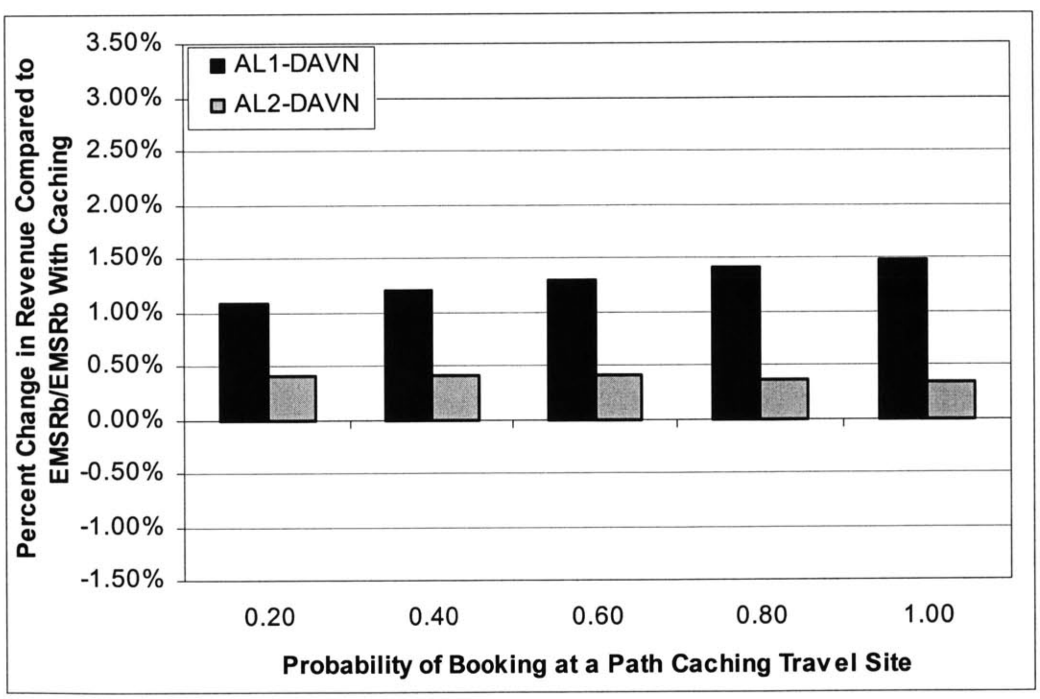


Figure 5.13:
Incremental Revenue Gains of Network RM When Both Carriers Use Network RM (Disutility Scenario)



5.5 Leg-Based Caching

5.5.1 Model Parameters

As with path-based Caching, the base case for the leg-based Caching experiments is both airlines using leg RM (EMSRb) with no website usage and thus no Caching. Simulations were run at the same five levels of website usage, and for all three combinations of RM-methods: EMSRb/EMSRb, DAVN/EMSRb, and DAVN/DAVN.

Although all three of the scenarios presented previously (Airline Accepts, No-Go, and Disutility) could be tested under leg-based Caching, we decided to focus our attention in certain areas. Specifically, we did not perform any simulations under the Airline Accepts scenario, and the No-Go scenario was primarily used to help define a boundary condition on the results. One batch of simulation runs was also devoted to developing a lower and more reasonable parameter value for the Disutility scenario.

Recall from Section 5.3.3 that there are two different options for how airlines can respond to connecting booking requests coming from the third-party caching sites. The airline can sell connecting inventory, or they can sell local inventory at the connecting fare. We tested both options, but gave both airlines the same response, i.e., both airlines sell connecting inventory or both sell local inventory. In addition, airlines have several options for how they record bookings and whether or not they make adjustments to their forecast to compensate for accepting connecting bookings after the connecting path has been closed. After some initial tests of the various permutations, two key combinations were selected and tested more thoroughly, as described in the next section.

5.5.2 Simulation Results

Sell Alternatives

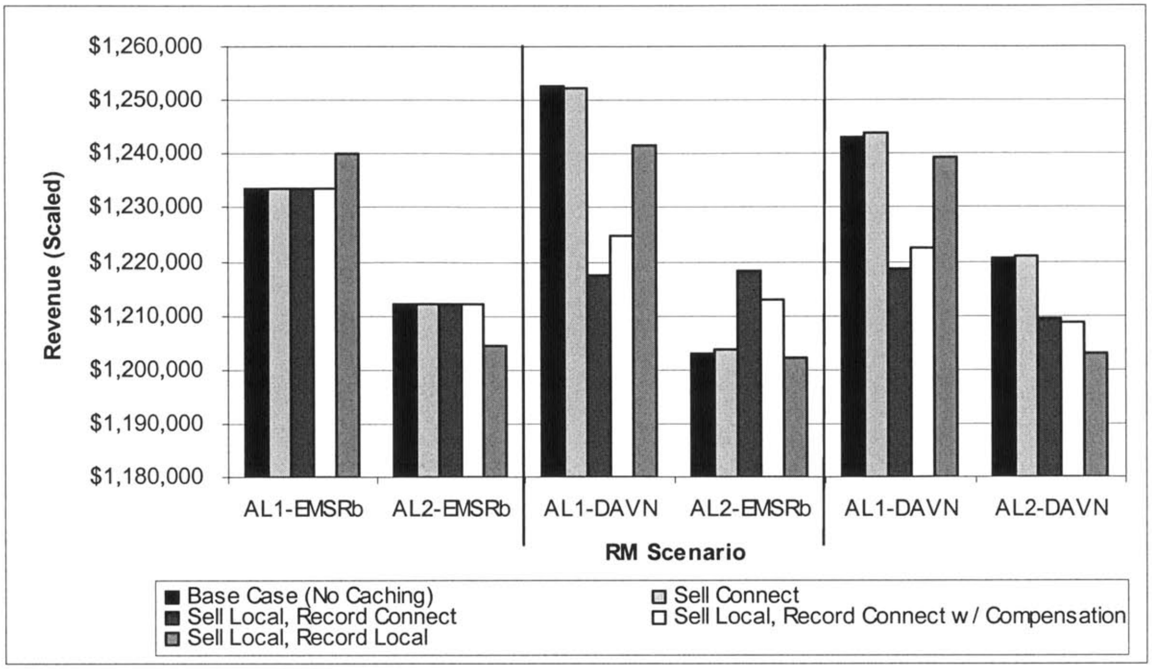
The first batch of simulations that incorporated leg-based Caching were an exploration of the various sell alternatives that each airline might choose to use when faced with leg-based Caching. In this case, both airlines were assigned the same response, essentially simulating what would happen under different RM method pairings if the entire industry were to choose the same selling strategy. In addition to a base case with no Caching in place, the following four scenarios were each tested at 100% use of caching websites:

- **Sell Connect**—In this case, the airline always checks connecting availability and will only sell the passenger a ticket if the requested connecting path-class is still open.
- **Sell Local, Record Connect**—Here, the airline sells the local inventory that corresponds to the passenger's request (assuming it is still open), but records the passenger as a connection for the purposes of forecasting. If the connecting path is already closed, the data would not be incorporated into future forecasts.
- **Sell Local, Record Connect, with Compensation**—This case is the same as the previous case, except that the airline now uses data about connects who arrive after the connecting path was closed to improve their forecast.
- **Sell Local, Record Local**—Here, the airline chooses to maintain internal consistency when selling local inventory by recording the passenger together with the other locals on each of the legs traversed.

Note that the passenger choice process was not modified for this batch of calibration experiments; there was no Disutility or No-Go behavior, so there was no penalty for a rejected booking. The revenue impacts observed are only due to errors caused by using a leg-based cache to order the choice set and select an itinerary for booking. The results of all five possibilities, tested in each of the three RM method pairings, are presented together in Figure 5.14. Note that the vertical scale of the chart has been adjusted to emphasize the differences between the scenarios.

The first observation from the figure is that there is very little difference among the sell options in the case where both carriers use leg RM. This is to be expected because the EMSRb seat protection algorithm does not distinguish between local and connecting inventories. However, the algorithm does rely on forecasts of local and connecting passengers in order to determine the total expected demand and revenue on each flight leg, so revenues are somewhat impacted when connecting passengers are counted as locals in the last sell option. As a result of the reclassification, the airline makes different choices about the booking limits in anticipation of the local passengers. As in previous experiments, increasing seat protection levels for the valuable local passengers tends to benefit Airline 1 and hurt Airline 2.

Figure 5.14:
Revenue Results for Five Selling, Recording, and Compensation Alternatives (100% Leg-Based Caching)



Turning to the cases where one or both airlines use network RM, we see much more variation depending on the sell option chosen. The strongest option is clearly the choice to sell connecting inventory, which either preserves or in some cases enhances the level of revenue achieved with the same RM methods under the base case with no Caching. This is due to the fact that some early arriving passengers who would have received their first choice under the base case end up with a different choice set when relying on the leg-cache data, and presumably end up selling up or spilling to the other carrier. This is another case where high levels of Caching appear to benefit the airlines, but the gains are more likely attributable to the two-carrier nature of the model in which passengers are simply spilled between the two carriers. These artificial benefits would not be realized in the more heterogeneous real-world environment.

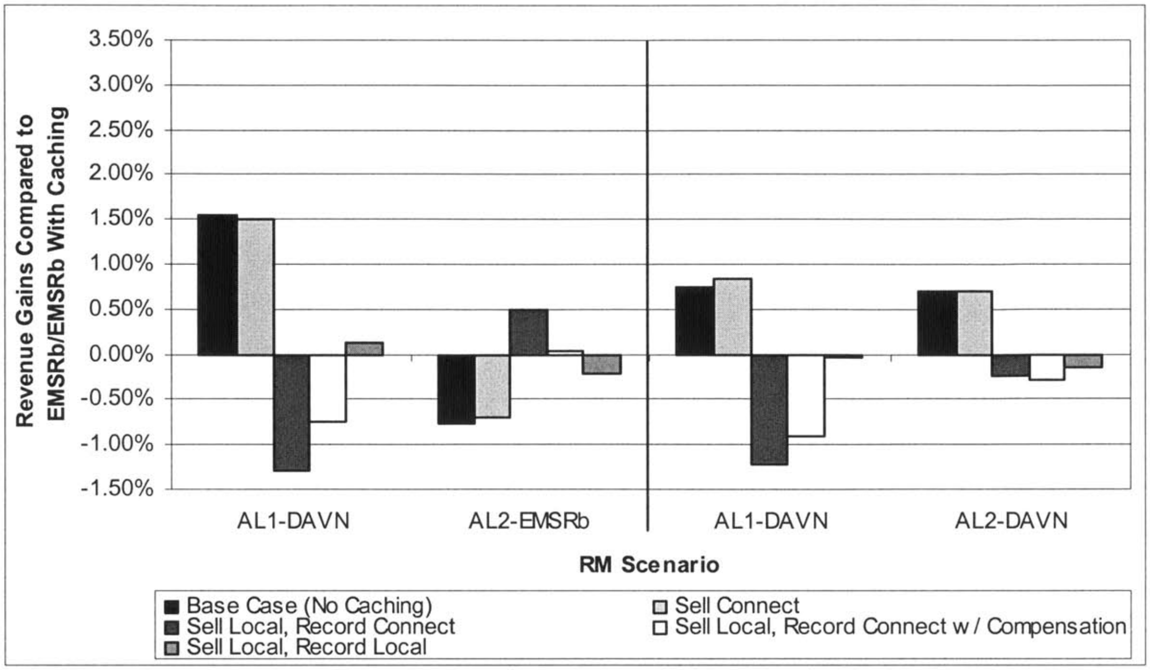
Recall that the Sell Connect option is a simplified representation of Journey Control, because a passenger is only able to obtain a ticket if the connecting path-class is open. As a result, there is no forecast error, and the minimal revenue impacts are simply due to slight shifts in which passengers are spilled and recaptured when they make their initial choice using a leg-based cache. Although more competitive tests would be required, these initial results would suggest that the implementation of Journey Control can provide strong revenue protection for an airline, particularly in light of the results of the other sell options, which are described below.

Of the three options for selling local inventory, there are distinct differences in revenues that depend on how the passenger is recorded. For the cases in which the passenger is recorded as a connection, there is a very noticeable negative impact to revenues that is only partially recovered with the use of a compensation method. By recording the passenger as a connection, the RM system builds a historical database that suggests other connections will arrive. At 100% website usage, all passengers are relying on caches, and because the network RM airline is selling only local inventory to them, these connections never materialize, leading to too many open seats in lower fare classes and decreased revenues.

The option of recording the passenger as a local appears to have much better revenue performance than the Record Connect options, but revenues are still down compared to the base case where there is no Caching. While forecast accuracy for the inventory allocation is not a problem here, recall that connecting passengers using local inventories pay the connecting fare, not the sum of the two local fares. As a result, the expected revenue calculations during the re-optimization process are inaccurate, and the airline cannot sufficiently protect itself from the same revenue losses that come from Connect-Closed Bypass, similar to the results discussed in Section 4.2.

We also examine the explicit revenue gains achievable using the network RM methods in the presence of leg-based Caching. Figure 5.15 depicts data for the same five scenarios, comparing the revenue gains from the use of network RM by one or both airlines to the leg RM case with the same level of Caching. In other words, the base case in the figure shows the benefit of using DAVN when no Caching exists, as previously presented in Figure 4.5. The other columns show the benefit of using DAVN compared to the use of leg RM in the presence of Caching under each of the other four sell alternatives.

Figure 5.15:
Incremental Revenue Gains of Network RM Methods Compared to Leg RM
(100% Leg-Based Caching)



Similar to the results shown in Figure 5.14, we see that selling connecting inventory is the strategy that is most comparable to the base case. On the other hand, choosing to sell local inventory and record the sale as a connect results in revenue losses across the board for carriers using network RM. Essentially, if the carrier chooses to sell local inventory but record the passenger as connecting, 100% leg-based Caching will completely reverse the revenue gains normally attributable to the use of a network RM method. Given these very sharp revenue losses when the network RM carriers use either of the Sell Local/Record Connect strategies, this alternative was not pursued further in simulation testing. Thus, the two sell alternatives that were tested in the remaining research are the Sell Connect and the Sell Local/Record Local (hereinafter referred to as simply "Sell Local") methods.

Disutility Calibration

The second batch of simulations in leg-based Caching was used to calibrate a reasonable disutility value for later experiments. Recall that a high value of the disutility would keep a passenger from booking on the rejecting airline unless it was their only available alternative. Anecdotal evidence suggests that passengers are actually more sensitive to price than to service quality associated with internet sites, and so the test values of disutility should represent only a minor penalty on the actual fares. Values of 10, 50, 100, and 150 were tested. Simulations were run with 100% leg-based Caching, and based on the results presented above both airlines were given the strategy of selling connecting inventory.

Results are presented in Figures 5.16 and 5.17, where lines have been added between the data points obtained by simulation. In the first chart, it can be seen that regardless of RM method pairing, most of the curves tend to flatten beginning at a disutility value of 50. Thus, higher values of the disutility do not influence the outcomes for two equivalent competitor airlines in the simulation. The second chart shows that the ability of network RM methods to create revenue gains is most significantly impacted around a value of 50, where the curves have their minima or maxima. Based on these results, a disutility value of 50 was chosen for all future disutility experiments.

Figure 5.16:
Impact of Variable Disutility on Revenue Results
(100% Leg-Based Caching)

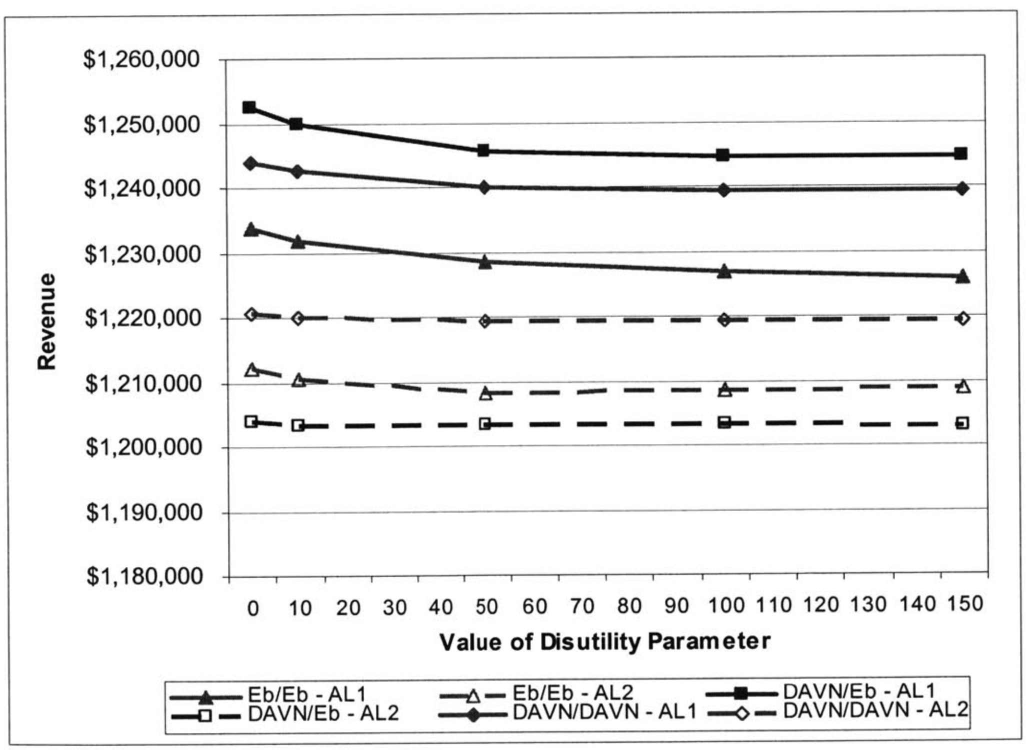
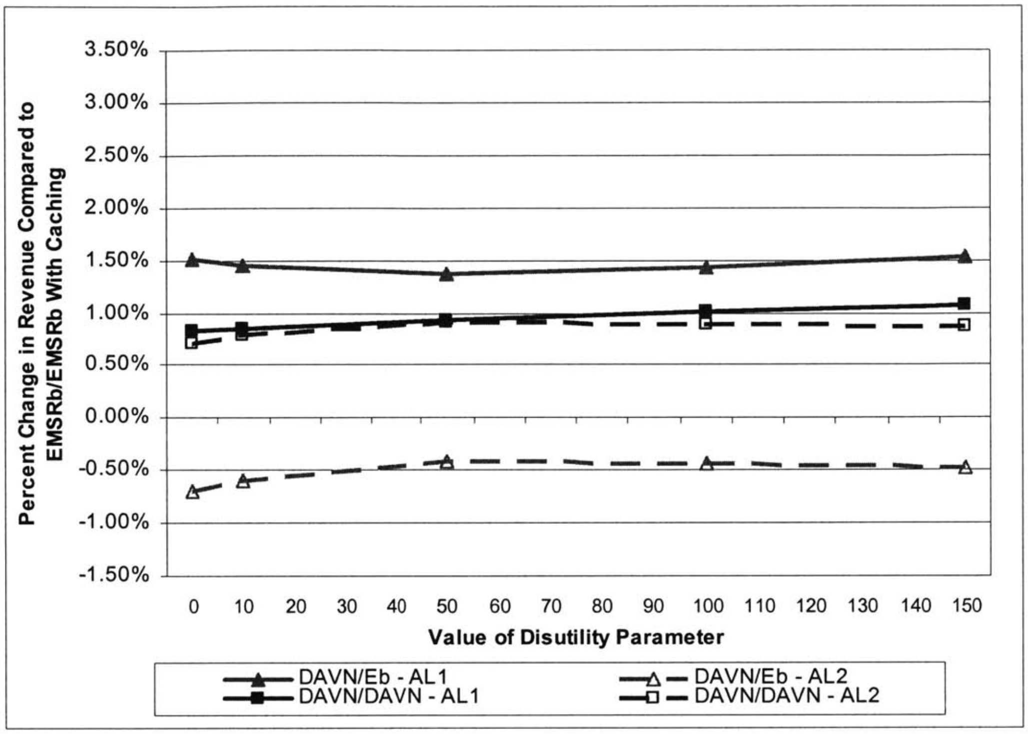


Figure 5.17:
Impact of Variable Disutility on
Incremental Revenue Gains of Network RM Methods
(100% Leg-Based Caching)



No-Go Scenario

For the purposes of benchmarking the difference between path-based and leg-based Caching, the next series of experiments was run under the No-Go scenario. As described for the path-based experiments, the No-Go scenario is an extreme case, but it could be useful for understanding the sensitivity of airline revenues to this parameter. Because the passenger becomes a No-Go immediately after being rejected, the application of a disutility is not relevant. Both airline sell options were evaluated (i.e., Sell Connect and Sell Local/Record Local), and the revenue results for all three RM method pairings are shown below in Figures 5.18 and 5.19.

First, we examine Figure 5.18, in which the airline chooses the Sell Connect option. The EMSRb/EMSRb case provides a useful baseline, because the leg RM method experiences the same effects from Caching regardless of how the cache is stored. The values in the figure for the case when both airlines use leg RM are identical to those in the path-based experiments presented previously in Section 5.4. Recall from

Figure 5.2 that under path-based Caching, the No-Go scenario involved declining revenues for all three RM pairings, but that the rate of decline was about the same for all three pairings. Thus, the incremental revenue benefit of using a network RM method was maintained regardless of the probability of using a travel website. In the case of leg-based Caching presented below, we see that when both carriers use network RM at high levels of website usage, revenues actually decrease below the levels that were achieved when both carriers used leg RM. One of the chief arguments for adopting network revenue management is that it provides a benefit even if all competitors in the industry use it. Here we see that at 80-100% use of caching sites, the use of network RM with the Sell Connect option could actually be a step backwards.

The results for the second option, in which the airline sells local inventory to website customers and records the bookings as locals, are presented in Figure 5.19. The EMSRb/EMSRb case provides the same baseline as before, revealing even sharper rates of revenue decline for network RM carriers than in the path-based Caching experiments. Here we see that at 100% leg-based Caching, all revenue benefits of using a network RM method have been eliminated—the DAVN method performs no better than EMSRb.

Figure 5.18:
Revenue Results for Three RM Method Pairings
(No-Go Scenario / Sell Connect Option)

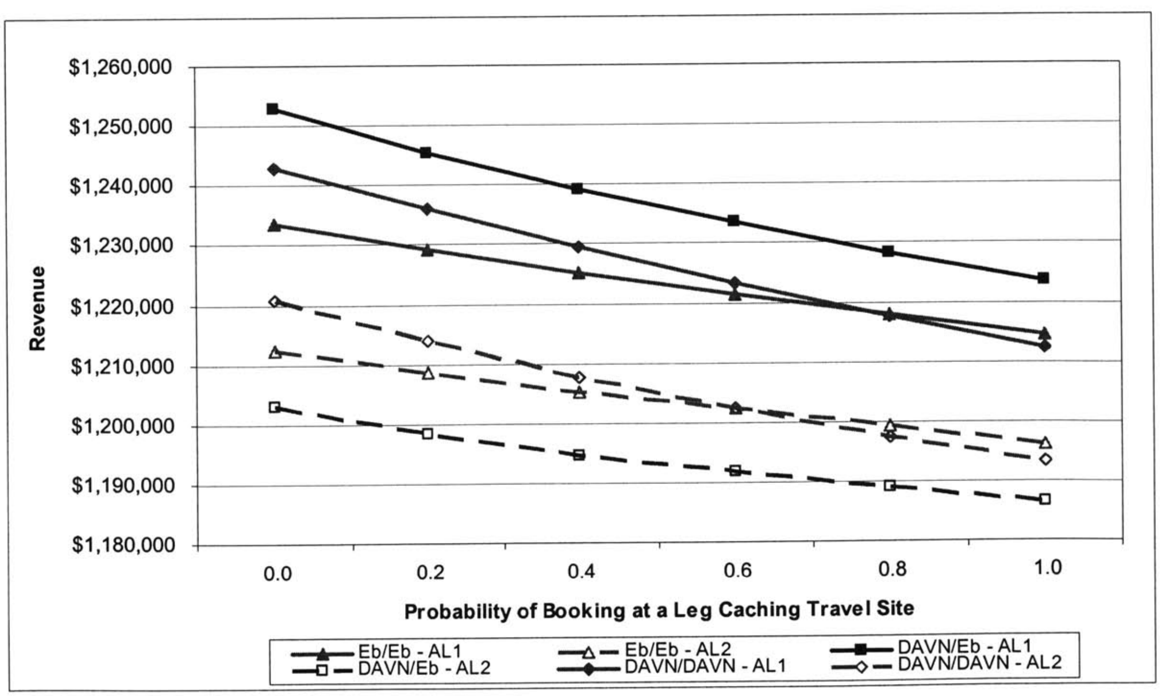
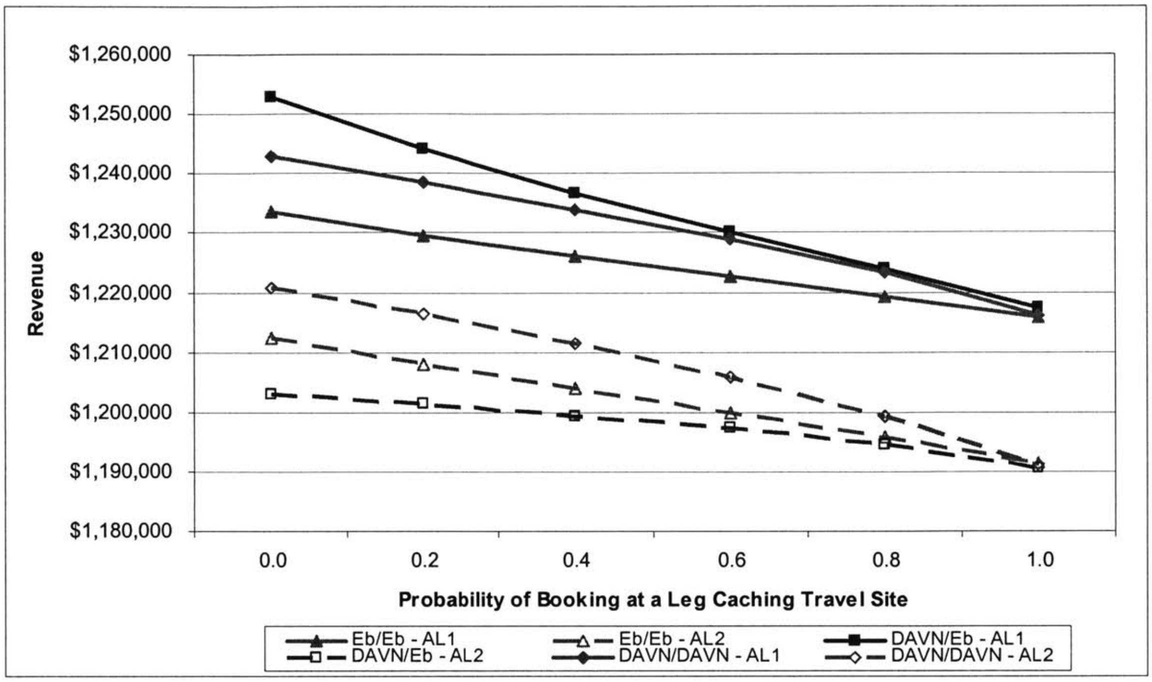


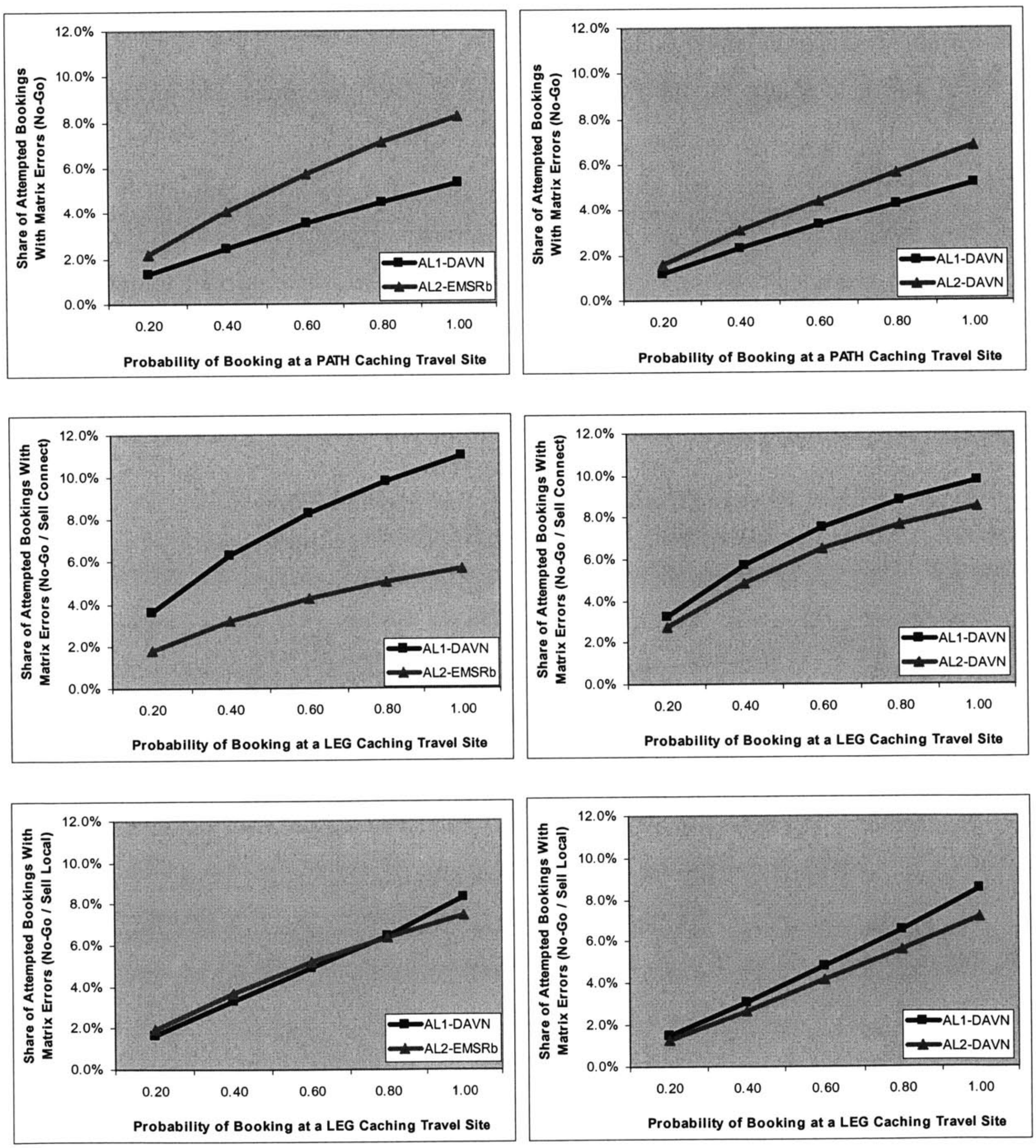
Figure 5.19:
Revenue Results for Three RM Method Pairings
(No-Go Scenario / Sell Local Option)



The cause of revenue decline in the No-Go scenarios relates directly to the effective error rate in the shadow matrix, or the number of times in which an error in the matrix actually causes a passenger to attempt to book in a class that has been closed and subsequently leave the booking process. Error rates vary depending on how the shadow matrix is stored and whether or not the airline manages its inventory in a way that corresponds to the shadow matrix. Error rates under several different scenarios are presented in Figure 5.20. The three charts on the left side show the DAVN/EMSRb pairings, and the charts on the right show DAVN/DAVN pairings. The first two charts are for path-based Caching, and it can be seen that Airline 1 has fewer errors than Airline 2. Also, the shape of the error curves is relatively linear so that as the probability of booking at a travel site increases, both carriers experience steady revenue declines. The next two charts show leg-based Caching where the airline sells connecting inventory. Now Airline 1 has more errors than Airline 2, and the higher error rates means that too many passengers are being turned away, leading to the cross-over effect observed in Figure 5.18, where DAVN is no longer a source or revenue gains. The last two charts show moderate and more similar error rates for the two airlines under the Sell

Local/Record Local option. In these cases, the leg-based cache still produces errors due to delays in matrix updates, but the airline selling response more typically matches what the cache has in memory, so it is less likely to generate a No-Go than the Sell Connect option.

Figure 5.20:
Effective Error Rates For No-Go Scenarios
Under Different Types of Caching and Airline Sell Responses



Disutility Scenario

After developing a better understanding of the consequences if passengers encountering cache errors leave the booking process entirely, we turn to the question of how a moderate disutility will affect choice behavior and revenue outcomes. As discussed above, the disutility value used is 50. Because this value is much smaller than the value used for the path-based Caching experiments, the results will not be directly comparable. Both the Sell Connect and the Sell Local/Record Local options were tested as airline responses to leg-based Caching.

Revenue results for the first option, Sell Connect, are presented in Figure 5.21. The revenue impacts of the disutility are not as extreme as leg-based Caching under the No-Go scenario. Here, the airlines are enforcing the discipline similar to Journey Control, so that they do not sell tickets to connecting passengers unless the RM optimizer has determined that accepting a connecting fare would help to maximizing total revenues. At the same time, there is a slight customer service impact from rejecting a passenger, and revenues generally decline as website bookings increase. These impacts are also shown on a percentage basis in Figure 5.22.

Figure 5.21:
Revenue Results for Three RM Method Pairings
(Disutility Scenario / Sell Connect Option)

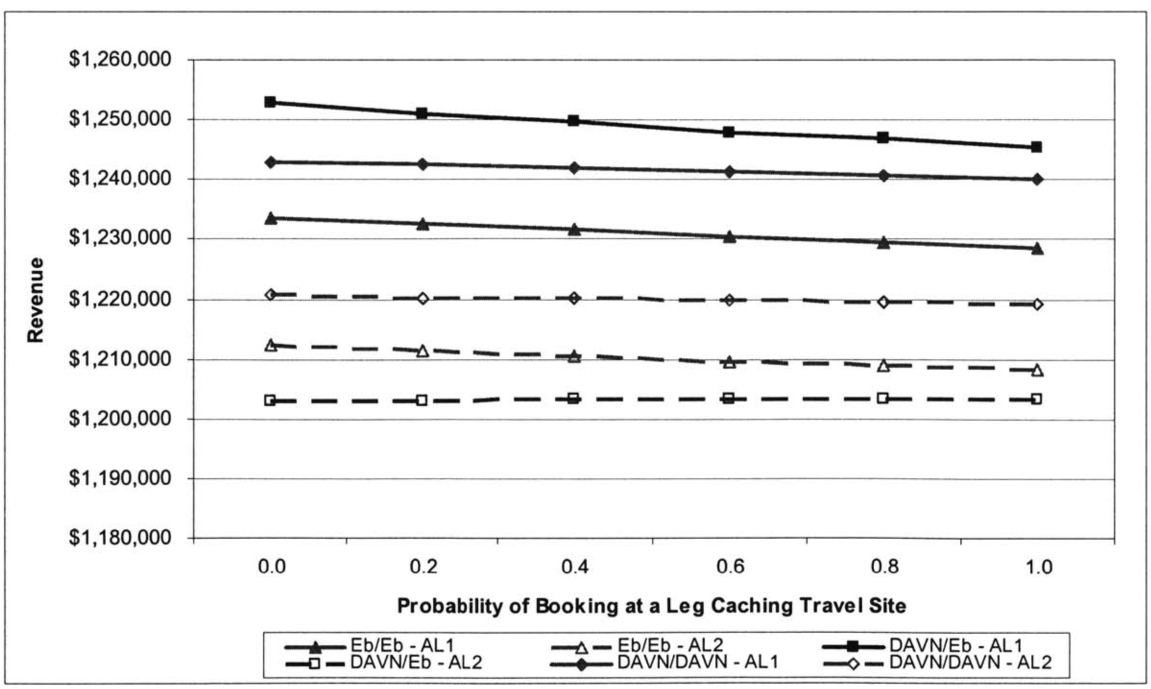
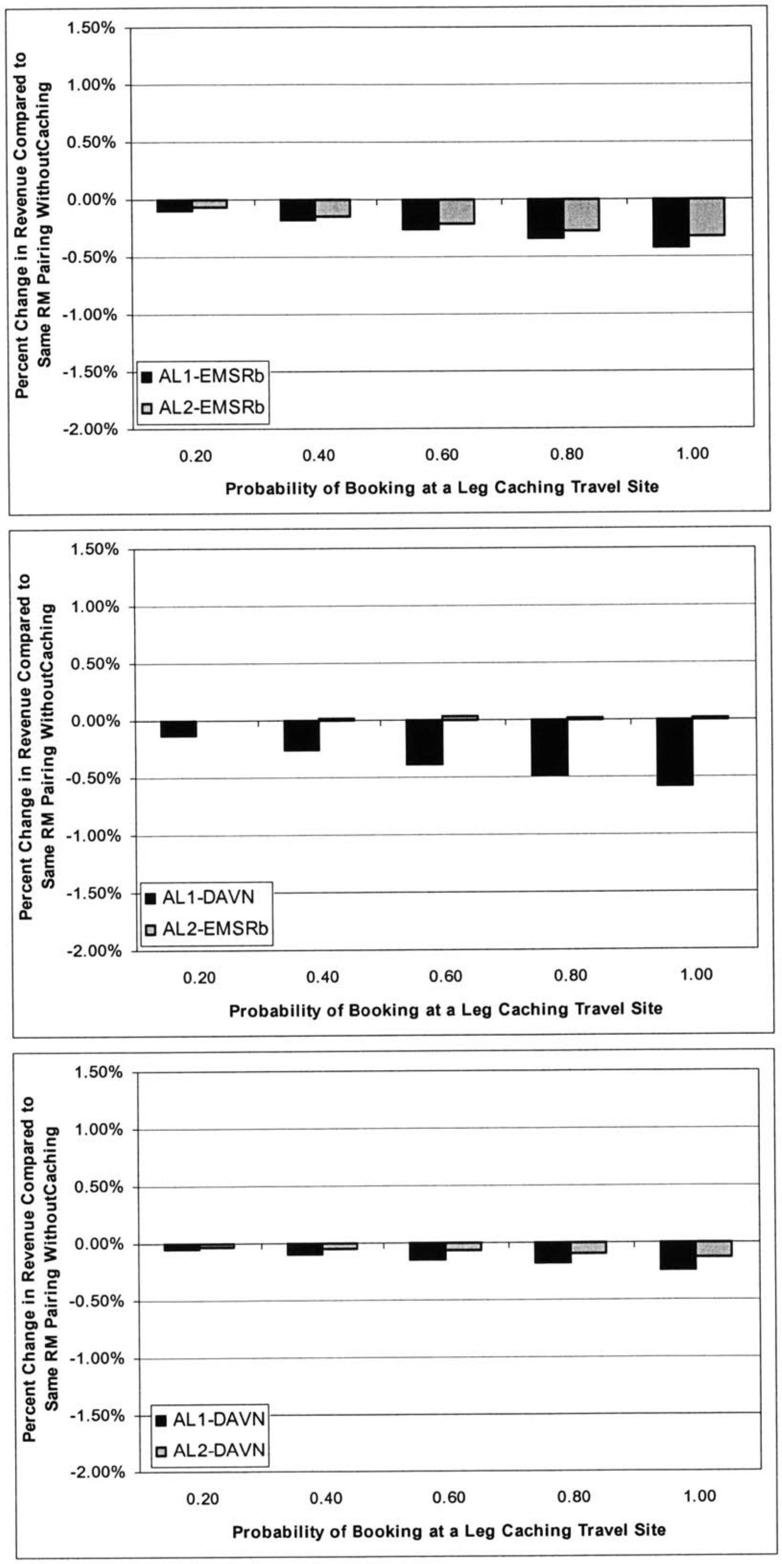


Figure 5.22:
Revenue Gains/Losses due to Caching for Three RM Method Pairings
(Disutility Scenario / Sell Connect Option)



The revenue declines in the Disutility scenario are relatively mild compared to previous scenarios that we have seen, but this is expected since the disutility value is small, and the carriers can take advantage of recovering the spilled passengers of their competitor. The one result which may be counter-intuitive is in the second RM pairing: DAVN/EMSRb. Here, the EMSRb carrier actually gains slightly, while the DAVN carrier experiences losses as large as -0.6%. This is largely because of the extreme discrepancy in error rates in this case, as was seen in the third graph in Figure 5.20 (middle row, left side). The errors for the DAVN/DAVN pairing were also high (middle row, right side), but because the two carriers are on an equal footing in terms of RM methods, the exchange of spilled passengers keeps them from experiencing losses.

In addition to overall revenue declines, leg-based Caching with a disutility also reduces the performance of network RM methods, as shown in Figures 5.23 and 5.24. This is in contrast to path-based Caching, in which revenues declined for all RM methods, but network RM was able to maintain its ability to generate incremental revenue gains, as was shown in Figures 5.12 and 5.13. In leg-based Caching with the airline selling connecting inventory, the cache is trying to estimate connecting availability based on local leg availabilities. If the connecting path is already closed, the airline rejects the booking, receives a disutility, and spills too many passengers to their competitor, an effect that increases as more and more passengers rely on the leg-based cache. If the competitor is using leg RM (Figure 5.25), then the network RM carrier will see incremental revenue gains diminish. If the competitor is also using network RM (Figure 5.26), incremental revenue gains are preserved and even slightly increased. Again, this is most likely a result of a closed experimental environment, and further experiments should be used to evaluate whether selling connecting inventory is able to protect airlines from revenue losses due to leg-based Caching.

Figure 5.23:
Incremental Revenue Gains of Network RM When One Carrier Uses Network RM
(Disutility Scenario / Sell Connect Option)

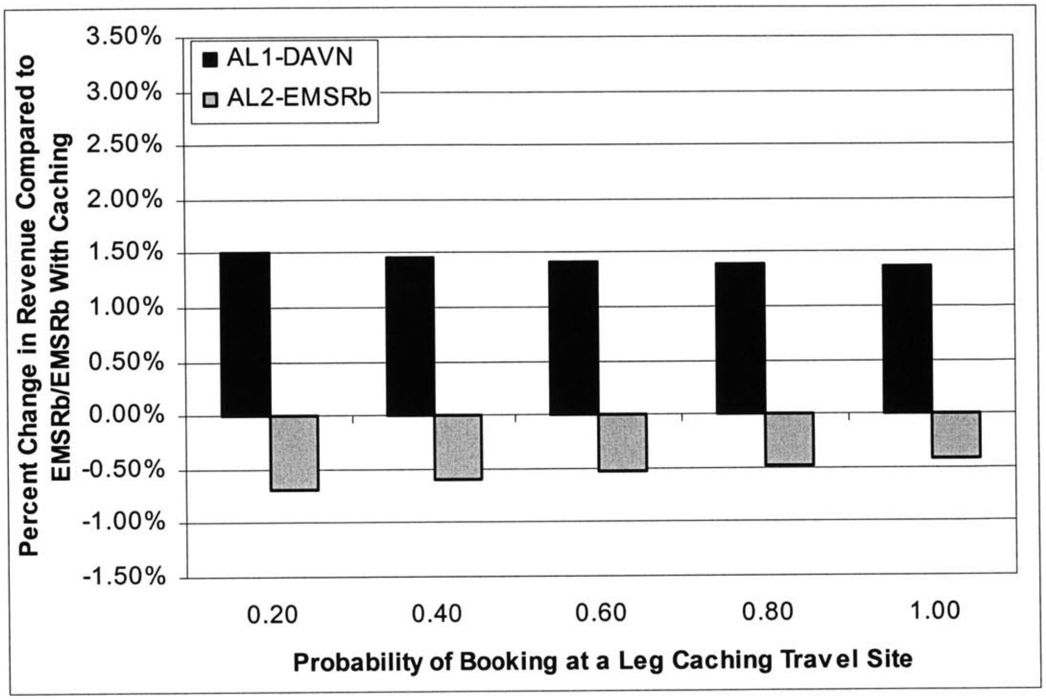
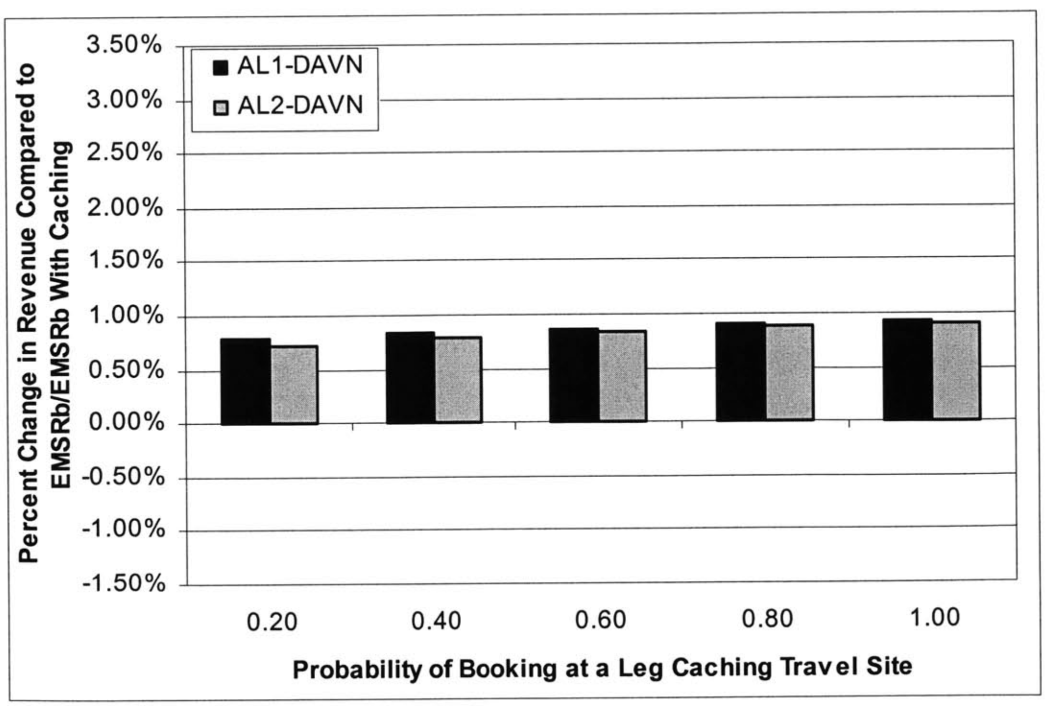


Figure 5.24:
Incremental Revenue Gains of Network RM When Both Carriers Use Network RM
(Disutility Scenario / Sell Connect Option)



The second of the two sell option examined under the Disutility scenario is the Sell Local/Record Local option. The revenue results displayed in Figure 5.25 show that declines in revenue are not as extreme as the No-Go scenario, but they are larger than under the Sell Connect option. This result is not surprising, because we have already seen in Figure 5.14 that Sell Local/Record Local does not perform as well as Sell Connect. Note that the earlier experiments were conducted at 100% Caching, which is actually somewhat of an anomaly in Figure 5.25 compared to results when the probability of booking at a website is 0.8 or less. At the highest levels of website usage, revenues for Airline 1 actually remain flat or increase while the revenues for Airline 2 decrease, largely due to the exchange of spilled passengers in the controlled two carrier experiment that we have seen previously. Only at the highest levels of website usage are enough passengers exchanged for Airline 1 to make up some of its losses, but this is result is not expected to occur in practice. Neglecting the data points at the far right of the charts, we see a similar "convergence" effect as in the No-Go scenario, where the network RM methods are losing their ability to outperform leg RM competitors. The revenue gains and losses for each RM method pairing are shown on a percentage basis in Figure 5.26.

Figure 5.25:
Revenue Results for Three RM Method Pairings
(Disutility Scenario / Sell Local Option)

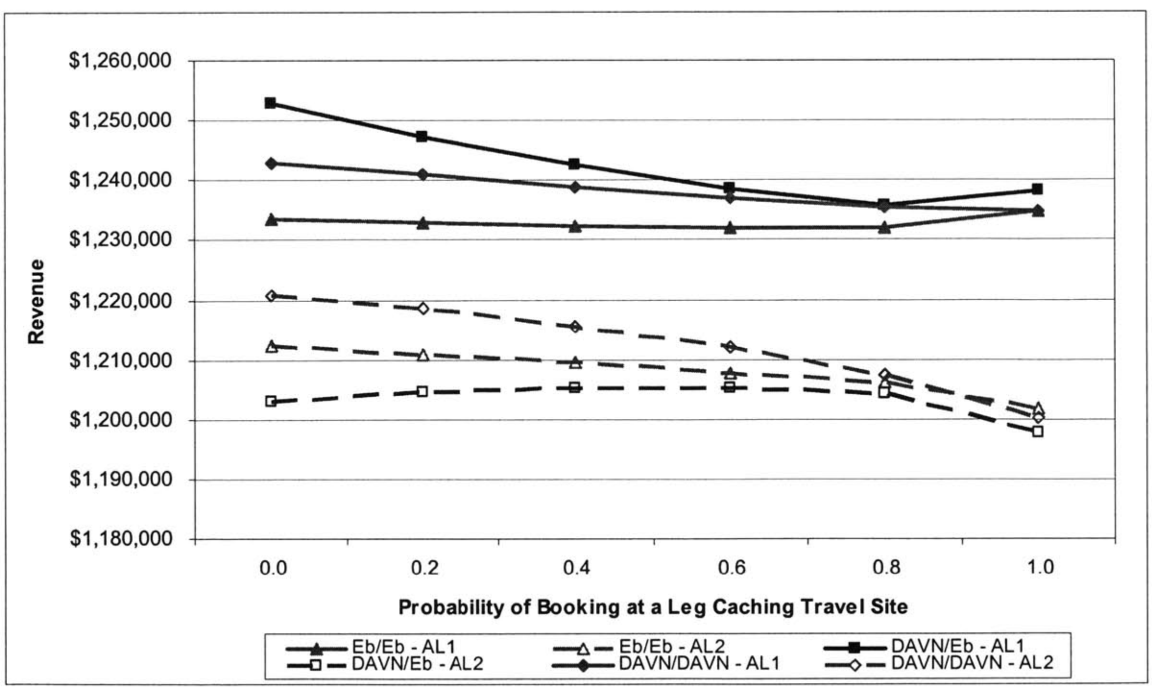
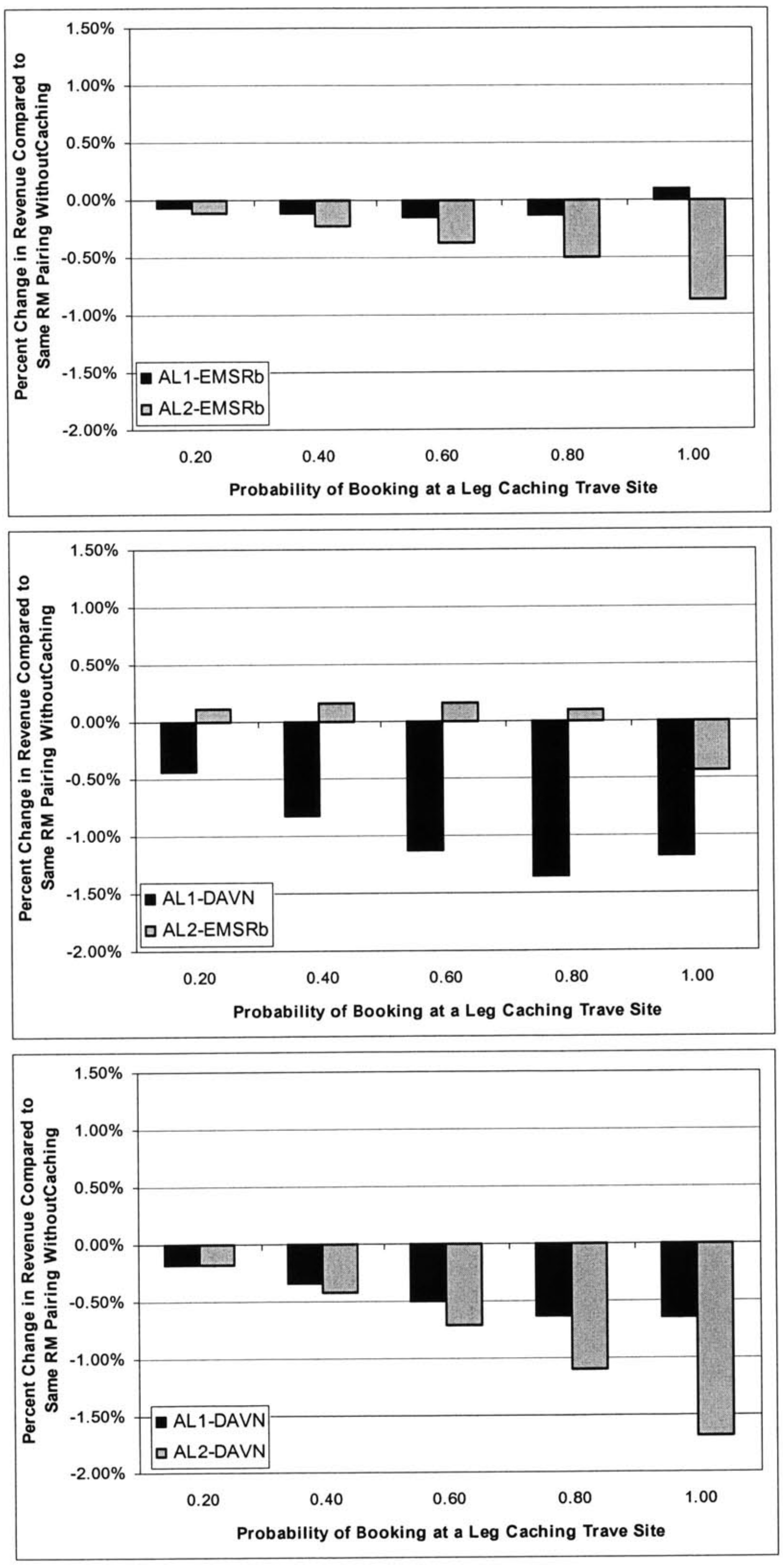


Figure 5.26:
Revenue Gains/Losses Due to Caching for Three RM Method Pairings
(Disutility Scenario / Sell Local Option)



The results shown in Figure 5.26 are very similar to the Sell Connect option, but with a larger magnitude of gains and losses. The pairing where both carriers have leg RM methods shows that leg-based Caching with a small disutility generally results in modest losses for both carriers. When the airlines use the Sell Connect option, the two carriers experience losses closer to the same level, but under the Sell Local option, Airline 1 seems to do slightly better than Airline 2. For the pairing where one carrier uses network RM, Airline 1 loses as much as 1.4% of revenues while the leg RM carrier (Airline 2) gains as much as 0.2% in revenues. Again, the carrier using network RM is worse off in this situation because of the fact that it normally performs better without Caching, and so there is more to lose when the passenger relies on mis-information in the leg-based cache. When both carriers use network RM, they again have more equal levels of revenue losses, and Airline 1 fares better than Airline 2 due to the strength of schedule.

To understand the magnitude of these losses, note that the cases above with at least one carrier using network RM produce revenue losses that are almost the same as under the No-Go scenario with path-based Caching. Thus, a small level of disutility combined with the errors inherent in the leg-based cache can have as strong an effect as the use of a more accurate path-based cache where rejected passengers leave the booking process entirely. However, a key difference between the No-Go scenario with a path-based cache and the Disutility scenario with a leg-based cache is that in the second case, the network RM methods no longer provide a revenue advantage over the use of leg RM.

The primary difference between the Sell Local and Sell Connect options is that the network RM methods can not make appropriate adjustments under the Sell Local option and cannot maintain their advantage over the use of leg RM. Figures 5.27 and 5.28 show the decline in performance of the network RM methods more clearly. While selling local inventory appeared to be a strategy that did not seriously damage revenues when there was no customer service impact, here we see that a small disutility can greatly reduce the ability of a network RM system to maintain a revenue advantage over leg RM. This deterioration is actually more serious than under the No-Go scenario, in part because the early No-Gos had left behind empty seats which could be filled by later arriving high-fare passengers. In the Disutility scenario, the airlines spill passengers to each other, but fill up too quickly to recover any revenues late in the booking process.

Figure 5.27:
Incremental Revenue Gains of Network RM When One Carrier Uses Network RM
(Disutility Scenario / Sell Local Option)

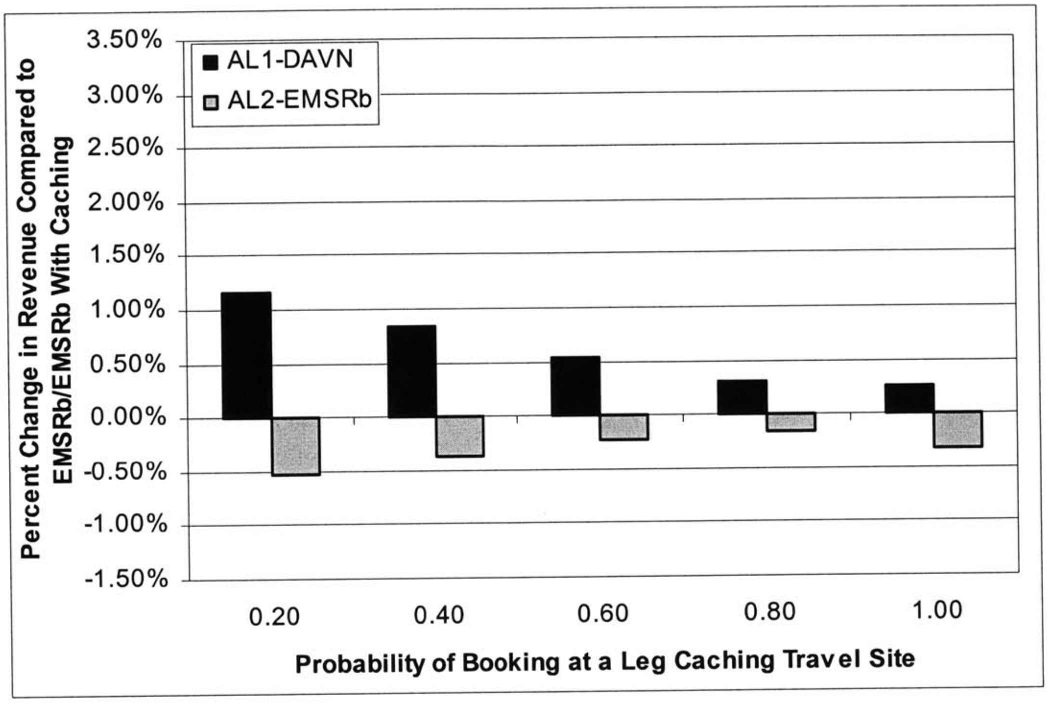
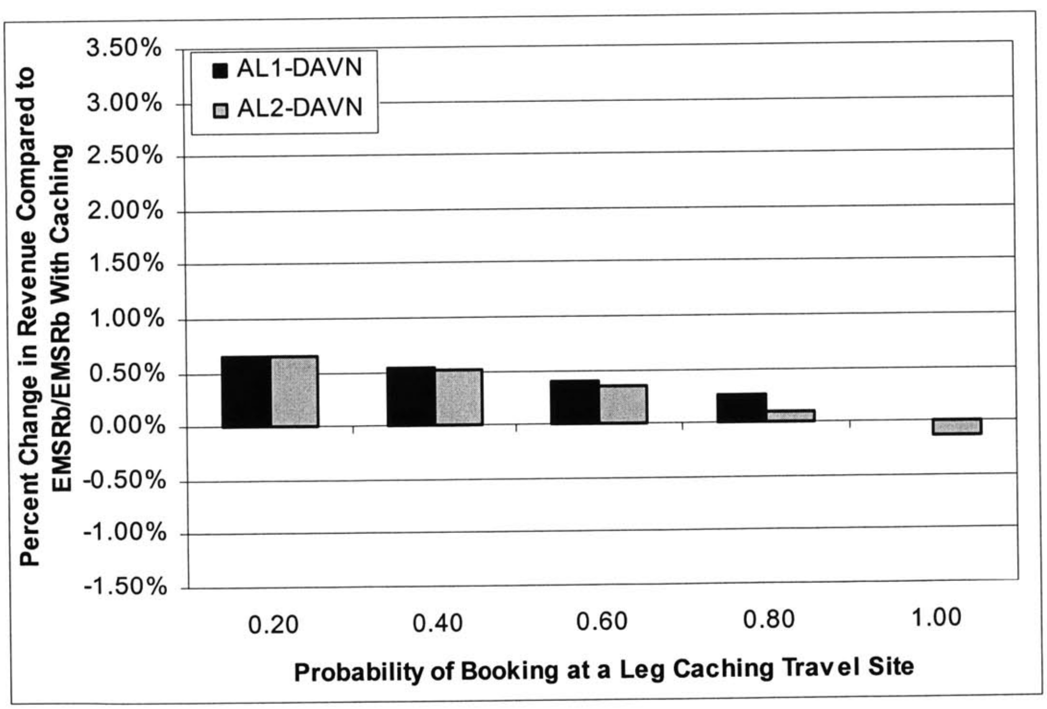


Figure 5.28:
Incremental Revenue Gains of Network RM When Both Carriers Use Network RM
(Disutility Scenario / Sell Local Option)



5.6 Summary

This chapter has introduced the subject of Caching, which is becoming a more and more significant factor in the distribution of airline tickets today. There are numerous variations in how Caching is implemented and in how passengers and airlines respond to Caching, and only a few of the options have been modeled here. While the experiments presented offer a highly simplified approach, several conclusions can already be made about which alternatives might be the most promising to pursue.

First, simulation results showed that delays in updating the more sophisticated path-based caches can impact airline revenues and that revenue results depend on the passenger behavior that results from errors in the cache. In the more extreme cases tested here such as No-Go and Airline Accepts, revenue losses exceeded -1.5% at 100% Caching, and losses were often larger than in previous Bypass experiments at the same level of website usage. In the Disutility case, the losses were on the order of -0.5%. The key difference between the Bypass results and the path-based Caching results is that the performance of network RM methods was maintained during the path-based Caching experiments, which was not the case in Bypass experiments. Under path-based Caching, network RM is still able to provide key incremental benefits over the use of leg RM.

The second set of simulation results covering leg-based Caching is a more accurate representation of current practice in the industry. Cache error rates are higher for a leg-based cache, and depending on the airline's choice for how to sell their inventory to website customers, these errors can have significant negative impacts on the revenue results. Experiments showed that under the No-Go scenario, the declines were so serious that network RM became a liability, causes revenues to decline even below leg RM levels at around 80% website usage. The No-Go scenario is an extreme case, but even one of the more moderate Disutility scenarios showed both revenue losses and an inability of network RM to maintain its revenue advantage.

The more realistic Disutility scenario produced two key results. In the case where the airlines sell connecting inventory, losses were generally smaller than -0.5% and network RM performance was only slightly affected. Adherence to the network RM systems booking limits for connecting passengers contributes to overall revenue performance, suggesting that the concept of Journey Control deserves further exploration.

In the Disutility experiments where the airlines sell the local inventory, negative revenue impacts were often as high as some of the more extreme scenarios, reaching -1.4% at only 80% website usage in one case and -1.6% at 100% website usage in another. At the same time, the revenue advantage of using network RM was also seriously compromised. This is not surprising since the sale of local inventory at connecting fares has already been shown to reduce revenues by 0.5% at only 50% website usage in the Bypass experiments. While the Caching experiments do not explicitly model passengers' attempts to bypass, it is clear that the revenue impacts of leg-based Caching are at least as serious as the impacts of Inventory Control Bypass. Unfortunately for the airlines, while there are technological changes that can eliminate Bypass, the airlines have little control over whether third parties initiate a caching system.

6 CONCLUSIONS

This thesis represents a first attempt to describe the interactions of various developments in the distribution of airline travel and to quantify the revenue impacts of such interactions. The work presented here has been but an introduction—much work remains to fully illuminate the most important issues, trends, and trade-offs. This final chapter presents a summary of what has already been learned and provides some ideas for the most promising avenues of research to pursue as next steps in the process.

6.1 Summary of Findings

6.1.1 Data Analysis

Chapter 3 presented a preliminary statistical study of one sample of ticketing data that has already yielded interesting results. This data indicated rather decisively that there are distinct differences in purchasing behavior of passengers who use different distribution channels to buy their airline tickets. Specifically, online customers typically purchase their tickets very quickly after acquiring their reservation, and they complete purchase further in advance of their intended date of departure than customers in other distribution channels.

These findings could have important consequences for airline managers, who depend primarily on historical data as the basis for their predictive models about future airline demand. As more and more tickets are purchased on the Internet, these historical models could be less and less accurate, unless airlines can successfully incorporate these new booking patterns and behaviors. For example, as the share of online ticket purchases increases, the number of bookings on hand which correspond to completed transactions will grow. This in turn reduces the rate of booking cancellations. Overbooking models should be reviewed to ensure that they capture such trends effectively.

While the data sample from a traditional GDS did not contain a large share of online tickets, website usage in developed countries is approaching a third of all tickets. In addition some estimates place the compound annual growth rate of internet usage for travel bookings at as much as 28% per year [Smith, 2004]. The transitions to new booking behaviors are happening very quickly, and airlines must be ready to adjust.

6.1.2 Inventory Control Bypass

The transition to a new type of booking behavior is particularly important, because we have seen evidence that internet bookings may have other consequences for revenue results. For example, online channels are more likely to use searching algorithms that engage in Inventory Control Bypass. The results in Chapter 3 show that passengers at online channels booked sooner than passengers at other channels, which implies that Bypass is more likely to occur early in the booking process. Thus, if an airline closes more low fare classes early, in an attempt to protect seats, they may be doing this at precisely the wrong time, because closed classes actually expose the airline to more potential Bypass, with both its direct revenue losses *and* the loss of forecast accuracy.

Chapter 4 also described some of the ideas for dealing with Bypass. Compensation methods for forecasting can correct only the indirect losses, so alternative methods that actually enforce booking limits across the entire ticketing transaction are currently being developed and implemented by some carriers. Some of the most effective methods require extensive coordination among parties in the distribution channel, so they are not likely to be implemented in the future unless a compelling case can be made for their revenue benefit. In the meantime, as computer search engines become faster and more thorough, it will be more and more difficult for airlines to maintain control over their inventory.

6.1.3 Caching

Although search engines are becoming highly sophisticated, the communications processes for making availability and inventory requests to a CRS is still somewhat antiquated. Hardware is being pushed to the limit, and websites have responded by building up caches of stored data to display to their customers instead of asking the airlines for actual status. Even in very simplified path-based scenarios without cancellations, this practice has been shown to reduce airline revenues because of the impact of poor service on passenger behavior. In the more realistic case where the cache is stored on a leg basis, the revenue impacts are even more serious. Not only is revenue lost due to Caching, but network RM loses its ability to create incremental revenue gains over leg RM.

The Internet is clearly the lowest cost distribution channel currently available to airlines, helping to drive the cost of selling a ticket down to as low as $0.25 per ticket [Leonard, 2004]. Unfortunately, because of Caching, the Internet also appears to be a lightning rod for revenue losses, particularly for those carriers with network RM systems. As it is unlikely that the Internet will disappear or be replaced anytime soon, airlines should be working to implement solutions, or be prepared to face continued and increasing revenue losses.

At the same time, there has been anecdotal evidence that Caching is actually contributing to one problem that it is ideally suited to solve: the increase in message traffic. As mentioned in Chapter 5, it is not possible to have a full O-D cache with the technology currently available, and so the third parties tend to make more availability requests to the CRS, not less, in an attempt to keep their stored data as updated as possible and reduce errors. As a result, Caching is currently driving the volume of message traffic higher, when the reality is that an accurate cache could actually reduce the number of real-time requests being made into the CRSs. For example, as with Selective Polling, airlines could work with the third parties to develop effective decision rules to determine which flights could confidently be cached, and which would require a real-time poll. If an appropriate algorithm can be found which does not have negative impacts on revenue, this could be a key innovation within the industry that would have major benefits for both airlines and websites.

6.2 Future Research Directions

This thesis set out to outline the various challenges facing airlines in relation to their ticket distribution options and strategies. Clearly much work remains to understand each of the issues and the most appropriate solution strategies. The following sections highlight some of the most critical tasks.

6.2.1 Journey Control

As mentioned in Chapter 4, it is relatively straightforward to model Journey Control in the PODS environment. When the connecting path is closed by the network RM system, passengers look for the "local" alternative, but this would only be available if the connecting path was still open. Future research should also consider the idea of offering the passenger the option of booking two local fares in case it is within their willingness-to-pay threshold. In addition, some airlines have indicated that there may be technical conflicts between the implementation of Journey Control and Married Segment Control and PODS would be ideally suited to a cost trade-off analysis between the two.

6.2.2 Price-As Booked

As with Journey Control, Price-As-Booked requires a modification to the passenger choice module in PODS to allow passengers to book two local legs at the sum-of-locals price to fully test the concept. Because this solution represents the total enforcement of inventory controls, Price-As-Booked actually corresponds to a PODS scenario without any Bypass. As a result, the benefits of introducing such a policy can be estimated by developing a "realistic" scenario that incorporates estimates of the levels of Bypass and website usage, together with fully calibrated passenger choice parameters such as no-go and disutility values. The challenge here is to understand the losses from both Connect-Closed and Local-Closed Bypass in a full-up competitive environment in order to quantify the total gains of mitigation measures such as Price-As-Booked.

6.2.3 Selective Polling

Although not fully implemented in PODS yet, there are a variety of studies which could be performed using Selective Polling. Future research could test and compare the different criteria for selecting requests that could be handled with AVS messages. In addition, the PODS format is ideal for measure risks in multiple settings, including the competitive advantage or disadvantage of choosing to implement Selective Polling, and the effects on different RM methods.

6.2.4 Caching

The experiments presented in Chapter 5 were a very small piece of the potential experiments with Caching that can be modeled with PODS. There are at least six areas that deserve further exploration. First, revenue results should be explored at different load factors. Because closed inventory is what leads to errors between the two matrices, and higher load factors imply more path-class closures, this could have a significant impact on the revenue results. Second, it has been suggested that the binary parameter in the No-Go experiments be changed to a variable between 0 and 1.0 that represents the probability of any individual passenger becoming a No-Go. This would help measure the sensitivity of airline revenues to customer service factors. A third key area to test is the update rate of the current implementation of the shadow matrix. One suggestion was to update the matrix after every single booking in order to isolate the effects of using a leg-based matrix to sell path-based inventory. This could actually be the dominant force in the revenue results, and if so, airlines may want to work cooperatively with third parties who cache to promote a transition to path-based Caching as quickly as possible. Fourth, the additional strategies for updating the matrix should be tested, such as the continuous sweep described in Chapter 5. These could help make the model itself more realistic. Fifth, PODS could be used to explore the idea of an airline deciding not to sell its inventory through distribution channels that use Caching. Then, website passengers would not find out anything about the non-Caching airline's inventory, but if the Caching airline has a lot of spill due to customer service, there may be some recapture benefit to the non-Caching airline. Finally, there are numerous potential competitive combinations, at a minimum including: the study of the how Caching affects bid-price RM methods; choosing to go ahead with "airline accepts" against a competitor who does not accept; and the various ways an airline can sell their inventory in response to a distributor who uses Caching.

6.2.5 Combination studies

Each of the suggested areas of research has merit alone, and they could also be interesting in combinations with each other. For example, because the pricing portion of the ticketing transaction is currently based on a final poll of the selected flights, there could be discrepancies between cached availability and the Price-As-Booked results, leading to complex revenue impacts.

As indicated in Section 6.2.2, one of the most important tasks before pursuing these topics will be to develop a realistic "baseline scenario" that represents likely current practice. Then, the revenue gains of eliminating Bypass or improving Caching can be measured from the current baseline. This should help airlines relate the cost to make these changes directly to their expected benefits. The key tasks in such an effort will be: estimating how often each type of Bypass and Caching are currently being used; determining which network is most appropriate for testing mitigation strategies; deciding which competitive interactions should be modeled; and determining which metrics will be used to measure effectiveness. Suggested measures include revenue gains (both for each individual airline and for the whole industry), improvements in forecast accuracy, and reductions in volume of bypass activity or caching error rates.

6.3 Summary

The work presented in this thesis covers many topics and shows that research areas which have historically been separate are being drawn together by the technological change occurring within the airline industry. Internet usage is changing the way that passengers book airline travel, both in terms of their outward behavior patterns, and in the way that computer systems carry out the requested transaction. The move to cut distribution costs by shifting to internet channels has opened the door for a variety of bypass and caching practices which have had unintended revenue-side consequences for the airlines. While the study of these consequences is only just beginning, the results presented here emphasize the serious nature of the problems being faced. As internet usage continues to grow, the ability of airlines to protect their already shrinking revenue will be further diminished. It will take time to fully understand the entire nature of the problems and develop robust solutions, but solving these challenges could go a long way towards ensuring that airlines have a stable financial future.

Bibliography

Barnhart, Cynthia, Belobaba, Peter P., and Odoni, Amedeo R. (2003) "Applications of Operations Research in the Air Transport Industry", *Transportation Science*, Volume 37, Number 4. November, 2003.

Belobaba, Peter P., (1998) "Airline Differential Pricing for Effective Yield Management", *Handbook of Airline Marketing*, First Edition, Section 3, Chapter 27.

Belobaba, Peter P., (2002a) "Airline Network Revenue Management: Recent Developments and State of the Practice", *Handbook of Airline Economics*, Second Edition, Section 2, Chapter 10.

Belobaba, Peter P. (2002b), "O-D Control Abuse by Distribution Systems: PODS Simulation Results", *AGIFORS Reservations and YM Study Group Meeting*, April, 2002.

Blank, Dennis, (1999) "Raising the Internet stakes", *Airline Business*, September, 1999.

Carpenter, Dave, (2004) "Number of consumers using Net for travel plans rises sharply", *Chicago Sun Times*, April 1, 2004.

Cusano, Andrew J., (2002a) "Variable O-D Control Abuse with Abuse Compensation Methods", *PODS Consortium Update*, April, 2002.

Cusano, Andrew J., (2002b) "O-D Control Abuse by Distribution Systems: Update – Impacts on DAVN at Lower ALF", *PODS Consortium Summit Update*, May, 2002.

Cusano, Andrew J., (2002c) "Selective Polling, RM and Advance Purchase Bypass", *PODS Consortium Summit Update*, June, 2002.

Darot, Jérémy F. J., (2001) "Revenue Management for Airline Alliances: Passenger Origin-Destination Simulation Analysis", S.M. Thesis, Massachusetts Institute of Technology, June 2001.

Fernandez de la Torre, Pablo E., (1999) "Airline Alliances: The Airline Perspective", S.M. Thesis, Massachusetts Institute of Technology, June 1999.

Flint, Perry, (1998) "Bigger than the Internet? Speech recognition technology may revolutionize airline ticket distribution, if hurdles can be overcome", *Air Transport World*, September, 1998.

General Accounting Office, (1986) "Airline Competition: Impact of Computerized Reservations Systems", GAO/RCED-86-74, released May 1986.

General Accounting Office, (1988) "Competition in the Airline Computerized Reservation System Industry", GAO/RCED-88-62, released September 1988.

General Accounting Office, (1992) "Computerized Reservation Systems: Action Needed to Better Monitor the CRS Industry and Eliminate CRS Biases", GAO/RCED-92-130, released March 1992.

General Accounting Office, (1999) "Domestic Aviation: Effect of Changes in How Airline Tickets are Sold", GAO/RCED-99-221, released July 1999.

General Accounting Office, (2001) "Aviation Competition: Restricting Airline Ticketing Rules Unlikely to Help Consumers", GAO-01-831, July, 2001.

General Accounting Office, (2003) "Airline Ticketing: Impact of Changes in the Airline Ticket Distribution Industry", GAO-03-749, July, 2003.

Lavere, Jane, (1998) "On-Line: a new web challenger", *Airline Business*, November, 1998.

Lavere, Jane, (2000) "Changing roles: travel agencies must adapt to stay afloat in the Internet era", *Airline Business*, October, 2000.

Lee, Alex Y. H. (2001) "Travel Agent / CRS Search Engine Abuse of O-D Controls", *PODS Consortium Summit XV*. November, 2001.

Leonard, Joe, (2004) CEO, AirTran Airways. Panel Discussion at Phoenix Aviation Symposium. April, 2004.

McGee, William, (2003) "Booking and Bidding Sight Unseen: A Consumer's Guide to Opaque Travel Web Sites", Consumer WebWatch. December, 2003.

McGill, Jeffrey I., and van Ryzin, Garrett J., (1999) "Revenue Management: Research Overview and Prospects", *Transportation Science*, Volume 33, Number 2. May, 1999.

Ratliff, Richard M., (2003) "Approaches to Availability Processing: The Increasing Problem of Shopping", *AGIFORS Reservations and YM Study Group Meeting*, June, 2003.

Reed Business Information, (1997a) "Web fever: Airlines on the Internet", *Airline Business*, February, 1997.

Reed Business Information, (1997b) "The Sales of the Century? - Airlines on the Internet", *Airline Business*, February, 1997.

Smith, Barry C., Leimkuhler, John F., and Darrow, Ross M. (1992) "Yield Management at American Airlines", *Interfaces*, Volume 22, Number 1, January-February, 1992.

Smith, Barry C., Gunther, Dirk P., Rao, B. Venkateshwara, Ratliff, Richard M. (2001) "E-Commerce and Operations Research in Airline Planning, Marketing, and Distribution", *Interfaces*, Volume 31, Number 2, March-April, 2001.

Smith, Mike, (2004) "Online Travel Growth", *PROS Revenue Management Conference*. Houston, TX. March 2, 2004.

Swartz, Karl L., (2004) Distance table provided via email communication, February 18, 2004. See Great Circle Mapper website, http://gc.kls2.com

Transportation Group International, L.C., (2002) "Consumer Attitudes and Use of the Internet and Traditional Travel Agents", report for the National Commission to Ensure Consumer Information and Choice in the Airline Industry, September 19, 2002.

Williamson, Elizabeth L. (1992) "Airline network seat inventory control: methodologies and revenue impacts", Ph.D. Dissertation, Massachusetts Institute of Technology, 1992.

Zeni, Rick, (2003) Personal Interview with Rick Zeni, Director of Revenue Management Optimization at US Airways, October 14, 2003.