

A MARKOV CHAINS BASED TRANSITION MATRICES APPROACH TO
FORECASTING AIRLINE SEAT DEMAND

By

Leslie M. Bobb

A dissertation submitted to the Graduate Faculty in Business in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2008

UMI Number: 3303799



UMI Microform 3303799

Copyright 2008 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2008

Leslie M. Bobb

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Business in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jan. 30th, 2008
Date

Prof. Emre A. Veral
Chair of Examining Committee

Date

Prof. Joseph Weintrop
Executive Officer

Dr. Emre Veral

Dr. Harry Rosen

Dr. William Millhiser

Supervisory Committee

Abstract of the Dissertation

A Markov Chains Based Transition Matrices Approach to Forecasting Airline Seat Demand

By

Leslie M. Bobb

Advisor: Professor Emre Veral

The key to the billions of dollars made yearly in the hospitality industry, particularly the airline and hotel industries, lies in its yield and revenue management systems. Yet very few people outside of the industry realize the complexity of those systems and what drives it. Any yield and revenue management system is only as good as its [over]booking engine. And if booking is to be considered the engine to those systems, then the forecasting module is the fuel that powers it.

The strength of a forecast is always measured by its accuracy but that can only be determined after the fact. Forecasts for airlines and hotels are subject to a myriad of factors including competitor actions, economic environments and natural disasters. As a result, forecasting in general and more specifically for the hospitality industries is

extremely complicated. This dissertation will investigate the Markov Chains approach to forecasting as a simpler solution.

The literature review within this dissertation provides an illustration of the various forecasting methods used. A Markovian based transition matrices approach is then proposed as an alternative method to improve demand forecasting. Its usage and performance criteria are then outlined and a methodology for testing its effectiveness under various simulated environmental conditions is described. A select few of these forecasting methods are compared to the proposed Markov Chain transitional matrices approach in terms of their forecasting accuracy.

Results indicate that a Markov Chains based transition matrices approach does indeed offer an improvement over current selected methods under certain conditions. Analysis also shows that capacity is irrelevant to forecasting accuracy and that the multiplicative pickup methods were the worst performers. Further investigation also reveals that an observation point scheme utilizing data up to 60 days prior to departure was best; and contrary to common statistical principles, a disaggregated fare class structure was the most efficient. In a self-fulfilling prophecy, it was also shown that the forecasting methods were more accurate for censored data.

DEDICATION

To my daughter and little gremlin, Meghna Izabella, for certainly being a joy and an inspiration.

ACKNOWLEDGEMENTS

I would like to express my sincerest thanks and appreciation to my primary advisor, Professor Emre A. Veral. This research would not have been possible without his expert guidance, insights and constructive advice. His unwavering commitment, dedication and encouragement were vital in bringing this research to fruition. His role as an academician and a thesis advisor is certainly unparalleled. For this, I am most certainly indebted to him.

I would also like to thank the other members of my advisory committee whose assistance were also invaluable. Professor Harry M. Rosen provided his academic wisdom and proficiency with Markov Chains being his area of expertise and Professor William Millhiser provided valuable comments in the fine-tuning of the dissertation at the most crucial stages. Professor Elsie Gottlieb from the Statistics and CIS department provided very useful assistance at the initial stages of this dissertation and I was also extremely fortunate to have her as a professor for a couple of semesters during my MBA and Ph.D years at Baruch College in Operations Research. Without her encouragement I would not have even considered a Ph.D program.

I would also like to express special thanks and my most heartfelt appreciation to Professor Georghios P. Sphicas, who as an initial member of my advisory committee, has been a great mentor for all students in the Management Planning and Information Systems area. He has always had an open-door policy for all of my ideas and has provided valuable advice over the years.

Undertaking a project such as this has its risks with one of them being the proprietary nature of firms' data. I therefore consider myself extremely lucky and I owe a

debt of gratitude to Dr. Richard Zeni, Vice-President of Revenue Management at JetBlue Airways, Queens, New York, for providing me with data he himself had used. Without his technical expertise, constant support and expert opinion, this research would have been impossible. It was certainly a delight working with him.

I am most grateful to each and every professor with whom I have had the pleasure to take courses with during my years at Baruch College and during my undergraduate and graduate years at The City College, The City University of New York. They have all prepared me very well for the culmination of my studies. Their role as teachers and academics has certainly inspired me.

Within my department, I have been very fortunate to have been surrounded by some exceptional fellow students. I would certainly be amiss if I failed to mention the friendships I have shared with Hung 'David' Chu, Tugba Cayirli, Ri Xia and Hong Deng Chen. Also, special mention must go to Linda Moore and Dularie Ramkishun, the two departmental secretaries who have assisted me on almost a daily basis. I would also like to express my appreciation to Alexis Aquino of the Statistics Consulting Lab who skillfully assisted me with the statistical coding and analysis.

Last but not least, I must certainly thank my mom and my uncle, Champa and David Seelochan respectively and my wife Radha for their unconditional support and tolerance throughout this arduous process. I would also like to thank my best friend Suresh Bisnauth who expertly assisted with the computer coding and all my other friends who have been there for me. Nothing can replace the years of friendship we have shared and their support has been invaluable. Finally, this work is dedicated to my daughter and little gremlin, Meghna Izabella Bobb. She is certainly an inspiration. Thank you.

Table of Contents

LIST OF TABLES		xiii
LIST OF FIGURES		xvi
Chapter 1	Introduction	
1.1	Introduction to Yield Management and Revenue Management	1
1.2	Importance of Yield and Revenue Management	6
1.3	The Organization of the Dissertation	12
Chapter 2	Literature Review	
2.1	Definitions	14
2.2	A Brief History of Yield and Revenue Management	17
2.3	A Survey of the Literature	19
2.4	Major Issues in Yield and Revenue Management	21
2.4.1	Objectives of Yield and Revenue Management	22
2.4.2	Forecasting	23
	2.4.2.1 Forecasting Models used in the Airline Industry	29
	2.4.2.2 Micro-Level Forecasting Methods Currently in Use	30
2.4.3	Market Segmentation and the Use of Price Fences	38
2.4.4	Aircraft Capacity	42
2.4.5	Inventory Control	43
2.4.6	Booking Limits and Policies	46
2.4.7	Overbooking	48
2.4.8	Demand Distributions Modeled	54
2.4.9	Cancellations, No-Shows and Walk-Ups including Stand-bys	59
2.4.10	Information Requirements	63
2.4.11	System Capabilities	65
	a) Human Interaction Concerns and Issues	65
	b) The Role of the Service Provider	66

	c) The Role of the Customer	68
	d) The Psychology of Revenue Management	70
2.5	Conceptual and Theoretical Models	72
2.6	Empirically Validated Research Models	75
2.7	Economic Approaches to Yield and Revenue Management	79
	2.7.1 Littlewood's Rule	81
	2.7.2 Expected Marginal Seat Revenue (EMSR)	82
	2.7.3 Threshold Curve Methods	84
2.8	Yield and Revenue Management by Sectors	85
	2.8.1 The Airline Sector	86
	2.8.2 The Hotel Sector	86
	2.8.3 The Restaurant Sector	86
	2.8.4 Other Sectors	87
2.9	Concluding Remarks and Directions for Future Research	88
	2.9.1 Summary of Salient Issues and Findings from the Literature	88
	2.9.2 Conclusions	90
2.10	Directions For This Dissertation	92
Chapter 3	Research Contributions	
3.1	Research Objectives	97
3.2	Proposed Markovian Transition Matrices Model	99
	3.2.1 Model Development	100
3.3	Environmental Settings and Limitations	102
3.4	Research Questions	105
3.5	Research Hypotheses	106
Chapter 4	Methodology	
4.1	Model Selection Criteria	107
4.2	Forecasting Models used in Experiment Design	109
	4.2.1 Models Used for Comparison	109
	4.2.2 Numerical Examples of Methods Chosen for Comparison	109

	a) A Classical Pickup Example	110
	b) An Advanced Pickup Example	112
	c) A Multiplicative Pickup Example	112
	d) A Markovian Transition Matrix Example	113
	4.2.3 Forecasting Implementation Issues	115
4.3	The Type of Demand Distribution	118
4.4	The Capacity of the Aircraft	119
4.5	Decision Factors	120
	4.5.1 Aggregation by Time Fences	120
	4.5.2 Aggregation by Fare Classes	123
	4.5.3 Censored and Uncensored Data	124
4.6	Measures of Performance	125
4.7	Data Properties	127
4.8	Experimental Design	129
	4.8.1 Phase I	129
	4.8.2 Phase II	130
4.9	Simulation and Computer Modeling	131
	4.9.1 The Number of Replications	132
	4.9.2 Limitations of the Model	133
Chapter 5	Results	
5.1	Phase I Analysis	135
	5.1.1 Capacity Analysis	136
	5.1.2 Fare Class Analysis – Overall	138
	5.1.3 Fare Class Analysis – Individual	140
	a) V class	140
	b) K class	140
	c) H class	140
	d) M class	141
	e) B class	141
	f) Y class	142

	g) Conclusions	143
	5.1.4 Summary of Phase I Analysis	144
5.2	Phase II Analysis	145
	5.2.1 ANOVA Results	146
	5.2.2 Main Effects	150
	a) Forecasting Method Used	150
	b) Observation Points Utilized	151
	c) Number of Fare Classes Used	152
	d) Type of Data Used	153
	5.2.3 Interactions between Main Effects	153
	a) Forecasting Method and Data Type	153
	b) Forecasting Method and Number of Observation Points Utilized	154
	c) Forecasting Method and the Number of Fare Classes Used	158
	d) The Number of Observation Points and the Data Type Used	159
	e) The Number of Observation Points and the Number of Fare Classes Used	161
	f) The Number of Fare Classes and the Type of Data Used	164
	5.3 Phase II Conclusions	166
Chapter 6	Conclusions and Directions for Future Research	
	6.1 Summary of Research Findings	167
	6.2 Conclusions	169
	6.3 Directions for Future Research	172
Appendix 1	Data Analysis Results	176
Appendix 2	Data Simulation Procedure	184
Bibliography		187

LIST OF TABLES	PAGE
1. Number of Papers Presented at AGIFORS by Subject Area	20
2. Mapping of Days Prior to Departure to Observation Points	100
3. Sample of bookings to date information	110
a. Sample of bookings to date including new information on Apr 30 flight	117
4. Calculated Pickup values from Table 3	111
a. Calculated Pickup values from Table 3a	117
5. Averaged Pickup Values	111
a. New averaged pickup values	118
6. Transition Probabilities (rates) calculated from Pickup values	114
7. Comparative Results between Pickup Methods used and MC methods	115
8. Observation Point Scenarios to be tested	122
9. Fare Class Scenarios to be tested	124
10. Base Case Scenario	129
11. Summary of the number of cases to be simulated	131
12. Rankings of the forecasting methods overall and by class	139
13. Revised summary of the number of cases simulated	146
14. Overall ANOVA results for Decision Factors	149
15. Rankings of Decision Factor Schemes According to Performance Measures	151
16. Crosstabs Table showing 2 nd order means and standard deviations for method and data type interactions	154

17. Tukey Test results for comparisons of FC methods and OPs	156
18. Crosstabs Table showing 2 nd order means and standard deviations for method and data type interactions	157
19. Tukey Test results for comparisons of FC methods by fare class	158
20. Crosstabs Table showing 2 nd order means and standard deviations for method and fare class interactions	159
21. Tukey Test results for comparisons of OPs by Data Type	160
22. Crosstabs Table showing 2 nd order means and standard deviations for number of OPs and data type interactions	161
23. Tukey Test results for comparisons of OPs by Fare Classes	163
24. Crosstabs Table showing 2 nd order means and standard deviations for number of OPs and fare class interactions	164
25. Tukey Test results for comparisons of Fare Classes by Data Type	165
26. Crosstabs results showing 2 nd order means and standard deviations for data type and fare class interactions	165
 Tables in Appendix 1	
1.1 ANOVA and Tukey Tests results for All Capacities according to the MAD statistic	180
1.2 ANOVA and Tukey Tests results for All Capacities according to the MSE statistic	181
1.3 ANOVA and Tukey Tests results for All Capacities according to the MAPE statistic	182

1.4 ANOVA and Tukey Tests results for All Capacities

according to the U statistic

183

LIST OF FIGURES	PAGE
1. The Position of Demand Forecasting in the Revenue Management Process	27
2. The Automated Booking Limit System	49
3. Capacity vs. MAD for all classes	137
Figures in Appendix 1	
1.1 Capacity vs. MSE for all classes	177
1.2 Capacity vs. MAPE for all classes	178
1.3 Capacity vs. U for all classes	179

Chapter 1

Introduction

1.1 INTRODUCTION TO YIELD AND REVENUE MANAGEMENT

Yield and Revenue Management are the two most important components of success for service companies especially in the hospitality industries such as airlines and hotels. These two techniques, at times very similar if not identical, and at other times differing in objectives form the core of profitability and continued success for service industries. In the introduction of this dissertation, the two terms are used synonymously until they can be differentiated.

An airline's or a hotel's daily existence is dependent on its ability to massage its yield management system to extract maximum total revenue from its customers. The ability to react to rapidly changing demands, availability, and competitors' actions are all functions of the yield management system. In fact, it has been shown in detail that the very introduction of a yield management system at American Airlines in 1998 resulted in a quantifiable benefit of \$1.4 billion over the next three years and they expect an incremental annual revenue contribution of over \$500 million to continue into the future (Smith et al., 1992). R.L Crandall, Chairman, President and CEO of AMR and American Airlines was quoted as saying: "I believe that yield management is the single most important technical development in transportation management since we entered the era of airline deregulation in 1979."

Yield management guides the decision of how to allocate undifferentiated units of capacity to available demand in such a way as to maximize profit or revenue. The problem then becomes one of determining how much to sell at what price and to which

market segment (Kimes, 1989b). In cases where the units of inventory cannot be actually differentiated, then the problem simply reduces to one of perishable inventory management. Simply put, the goal is to maximize the number of customers who purchase at higher price levels as well as the number who purchase at all, instead of losing customers who find the average price too high and foregoing revenue from customers who would be willing to pay more (Dyson, 1989).

Researchers such as Kimes (1989a,b), Weatherford and Bodily (1992), McGill and Van Ryzin (1999), and Murtagh and Mitra (2002) have identified the conditions under which yield management techniques are applicable as: (1) a relatively fixed capacity or instances such that the capacity can only be extended at a very large cost, (2) the basic inventory unit is perishable, (3) the demand can be segmented or separated in classes, (4) the inventory unit can be sold well in advance, and (5) the demand is stochastic.

American Airlines was the first to implement yield management techniques with the advent of its SuperSaver fares in 1977, shortly before the deregulation of the U.S domestic and international airline industry (McGill and Van Ryzin, 1999). However, it was Littlewood (1972) who laid the foundation for yield management when he proposed the foundation of the expected marginal seat revenue (EMSR) rule in which he said that discount fare bookings should be accepted as long as their revenue value exceeded the expected value of future full fare bookings. This was a simple two-class seat inventory control rule which became known as Littlewood's rule. Now, yield management solution techniques employ much more complicated approaches due to the increasing

sophistication of computers, powerful mathematical techniques and an enormous amount of readily available data.

Intrinsic to yield management is the idea of price discrimination which is based upon customers' perceived willingness or ability pay a particular price. But this perceived willingness is conditioned by the fact that information must be available to the customer as to why this discrimination is being made. If the customer feels that the discrimination is unfair then the entire sale and subsequent revenue is lost. However, yield management comes into play when the first sketchy evidence arises that a company's fixed capacity will not be fully utilised for a part of the company's business, even if only for a brief time. The decision to lower or raise a price must often be made quickly and sometimes within minutes of the arrival of new information – and that decision will be short term. Moreover, the special offer can be withdrawn just as quickly as it was made (Weigand, 1999).

In the hotel industry, Jauncey et al. (1995) have defined yield management as an integrated, continuous and systematic approach to maximizing room revenue through the manipulation of room rates in response to forecasted patterns of demand. In a keyword meaning analysis of 12 publications referenced by the same authors to identify the essential components of yield management, the following terms (listed in order of greatest frequency) were found:

- variable room rates/price structures
- maximization of revenue
- structured process, way or procedure
- management of a fixed and perishable commodity

- understanding patterns of demand and customer behavior
- forecasting future patterns of demand
- developing strategies and tactics
- improving customer service
- improving profits
- information technology based
- involves organization change
- replaces traditional experience and knowledge
- requires policy and procedural support
- involves constant feedback and support.

It should be noted that while much more research has been done in the airline industry than in the hotel industry with regards to yield management, as evidenced from the keywords identified above, there are definite similarities and common issues to both industries with respect to yield management.

Yeoman and Watson (1997) describe yield management as a human activity system (HAS) based on three interactive sub-systems: people, forecasting and strategy:

- People – the process of managing yield is basically a human activity. A commitment in terms of resources and authority to the principles of yield management is required and the organisation should have a designated team to manage the process. This policy has been reiterated in the work of Donaghy et al. (1997) who state that “organizational and individual benefits of yield management should be highlighted to make staff sufficiently dissatisfied with current operations to be motivated towards change.”

- Forecasting – the ability to forecast accurately is an enshrined principle of yield management. Factors such as weather, competition, risk taking behaviour, booking patterns, and overbooking policy influence the decision maker’s ability to forecast yield. These factors are combined with historical and consumer profile information to present a more accurate forecast. In fact, data handling needs for yield management have grown from simply putting some data into a spreadsheet to running a fully automated, up to the minute data collection and processing system. Computer technology and network implementations have matured to the point that system-wide data collection and processing can be done in real time (Baliga, 1998).
- Strategy – clear guidelines of policy are required in terms of overbooking and market segmentation. Options and tactics that can immediately reply to varying demand conditions because of the turbulent environment should be identified. Critics of yield management software will argue that it often removes the human element of experience and intrinsic knowledge but proponents will counter with the fact that the yield management software is a tool to enhance the decision making process. Practitioners have encountered similar situations in the information systems and information technology (IS/IT) fields. A firm will install an IT package only to find itself deviating from its core strategies and business model as it becomes increasingly reliant on its data processing. As a result the focus of its decision-making has been changed from using the IS to augment its strategies to one in which the IS dictates the strategies. (Yeoman and Watson, 1997).

In a somewhat simplistic proposal, Harris and Peacock (1995) delineated a ten-step procedure for yield management success as follows: (1) determine whether market segmentation is viable, (2) specify booking classes, (3) establish the number of days the system will look ahead, (4) specify class membership conditions, (5) set prices, (6) identify service differentiators, (7) collect the data, (8) specify booking class sizes, (9) establish threshold curves for booking class sizes and, (10) develop a competitor surveillance system. This ten-step approach could be viewed as directly addressing the characteristics of any service industry to which yield management is applicable.

1.2 IMPORTANCE OF YIELD AND REVENUE MANAGEMENT

Historically, yield management started as an operations function, focusing only on capacity allocation given exogenous demand estimates (Gallego and Van Ryzin, 1997). Now, yield management and revenue management applications are becoming more ubiquitous in service industries in today's changing global economy. Originally, it was the domain of the airline industry only. It was then adopted by the lodging industry, which is comprised of hotels, motels, long-term and short-term stays properties. Now it has become prevalent in service industries such as car rentals, TV advertising, broadband networks, holiday shopping, and rail and cruise line companies.

Yield and revenue management's importance arises when a firm has to sell the right perishable inventory unit to the right customer at the right time and for the right price (Smith et al., 1992). In the billion-dollar hospitality industry, an increase in yield or revenues of even one percent is considered a huge advantage in terms of the survival of any firm. This highly significant impact is primarily due to the high fixed or large

overhead costs and the very nature of the service industry. As a consequence of these high-fixed costs, ways have to be determined to increase profitability, by increasing revenues or by lowering costs. Arbitrarily increasing prices could possibly lead to an increase in profitability and/or revenues but this may eventually have a negative effect as the commodities in these industries do have some level of price elasticity.

For example, a cheaper hotel room or a cheaper airline seat could be easily obtained by shopping around. Lower costs in some cases often lead to a degradation of service levels and this may ultimately backfire as customers could once again find a substitutable product. On the other hand, if the company has managed to lower their costs while maintaining or increasing their service levels, this would definitely be a boon to business. As such, to increase revenues, companies rely extensively on their yield management systems that are often very closely guarded trade secrets. In fact, it is their yield management systems that give them a competitive advantage.

Trade secrets are defined as proprietary processes, formulas, devices, or compilations of information that are used by particular businesses (Kimes and Wagner, 2001). Patents and copyrights are often ways of protecting trade secrets and as such become extremely coveted material. Yield and revenue management systems in the hospitality industry should therefore be considered trade secrets as they contain proprietary information. The only way that these systems are somewhat exempt is if they are in fact licensed out to other users.

The proprietary information contained in yield management systems is most often the company's historical data and overbooking policies. As a result, companies should be ever vigilant in protecting their yield management systems and software from external as

well as internal elements. It has generally been assumed that those airlines or hotels with the better yield management package would be those that would tend to have the best revenues.

There are a couple of dominant market players in yield management systems for the airlines and hotels. The Sabre Group (www.sabre.com), a world leader in the electronic distribution of travel and travel related services around the globe, is the leading provider of information technology solutions for the travel and transportation industry, including customised software development and software products, transaction processing, systems integration, and total information technology outsourcing. This pioneering yield management system was developed by American Airlines in response to the deregulation of the airline industry in the late 1970's. While the Sabre Group is the dominant force in the airline industry there is no one such force as yet in the hotel industry.

However, HotelBANK Inc. (www.myfidelio.net), a subsidiary of MICROS Systems Inc. (www.micros.com), provides an advanced e-commerce solution that connects individual hotels, hotel chains, corporations and travel portals allowing them to conduct travel related e-commerce, including reservations, electronic RFP solicitation and response, GDS connectivity, yield management, WAP connectivity and e-commerce targeted marketing to corporate employees. HotelBANK also provides the Internet booking engine behind several hotel chain websites and various portal sites and MICROS provides property management, central reservation and customer information solutions under the brand MICROS-Fidelio for more than 10,000 hotels worldwide.

With regards to practical applications of yield management, there are many instances of success stories as well as some failures. For example, Virgin Express, a Brussels based airline, had indicated that it hoped to report a profit in 2001. Recent financial information indicates a doubling of their third quarter net profits for 2004 to \$4.6 mil. The low-fare carrier had lost money in both 1999 and 2000 and managing director John Osborne had said the airline would boost revenue by selling fewer cheap seats and discounted tickets. It was also indicated that the carrier had been misusing yield management software that was designed to increase fares as services filled up, but the problem was now resolved. By simply and correctly applying yield management techniques, the transformation from losses to profits was easily obtained in 2001. Delta Airlines has estimated that selling just one seat per flight at a full fare rather than a discounted fare can add over \$50 million to its annual revenues (Belobaba, 1987c).

Alaska Air Group comprising Alaska Airlines and Horizon Air together, which makes 1,000 flights each day, has also dramatically improved their revenues with the usage of two new yield management and revenue management packages. They have replaced their existing manual system with the Sabre's Decision Technologies, Inc.'s Airmax revenue management system and Tivoli Systems Inc.'s Maestro automated job scheduling systems. The result was up to a 10% drop in overbookings and an overall 1% to 2% return on investment in the past 9 months (DiDio, 1998). The most recent financial information available for year ending 2003 indicated a tripling of profits as compared to year ending 2002.

In a 1997 study done by Norman and Mayer on Las Vegas casino hotels in which they surveyed seven Las Vegas casino hotels, five of them indicated that they employed

yield management techniques and systems. Depending upon the degree of implementation of the yield management systems, some impressive revenue gains of over \$1 million dollars annually were reported.

In the semiconductor industry, yield management has extensive applications particularly in the area of producing dies from wafers. In a demonstration example, modeled by engineers from KLA-Tencor, SEMATECH, AMD, and Stanford University, a yield management investment of \$2.2 billion showed a six-month payback period with the one-year return on investment (ROI) of 196%. In this example, the expected revenue increase because of a reduction in the percentage of excursion lots (non-usable lots) was \$379,000 per month (Peters, 1998).

In July 1998, the Jan De Wit Company began work on implementing a yield and revenue management application in their lily flower production system. The benefits were astounding. Between 1999 and 2000, company revenue grew 26 percent with a forecasted increase of 32 percent in 2001 and return on equity went from 15.1 percent in 1999 to 22.5 percent in 2000. The mainstay of this application was the use of linear programming as a production-planning and trade tool for the management of the business (Caixeat-Filho et al., 2002). Using yield and revenue management, the company eliminated unprofitable varieties, saved time and money, and shifted demand to off-peak times.

The Internet with its far-reaching powers has become an important tool and accessory to yield management. It poses both opportunities and threats to pricing, but a creative strategy can help avoid the intense pricing competition that obliterates profits (Pitt et al., 2001). Primarily due to its anonymity, buyers and sellers are engaged in a

global dance of price haggling, the perfect example of this being the auction website E-Bay. The same product or service could easily have a myriad of different prices based on time of purchase or class of purchase by customizing the interaction between the seller and buyer.

This customization is done by the use of cookies which identify a particular user to a website by means of a text file located on the user's computer. This implies that the same customer attempting to purchase the same service but at a slightly different time (for example, minutes later) may be charged a different price because of a change in demand for that service. This customization allows for the ultimate in price differentiation in that no two customers may pay the same price for the same service due to a change in demand. At the same time, it allows for the customer's belief that they are getting the best bargain because of the power of the Internet and the fact that the price comparisons can easily be made.

Additionally, this price customization makes the customer feel like an individual and that they are experiencing a high level of service, which they naturally translate into great quality. Buoyed by the experience, they will return to that website again, drawn in by the undercurrent of the yield management system. This practice of yield management using the Internet has so far not exploded in growth as non-hospitality firms grasp the rudiments and advantages of yield management.

But the use of yield management has not always translated into success. For example, in October 1999, Coca-Cola the soft drink giant quietly began testing a vending machine that automatically raised the price of its beverages when the external temperature increased and during the summer months when it was hot. While the

rationale would seem extremely unfair to a thirsty customer or even a regular customer, it was a simple practice of yield management (NY Times, Oct.28, 1999). The idea of adjusting prices based on levels of demand is the underlying concept of yield management. This concept however was completely lost upon the consumer in this particular industry and as a result the practice was abandoned. But at the same time, what might have been construed as an unfair practice could have been advantageous to the customers because prices would have been lowered by the same vending machines during off-peaks hours or in low traffic areas.

Yield and revenue management has far reaching uses ranging from the well-known airline applications to up and coming applications in areas such as the non-profit sector. In fact, within the last dozen years research has been conducted in diverse areas such as: automobile rentals, broadcasting, cruise lines, the Internet, the non-profit sector (Metters and Vargas, 1999); passenger railways (Kimes, 1989b; Ciancimino et al., 1999); golfing (Kimes, 2000); and the flower industry (Caixeat-Filho et al., 2002). With usage in such areas, it will not be surprising if yield and revenue management software and applications become common knowledge and the practice of yield and revenue management becomes much more readily acceptable in all service sector areas.

1.3 The Organization of the Dissertation

Chapter 1 gives an introduction to Yield and Revenue Management and also illustrates the importance of Yield and Revenue Management. The remainder of this dissertation is organized as follows: Chapter 2 explores the definitions of yield and revenue management, issues affecting yield and revenue management, and discusses the

traditional and modern approaches to yield and revenue management. It also examines yield and revenue management by sectors highlighting the differences and similarities and ways that the same techniques can be applied in other areas and also summarizes the salient issues and findings. Based on these salient issues and findings, Chapter 3 outlines the proposed research objectives and contributions of this dissertation. Chapter 4 describes the methodology for empirically testing a Markov Chain based forecasting algorithm approach to yield and revenue management. Chapter 5 provides the results obtained and Chapter 6 gives the conclusions and directions for future research.

Chapter 2

Literature Review

2.1 DEFINITIONS

Several authors use the terms yield management and revenue management interchangeably (for example, Weatherford and Bodily, 1992; McGill and Van Ryzin, 1999; and Kimes, 2002). This dissertation utilizes the same approach. Additionally, the terms service industry and hospitality industry have been used interchangeably throughout this dissertation. The type of service industry being referred to, namely, the airline or hotel or restaurant industry, seems to have some influence on the choice of terms used.

Researchers prefer to use the term yield management when referring to hospitality industry comprising of the airline, hotel and car rental industries, and the term revenue management when referring to the restaurant industry and other types of hospitality industries. The top-tier journals in the field of yield and revenue management research do tend to pick a single term and use it consistently after an initial explanation as to the similarity of both terms (for example, *Journal of Operations Research*, *Operations Research and Transportation Science*). Journals in the field of hospitality management such as the *International Journal of Contemporary Management* and the *Cornell Hotel and Restaurant Administration Quarterly* are somewhat more lenient in their approach as very often the type of articles presented here are qualitative even though yield is associated in the mathematical sense.

The most wide ranging definition of *yield management* is that it is a micro-economic concept which uses such data as the hotel or airline booking history, current

market conditions, competitive room rates or fares, booking cancellation, and seasonal events to determine the maximum rate that can be charged while aiming for maximum occupancy (O'Leary, 1989).

Cross (1995) attributed the original term of yield management to Robert L. Crandall when he was the Senior Vice President of Marketing (later CEO) at American Airlines (McGill and Van Ryzin, (1999). The most recent definition used is that: yield management is a method that can help a firm sell the right inventory unit to the right customer at the right time and for the right price (Kimes, 1989a). It guides the decision of how to allocate undifferentiated units of limited capacity to available demand in a way that maximizes profit or revenue. The question is how much should one sell at what price and to which market segment? (Kimes, 2002)

Yield management has also been defined as an intelligent approach to the dynamic reservation control and pricing of a perishable asset across customer types. Reservation control is the process by which perishable assets are made available/unavailable to customers. For example, a hotel has a fixed set of standard-sized rooms to rent to the various customer types (Collier, 1987). These different customer types (tourists, businesses, VIP, etc.) pay different rates, purchase different amenities, stay different lengths of time, and make reservations at different times in advance of the first stay-over. The essence of reservation control is to systematically accept-reject reservation requests to maximize benefits to the hotel. Reservation control becomes necessary when total room utilization in the hotel or seat utilization on the airline is high.

In fact, it was Weatherford and Bodily in their comprehensive 1992 taxonomy who coined the all inclusive term, *perishable-asset revenue management* (PARM). They

defined it as the optimal revenue management of perishable assets through price segmentation. There have been many adaptations of the definition since then. Some of the more common definitions in use are described as follows.

Yield management has also been defined as the dynamic pricing, overbooking, and allocation of perishable assets across market segments in an effort to maximize short-term revenues for the firm (Baker and Collier, 1999). Mathematically, yield has been defined by Orkin (1988) as 'actual revenue' divided by 'potential revenue.' Actual revenue is defined as the total sales receipts and potential revenue is defined as the income which could be earned if 100% of the inventory was sold at its full rack rate.

Lieberman (1993) defines revenue management as "systems & procedures to maximize results from the sale of a product or service in more or less fixed supply whose revenue producing ability diminishes with time."

Revenue management has also been defined as the art and science of predicting real-time customer demand at the micro-market level and optimizing the price and availability of products (Cross, 1995).

Almost all of the articles with a mathematical focus, that is, those articles proposing a model or an algorithm, utilize the term yield management even if it is a recent publication. McGill and Van Ryzin (1999) have suggested (based on their research of the forty-year history of yield management) that the term *revenue management* has largely replaced the term yield management. In checking their references used, the earliest mention of revenue management was by Littlewood in his 1972 paper. However, to preserve the integrity of the original research works, the two terms will continue to be used interchangeably.

2.2 A BRIEF HISTORY OF YIELD AND REVENUE MANAGEMENT

The origin of yield and revenue management can be traced back to its start when almost all quantitative research in reservation control in the airline industry was focused on controlling overbookings. Overbooking was the process of predicting the number of passengers to appear for boarding at flight time and then double-selling the seats of those who were predicted not to show. Beckman and Bobkowski (1958) did the first statistical modelling of passenger overbooking, cancellation and no-show behaviour. They compared 3 different probabilistic models and determined that the Gamma distribution was the best fit for airline data.

In the 1960's, American Airlines started to use Operations Research models for revenue management decisions. These models were based upon using data from its reservations inventory to influence variable pricing models. They found this to be a huge task, given the limitations on technology. Therefore revenue management decisions were based upon overbooking, discounting and traffic management. This process enabled a relationship between these flight parameters and price (www.iymrs.com).

In the 1970's, once again American Airlines were the pioneers. They developed the SABRE (Semi-Automatic Business Research Environment) computer system into which all sales were entered, automatically updating the availability and prices for flights, and enhancing the ability to make infinite pricing decisions. Detailed information on every flight was recorded. With the deregulation of the US airline industry in 1970's, Peoples Express, one of the new airlines, emerged, offering customers a low priced ticket with minimal amenities. Major airlines such as American were forced to compete with

Peoples Express by offering a few seats at even lower fares while maintaining higher fares on the remainder of their seats (www.iymrs.com). American was able to offer these new low fares because of the advantage offered by their yield management systems and by maintaining their service levels American enticed customers to switch back to them.

With improvements in technology in the 1980's, many airline companies benefited from utilization of quantitative approaches to revenue management. These ranged from simple rule based heuristics to sophisticated mathematical programs, aiding management with the problem of seat allocation. Technology that was fast, accurate and relatively inexpensive enabled management science models to solve problems repeatedly. The 1980's saw revenue management become a robust and workable system for solving the problems of fixed capacity, time-varied demand, segmentation, perishable inventory and high fixed costs, bringing a practical solution. Belobaba's PhD thesis work, Air Travel Demand and Airline Seat Inventory Management in 1987 was a significant contribution to this solution (www.iymrs.com).

The era of the 1990's saw even more proliferation of revenue management in the airline industries, but also the growth in the hotel and other industries that faced similar problems such as car rental, cruise lines, sports stadiums, and rail transportation. Revenue management was being established as a core management practice in these areas, becoming an essential tactical and strategic tool for companies to sell the right inventory unit to the right customer at the right time and at the right price. This proliferation was based upon enhanced management science models and greater accessibility to technology in addition to the acceptance of the guiding principles of revenue management in enhancing the bottom-line.

Currently, there is a growing acceptance of yield and revenue management and its place as a core management activity in most areas of the hospitality industry by consumers. Its use as a strategic weapon has come down to which company has the better algorithm or model, as an increase of even 1% is considered a huge success in multi-billion dollar service industries. Yet to be fully exploited are the explosion of e-commerce and the burgeoning use of expert systems.

2.3 A SURVEY OF THE LITERATURE

The origin of yield management can be traced back to late 1950's when research began in the area of overbooking control for passenger reservation systems. This was the process used by the airlines to estimate what percentage of the passengers would be no-shows and then "double-sell" those seats. In the fifty years since the first publication on overbooking control, passenger reservation systems have evolved from low-level inventory control processes to major strategic information systems (McGill and Van Ryzin, 1999).

The research on overbooking control, forecasting and single-leg seat inventory control in the airline industry has been extensive and is still being continued to this point. The research in origin-destination (O-D) control and pricing has not been as extensive as it has only been done for the last 35 years in the case of pricing and the last 25 years in the case of O-D control. In the hotel industry, research has not been as extensive as in the above mentioned areas but has been mostly conceptual often centering on the use of heuristics.

The central issues to any yield management system, whether it is being used in the airlines or any of the other service industries have been identified as accuracy of the forecasting methods, the overbooking models used, seat inventory control (or room inventory control as the case may be for hotels), the pricing structure or use of price fences, the ability to segment the market cancellations, no-shows and walk-ups, and the role of information.

Since 1961, the Airline Group of the International Federation of Operations Research Societies (AGIFORS) has been the body responsible for the applications of operations research to the field of air transport. Analysis of the number of papers presented at their conferences is shown in the following table. The data was obtained from their website (www.agifors.org) and from Richter (1989).

Period	Total Papers	Forecasting	Maintenance & Inventory	Yield Control	Other
1956-1960	52	8	12	6	26
1971-1975	129	36	18	1	74
1981-1985	88	15	5	10	58
2000-2006	N/A	N/A	N/A	155	N/A

Table 1. Number of Papers Presented at AGIFORS by Subject Area

As Table 1 illustrates, there has been more than a fifteen-fold increase in the amount of research done in the area of yield control over the past six years. The most plausible reason for this increase in attention to yield control (as it was called then) during the period from 1981 to 1985, and then from 2000 to 2006 could be attributed the fact that the airline industry was deregulated in 1979 and competition was dramatically increased. In the most recent period from 2000 to 2006, the resurgence of interest in the field could be attributed to quite a number of factors of which an increase in air travel and an increase in the number of hotel reservations are probably the most important.

2.4 MAJOR ISSUES IN YIELD AND REVENUE MANAGEMENT

One of the sticking points in utilizing any yield management system is to define the objective of the system. Is it to maximize yield or revenue or occupancy? Or is it to minimize overbookings? Kimes (1989b), Relihan (1989), and Lieberman (1993) have implied by their definitions that profit maximization should be the primary objective of any yield management system. Other researchers such as Belobaba (1987b), Gallego and Van Ryzin (1997), and Talluri and Van Ryzin (1998) have emphasized pricing controls as the objective of yield management systems.

Jones and Hamilton (1992) have also identified the various types of information (such as property location, demographic makeup, sources of demand, and the relative advantages held by each competitor) to be collected and analyzed as being important to any yield management system. Jauncey et al. (1995) also emphasise the importance of information technology.

Another issue facing the success of any yield management system is the lack of understanding as to what the system is capable of doing. Andrew et al. (1990) and McEvoy (1997) have also identified this issue as a major stumbling block in accurately measuring the impact of any yield and revenue management system.

In addition, previous discussion has already identified forecasting, overbooking, inventory control, price fences, market segmentation, walk-ups, no-shows, and cancellations as relevant issues affecting the success of yield and revenue management systems.

The following sections examine these issues in more detail and highlight the problems associated with each.

2.4.1 Objectives of Yield and Revenue Management Systems

As delineated in an earlier section, and according to McGill and Van Ryzin (1999), almost all quantitative research in reservation control before 1972 focused on controlled overbooking. The objective of overbooking (which will be discussed in detail) was to predict the probability distributions of the number of passengers who actually show up for a flight. Predicting the no-shows or passengers with reservations who are not at the boarding gate at the time of departure could greatly enhance the revenues of the airline. Since these numbers of passengers were predicted not to show, then their seats could be double-booked, thus increasing the revenues for that particular flight.

Kimes (1989a, b) defines that the objective of yield management is to maximize the revenue or yield of the firm. Relihan (1989) defines yield management as a proven technique for maximizing revenues. Similarly, Lieberman (1993) defines yield management as the practice of maximizing profits from the sale of perishable assets, such as hotel rooms, by controlling price and inventory and improving service. It can be easily seen that all of these definitions converge to a single objective of maximizing revenues or profits. However, this does not, by any means, imply that it should be the pervasive definition used, as there are other objectives defined for yield management systems.

Belobaba (1987a) asserts that to prevent revenue dilution – whereby the airline sells too many seats at low fares – requires effective yield management, which includes both pricing and seat inventory control. Gallego and Van Ryzin (1997) claim,

“historically, yield management started as an operations function, focusing only on capacity allocation given exogenous demand estimates.” They then further elucidate that, “one can view allocation, at least conceptually, as a pricing decision.” This can lead to the inference that yield management is conceptually a pricing decision.

Talluri and Van Ryzin (1998) explore the concept of bid pricing as a method of controlling the sale of inventory in revenue management applications. Bid pricing is the process by which a threshold price is set for the particular unit of inventory. Weatherford and Bodily (1992) in their comprehensive taxonomy define perishable-asset revenue management as the optimal revenue management of perishable assets through price segmentation. Again, the similarity in findings can be noted here, whereby pricing is used to control the yield management system.

It may appear that even though there are sometimes conflicting statements in the objectives of yield and revenue management and the pathway to achieve those objectives, there is no mistaking the final outcome is one of obtaining the greatest reward possible for providing services which aim to extract the maximum price paid by each customer.

2.4.2 Forecasting

Forecasting can be defined as the art and science of predicting future events. It may involve taking historical data and projecting them into the future with some sort of mathematical model. It may be a subjective or an intuitive model (Heizer and Render, 2001). It is an important component of planning in any enterprise; but it is particularly critical in airline revenue management because of the direct influence forecasts have on the booking limits that determine airline profits (McGill and Van Ryzin, 1999). Even

though McGill and Van Ryzin's statement is directed to airline revenue management, the importance of forecasting in service industries cannot be understated.

The difficulty that arises when utilizing forecasts in the service industries and more specifically, the hospitality industries is that there is no output to be measured as compared to a manufacturing firm. However, in some sense, in the hospitality industry, there is a fixed output (as in the number of airline seats or the number of hotel rooms) and what is being forecasted is the degree of utilization or occupancy. A comparable scenario in the manufacturing sector would be forecasting the demand for a particular product and then further forecasting what proportion would be actually consumed. This of course, intrinsically compounds the problem.

In the hospitality industry, there is the need to concurrently forecast demand, capacity, and prices. Even though they may seem to be independent of each other, they are in fact correlated. Demand includes all the reservations made and it is dependent upon factors such as time of day, origin-destination, and day of the week. Demand could also be manipulated by pricing controls. The inventory controls established by this method will only be as good as the forecast of demand for these fare classes.

In general, there are two consequences to a poor forecast. If the forecast for the business passengers (or the higher fare paying) is too high, then too many seats will be protected for these passengers. If this happens, then the airline runs the risk of departing with empty seats which could have otherwise been sold to leisure type (or the lower fare paying) passengers. If the forecast for the business passengers is too low, then too few seats will be protected. The airline would lose revenue in this case since more seats would have been sold to the lower fare paying customers that could have been sold to

higher fare paying customers (Zeni, 2001). If there is low demand then discounts could be offered or prices could be reduced to stimulate demand and possibly increase revenues. Likewise, if there is a significant upswing in demand then prices could be increased to generate more revenue. In reality, demand is always tempered by capacity.

Capacity utilization is associated with the number of show-ups. Show-ups are the actual demand without cancellations or no-shows and include those persons who walk-up without reservations. In certain situations, capacity may actually equal demand. For example, cancellations allow sufficient time for the airline or hotel to re-book the seat or room, while no-shows do not allow the airline or hotel sufficient time for re-booking but walk-ups could counter-balance this loss thus adding up to the original demand amount. Modeling cancellations, no-shows and walk-ups require a good understanding of consumer behavior and accurate data collection for use in stochastic models. This issue will be discussed in detail in a later section of this dissertation. Since overall capacity is fixed, the problem is with fare-class capacity allocations.

Forecasts can be biased because of permanent changes in passenger behavior that the time series models do not rapidly incorporate (Smith et al., 1992). Regression and time series techniques have proven of some use for forecasts of aggregate demand, but not at the disaggregate level (McGill and Van Ryzin, 1999). Aggregate demand represents a collective version of the demand incorporating all the various factors such as times of day, flight numbers, and origin-destination pairs for example. However, the trend is towards disaggregation as this provides much more tractable and meaningful solutions. Disaggregation could be done by factors such as week, month, season, market, and even the booking class; but at the expense of forecast accuracy.

Even with disaggregation techniques available, the dynamic nature of yield and revenue management increases the complexity of accurate forecasting. One of the primary causes of this is the constantly changing nature of prices which may occur more than 100,000 times a day per airline. Additionally, the models for demand distributions and the arrivals process are stochastic (see McGill and Van Ryzin, 1999; Lautenbacher and Stidham, 1999 for example) and are not easily modeled using only the mean.

Forecasting relies entirely on historical data and although the airline industry – other than the stock market - has the best data with regards to volume, duration, and quality; there exists a major deficiency with most yield management systems. They are not capable of capturing data or information on rejected demand and as a result this is where censoring or constraining of the data occurs. In addition, policies relating to the handling of repeat and valued customers (who may be charged a standard or lower rate throughout the year) need to be developed and communicated via the application (Jauncey et al., 1995). The choice of the amount of historical data to use is therefore often subjected to the biases of the forecaster as well as the model to be used.

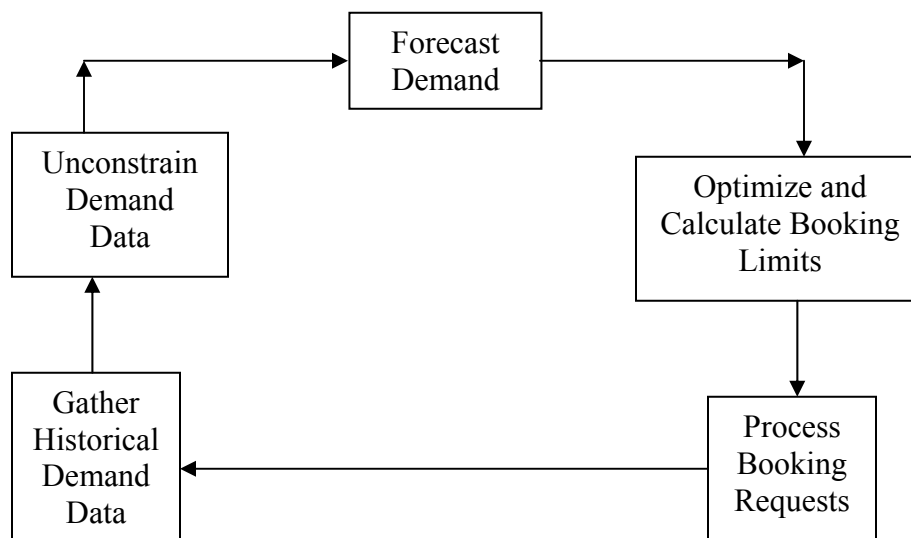


Figure 1. The Position of Demand Forecasting in the Revenue Management Process (Zeni, 2001)

For example, Lee's (1990) dynamic estimation model utilizes a small window of historical data, as little as 8 to 17 weeks of data, and re-estimates the parameters each week as the most recent historical data becomes available. This is analogous to the simple moving average method. His static estimation model uses a larger window of historical data, from 26 to 52 weeks of data but re-estimates the parameters only after a suitable period of time (for example, every quarter or semi-annually). However, these two models are polar opposites of each other as the static model preserves the long-term trends in the data, and the dynamic model captures the short-term trends more accurately.

Forecasting could be considered as the backbone of yield management systems and with exponential increases in computational power and better designs of information systems and utilization of information technologies, its role could prove to be even more indispensable. But, no matter how powerful the technology, there will always be some error in the sales forecasts, because consumer behavior will never be 100% predictable (Chase Jr., 1999). With the advances in computational power, neural networks are now

playing a much more important role in forecasting mainly because of their dynamic nature.

The power of neural networks stems from parallel processing operations that allow multiple computations to occur at once (Bermar, 1988). Neural networks simulate the way the human brain recognizes patterns to identify information and how we are able to learn from our mistakes and from experience. It is a huge component in artificial intelligence use and is responsible for a wide range of applications from banking to airlines. In 1994, USAir introduced a neural network-based revenue management system called Excalibur. So far, they have been the only airline to utilize this type of system which was developed by BehavHeuristics Inc. Since January 1994 when it was installed, the system is said to have generated \$120 million in net revenue improvement. It is believed to be one of the world's largest single applications of neural nets (Sweetman, 1996).

In yield management, neural networks are dependent upon the fact that the number of inputs has to be finite and consistent and some of them must be related to the output. Standard inputs such as the number of days until flight departure, origin and destination, aircraft type and time of day data, which influence the output of maximizing revenues, are required initially. The subsequent step is to train the network (giving it experience) by loading it with some amount of data. This training is necessary so as to minimize the amount of human interference needed. In the initial two phases, the output is compared to the actual demand and the system is adjusted accordingly. A third set of data is also supplied for an "online simulation" which demonstrates how well the system is performing.

An advantage of this type of approach to yield management is that neural networks automatically adapts to marketplace changes or even changes in the airline schedule making it truly dynamic. Since the neural net forecast is based on the flight characteristics and not a particular flight number, changing the schedule or the flight number does not affect the forecasts. However, a “disadvantage” of neural nets is that they are considered to be “black-box” type operations. There is no real simple understanding of the process in between the input and the output.

The following sub-sections discuss the forecasting techniques currently in use by the airlines and some of their advantages and disadvantages.

2.4.2.1 Forecasting Models Used in the Airline Industry

Forecasting in the airline industry could be broken down in two broad categories - macro-level and micro-level forecasting. Macro-level forecasting, as the name suggests, is on a much larger scale to be of practical use to revenue management. It is primarily concerned with the aggregate demands of airline travel and is associated with a longer time horizon than would be beneficial to revenue management. Examples of macro-level forecasting would include forecasts for the total annual demand for air travel, as well as forecasts for a particular route over the next five years to enable long-term planning.

Micro-level forecasting on the other hand is the focus of yield and revenue management; since it is at this level, passenger demand could be broken down by flight number, date, time, and fare class. Researchers (Kanafi, 1983; Lee, 1990; Zeni, 2001) have argued that if micro-level forecasting is further disaggregated, then it becomes passenger

choice behavior modeling, which is the study of an individual's behavior based on socio-economic factors and the characteristics of alternative travel options.

There are five micro-level forecasting techniques used frequently by airlines in their revenue management systems today. They are the: exponential smoothing, moving average, linear regression, additive (or pickup), and multiplicative models (Weatherford, 1999). These models will be described briefly in the following sections for comparative purposes.

Since forecasting is the central input to any revenue management system, the objective of these models needs to be clearly defined. These models forecast the total demand for seats at departure time (DP_0), given the current state and the historical bookings data (HBD). The current state is defined as the (cumulative) bookings to date (BTD) at the time of the forecast. The HBD is represented by a booking profile which is a graphical representation of the bookings history. These booking profiles are usually not necessarily strictly increasing as time to departure nears. Since they represent BTD (at time t), it is entirely possible that the snapshot at the next BTD (at time $t+1$) could be less if the number of cancellations at time $t+1$ are greater than the number of new bookings at time $t+1$.

2.4.2.2 Micro-Level Forecasting Methods Currently in Use

All forecasting models possess the same objective, which is to forecast total demand for seats or bookings at departure time given the current state and historical bookings data. Of the five micro-level forecasting methods used in airline revenue management the most commonly used are the pickup methods and linear regression in the Passenger Origin-Destination Simulator (PODS) developed at Boeing (Gorin, 2000).

In this section, each of these models will be described along with their relative merits and shortcomings.

1) Exponential Smoothing

Exponential smoothing methods are used in revenue management practice because they are simple and robust and generally have good forecast accuracy (Talluri and Van Ryzin, 2004). In simple exponential smoothing, only one parameter needs to be estimated. The simple exponential smoothing model is defined as:

$$F_{t+1} = \alpha D_t + (1-\alpha)F_t,$$

where F_{t+1} is the new forecast, alpha (α) is the smoothing constant, D_t is the current demand and F_t is the current forecast.

Small values of alpha result in relatively stable forecasts since the forecast responds slowly to changes in the data. Large values of alpha causes the forecast to respond quickly, but the rapid changes in the data maybe more symptomatic of an underlying new trend or random fluctuations. As a result, a balance must be struck in choosing the appropriate value of alpha. The choice of alpha is therefore done by careful analysis of historical data. In practice, alpha is typically set between 0.05 and 0.30 in RM applications (Talluri and Van Ryzin, 2004).

As the mathematical formulation demonstrates, this model is indeed a favorite of the airline industry as defined by the model selection criteria (discussed in Section 4.4). The simplicity and data requirements are very minimal since all that is required are the values of alpha, the current demand and the current forecast.

2) Moving Average

Moving averages or simple moving averages (SMA) determines a forecast by averaging the previous n -periods of historical data. Smoothing of the forecast takes place when the oldest observation is dropped and the newest one is incorporated. The SMA forecast model is defined as:

$$F_{t+1} = \frac{1}{n} \sum_{i=t-n+1}^t D_i$$

or equivalently
$$F_{t+1} = F_t + \frac{D_t - D_{t-n}}{n},$$

where the second formulation has been shown to be computationally faster. Similarly to exponential smoothing, SMA only requires the estimation or determination of a single parameter.

If n is small, then the forecast responds more quickly, but it is more sensitive to random fluctuations in the data, similar to a large alpha in exponential smoothing. In practice n may range from 3 to 15 depending on the data characteristics and the units used for the time intervals (Talluri and Van Ryzin, 2004).

As the mathematical formulation again demonstrates, this model is computationally simple, but it requires more data storage than exponential smoothing since it needs all of the n latest observations to be stored. As previously described, exponential smoothing and moving averages do not react well to recent trends. Since airline demand is very seasonal and does exhibit trends, both of these models are at a significant disadvantage when used as forecasting methods unless the value of alpha or the number of periods is carefully chosen.

There is a paucity of research on weighted moving averages. In fact, there was only one study by Weatherford et al. (2003) which used weighted moving averages as a comparative method. As their results showed, the weighted moving average technique gave some of the worst results of the methods used in comparison to neural networks. This could be attributed to the fact that while the method may be simple mathematically, the requirements make it infeasible for use in revenue management forecasting. Firstly, the number of weights must be determined. Secondly, their values have to be calculated by some other mathematical technique (possibly multiple linear regression). And thirdly, the assignments of these weights to the preceding data points must also be determined. Evidently from the requirements, there is at least a minimum of 3 parameters to be determined before they can be applied. If a simpler method is required then linear regression could be utilised.

3) Linear Regression

Linear regression models calculate a forecast (the dependent variable) by a linear combination of one or more independent variables. The number of independent variables to consider is based upon the available data and their predictive significance. The parameter or parameters are estimated by the least squares error (LSE) method or other relevant methods. A three variable model is defined as follows:

$$F_t = \alpha_0 + \alpha_1 D_{t-7} + \alpha_2 D_{t-14} + \alpha_3 D_{t-21},$$

where F_t is the forecasted number of departures (or bookings) at time t ; D_{t-7} , D_{t-14} and D_{t-21} representing the total number of bookings at 7,14 and 21 days before departure respectively; and α_0 , α_1 , α_2 , and α_3 are the parameters to be estimated. It should be noted

that one of the independent variables could be chosen to represent the total number of bookings in a higher fare class as all fare classes are inter-related (Wickham, 1995).

As is evidenced from this formulation, this model is mathematically more complicated and while it may not require as much data storage as an n -period SMA, the computations to determine the parameter values may be a time-limiting factor. Another significant problem is the underlying assumption of linearity. Also, it should be noted in such formulations, there is no consideration for any environmental or economic variables which always have an impact on the number of bookings. Even if consideration is given to such variables, determining which one or which ones to choose could be an enormous task.

Box-Jenkins' Auto-Regressive Integrated Moving Average models known as ARIMA models are a complex, powerful and flexible technique which are special cases of linear regression or moving averages models. But, they are not easy to use and require a great deal of expertise and experience. Since the number of parameters (to be estimated) of each kind is usually about 2, it is often practical to try alternative models on the same data. Parameters need to be estimated for both the auto-regressive and moving averages components resulting in at least 4 values to be determined – a significant number when dealing with the data intensity of airline reservation forecasting.

In his 1987 dissertation, Sa presented two ARIMA time series models for a single fare class on a single flight number. The results of his models were not encouraging and he saw no further benefits of time series modeling. He also developed a general structure model via regression analysis which proved to be more useful when compared to time

series analysis or historical averages. However, he did not include the effects of booking limits into his analysis.

Ben-Akiva's 1987 research on a combined model using an ARIMA time series component and a regression component showed that when the components were run separately the correlation coefficient between the actual and forecasted observations declined but the regression model by itself proved to be a better predictor. As both of these studies showed, and as is evidenced by the literature, the complexity associated with ARIMA models makes it a difficult and limited tool to use for airline reservation forecasting.

4) Additive Pickup

The Additive Pickup or the Classical Pickup Model as defined by Duncanson in 1974 uses data from departed flights only. Pickup forecasting is one level of detail higher than simple time-series forecasting. The pickup forecaster includes more information and averages not just the final bookings for the flight, but actually the number of passengers picked up in the intervals preceding departure (i.e., the average incremental booking received in each time interval before departure) (Zickus, 1998). This implies that the final bookings are a function of current bookings on hand and the amount picked up between the current day [DP_x] and the day of departure [DP_0] (Zeni, 2001).

Borrowing the notation and formulation from Wickham (1995) we have,

$$\overline{PU}_{day(X,0)} = \overline{Bkd}_{day0_n} - \overline{Bkd}_{dayX_n} \quad \text{and} \quad \overline{Bkd}_{dayX_n} = \frac{1}{n} \sum_{k=0}^{-n} Bkd_{dayX_k},$$

where $\overline{PU}_{day(X,0)}$ = the average pickup between day X (day X being one of the observation points, for example 21 days prior to departure – DP_{21}) and the day of

departure, \overline{Bkd}_{day0_n} is the average final bookings for n departures and \overline{Bkd}_{dayX_n} is the average bookings on day X for n departures. It should be noted here that there are no parameters to be estimated here. If the averages are desired such that more emphasis is placed on recent flights (that is, a weighted mean) then using formulation by Zickus (1998) we have:

$$X_{M,i} = \left[\frac{1}{t} \sum_{j=1}^t W_j \cdot (X_{j,i} - X_{j,i-n}) \right] + X_{M,i-n}$$

where $X_{M,i}$ = total bookings after interval i for flight M ,

M = flight on which forecast is desired,

n = number of time frames over which the pickup is calculated,

t = number of flights upon which forecast is based,

j = flight index of flights upon which the forecast is based, and

W_j = weighting value applied to individual flights (if desired).

While this model may seem to be computationally more complex than the preceding models, this turns out to not be the case, primarily because of the way the airlines store data. Airlines usually store data in the form of cumulative bookings or bookings to date (BTD) for any given flight and as a result, the pickup between any two intervals could easily be obtained by subtraction. It can therefore be easily deduced as to why this method is a favorite among airline forecasters and RM practitioners in the industry. The one major disadvantage to this model is that it does not incorporate data on flights yet to depart even though such data is available. The following model addresses this deficiency.

5) The Advanced Pickup Model

A relative weakness of the Classical Pickup method is that it could be influenced by a single high-demand or a single low-demand flight (similar to the effect of an outlier point in regression analysis). The Advanced Pickup Model developed by L'Heureux (1986) at Canadian Pacific builds upon the Classical Pickup Model by combining information about flights which have departed and available, partial information about flights yet to depart. L'Heureux argued that by exploiting this data, his model would be better able to handle irregular situations such as holidays. Additionally, if a new trend is exhibited, this method would be more suited to deal with it. The adjustment to include data from flights yet to depart is very similar to using a weighted average with more emphasis on the recent data.

The model is therefore formulated as follows:

$$PU_{DP(X,X-7)} = BTDDP(X-7) - BTDDP(X),$$

where $BTDDP(X)$ = is the number of bookings to date as of day prior X as defined previously. As can be seen, this model is as simple computationally, as the Classical Pickup Model. The key to the Advanced Pickup Model involves estimating the aggregate pickup by summing estimates of the pickup over smaller disaggregate intervals. Therefore, before the advanced pickup model can be employed, a pickup sub-matrix must be generated (Wickham, 1995).

6) The Multiplicative Pickup Method

Another pickup method used occasionally is the Multiplicative Pickup (MP) method. This method multiplies current bookings to date by an average pickup ratio for a specific period prior to departure. The average pickup ratio for 14 days prior to departure for example, is calculated as follows:

$$\overline{PUR}_{DP14,DP0} = \frac{\overline{BTD}_{DP0}}{\overline{BTD}_{DP14}}$$

The forecast of final bookings at departure is then calculated as follows:

$$FC_{DP0} = \overline{BTD}_{DP14} \times \overline{PUR}_{DP14,DP0}$$

It should be noted that two methods could be used to calculate the average pickup ratio. The data from all departed flights or the data from flights yet to depart as well as departed flights can be used to determine the average pickup ratio (similarly to the difference between the Classical and Advanced Pickup methods).

2.4.3 Market Segmentation and the Use of Price Fences

In order for a yield management program to be effective, the firm must segment its market into different types of customers (Kimes, 1989a, and b). The common mechanism used to segment customers in yield management situations is the time of purchase, that is, the less price-sensitive customer generally waits until the last minute to make reservations. On the other hand, people who make their reservations early are generally more price sensitive; they are willing to trade away some flexibility for a reduced price (Weatherford and Bodily, 1992).

The sizes of the market segments in the hospitality industry are usually measured in terms of buckets. Buckets are defined as clusters of different fare classes that are

grouped together for control purposes in a virtual nesting system (McGill and Van Ryzin, 1999). Also associated with the term “buckets” is the term “booking limits” which can be simply defined as the capacity of the bucket. The concept of market segmentation could be validated by Pareto’s argument that the general welfare is increased if the welfare of some individuals is improved and no one’s welfare is diminished or harmed.

Segmentation by customer type also occurs with a noticeable similarity existing in inventory management with the Pareto Principle also known as the “80-20 rule.” This statement attributes itself to the rule of thumb that 80% of a firm’s revenue comes from 20% of its customers or 80% of a firm’s income is derived from 20% of its inventory items. Nurturing these customers or being extremely attentive to those items is therefore of utmost importance to any company. In the airline and hotel industries, these “20%” type customers or “cash-cows” as they would be termed in marketing are none other than the business travelers who must often make reservations on short notice. However, these “20%” type customers in the hotel industry present somewhat of a different problem than in the airline industry. Due to the demand elasticities and the greater number of competitors in the hotel industry, switching is much easier than in the airline industry and as a result, managerial judgment or the human factor approach to yield management plays a more pivotal role when dealing with this customer type.

Knowing the target market has often been the precursor to market segmentation. Therefore, certain questions need to be answered before the market could be actually segmented. Questions such as: (1) what proportion of customers will order in advance or during off-peak periods at a reduced price? (2) what proportion of customers would agree to purchase units of inventory for a lower per-unit price (group or volume discounting)?,

(3) what proportion of customers will commit to repeated purchases over time for reduced per-unit prices (frequent flyers or guests)?, (4) which customers will trade ancillary services for lower prices (budget airlines or hotels)?, (5) can discounts be promoted cost effectively (cost of advertising)?, (6) who are the less price-sensitive customers and can they be prevented from purchasing at discount, thereby diluting revenue (Harris and Peacock, 1995)?

Price fences could also be used as a market segmentation tool in a yield management system. Price fences are designed to allow customers to segment themselves based on their willingness to pay, their behavior, and their needs (Hanks et al., 1992). There are two types of price fences: physical and non-physical.

Physical price fences such as: location (for example: window or aisle seat on an airline, beachfront or penthouse suite in a hotel), amenities (for example: TV in the first class cabin or a personal attendant on a cruise line), and furnishings (for example: more leg room in business class or a king sized bed in the hotel room) are all used as some form of price discrimination to segment a market.

Non-physical price fences such as: time of usage (for example: early bird specials), transaction type (for example: advance purchases, minimum purchases, and cancellations), and consumer type (for example: transients or frequent users) can also be used as means of price-discrimination to segment a market.

Price fences only work if they are transparent to the customer and they cannot easily be circumvented. Most airline passengers know that it is unlikely that any two persons on the same flight would ever pay the same price for their tickets, but this is an accepted practice. In the hotel industry and other service industries such as the restaurant

industry, this type of justification is a little more difficult. In the airline industry, price fences are an easily accepted practice because the industry is very transparent about its policies.

In the hotel industry the price fences are somewhat more negotiable because of a greater amount of competitors and human interaction even though the hotel's policies may be transparent. This is also due to the fact that the hotel industry has greater price elasticity than the airline industry mainly because it is much easier to shop around for another hotel in the nearby vicinity as compared to shopping around for an airline ticket at the airport. However, with websites such as Expedia.com this is not entirely true as you can easily customize your search and resulting booking by evaluating many air and hotel combinations simultaneously.

Different fare classes or price fences are utilised by the airlines and other revenue management practitioners as a means of controlling availability of inventory and to encourage particular consumer purchasing patterns. For example, the many restrictions placed on the cheaper fares persuades customers to purchase these fares well in advance. However, in order to maximize revenue, the airline must set aside or "protect" seats for the higher fare paying customers who tend to book at the last minute. To determine the amount of seats needed to be set aside, the airline must have accurate forecasts as input to its inventory control system.

A 1997 study by the Mackinac Center for Public Policy found that airline travel in the short run had an estimated elasticity of demand of 0.1 and in the long run, this estimated elasticity was 2.4. In the restaurant industry, price fencing is not as easily accepted since it may be somewhat more difficult for a restaurant to justify charging

more for a meal because the table faces a window and negotiating the price of a meal is not a very common practice.

2.4.4 Aircraft Capacity

Very few studies have actually investigated the effect of the capacity of the aircraft on the performance of the forecast model or vice versa. In fact, only a proof-of-concept study by Berge and Hopperstad (1993) addresses the relationship between accurate forecasting and the capacity of the aircraft. However, their Demand Driven Dispatch (D^3) concept is based on the critical assumption that the forecasting errors at the midpoint of the booking process are relatively small and thus, enables the correct aircraft assignment. They also base their concept on the assumptions that there are no cancellations but they allow for no-shows with the no-show rates being the same for all flight legs and all fare classes. In reality of course, these are not valid assumptions.

The advantage offered by the D^3 concept is that it allows for far more dynamism than is currently exhibited by aircraft fleet assignments. Fleet assignments are typically done 30-120 days prior to flight time and are based on crew ratings (the ability to fly a particular family of aircraft), and the accuracy of sophisticated reservation and yield management systems. The results of Berge and Hopperstad's study indicated that utilizing aircraft with 5% fewer seats (using the D^3 assignment concept) produces as much revenue as the full complement of seats without D^3 . Achieving this 5% reduction in seats is of course attributable to using smaller aircraft (based on forecasted demand) which results in a capital expenditure savings of 3% (Berge and Hopperstad, 1993).

At the time of the study, only two airlines: KLM (Royal Dutch Airlines) and Australian, regularly swapped aircraft assignments in response to demand variation (Berge and Hopperstad, 1993). Fan's (2002) study also demonstrates the impact of optimising aircraft fleet sizing, based on fare, demand, and operating cost distribution. His findings indicated a 0.6% - 21% improvement in operating profit. He also found that the profit maximizing framework tended to recommend smaller aircraft size in view of sharply decreasing expected marginal seat revenue. It should be noted that standardized fleets offer big savings in purchase, maintenance, parts, and operator training.

2.4.5 Inventory Control

Inventory control plays a vital role in any yield and revenue management system particularly because of the fact that, in the service industry, the unit of inventory is perishable and the total inventory cannot be easily increased as demonstrated by the previous section. The airline which takes off with an unfilled seat cannot recover lost revenue for that seat at a later time. Similarly, an unused hotel room cannot retroactively be assigned guests. As a result of the perishable nature of the inventory unit, firms must have a fast and efficient method of matching inventory with prices in the most profitable manner.

Additionally, units of inventory cannot be easily produced. For example, it would be infeasible to fit another seat on a plane or to build another room in a hotel on a moments notice. As a result; inventory is relatively fixed. However, in some cases the existing inventory could be reconfigured. For example, Scandinavian airlines have a system whereby a partition could be moved between the first class and economy sections

on a plane to increase the number of first class seats available. One of the major drawbacks here of course is the seat comfort level. Other drawbacks are the differences between the amenities available in the first class section such as an in-built TV, cup holders etc. and other sections.

Another issue arising with the role of inventory control in yield and revenue management systems is the availability of inventory in certain classes. At issue is whether a business can ever turn down a request for a full-price unit if discount units are available (Weatherford and Bodily, 1992). They found that the business may deny service to the full-price customer by using strategies such as first-come first-served or an auction to get people to give up their reservation. But in the airline industry, the counter argument to the above is: selling too many seats at low fares however, can cause per-passenger revenues (yields) to decrease, and diversion of high-fare passengers to more readily available low fares can lead to lower total revenues (Belobaba, 1987a). One solution to this situation on an airline and similarly in a hotel is nesting of the fares or prices.

In a nested system the fare classes are structured in such a way that all seats are always available to the higher fare passengers. That is to say, a higher fare class request is never refused as long as there is any seat available (Murtagh and Mitra, 2002). While this may seem counter-intuitive, there are customers who would require more than standard amenities usually available with an economy or standard unit of inventory. For example, business travelers may require the use of an Internet connection or a fax machine on-board an airline or in a hotel room, facilities which may not normally be available with a standard seat or room.

The reverse scenario, whereby a request is made for a discount unit and only full-priced units are available poses much more of a problem. Accepting every request on a first-come-first-served basis would lead to the early purchase of all lower priced inventory units. In the airline industry, and especially in the hotel industry, managers need to be extremely careful in their focus on the strategic objectives of their firm. In the hotel industry, occupancy is often used as the barometer of performance, but maximizing occupancy does not imply maximizing revenues.

This is where approaches such as the mathematical or the human factor approaches can be utilized to deal with the issue of overbooking and inventory control. A good manager should be able to, in conjunction with his or her yield management system, identify the type of current demand and determine if it is worthwhile to fulfill that demand as opposed to waiting for higher revenue to be gained from future expected demand. Supplementing this point is the statement by Orkin (1988) that, “the displacement of future, high-rate transient reservations by groups that are paying lower rates must be a carefully considered decision.”

Inventory control by itself is an extremely extensive field and a complete review is beyond the scope of this dissertation. Pricing has also been extensively studied. These approaches vary from mathematical methods developed by Feng and Gallego (1995) who determine the optimal timing of a single price change from a given initial price to either a given lower or higher second price; to Dana (1998) who demonstrates that customers are heterogeneous in both their valuations and demand uncertainty. Dana also shows that the pattern of advance purchase discounts is such that customers with low valuations are more likely to buy in advance.

More recently, You (1999) addresses the problems facing an airline as to what are the suitable prices for the opened booking classes and when to close those open booking classes. He developed a dynamic pricing model as a discrete time stochastic process and showed that the booking policy can be reduced to a set of critical decision periods which eliminates the need for large amounts of data storage.

2.4.6 Booking Limits and Policies

The objective of a revenue management system is to set booking limits at different levels of control to maximize revenue. First, a forecast of demand is made, and then optimization is performed to calculate protection levels (Zeni, 2001). These protection levels are the numbers of seats that are prevented from being sold to lower fare-paying customers. For example, if it is forecasted and then optimized that a total of 40 business class seats are expected to be reserved, then the protection level for the business class would be 40. A problem then arises if more than 40 reservations are requested. Certainly the airline would not reject demand for its highest fare-paying class. Therefore, the booking limits are nested. Nesting of the booking limits is whereby any seat at a lower fare is automatically available to be sold at a higher fare (within the same cabin class). Thus, for example on a 90 seat capacity aircraft, even though the booking limit for the highest fare class may be 40, all 90 seats are available to be booked at the highest fare in the event that they are needed. This occurs with all the fare classes descending down to the lowest.

Airline seat demand falls into 2 categories: unconstrained and constrained demand. Unconstrained demand is only observable in the absence of any constraints such

as rate controls, stay controls and capacity limitations (Liu et al., 2002). It could therefore be argued that no airline is actually free of constraining since the booking limits serves as censor points. However, these limits are necessary since if there were none, then all reservations would be accepted. This in turn implies that the aircraft capacity would vary as the demand fluctuates. It may be greater than is optimal for that particular market, in which case, there is unused capacity and lost revenue. Or, it may be lesser than is optimal for that market, in which case, the flight may be always sold out. Revenue is then lost by having to turn away passengers (who may be willing to pay a higher fare) because of a lack of capacity. This aircraft capacity allocation problem is indeed a very complex one and has only been addressed dynamically by very few researchers as illustrated in an earlier section.

Brummer et al. (1988) uses the unconstrained total bookings in their distribution modeling. Wickham (1995) provides a method for estimating the unconstrained bookings for the closed booking classes because their limits have been reached. Hopperstad (1995) in contrast to Wickham's (1995) deterministic method, develops a probabilistic approach for obtaining the unconstrained booking class demand which is constrained by the booking class limits. Belobaba and Farkas (1999) utilizes the probability density function of the unconstrained demand to calculate expected spill rates.

Wickham (1995) also tested the effect of constrained versus unconstrained data under a couple of different test scenarios. Skwarek (1996) performed a simulation study comparing the performances of one airline using an unconstraining method versus the performance of another airline using no un-constraining. He found that with a demand factor of 90% (measured as: forecasted demand/aircraft capacity), the airline that did not

unconstrain demand achieved 3.5% less revenue than the airline which did constrain demand. When the demand factor increased to 120%, the revenue difference increased to as much as 50%.

Since there are many areas in which demand forecasting is used by the airlines (for example, revenue management, galley or kitchen requirements, and the scheduling department) it is best for all of them to use the same unconstrained data set. The primary reason for this is that there are very few forecasting models which explicitly deal with censored data (Zeni, 2001).

2.4.7 Overbooking

Historically, overbooking research has been the foundation for the study of yield management. According to Davis (1994), overbooking is a conscious attempt to balance the increased revenue obtained from flying a full aircraft against fewer, but higher paying customers. As a result, some sectors of the hospitality industry mistakenly measure or use capacity as being directly proportional to revenues. This can easily be shown mathematically to be false. The objective of most of the early quantitative research on airline overbooking was to control the probability of denied boardings within limits set by airline management or by external regulating bodies (McGill and Van Ryzin, 1999). This is very similar to the concept of service levels in inventory management with the service level being defined as the complement of the probability of a stockout. Thus, for example, if the probability of a stockout were 5%, then the service level would be equivalent to 95%. In making this comparison, denied boardings (unavailability of a seat – the unit of inventory) would be the same as having a stockout.

If forecasting is to be compared to the backbone of yield management system, then overbooking could be thought of as the skeleton. Typically, most airlines open a flight for bookings almost one year before departure, and during that time useful reservation data is accumulated (Zaki, 2000). Plotted over time, these reservation data constitutes a booking profile and actual booking patterns are compared to forecasts. If the forecasted demands were high then the overbooking limits could possibly be high also because there may be a greater chance of cancellations or no-shows but improving the demand forecasts does not necessarily imply an improvement in the overbooking decision since it is primarily dependent upon the no-show rates and the overbooking model chosen. Figure 2 below illustrates the relationship between the forecasting model, the overbooking model and other components of the booking process.

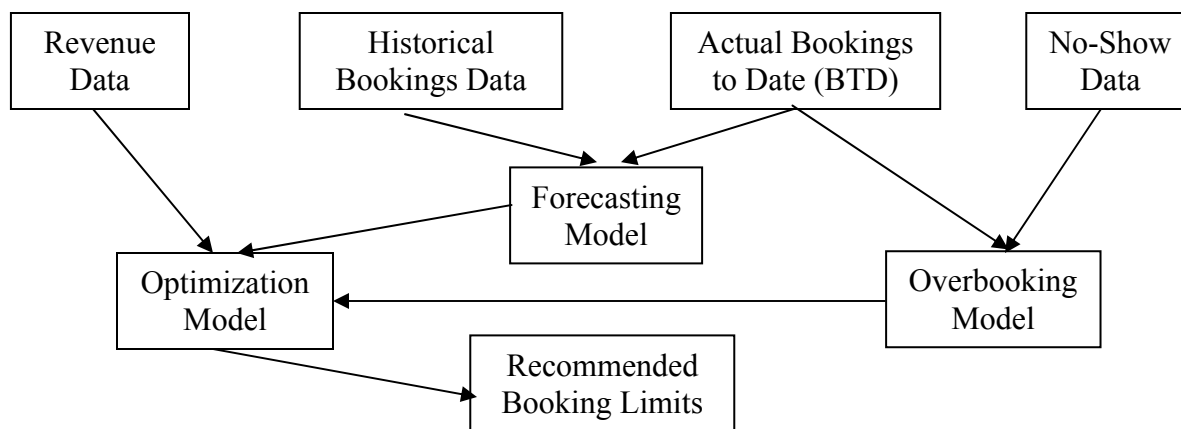


Figure 2. The Automated Booking Limit System

Overbooking is not illegal and airlines originally overbooked their flights according to some fixed standard, for example, one overbooking per 2000 reservations. Initially, it was termed “controlled overbooking” and wasn’t even acknowledged in many cases. Airline overbooking is based on the argument that since the passenger has the option to cancel or even to “no-show” without economic penalty, the airlines have the

compensatory right to take more reservations and sell more tickets than there are physical seats on the aircraft (Rothstein, 1975). This of course was in the era of regulation and the airlines' argument is now modified in such a way that the penalty imposed for a cancellation or no-show varies according to the fare class. The consequence of the decision of overbooking however has generally remained the same, namely, what compensation should be provided to a passenger with a reservation who has been denied boarding or has been bumped?

From an airline or hotel point of view, as previously mentioned, the incremental or marginal cost of an additional customer is so small as compared to the additional or marginal revenue that it is no wonder these industries are willing to gamble on overbooking. Most of the early work on overbooking was based on static one-period models. There has been some relatively recent work by McGill (1989) who presented a static formulation of the overbooking problem and shows that it is a special case of a general model of the two fare class seat allocation problem.

The issue of overbooking and its consequences can be addressed in many ways. In the airline industry, travelers can, simplistically, be separated into two categories: those with urgent travel needs and those who are primarily concerned with their ticket price. Inferentially, those travelers with urgent travel needs would be more willing to pay a higher price for the convenience of their choice of flight. If a flight is oversold, that is more people show up for the flight than there are physical seats, then the airline will look for volunteers to give up their seats to those with more pressing travel needs.

The major consequence of overbooking then arises as to what is "fair compensation" to those passengers who have voluntarily given up their seats or who

involuntarily have been bumped. Passengers are involuntarily bumped or denied boarding if there are no volunteers to give up their seats on an oversold flight. The airlines' bumping procedure is of course directly related to the purchase price of the ticket. Those customers with lower priced or discount tickets would inevitably be the first to be bumped. Calculation of this "fair compensation" is rule based and varies by airline; length of delay and even in some cases the demeanor of the passenger.

Another consequence of overbooking is the cost of a "walked" customer. A "walked" customer, the technical term for a lost customer, is used by the industry for customers not accommodated due to a lack of an inventory unit. In the airline industry, the cost of walking a customer may sometimes be as high as 200% of the original price of the ticket. This may include the fact that the airline has to make alternative travel arrangements for the inconvenienced passenger and accommodations for any interim time. A number of overbooking models have been proposed as solutions to walking a customer. For example, Hadjinicola and Panayi (1997) propose a simple overbooking technique which selects an overbooking policy based on minimum expected loss of revenue for hotels with multiple tour operators.

Even if some monetary compensation is agreed upon, there always exists the problem of goodwill costs. This type of costs is extremely difficult to measure as it incorporates tangible as well as intangible aspects of walking a customer. The tangible aspects could be measured by the amount the airline has to pay to compensate the customer and any other costs which is the agreed upon monetary compensation. However, the intangible aspects such as the inconvenience to the customer in terms of the customer's time and expectations and word of mouth advertising are very difficult to

measure. Numerous marketing and economic studies have been done to arrive at some values for goodwill costs, but to date, no conclusive evidence has been obtained.

There are two types of overbooking models normally used in the airline industry. The first is a basic econometric model which sets the overbooking limit at the point where marginal revenue equals to marginal cost. The second and more popular method is based on a desired Level of Service (LOS) model similar to that used in determining safety stock for inventory management. This method will be used where applicable for this research.

The LOS method works on the following principle: 1) the airline pre-determines the percentage of denied boardings that they are willing to accept (for example, 3% denied boardings), 2) given a historical no-show rate (percentage) for similar flights, 3) combine these two rates and accept reservations up to the capacity of the aircraft plus these combined rates. For example, given a historical no-show rate of 5% for a particular fare class and that the airline is willing to tolerate 3% denied boardings for that same fare class, and for a capacity of 50 seats in that fare class, the airline would accept up to 54 reservations: $([1+0.05+0.03]*50 = 54)$.

Bodily and Pfeifer (1992) proposed simple optimal rules for overbooking that are well matched to the data available in service firms such as the airline industry. However, the majority of work has been done in dynamic or multi-period modeling. For example: Rothstein (1971) constructs a general model for determining airline overbooking policies; Hersh and Ladany (1978) present a model for finding the maximum allowable allocation of seats on the various spans at each time prior to departure such that a maximum expected contribution to profit for the flight is obtained; Alstrup (1986) develops an

overbooking model for a fixed non-stop flight with two types of passengers which treats the booking process as a Markovian non-homogenous sequential decision process; and Wollmer (1992) presents a model of when to refuse booking requests for a given fare level to save the seat for a potential request at a higher fare level.

There have been many mathematical models proposed on overbooking research including work done by many notable researchers such as Alstrup et al. (1986) who formulated a two-dimensional stochastic dynamic programming model whose objective was to minimize the difference between the maximum gain obtainable and the actual gain; Belobaba (1987a) who incorporates probabilistic demand levels into the seat inventory control problem, and Chatwin (1998) who gives conditions (for example, the booking limits fall toward flight-time) under which the booking limit policy is optimal. In addition, Subramanian, Lautenbacher and Stidham (1999) formulate and analyze a dynamic programming model for seat allocation which accounts for cancellations, no-shows and overbooking.

The research on overbooking still forms a core component of yield and revenue management studies as can easily be seen by the continued production of academic articles. In the literature review, there has been no evidence of any comparative study on measuring the amount of overbooking or the compensation levels in the airline industry versus the hotel industry, but anecdotal evidence would indicate that overbooking is much more of an issue in the airline industry than in the hotel or any other service industry where yield and revenue management is practiced. Even though someone who cancels or no-shows for their flight may also cancel their hotel reservation, it is entirely possible for them to have alternative travel arrangements to get to their destination.

Additionally, the walk-up rates for the hotels may be greater than for the airlines simply because of the elasticity of demand.

2.4.8 Demand Distributions Modeled

Various distributions have been used to model the demand for reservations. The three most commonly used are the: Truncated Normal, Censored Poisson, and Adjusted Gamma. It is important to evaluate the performance of the [forecasting] methods across a variety of distributions. Otherwise, conclusions may be drawn with regards to well-behaved data that is not a fair representation of reality (Zeni, 2001).

The earliest known modeling of the demand distributions was by Beckmann and Bobkowski (1958) who compared the Poisson, Negative Binomial and Gamma models to determine that the most reasonable fit was offered by the Gamma distribution. Shlifer and Vardi (1975) determined that the normal distribution provides a good continuous approximation of the distribution of demand. Harmer (1976) also utilizes a normal demand distribution for his research into flight spill rates. In his 1987 research, Belobaba utilised the normal distribution as his choice of a demand density function. Customer demand during a schedule period (for a given flight leg) is assumed to be homogenous when modeling airline fleet assignment and researchers such as Soumis et al. (1981), Abara (1989), and Hane et al. (1993) all utilise this assumption in their models.

Lee (1990), Subramanian et al. (1994) all represent the demand as the normal distribution in their models. Berge and Hopperstad (1993) assume that the booking demand is normally distributed but truncated at zero. Gorin (2000) also uses the normal distribution to model the demand pattern. Zeni (2001) found that the normal distribution

provided a better fit for the high demand markets. In a small empirical study, Brummer et al. (1988) proposes that the underlying demand for bookings is log-normally distributed. This is unique in the sense that the log-normal distribution offers several advantages. These include the following properties: it is flexible and can take the skewness of airline data into account. Also, it is truncated at zero, and it can easily be transformed into the normal distribution by taking the simple logarithm of the data.

In contrast, Rothstein (1971) used a truncated Poisson distribution to model customer arrivals. Williamson (1992) as well as Belobaba and Weatherford (1996) also uses the Poisson distribution to model customer arrivals within each booking period. Belobaba (1985) extensively tested observed airline data to determine demand distribution patterns. His conclusions, slightly modified by more recent results (Lee, 1988) are that significant positive skewness of the booking distribution may occur at low levels of demand, a spike in the distribution may occur at high levels of demand, and an assumption of normally distributed booking data seems to hold for moderate levels of demand (Lee, 1990). Lee (1990) also effectively demonstrates that the booking process can be modeled as a censored Poisson process with a modified request rate.

Smith et al. (1987) from American Airlines found that the Gamma distribution was the best choice for the demand distribution. Smith and Penn (1988) also from American Airlines, reiterated these findings when they found that the underlying demand for airline bookings is Gamma distributed. In his 2001 research, Zeni reported that the Gamma distribution could not be rejected and that it provided a slightly better fit for low levels of demand in both business and leisure markets. There are two distinct advantages offered by the Gamma distribution: it has two parameters as compared to the one

parameter for the Poisson distribution, and depending on the value of these parameters; the variance could be greater than or lesser than the mean of the demand. However, there are also two disadvantages: one is that the Gamma distribution is continuous while the bookings are discrete, and the other is that the Gamma distribution is undefined at zero.

McGill and Van Ryzin (1999) have identified the following environmental parameters of customer behaviour which influence demand patterns:

- **Seasonality, day-of-week variation** – airline travel is highly seasonal and there is also day-of-the-week variation patterns. These patterns are especially more evident between city pairs where there are business centers and where business travel is a large component of certain flight times.
- **Special events** – special events such as a conferences or festivals in another can temporarily affect the demand patterns for airline travel. Again these events can be identified beforehand and eliminated from the data analysis and forecasting.
- **Sensitivity to pricing actions** – airline pricing is highly price elastic for the same flight legs but there are different price elasticities for different fare classes. Thus, if the competitor slashes prices or offers a low fare for a particular O-D pair then all other airlines in that market must be prepared to follow suit. To illustrate these reactionary fare changes for example, in 1989, there were 30,000 daily fare changes in the US domestic airline industry alone (Williamson, 1992). As a result, forecasting for particular sale days becomes more difficult.
- **Demand dependencies between booking classes** – the theoretical assumption that demand for one booking class is statistically independent of another is usually not satisfied in practice. To accurately determine this, the recapture (the re-

booking of a passenger who is unable to get a reservation of choice onto an alternative flight with the same airline) and defections (occurs when a passenger with a confirmed reservation switches to a flight with another airline for reasons of convenience) probability must be correctly modeled. While the recapture probability could be determined with some certainty, the defections probability requires the much more complicated determination of the probability distribution for departure delays.

- **Diversions** – occur when for example, a business traveler has sufficient advance notice of a trip which enables them to purchase a lower priced or discount ticket. To counteract diversions, airlines usually impose restrictions on the purchase of discount tickets. But knowing if a business traveler would have purchased a business class ticket instead of a lower price discount ticket involves an enormous data collection exercise which is beyond the scope of this study and has not been done so far.
- **Go-Shows** – are similar to no-record passengers but are allowed to board because there are seats available. No-records are passengers who appear at flight time with a valid ticket (such as an open date ticket) but for whom there is no reservation record. This variable is difficult to model as it would involve having data as to who exactly boarded and not only those who had reservations.
- **Group bookings** - could be considered as separate or individual bookings all arriving at the same time. If the interval is further delineated, these arrivals could be thought of as incremental arrivals. As such, they are not explicitly considered as group bookings since this only complicates any analysis.

- **Recaptures** – are defined previously, but it is also possible for passengers to travel on an entirely different flight than they originally intended.
- **Upgrades** – work in two ways: bumping up to a higher fare class of a passenger who may have been otherwise denied boarding, and the choice by a passenger to pay a higher fare than originally planned if their original fare requested is sold out. This parameter is influenced by the dynamic nature of bookings at flight time and dependence on no-show probabilities.
- **No-shows** - no-shows are passengers who fail to show up at the time of departure. This does not allow the airline to re-book or resell this seat through the normal reservation channels. The only way the airline could earn additional revenue from this seat is by means of a walkup or standby passenger or by overbooking. No-shows are very common amongst full fare passengers whose tickets are usually without any penalty conditions. The standby factor would require knowledge of the exact number of passengers who boarded for each fare class.
- **Walkups** - walk-ups are passengers who show up at flight times without reservations but wish to travel on a particular flight. Additionally, in some instances, the term no-record is used to describe walk-up passengers with reservations, but for whom the airline has no record. Both of these cases affect the revenue of the airline. Again, this parameter requires knowledge of pre-boarding data which is usually unavailable. For simplicity, a walkup rate of zero percent is usually assumed unless it is explicitly being modeled.

After considering the properties of each probability distribution, it becomes clear that the “best” choice depends on the characteristics of the particular sample of booking

data under consideration (Lee, 1990). The process in determining the best fit usually involves some exploratory data analysis. This may include the Chi-Squared goodness of fit test which is used to choose a specific distribution. Other tests include the Anderson-Darling, Cramer-von Mises, and the Kolmogorov-Smirnov tests. In-depth regression analysis is then done and a visual inspection of residual versus fitted value plots are examined to determine any significant deviations from the selected distribution.

2.4.9 Cancellations, No-Shows and Walk-Ups including Stand-Bys

One of the biggest issues facing yield and revenue management is how to deal with cancellations, no-shows and walk-ups. Cancellations, no-shows and walk-ups are a major issue because it affects and determines the overbooking policies and is responsible for a significant portion of costs to the system. While the consequences and mannerisms of cancellations and no-shows may be the same in that the customer does not use the unit of inventory, their management is somewhat different. Whether the original booking is cancelled prior to departure or the passenger simply fails to appear at departure time, the outcome from the airline's perspective is the loss of a revenue passenger for that particular flight (Belobaba, 1989b).

Cancellations are returns or changes in bookings that occur early enough in the booking period to permit subsequent rebooking through the reservation system (McGill and Van Ryzin, 1999). Factors such as schedule changes, price changes, and sale pricing all affect cancellations. And as Chatwin (1998) explains: "the number of cancellations depends only on the number of reservations on hand, and not, for example, on the length

of time for which each reservation has been on record.” This, he attributes to the forgetfulness property of the distribution of the cancellations.

Passengers cancel their airline reservations and hotel reservations for a myriad of reasons. These reasons range from personal to business to acts of nature (such as bad weather) to unforeseen circumstances. Rates of no-shows and cancellations will vary according to the period and market segment. Unless these no-shows and cancellations are accounted for, the level of forecasted business will be unrealistic (Jauncey et al., 1995).

Another factor affecting the number of forecasted reservations and show-ups is passengers who are termed no-shows. No-shows are passengers who fail to show up at the time of departure. This does not allow the airline to re-book or resell this seat through the normal reservation channels. The only way the airline could earn additional revenue from this seat is by means of a walkup or standby passenger or by overbooking. Unlike cancellations, passengers who are no-shows are much more of a revenue problem to airlines and hotels. No-shows are usually most common amongst passengers who have full-fare tickets that are fully refundable under most circumstances.

Conventional forecasting methods predict the number of no-shows using time-series methods such as seasonally weighted moving average of no-shows for previous instances of the same flight leg (Lawrence, 2003). Since some customers derive benefits from holding reservations that go unused (i.e. no-showing), for example, to provide for flexible travel plans; an efficient penalty (one commensurate with the external costs that no-shows impose on others) will generally not reduce the no-show rate to zero (Snow, 1982). Flight delays and cancellations may also cause passengers to be no-shows on connecting flights for example.

In the literature, empirical data collected on no-show probabilities (p_N) are in the range from 3% - 9% for the leisure class and about 4% - 14% for the business class (Alstrup, 1986). It is well known that the business class or first class fares generally have much higher no-show and cancellation rates than the economy classes simply because of the volatile nature of their travel. Also, combined with this volatility, is the fact that these business or first class fares are usually fully refundable. Most studies however, acknowledge this fact; but at the same time, assume that the no-show probabilities are the same for all fare classes.

No-show probabilities would of course vary considerably amongst various airlines and various markets. Toh (1979) calculated a 23.6% no-show rate amongst full-fare passengers in what is the first available data for that statistic. Alstrup (1989) found that the typical range in Europe is five to 20 percent, and for US carriers, experience suggests a no-show range of 15 to 30 percent. These relatively high no-show rates could possibly be attributed to non-deterrent penalties at that time.

In contrast, Snow (1982) calculated the no-show rate to be approximately 11%. Ringbom and Oz (2002) state that in the US, the average no-show rate is between 10 and 20 percent. Also, in concurrence with Alstrup's 1989 findings, they indicate that these figures are lower in Europe. In his 2003 study, Lawrence found no-show rates averaged around 10%. Toh and Raven (2003) note that the no-show rate may exceed 20% in some markets (for example, Las Vegas). Suzuki (2006) estimates around a 10% - 15% no-show rate, leading to an average of 12.5% for all travelers. No-shows are problematic for the very reason that the airline or hotel has no chance of rebooking or utilizing that lost customer revenue unless there are walk-ups or stand-bys.

Walk-ups are those travelers who usually arrive with no reservation but wish to purchase one or travel on a particular flight. Very often they are business travelers who are forced to utilize a particular service for the very same reasons (such as those mentioned above) that another traveler may use to cancel his or her reservation. Additionally, in some instances, the term no-record is used to describe walk-up passengers with reservations, but for whom the airline has no record. Both of these cases affect the revenue of the airline.

In the first instance of passengers with no reservations but wishing to travel, they represent the elite income group for the airline. In most cases, these are business class or first class travelers; but in some cases it may be passengers with travel emergencies. These emergencies may be bereavement situations, medically related, or even military personnel for whom discounted fares are available.

The second instance represents a more worrisome situation for the airline. It involves either: displacing a currently booked passenger, or rejecting the “no-record” reservation. In both cases an appropriate compensatory amount has to be decided to reward the “denied boarding” passenger so as to possibly recoup future expenditure from them. For US airlines, these compensatory amounts typically average around \$300-400 up to a maximum of \$1000 (Suzuki, 2002). No-records could also include passengers who have bought an “open date ticket” and thus have no reservation for a particular flight time. For simplicity, they will be treated in a similar manner to a stand-by or walk-up passenger.

It may be more prudent from a customer service point of view to reject the Type 1 walk-up passengers (those with no reservations) than the Type 2 passengers (those with

reservations but no-record) since the Type 1 passenger may be more of a “mercenary” type buyer. In the only statistic found to date, Toh (1979) utilizes Pre-Post Departure Records and determines a 5.5% average Type 1 walk-up rate.

Even if there are successful implementations of policies controlling the no-show rates, the walk-up rate could still be expected to remain in the same range as is traditionally observed. This could be attributed to random factors and transactional errors. Also, it is presumed that since the walk-ups fall into either Type 1 or Type 2, these passengers will want to board provided that there is capacity available. This is usually advantageous to the industry because this segment could be charged a premium price. Another segment that could also be charged a premium price are stand-bys.

Stand-bys are those categories of travelers (e.g., youth or military service personnel) who are willing to wait for a flight that would otherwise depart with empty seats, and who may or may not have an existing reservation but wish to utilize a particular service time. The amount of forecasted cancellations, no-shows and walk-ups all go into determining the overbooking policy. Therefore, cancellations, no-shows, and walk-ups are all vital issues to any yield and revenue management system.

2.4.10 Information Requirements

The biggest problems facing a firm contemplating adoption of a yield management program are data availability and accuracy. A completely computerized management information system is not required; what matters is the type and accuracy of the data collected (Kimes, 1989b). Conversely, in a survey carried out by Donaghy et al. (1997), seven out of eight hotel managers interviewed believed they had adequate

software for storing and sorting data but noted that the efficiency of their yield management systems was severely inhibited as integrated software was not available to handle all the requirements of yield management.

Customer information is a precious company asset (Relihan, 1989) and utilizing a previous analogy, if forecasting is the engine for yield management and overbooking is the accelerator, then without question, information is the fuel. In a demonstration of the power of information, the Walt Disney World studied guests' stay-patterns at their Orlando resort hotels and altered their reservation policy by imposing minimum-stay restrictions during certain periods. The result was an increase in room revenues of over \$1.5 million in the first week of January after New Year's Day – a traditionally low-occupancy week. Disney was not paralyzed by the fact that it did not have perfect information or that it was not in a position to maximize revenue with “optimal” inventory controls. Instead, Disney's revenue managers made use of the information available to them to achieve a measured improvement in revenue while minimizing their risk (Quain et al., 1999).

One of the major shortcomings of yield and revenue management systems which is only now being rectified is failing to keep or obtain information on lost customers. These lost customers who may be cancellations or no-shows offer vital information for use as input into the forecasting and overbooking models. Identifying and analyzing the various reasons for cancellations and no-shows would definitely enable more accurate forecasting and overbooking controls, which in turn enables better yield and revenue management.

There is no question that information technology (IT) and yield management software can sort, categorize, organize, and present data with speed and precision (Jones and Hamilton, 1992). But, as illustrated in an earlier section, this is just the mathematical aspect of the yield management system. Accordingly, information should only be used as a tool to aid in, and not, as the sole decision making criteria.

Information plays a crucial role in any yield management system but the manner in which it is utilized goes a long way in determining the success or failure of the system.

2.4.11 System Capabilities

The following sections identify and discuss the system capabilities for a qualitative perspective. While revenue management is certainly a quantitative field there is an overlooked qualitative side to it.

a) Human Interaction Concerns and Issues

Even though the terms yield management and revenue management are associated with the mathematical techniques of increasing or improving yield, there is a much under-researched human side. Forecasting or the prediction of future demand is the linchpin in yield and revenue management, but it is the consumer as well as the service provider's behavior that predicates the demand patterns. There have been many models for the various demand distribution processes ranging from the Poisson to the Negative Binomial to the Gamma distributions. There have also been many similar models explaining the arrivals processes ranging from homogenous Poisson processes to

stuttered point Poisson processes to general point processes, but all of these hinges upon human behavior.

Pricing could be used to stimulate consumer behavior, which in turn affects demand and forecasting. However, the non-mathematical approaches to pricing, from the role played by the service provider to the roles played by the service receiver or the customer have not been totally investigated. The following discussion outlines what has been done so far on both fronts and the collective impact on the psychology of yield and revenue management.

b) The Role of the Service Provider

“Yield and revenue management requires knowledge of guests’ expected behavior, plus an understanding of which business is beneficial to a hotel – but it does not require high-power computers” (Quain et al., 1999). This statement serves to partially nullify one of the more common myths that yield management is a computer system. Among the many problems facing a firm practicing yield and revenue management, factors such as a decrease in employee morale, customer anger and confusion, and a shift in competitive strategy can be attributed to the role of the service provider.

Resistance to change has been well researched and documented at various levels and it has been found that employees fear replacement by computers or technology to an extremely high degree. As a direct result of this perceived job threat, the employee will harbor some resentment and may be less productive or distrust the yield management system. The employer should reiterate the fact that the yield management system is a tool

and not a substitute. Simply put, yield management tools do not replace employee decision-making and control (Lieberman, 1993).

As was previously illustrated in an earlier section, consumers became irate upon learning that the soft drink giant Coca-Cola planned on using yield management to raise prices when the weather became hotter. Since airlines have been practicing yield management for a relatively long time, it has become an accepted practice, but in industries where there is much more competition – even within the hospitality industry – customers will simply take their business elsewhere. For example, in certain areas where hotel or motel density is high, switching a service provider may be as easy as walking across the street.

The airlines have succeeded in utilizing yield management techniques effectively due to an education campaign. Customers are made aware of the fact that making a reservation two weeks in advance entitles them to some form of discount and for those who wait until the last minute; expectation of a higher price is implied. This argument could have complications when dealing with regular customers who become accustomed to a certain pricing levels and even discounts. Therefore, even when demand is unusually heavy for an upcoming date, it is unwise to refuse a loyal customer's request for a member's discount rate as that guest represents future sales (Lieberman, 1993).

Many critical decisions are made by a casual continuation of past practices, rather than an analysis of and choice for the future (Orkin, 1988). Various researchers (Kimes, 1989a; and Jones and Hamilton, 1992 among others) have demonstrated that companies often put short-term profits ahead of long-term goals such as customer service as an unintended consequence of using a yield management system. A company's strategic

positioning may be such that it is the lowest-cost provider in its industry. However, in utilizing a yield management system, it may be determined that revenue or yield could be improved by increasing prices all round. It is in such a scenario that managerial experience plays a pivotal role because if the service provider were to increase prices, then they would have deviated from their strategic objective of being a low-cost provider.

Thus, utilizing any yield or revenue management system needs to be carefully balanced against the strategic objectives of the company. This balance could be accomplished by human intervention.

c) The Role of the Customer

Customers also play an important part in yield and revenue management. In fact, they are the only source of revenue. Consumer behavior has been analyzed and researched extensively, a review of which would be beyond the scope of this dissertation. (For a most recent review, see *Consumer Behavior – A Literature Review*, by Pachauri, 2002). An important aspect of any yield and revenue management system is its overbooking model. The formulation of this model determines the booking limits for each time period before the service usage that maximizes expected revenue, such that allowances are made for the dynamics of cancellations, subsequent reservations and penalties for oversold inventory. As a result, understanding consumers' behaviors is an integral part of this process. The extent to which customers use the services they have paid for determines whether they will repeat the purchase (Gourville and Soman, 2002).

The behaviors which have the most impact are those of no-shows and cancellations. Customers may no-show or cancel for many different reasons. Very often

airlines and hotels find it difficult to obtain the reasons for no-shows or cancellations, but every effort should be made to collect information as this would greatly enhance the accuracy of the overbooking process. The most costly scenario to an airline or hotel is when a customer who has a reservation is denied an inventory unit because of overbooking. Measurement of this denial of service cost has not been done effectively by any researcher.

Customers believe that value to the firm should equal the value to the customer (Kimes, 2002). If the firm arbitrarily raises prices with the intention of generating revenue, this will be viewed as an unfair business practice and the fallout from loss of business could be significant. An educated or informed customer would be much more desirable than one who sees no justification in the differentiation of a service. If it can be justified to the customer the reason for the price differential in providing the same service but with different conditions then the usage of yield management from the customer's point of view is satisfied. However, in some cases achieving the balance between the asking price and the selling price often comes down to the knowledge and experience of the reservation clerk.

Customers usually have some reference transaction by which they measure the value of the service they are about to receive. The customer views this reference transaction as such that there is a standard inventory class – economy class on an airline or a standard room in a hotel – for a standard price and that price could be negotiated. If the perceived difference or actual difference between the reference price and the asking price is unsatisfactory, then that customer will continue to shop around. They are willing to trade some restrictions for a discount of some form but, too severe restrictions will

result in lost transactions; too little restrictions will result in the service provider losing revenue.

It is therefore imperative that the customer be made to understand that some “rack rates” are non-negotiable due to current demand or forecasted demand. This shapes the customer’s behavior so that rather than “walking” or totally losing that customer, the service provider might in fact be able to sell a higher cost hotel room or airline seat simply by the human art of negotiation.

d) The Psychology of Revenue Management

When they were asked what were the problems regarding yield management, nearly every manager responded with issues relating to people (Jones and Hamilton, 1992). Five years later in 1997, Yeoman and Watson again asked a similar question, why do yield management systems fail? They attributed it to people failures of two types: errors and violations. Errors were defined as mistakes in judgment and violations were defined as acts which were designed to obstruct operating procedure.

Since yield management is a complex tool involving both mechanical (or computerized) and human systems, understanding the psychology or the workings of the human aspect is also of some relevance. Yield management systems should not be designed in such a manner as to remove total decision-making actions from employees. As mentioned in a previous section, this can cause alienation of the employees and lead to an inefficient use of the system. Instead, a yield management organizational culture should be established. The organization should be structured in such a way that every employee has some input into the process.

The yield management system should be such that it is an essential part of every employee's work and that they can see the results or outcomes of their decision-making aided by the computerized part of the system. For example, at the end of the month or the end of the year a report could be compiled showing where and how the employee impacted the revenues for the system. This could be helpful in determining compensation. Employees are always striving for some type of recognition be it financial or psychological; and very often, if they feel that they are under-appreciated, productivity will drop.

In the hospitality industry a frequently overlooked aspect of yield management is ancillary spending or the bundling of services. Luring guests or travelers by the use of incentives such as package discounts or so-called "free extras" has always been a staple of the industry. Everyone likes to think that they are getting "something for nothing" and it is with this in mind that a whole new market segment and consequently new demand could be exploited. Transient or infrequent users could be enticed to boost the revenues or yield of a particular firm during slow or low demand periods by the use of this particular strategy. For example, offering a free \$100 gift card to be used on affiliated merchandise or in affiliated locations could boost revenues because it ultimately has the effect of keeping the money "in-house."

Since yield and revenue management calculations are normally optimized based upon the inventory unit price only and exclusive of ancillary services, bundling it with the ancillary services could increase sales of higher-priced inventory units. In effect, the cost of the ancillary services which may be breakfast in a hotel or a movie on an airline, are essentially buried by the cost of the higher-priced service. The marginal revenue

generated exceeds the marginal cost simply because of the psychological satisfaction of getting something for nothing that is thought of by the guest. The add-on service is so attractive that it is purchased with little regard for the open fact that it is not discounted (Weigand, 1999). An ideal example of this are the Las Vegas and Atlantic City casino and hotel industries.

Bartering and bargaining for goods and services is an age-old practice and in some places it is part custom and part culture; and in some sense it could be seen as the precursor to yield management. It centers on the simple fact that the seller wishes to obtain some reward for providing the service while the buyer wishes to minimize the amount of reward offered for the service. Arriving at a compromise is all a matter of psychology as to who outwits whom. Yield management with all of its mathematical techniques and algorithms would be completely useless were it not for the psychological battle between the buyer and seller.

2.5 CONCEPTUAL AND THEORETICAL MODELS

Even though yield and revenue management is undeniably very mathematical in nature, there exist some underlying theoretical principles, concepts and techniques. Of the literature reviews and survey articles done so far, one by McGill and Van Ryzin (1999) offers a mathematical perspective; another by Weatherford and Bodily (1992) offers a comprehensive taxonomy, and another by Kimes (1989b) serves more as a user's guide. However, there are other authors and various other articles, which propose certain conceptual and theoretical ideas which will be discussed in this section.

While mathematical approaches and techniques could be thought of as being in the developmental phase, the theoretical aspects are definitely mature. Weatherford and Bodily in their landmark 1992 study proposed the term perishable asset revenue management (PARM) which combined the areas of yield management, overbooking, and pricing for perishable assets. In their study they offered a comprehensive taxonomy consisting of 14 different elements that can be used to identify and describe PARM type problems.

Their study was highly beneficial in that it isolated the characteristics of any yield management system and the descriptors or the possible values of those characteristics. Their objective was to provide a framework for categorizing future problems in the three areas mentioned above. In addition, they suggested some important areas of future research which can be used to help bridge the gap between theory and practice.

In her 1989 paper, *Yield Management: A Tool for Capacity Constrained Service Firms*, Kimes, similarly to Weatherford and Bodily's study, describes the appropriate situations for effective yield management application and then identifies the components of the problem. Various mathematical solution techniques and specific problems associated with these techniques, and managerial implications of yield management systems are then discussed and she concludes with directions for future research. Among the many points she makes in her directions for future research; integration of multiple-phase services (such as multiple flight legs, multiple night stays), analyzing of demand distributions and correlations, analyzing booking rates over time, and incorporating pricing decisions into the solution methods (many of approaches used assume price is a given) are notable.

Yeoman and Watson in their 1997 paper take a systems theory approach in explaining yield management as a human activity system. They argue that most research has converged on the forecasting and strategy elements, neglecting the people element of yield management. In their paper, they argue that the people element is the cornerstone of any successful YMS and suggests a matrix framework for an organizational, team and individual approach to yield management. Jones and Hamilton (1992) have also studied a similar model with people as an important component of yield management systems. However, they propose primarily, that yield management should be an integral part of the organization's culture.

Desiraju and Shugan (1999) investigate the concepts underlying yield management systems (YMS) and the conditions necessary for YMS to improve profits. They do this investigation in terms of the pricing strategies such as early discounting, overbooking, and limiting early sales for capacity constrained services. Their findings suggest that YMS works best when price-insensitive customers prefer to buy later than price-sensitive consumers. This reinforces an earlier point made in this dissertation about the acceptance of the worth of one more expected higher-paying customer as opposed to accepting a current lower-paying customer. It is worthwhile noting also that their focus was on the strategic applications of YMS on long-term objectives such as profit maximization rather than short-term objectives such as filling capacity.

Reece and Sobel (2000) present a diagrammatic solution to the firm's profit-maximizing price discrimination problem in the face of capacity constraints. This is a conceptual model because no data validation is done and they demonstrate that the

conditions required in this situation are very different from those of the standard model where the firm is not constrained by capacity.

Other authors such as Lieberman (1993) and Harris and Peacock (1995) do not propose any theoretical models, concepts or techniques but instead offer a prescription for success in using YMS. Lieberman identifies ten common myths about yield management and then proceeds to debunk them. Peacock and Harris are much briefer in their paper and simply delineate ten steps for success when implementing any yield management program.

As the existing research shows, the literature on conceptual and theoretical models is somewhat sparse but there exist some common underlying principles and techniques in identifying the concepts and theories used in yield management systems.

2.6 EMPIRICALLY VALIDATED RESEARCH MODELS

There are numerous empirically validated research models in yield and revenue management which warrant discussion. In a previous literature review by McGill and Van Ryzin (1999), articles were grouped according to research in forecasting, stochastic process models for arrivals, overbooking, seat inventory control, origin-destination control, and pricing research.

In the area of forecasting, models were grouped according to their demand distributions, arrivals processes and type of forecasting: aggregate or disaggregate. The various types of models for demand distributions were identified as Poisson, Negative Binomial, and Gamma. Of these, the Gamma distribution was determined to be the best fit for airline data by researchers such as Beckman and Bobkowski (1958), Beckman

(1958), and Lyle (1970). In their 2000 paper, Feng and Gallego, propose a model for PARM with Markovian time dependent demand intensities. The purpose of this model was to determine the optimal timing of price changes.

For the arrivals processes, variations of the Poisson distribution such as the stuttered Poisson, the compound Poisson and the Censored Poisson have been proposed. Researchers such as Alstrup et al. (1986), Lee (1990), Lee and Hersh (1993), and Subramanian et al. (1999) have all developed and solved models for the arrivals process using stochastic dynamic programming on a single leg flight with multiple fare classes. In the hotel industry, Weatherford et al. (2001) present a model for testing aggregation versus disaggregation when used to forecast revenues. Their results indicated that full disaggregation by rate category and length of stay produces smaller forecast errors and that managers should at least track arrivals by rate or price category.

From the literature, it can be seen that overbooking research has the longest history of any component associated with yield and revenue management. Most of the early research by Beckman (1958), Taylor (1962) and Rothstein (1971) focused on the problem of controlled overbooking and on the reduction in the number of denied boardings. Beckman's model yielded the upper bound for reservations which minimized the expected costs. Taylor's model obtained bounds on the proportion of passengers who are denied boardings. All of these models indicated at the time that the prevailing thought was that overbooking should be the focus instead of more accurate forecasting.

Rothstein's 1971 model proposed a new approach to overbooking. He used a discrete time approach with the reservation or arrivals process being viewed as a Markovian sequential decision process. While this model refers to a single fare class and

a single-leg flight, Shlifer and Vardi (1975) proposed a model to allow for two fare classes and a two-leg flight. Recent research, with similar conditions to Rothstein's, has been done by Chatwin (1998), in which he proposes two multi-period overbooking models for a single fare class and a single-leg flight. In these models, he determines the optimal booking limits and shows that they are monotone in the time to flight departure. Monotonic booking limits are defined such that the booking requests arrive in strict fare sequence, generally from lowest to highest as flight departure approaches. However, the existence of low standby fares violates this finding. Subramanian's et al. (1999) research builds upon the foundation of Rothstein's work by formulating a Markov decision process model for seat allocation on a single-leg flight but with multiple fare classes.

Littlewood's (1972) rule mentioned in a previous section was basically the pioneering research in seat inventory control. Bhatia and Parekh (1973) followed by Richter (1982) give derivations of Littlewood's rule. Belobaba (1987) extended Littlewood's rule to multiple fare classes. In the years following major contributions to this area were made, particularly by Weatherford. His 1991 thesis formulates and solves the discrete, generic PARM (Perishable Asset Revenue Management) problem, while his 1993 work presents a model for customer arrivals that allows for evaluation of different decision rules in PARM situations. In 1994 he derived the closed-form analytical results for the expected profits for the single-period inventory problem both with and without advance bookings (akin to single-leg reservations), and in 1996 he presents a new heuristic approach that incorporates diversion and could be used by airlines and other relevant industries to achieve improvements over current methods. Most of the models

presented deal with single-leg seat inventory control and some demonstrate the generalization of the yield and revenue management to other areas.

Of particular importance in the area of origin-destination (O-D) control - an approach to revenue management that accounts for all possible passenger itineraries between origins and destinations in the airline network, at all fare levels - are the models describing bid pricing. Bid pricing involves the use of dual prices from a deterministic LP model to determine marginal costs for different legs in an airline network. These threshold values determine whether or not a fare is sold. If the fare on a single leg exceeds the sum of the fares along the path then that fare is sold. Talluri and Van Ryzin (1998) provide an excellent analysis of bid-price controls for network revenue management by proposing a general model for O-D control problems and solving it via dynamic programming.

There is a significant amount of research done on pricing in various fields such as yield management, marketing, and economics. Since the concern here is with yield and revenue management, only articles with those foci will be discussed. Dana (1998) explains the empirical pattern of advance-purchase discounts and price discrimination in the airline industry, without assuming that individual firms do not have market power. He demonstrates that advance-purchase price discounts is a plausible profit-maximizing response to conditions of demand uncertainty and high inventory holding costs. Feng and Xiao (2000a) and Gallego and Van Ryzin (2000) both propose models covering multiple products with multiple predetermined prices. Both of their results demonstrate that the optimal solution exists in closed form and the ease of applicability of their solutions.

Ciancimino et al. (1999) demonstrate the advantage of using yield management as a solution in the lucrative European passenger railway industry. They proposed a deterministic LP model and a probabilistic NLP model to analyze the overall expected revenue gains and CPU times when compared to standard algorithms. A large improvement in overall revenue is observed when the NLP model is used but it requires more CPU time.

Bitran and Mondschein (1995) also offer insight into the empirical use of yield management in the hotel industry. They developed heuristics to solve the problem of finding optimal strategies for renting hotel rooms when there is a stochastic and dynamic arrival of customers from different market segments and considering multiple-day stays. Their results show satisfactory performance of the heuristics using real hotel data.

The literature on empirically validated research in yield and revenue management is indeed rich and even though several studies support and build upon each other's findings, there are some which are contradictory. Most of the studies are somewhat narrow in focus, but this is due to the particular nature of yield management being made up of different components. As the evidence indicates, there has been no real study linking the various components of yield management.

2.7 ECONOMIC APPROACHES TO YIELD AND REVENUE MANAGEMENT RESEARCH

The study of yield and revenue management began in the late 1950's - early 1960's when American Airlines started to use operations research models to make revenue management decisions. The aim of most Operations Research / Management

Science (OR/MS) work is to improve the performance of an organization, in some respect, through the application of science. Revenue management was seen as embodying espoused techniques of OR/MS, that of improving revenue, through modeling, simulation, forecasting and inventory control (Littlechild and Shutler, 1991).

So far, there have been two types of approaches to yield and revenue management which are discussed in this chapter: quantitative and qualitative approaches. The quantitative techniques include mathematical programming involving linear and dynamic programming, economic approaches including Littlewood's rule, EMSR, threshold curve methods which give upper and lower bounds, and more recently, neural networks heuristics. These will be discussed in the upcoming sections. The qualitative approaches include the role of the service provider, the role of the consumer, and the psychology of yield management which were discussed previously.

Economic approaches to yield management become more and more important as firms discover the extent to which they can match demand and supply using their yield management applications. Indeed, the incidence of firms segmenting markets and tailoring marketing efforts to meet the needs of specific customers has increased during the past decade (Noone, 1997). The best known applications of yield management: overbooking and discount control are most useful when excess demand exists and the service provider has the luxury of choosing the type and amount of demand it will satisfy (Lieberman, 1993).

However, there is also the perception that yield management is of little or no use when supply is greater than demand. It is at this juncture that economic incentives such as group discount and low-price advance-purchase rates could be utilized to stimulate

demand. It should be noted that these economic incentives are not creating new demand, but merely emphasizing or exposing a previously under-exploited market. At the same time, there must be some lower price limit below which the inventory unit will not be sold, quite possibly where the marginal profit is negative or zero.

Utilized correctly, yield management can entice new customers to a service being offered at a “new and affordable” price. It can also shift demand away from a time when the firm’s capacity is over-extended to a time when it is under-utilized. This concept is also known as price discrimination. Matinees, off-peak rail fares, and special day of the week pricing are all examples of using yield management to shift demand when peak times are known. Economists would explain that every additional sale made is profitable if the incremental or marginal revenue exceeds the incremental or marginal costs associated with the sale. In other words, managers more likely would think that the additional income is better than nothing (Weigand, 1999). It is upon this very concept that Littlewood’s rule and EMSR is based.

2.7.1 Littlewood’s Rule

In 1972, Kenneth Littlewood proposed the discount allocation process which is based on the same idea of marginal revenue balance that sets overbooking levels at British Overseas Airways Corporation (BOAC), now British Airways.

His approach consisted of equating marginal revenues in each of two fare classes (one lower and one higher, compared to each other) to find the revenue maximizing seat allotments for a single flight leg inventory control problem. This application was used in a dynamic reservation context. He suggested that the total flight revenues would be

maximized by “closing down” the lower fare class to additional bookings when the certain revenue from selling another low-fare seat is exceeded by the *expected* revenue of selling that same seat to a potential higher fare-paying passenger. That is, lower fare-paying passengers paying f_2 should be accepted as long as:

$$f_2 \geq P_1(S_1) \cdot f_1 \quad \text{where } P_1(S_1) = \text{the probability of selling all remaining seats}$$

to higher fare-paying passengers and $f_1 =$ the higher fare-paying level. The smallest value of S_1 , which satisfies the above condition, is the revenue maximizing booking limit on class one (Belobaba, 1987a).

Further work on Littlewood’s Rule was done by Bhatia and Parekh (1973) and Richter (1982) who developed similar models for the same problem and derived equivalent results. Mayer (1976) found that lower expected revenue results as the ratio of low to high fare decreases and when controls deny too many low fare passengers. Titze and Griesshaber (1983) tested Littlewood’s model by simulating realistic passenger behaviours for a four-period model.

2.7.2 Expected Marginal Seat Revenue (EMSR)

Buhr (1982) while at Lufthansa, looked at the two-leg problem allowing for passengers boarding at the intermediate node and postulated that total flight revenues would be maximized when each individual flight-leg revenue is maximized. His formulation was extended to more complicated situations by Wang (1983) at Cathay Pacific who looked at multiple origin-destination (O-D) pairs where all feasible O-D and fare combinations are ranked according to their expected marginal revenue. Wang then

used stochastic dynamic programming (SDP) to solve this problem. These were two of the works which formed the basis of Belobaba's (1987) thesis.

Belobaba in his 1987 PhD thesis, extended Littlewood's rule to multiple fare classes by introducing the term expected marginal seat revenue (EMSR) for the general approach. This model recommended fare class booking limits taking into account the probabilistic nature of future demand for a flight. But, it does not produce optimal booking limits except for the two-fare case. Its acceptance is due primarily to its ease of implementation. $EMSR_i$ is defined as the expected marginal seat revenue for class i when the number of seats available to that class is increased by one. The expected marginal seat revenue of the S_i th seat in fare class i , $EMSR_i(S_i)$, is simply the average fare level in that class multiplied by the probability of selling S_i or more seats:

$$EMSR_i(S_i) = f_i \cdot \bar{P}_i(S_i) \quad \text{where } \bar{P}_i(S_i) \text{ is the probability of spill}$$

occurring (that is, the probability of receiving more than S_i requests for fare class i), S_i is the number of seats allocated to a particular fare class and f_i is the average fare for class i (Belobaba, 1989). A basic performance application of this technique done at Western Airlines in the first quarter of 1987 showed an increase in total flight revenues of 6.2%.

Dana (1999) in his paper presents a model that suggests yield management may be an efficient equilibrium response to uncertainty about the distribution of consumers' departure time preferences in the airline industry. It is not intended to be a general model of yield management but it does offer an interesting economical approach: his findings were that uniform pricing is inefficient because of the need to allocate uncertain demand across different departures, even when peak-load pricing is allowed. When firms face foreseeable demand fluctuations, traditional peak-load pricing is both effective and

encouraged but when firms face unforeseeable demand fluctuations, equilibrium price dispersion can efficiently shift demand and lower capacity costs, particularly when the costs of changing departure times are high (Dana, 1999).

Dana's results and conclusions demonstrate the effect of a high switching cost that is the equivalent of price fences used in yield management to differentiate fare classes. The existence of differential pricing for airline seats is the starting point for revenue management, and price is generally the most important determinant of passenger demand behavior. Passenger demand behavior or the demand distribution is defined as the probability of a reservation being booked in each class. Pricing can therefore be used to shape passengers' behavior, but at the same time the passengers' behavior can also shape the pricing as discussed earlier.

Reece and Sobel (2000) give a compact diagrammatic solution to the firm's profit maximizing price discrimination problem in the face of capacity constraints, which is an underlying characteristic of yield management. However, their paper is tempered by the fact that modeling demand uncertainty requires complex mathematical programming which is not easily explained by a diagrammatic approach. Even though their intended audience is classroom usage in intermediate micro-economics, industrial economics, and applied subjects such as the economics of transportation or tourism, it does have applications to the service industry and relevance to this dissertation.

2.7.3 Threshold Curve Methods

Threshold curve methods developed by Cross (1985) and Relihan (1989) are also used in economic approaches to yield management. With threshold curve methods, 1)

historical data on the trend of demand over a period of sixty to ninety days prior to the date of the actual flight, and 2) the number of reservations versus the number of days before departure are plotted. Actual bookings are also plotted against the past data and the nature of the curves keeps on changing as the time to departure nears. The curves are constructed by calculating an acceptable range of variation using the past mean and standard deviation (Murtagh and Mitra, 2002). Usually, the curves represent between ± 0.25 and ± 2 standard deviations of the demand (Kimes, 1989b). This threshold curve method is simple but it does not give optimised booking limits.

Wang (2001) extends and improves the threshold curve model by proposing a hybrid model which combines neural networks and dynamic programming. The model works by utilizing the neural networks to generate threshold curves in a high-dimensional space based on historical data and then dynamic programming is used to find the optimal pricing scheme based on the threshold band represented by the neural network. Some advantages of this type of hybrid modelling are that no strict assumptions about the demand distribution functions are required and that the model is very adaptive to most yield management systems.

2.8 YIELD AND REVENUE MANAGEMENT BY SECTORS

The following sections offer a brief comparison of the literature findings by sectors in the service industry with regards to yield and revenue management. While there may be noticeable overlap to previous material mentioned earlier in this dissertation, there are nevertheless some unique perspectives from each sector.

2.8.1 The Airline Sector

In the airline sector, yield and revenue management has become a widely accepted practice even in the cases where such practices are not transparent. Since the use of yield management systems has had the longest history in this sector, the effect has become somewhat normalised and there is less of an adverse reaction to its use. The correct use of yield management techniques has contributed immensely to the bottom line of the airline industry. With competition becoming boundary-less in certain markets, efficiently utilised yield management systems could be the difference between success and failure.

2.8.2 The Hotel Sector

It is in this sector that yield management systems are second in prominence only to the airlines but their usage began only in 1983 compared to early 1960's for the airline industry. However, there is slight complication in the usage of yield management because of the greater elasticity in demand and supply of the hotel industry. In very few cases is the price of an airline ticket negotiable at the time of purchase, but this is more widespread of an occurrence in the hotel industry. Therefore it is not much of a surprise in finding that there have been no studies in determining the adaptability of the algorithms, models and heuristics from the airline industry to the hotel industry.

2.8.3 The Restaurant Sector

Even though yield and revenue management could be practiced in the restaurant sector because of the similarity in characteristics to those possessed by the airline and

hotel industries, its use must be carefully examined before implementation. As the literature shows, the application of actual revenue management techniques has only been studied within the last nine years by Kimes (1998). Theoretically, because of the similarity in the requirements for yield management to be practiced (for example, fixed inventory), it provides a good fit. But practically, it may prove to be a nightmare primarily due to the inability to justify the price discrimination. It may be easy to charge more at different times of the day, but customers may find it inconceivable to pay more for a table because it has a view or may be further from the kitchen.

2.8.4 Other Sectors

Other sectors in which yield and revenue management has gained attention are for example; cruise lines, golf course, broadcasting, and the Internet. Another promising area of application is the cell phone industry and its various pricing plans. The most important characteristics of these industries which make them attractive to yield management techniques are the perishable nature of the inventory and the fixed capacity of inventory. Research articles in these areas have been limited and have only been conducted within the last five years. It is expected that with the rise in need to exploit full revenue potential, yield management techniques such as price discrimination will be sought.

2.9 Concluding Remarks and Directions for Future Research

During the course of this literature review, various issues pertaining to yield and revenue management were identified and are highlighted in the following sections.

2.9.1 SUMMARY OF SALIENT ISSUES AND FINDINGS FROM THE LITERATURE

There is a vast amount of academic and practitioner studies done in the areas of yield and revenue management but the majority of research in these areas has been focused on the airline industry and its applications to other service industries are only now being exploited. This is evidenced by the burgeoning research in areas such as lodging and hospitality, automobile rentals, broadcasting, cruise lines, Internet service provision, non-profit sectors and passengers railways. The common identifying characteristic in all of these industries is their fixed capacity. Even though in some of these sectors such as automobile rentals, and Internet service provision it may not be prohibitive to expand capacity, they are generally considered to have fixed capacity at least in the short term.

The existence of the human factor approaches is becoming more recognizable as an attribute to yield management in the service industry. In the more customer interactive sectors such as lodging and hospitality, restaurants and Internet service provision, researchers are paying more attention to the underlying issues. Authors such as Kimes (1989a,b), Jones and Hamilton (1992), Yeoman and Watson (1997), and Gourville and Soman (2002) all emphasize the people aspect of yield and revenue management.

However, as the literature demonstrates, there is no binding framework connecting areas such as behavioral and general business aspects of yield management. In the airline sector there has been no study analyzing attitudes towards, and perceptions of, yield management systems such as the neighboring seat on an airline costing half as much as your seat. There has been one such study done by Kimes (2002) in the hotel sector. It could be presumed that with the transparency of the airline sector with regards to the use of yield management, the findings of a similar study may validate the findings from the hotel sector.

As the literature also shows, no studies demonstrating the effect and impact of information systems and information technology on yield management systems have been done. Such a study may demonstrate that in the instances where yield management does not work as effectively as it should, that it could probably be traced to ineffective usage and perceptions about information systems. Lieberman (1993) has alluded to this by addressing the fact that yield management is not a computer system. However, no empirical study has been done to measure the impact or the strength of this perception.

New techniques and approaches can always be found to improve existing algorithms and heuristics in yield management systems. For example, comparisons of heuristics by Baker and Collier (1999), and comparisons of models by Belobaba (1987), have all shown the impact of improvements in mathematical techniques. In the ultra-competitive era of the global economy, the multi-billion dollar hospitality industry would gladly accept any gains of even one percent in revenues and yields. However, while the mathematical approaches are easily measurable, no attempt has been made to address the increases in yield or revenue attributable to the human factor approaches.

Another finding in the literature that is gaining some prominence now is the compatibility of yield management systems with good customer service (Lieberman, 1993 and Jauncey et al., 1995). In the current era in which the customer is more demanding and not as easily satisfied, customer service is of utmost importance in the hospitality industry. While yield management systems dictate prices and availability according to market conditions, customers generally rely on some reference transaction and may not understand the need to be charged a higher price at a different time for the same service. Firms that practice good customer service could easily explain away this difference in price without reference to its yield management systems.

As is evident from the literature, there are indeed many important findings to be considered. Of particular relevance is the transition of yield and revenue management from the eminent domain of the airline industry to widespread use in other service industries such as hospitality, restaurants and recreation. The growing numbers of mathematical techniques and approaches as well as untapped theoretical approaches also make this topic an area for fruitful research.

2.9.2 CONCLUSIONS

Yield and revenue management has played an important role in the hospitality sector especially in the airline and hotel industries. Its importance cannot be underestimated nor downplayed. The immense monetary benefits derived from the manipulation of mathematical techniques as well as psychological factors can make or break a firm. As Robert Cross says in his 1998 book, *Revenue Management: Hard Core Tactics for Profit-Making and Market Domination*: “Revenue management has

contributed millions to the bottom line, and it has educated our people to manage their business more effectively.”

Chapter 1 of this dissertation traced through the advantages enjoyed by companies practicing yield and revenue management correctly and highlighted the inadequacies and failings by those companies with improper applications. It showed the highs and lows that can be experienced and financial impact of yield and revenue management.

Chapter 2 outlined the history of yield and revenue management and its transition phases. It also provided an outline of the types of approaches used in the practice of yield and revenue management and assessed their relative impacts. It then introduced the most recent mathematical approach of neural networks and then shifted to illustrating the importance of the human factor approaches in yield and revenue management.

Chapter 2 also gave a survey of the literature by identifying the major issues in yield and revenue management including forecasting which will be the focus of this research, and then proceeded on to the conceptual and theoretical models, followed by the findings of selected empirically validated research models. It also gave a brief outline of yield and revenue management by sectors followed by the overall findings of the literature.

Airlines are not the only business to benefit from yield management. Other industries with better data to work from and more competition to work against are also starting to adopt the practice. It makes sense anywhere that an asset goes to waste if not sold for a particular period of time (Dyson, 1989). With their enormous resources and armed with their mathematical techniques, the power has always resided with the service provider. But, with the power of the Internet - primarily through readily available

information and its ease of accessibility – the balance may be slowly shifting back to the consumer.

Yield management can be overdone. In particular, lower prices offered at too many market segments can often fill hotels, golf courses, rental cars, cruise ships, and airplanes. But when such facilities are filled too quickly, the yield manager must ask the question, “Are we maximizing revenue? (That’s good!)” or “Are we just operating at full capacity? (That’s not so good.)” (Weigand, 1999). As outlined in a previous section concerning the issues surrounding yield and revenue management, the obstacles in determining the objectives of yield and revenue management need to be hurdled before the full potential of the system can be realised.

2.10 DIRECTIONS FOR THIS DISSERTATION

The primary goal of yield and revenue management is to maximize yield or revenue. While to some purists it may seem that the cut and dry answer would be that the mathematical approaches are the most logical solution, others will argue that the human concerns and the mathematical approaches are intertwined. Yield management applications and software cannot and should not be seen as only a computerized solution but rather as a tool to aid the human decision making process. In fact, a hybrid system may prove to be the most effective economically as well as managerially.

Arguments could be made for which type of mathematical approach is better in which situation and even for considering the customer to be more important than the service provider, but exclusively using one approach would be a recipe for disaster. Since the overwhelming majority of yield management applications are in the service

industries, emphasis should be equally placed on the people aspects of the applications as well as the computerized aspects. Even though the breadth of research in yield and revenue management is indeed wide, there still remain unanswered and untapped areas.

It has been widely acknowledged that the majority of research has been conducted in the airline industry with regards to yield and revenue management. And while it may seem that research has reached a plateau in that industry, it is definitely not the case in the other hospitality areas such as the hotel and restaurant industries. As this literature review shows there is a rich pool of resources to draw from to integrate into these other sectors.

As mentioned earlier, there is no comprehensive framework which models the importance of the human factors in yield and revenue management. Such a framework, with borrowed material from marketing, economics and psychology could lend an important understanding in the context of human behaviour and attitude towards the practice of yield and revenue management. There may come a time when the mathematical techniques may yield only so much more improvement in revenue. The question will be then: “Where or what else can be exploited for that minute gain in yield and revenue?”

The comparative wealth of research in the demand distributions for the airlines could be applied to the hotel industry where it may be much easier to obtain unavailability or denial of service data. This could enable better forecasting and overbooking models to be utilised and consequently improve yield or revenues. At the same time, the normal hotel industry measurement of capacity utilization needs to be addressed. As previously pointed out, yield and revenue management should not be considered as a tool to fill only capacity.

The foundation of most yield management systems lies on its data history. It is primarily due to this reason that the airline industry is the leader in yield management applications. This data history is used in the forecasting, overbooking and pricing models. Time series analysis has been the method of choice for forecasting as it offers longitudinal use. Noticeably, there is a very sparse amount of literature on modeling the reservations forecasting as a stochastic process. Taylor (1962) developed a model for the cancellation process and Rothstein (1971) developed a non-homogenous Markovian model of the booking process and his objective was to determine optimal levels of overbooking and not to forecast bookings. In a later paper by Rothstein (1985), he states that it is not clear whether this non-homogenous model has ever been implemented.

However, another technique that has not been tried is the use of Markov chains. With their state transitions and probabilities, they can be used to model the dynamic nature of forecasts in yield and revenue management. In his 1990 PhD thesis, Lee found that forecasting accuracy has a significant effect in revenue generated from revenue management systems. A 10-percent improvement in forecast accuracy on high demand flights corresponded to an increase of 1.5 to 3.0 percent in revenues.

Additionally, even with techniques such as disaggregation available, the dynamic nature of yield and revenue management increases the complexity of accurate forecasting. This dynamism is due to the constantly changing nature of prices which may occur more than 100,000 times a day for an average flight in the airline industry and to a lesser amount in the hotel industry. Using Markov transition matrices could better help to conquer this dynamic problem. Since the models for the demand distributions and the arrivals processes are stochastic (e.g. McGill and Van Ryzin, 1999; Lautenbacher and

Stidham, 1999), it is plausible that a stochastic forecasting technique such as Markov Chains Transition Matrices would be of value.

As Lee (1990) argues in his dissertation, the booking process can be described as a discrete state, continuous time stochastic process which can ideally be represented as a Markov Chain provided that the discrete states are clearly and easily defined. Simple evidence of this is demonstrated by the fact that the arrival rate of additional requests for reservations (demand) does not depend on the number of bookings currently in the system. This property is suitable for a Markovian process.

In their 2001 study, Weatherford et al, showed that a purely disaggregated forecast (forecasting based on smaller data subsets) strongly outperformed even the best aggregated forecast. This disaggregation was done according to rate classes and length of stays for the hotel industry. However, disaggregation by customer type was not studied. The fact that the case for aggregate long-term forecasts seems inadequate does not, however, preclude the possibility that the Markov formulation maybe be useful in forecasting short-run behaviour of individual classes or customer types (Armstrong, 1969).

The power and impact of Internet auction websites such as Priceline.com and lastminutedeals.com also have to be taken into consideration. No study has been done as to how this type of bargaining has affected the forecasts and ultimately the yield and revenue gains of the airlines and hotels. Again the question arises as to whether or not they are filling capacity or increasing revenues. In addition, the applicability of price fences to other service industries needs to be discussed. Could the Internet be used to create these price fences and increase demand for other industries?

The use of cross-sectional studies in the various sectors of the hospitality industry has also not been researched. Issues such as the demand characteristics by market segment (business or leisure use) in the airline industry as compared to the hotel industry may provide a better understanding of demand patterns and provide better forecasts. Similarities in behaviour patterns could be used to determine which particular overbooking model to use. For example, is the behaviour of the business class flyer replicated by the behaviour of the guest in the executive suite?

As this literature review demonstrates, the continued use of both theoretical and empirical research studies are needed to ensure that the competitive and strategic advantages gained by the use of yield and revenue management systems are sustained. The future for academic research also looks bright, as there is the need to develop better forecasting techniques since this is one particular area that has been for some time, untouched.

Chapter 3

Research Contributions

3.1 RESEARCH OBJECTIVES

The primary objective of this study is to explore the viability of Markov chains and Markov transition matrices as forecasting tools for airline demand rates. These demand rate forecasts are integral to revenue management systems.

Among the most common mathematical techniques used in forecasting for yield management are regression and time series analysis. But forecasts can become biased because of permanent changes in customer behavior that time series models do not rapidly incorporate (Smith et al., 1992). Given the recent trends towards disaggregated forecasts which provide much more accurate results than aggregated forecasts but require more data; both regression and time series forecasting techniques have their drawbacks.

While it is statistically counterintuitive for disaggregated forecasts to be more accurate than aggregated forecasts, it is possible, because of differences in model specifications, or offsetting errors as data are aggregated or by aggregation and pooling bias. Additionally, if the demand for different rate categories and lengths of stay (in the hotel industry) are highly correlated, then the interpretation of results may be difficult (Foekens et al., 1994).

As a review of the literature shows, the majority of studies done on the forecasting module of yield management have focused on time series analysis. The objective of this research is to go in a slightly different and a possibly more relevant direction by utilizing dynamic relationships, which differ from those currently in use.

Current airline practices utilize disaggregation of the booking horizon by semi-regularly spaced time intervals and disaggregation of fares by up to a dozen classes. However, no study to date has been done to determine if this partitioning is optimal. While most airlines disaggregate the booking horizon for the short term in approximately the same way, the long run disaggregation methods usually differ. Given that data storage and manipulation has never been an Achilles' heel for the airlines, it is worthwhile to investigate if a more efficient method could be utilized than what is currently employed. A recent trend has shown the simplification (aggregation) of fare classes by many low-cost airlines, but again no study has been published demonstrating its forecasting accuracy and net effects on revenues. The second objective of this study is therefore, to investigate the changes in forecasting accuracy when the aggregation is done by reducing the number of time-periods and/or fare classes.

One problem this research will not address however is the effect of cancellations and no-shows on the system. Cancellations are not modeled because only the net bookings changes are given by the data available. In previous attempts at forecasting, no-shows have proven to be the most problematic to model simply because the airline or hotel has no chance of rebooking or utilizing that lost customer revenue unless there are walk-ups or standbys; and because in the hospitality industry, post-departure data is not easily accessible to researchers. However, it should be noted that Markov chains with their absorbing states could easily take into account the modeling of cancellations and no-shows, if data were available.

This research study will be conducted using actual as well as simulated airline data, which represents the dynamic complexities of the hospitality industry. The dataset used in this research is described in Section 4.7.

3.2 Proposed Markovian Transitions Matrices Model

There is sparse literature on modeling the reservations forecasting as a stochastic process. Taylor (1962) developed a model for the cancellation process and Rothstein (1971) developed a non-homogenous Markovian model of the booking process. His objective was only to determine optimal levels of overbooking and not to forecast bookings. Since then, there has been very little work done on forecasting the bookings using a Markovian stochastic process such as transition matrices.

Primarily due to their ability to incorporate many different states and the movement(s) between those states, Markov transition matrices could be viewed as a dynamic tool to create bookings forecasts. A single unit of inventory (a seat) which is recorded by the bookings or reservation data can be viewed as existing in any one of the following states: available, booked, cancelled, departed, no-showed, or walkups. Of these states, departures and no-shows are absorbing states; while available, booked, cancellations and walkups are transient states.

While the dataset utilized in this research only conveys information on available and booked states, the methodology can easily be adapted to account for the other states. The other Markovian properties such as stochastic independence and steady state conditions also lend themselves to Markovian modeling assumptions. It is within this

context that the rationale for using transition matrices as a forecasting tool is proposed in this study.

3.2.1 Model Development

At any point in time t prior to and including the time of departure, a seat (the unit of inventory) is classified as state A (available) or B (booked). The t value used is dependent upon the data collection method used by the airline. For example, $t = 0, 1, 2, 3, 5, 7, 10, 13, 18, 23, 30, 40, 60, 90, 120,$ and 180 days prior to departure is the data available and used in this research. While these intervals are unevenly spaced, they can be mapped or transformed to observation points (OP) in time and indexed as illustrated in Table 2.

# of Days prior to Departure (DP) t=	0	1	2	3	5	7	10	13	18	23	30	40	60	90	120	180
Observation Point (OP)	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

Table 2. Mapping of Days Prior to Departure to Observation Points

No-shows and walkups can only be identified at boarding time and thus can only occur at the transition from DP_0 to “Boarded.” Since the available data does not convey this information, they could be modeled separately using historical rates. Over the period of observation, a reservation can move from state to state. Eventually, it ends up as either a cancellation (meaning the seat is available at the next OP), or, it remains booked. Based on these states, a transition matrix E can be constructed to display the probability of moving from state i at observation point OP_t to state j or remaining in state i at observation point OP_{t+1} .

For the cells in Matrix E it is then easy to determine the transition probability p_{ij} by the historical probabilities from OP_t to OP_{t+1} . Since airline data only shows net bookings at each stage, we could have either of two transition matrices as follows:

	A_{t+1}	B_{t+1}
A_t	$1 - PPU$	PPU
B_t	0	1

or

	A_{t+1}	B_{t+1}
A_t	1	0
B_t	NPU	$1 - NPU$

Matrix 1 – Positive Pickup

Matrix 2 – Negative Pickup (net

cancellations)

where PPU is the probability of a positive pickup and NPU is the probability of a negative pickup or net cancellations.

The forecasted calculations using these transition matrices are then as follows:

$$B_j[E_{j,j+1}] = F_{j+1}$$

where, j = a data collection point (DP_x as defined previously)

B_j = the bookings vector which shows the BTD at end of period j

$[E_{j,j+1}]$ = the transition matrix from period j to $j+1$

F_{j+1} = the forecasted bookings vector at the end of period $(j+1)$

$b_{j,k}$ = the k^{th} element of B_j where k = Available or Booked

$f_{j,k}$ = the k^{th} element of F_{j+1}

The main idea here is to determine how many bookings go from state i at OP_t to state j at OP_{t+1} with the difference being the number of reservations picked up or lost and then converting that to a transition probability.

3.3 Environmental Settings and Limitations

This study does not explicitly model any of the environmental variables listed below relating to customer behavior and consequent demand patterns. These variables cannot be modeled because of the extreme accuracy needed to model them and the general lack of data availability. But they need to be mentioned because they indeed influence the demand patterns. (McGill and Van Ryzin, 1999)

- **Seasonality, day-of-week variation** – airline travel is highly seasonal and there are day-of-the-week variation patterns. These patterns are especially more evident between city pairs where there are business centers and where business travel is a large component of certain flight times.
- **Special events** – special events such as a conferences or festivals in another can temporarily affect the demand patterns for airline travel. Again these events can be identified beforehand and have been eliminated from the data analysis and forecasting.
- **Sensitivity to pricing actions** – airline pricing is highly price elastic for the same flight legs but there are different price elasticities for different fare classes. Thus, if the competitor slashes prices or offers a low fare for a particular O-D pair then all other airlines in that market must be prepared to follow suit.
- **Demand dependencies between booking classes** – the theoretical assumption that demand for one booking class is statistically independent of another is usually not justified in practice. To accurately determine this, the recapture (the re-booking of a passenger who is unable to get a reservation of choice onto an alternative flight with the same airline) and defection (occurs when a passenger

with a confirmed reservation switches to a flight with another airline for reasons of convenience) probability must be correctly modeled. While the recapture probability could be determined with some certainty, the defections probability requires the much more complicated determination of the probability distribution for departure delays.

- **Diversions** – occur when for example, a business traveler has sufficient advance notice of a trip which enables them to purchase a lower priced or discount ticket. To counteract diversions, airlines usually impose restrictions on the purchase of discount tickets. But knowing if a business traveler would have purchased a business class ticket instead of a lower price discount ticket is beyond the scope of this study.
- **Go-Shows** – are similar to no-record passengers but are allowed to board because there are seats available. No-records are passengers who appear at flight time with a valid ticket (such as an open date ticket) but for whom there is no reservation record. Again this variable is difficult to model as it would involve having data as to who exactly boarded and not only those who had reservations.
- **Group bookings** - could be considered as separate or individual bookings all arriving at the same time. As it is, the transition matrix cannot model this since we are looking at discrete arrivals. Instead of having a small probability assigned to a large increase in demand, a sudden greater increase in demand, such as a group booking, can easily be accounted for by a much larger probability for the increase in demand. If the interval is further delineated, these arrivals could be thought of as incremental arrivals. As such, they are not modeled explicitly as group

bookings since there is no evidence in the data to suggest that there were group bookings.

- **Recaptures** – as defined previously, could be modeled since we are looking at the transitions between only 2 consecutive OP's. Thus it is possible that the demand which is treated as a new reservation for the current flight at some OP, is also treated as a cancellation at another OP for another flight. But we are not going to explicitly model recaptures since it is also possible for passengers to travel on an entirely different flight than they originally intended.
- **Upgrades** – work in two ways: bumping up to a higher fare class of a passenger who may have been otherwise denied boarding, and the choice by a passenger to pay a higher fare than originally planned if their original fare requested is sold out. This variable cannot be modeled here because of its dynamic nature at flight time and dependence on no-show probabilities.
- **No-shows** - no-shows are passengers who fail to show up at the time of departure. This does not allow the airline to re-book or resell this seat through the normal reservation channels. The only way the airline could earn additional revenue from this seat is by means of a walkup or standby passenger or by overbooking. No-shows are very common amongst full fare passengers whose tickets are usually without any penalty conditions. The standby factor cannot be modeled explicitly here, as it would require knowledge of the exact number of passengers who boarded for each fare class and as a result, it is not considered to be an environmental variable in this study.

- **Walkups** - walk-ups are passengers who show up at flight times without reservations but wish to travel on a particular flight. Additionally, in some instances, the term no-record is used to describe walk-up passengers with reservations, but for whom the airline has no record. Both of these cases affect the revenue of the airline. Again, since this variable requires knowledge of pre-boarding data which is unavailable, it cannot be modeled here explicitly. For simplicity, we will assume a walkup rate of zero percent.

3.4 RESEARCH QUESTIONS

Given the limitations which are described in the previous sections, this current investigation is concerned with the following research questions:

a) Primary Question

- Does the use of Markov Chains based transition matrices as forecasting tools for airline revenue management offer an improvement over selected current methods?

b) Secondary Questions

- Does aggregation levels by fare classes affect forecast accuracy?
- Does aggregation levels by time periods affect forecast accuracy?
- Is the choice of a forecasting method affected by aggregation levels of (i) fare classes or (ii) observation points utilized?

3.5 RESEARCH HYPOTHESES

a) Primary Hypothesis

- Markov Chains based transition matrices are better forecasting tools for airline seat demand compared to the selected existing methods.

b) Secondary Hypotheses

- Aggregation levels of fare classes impacts forecast accuracy.
- Aggregation levels by time periods impacts forecast accuracy.
- Model performance depends on (i.e. interacts with) aggregation levels of (i) fare classes and (ii) observation points utilized.

Chapter 4

Methodology

This chapter develops the framework within which the research hypotheses will be operationalized and tested. These hypotheses will be tested by means of a computer simulation model using actual airline data. First, the model selection criteria are identified then the selected models for comparison are outlined along with a simple worked example of each. Next, the experimental and decision factors involved in these models are discussed followed by a description of the simulation modeling. The measures of performance that will be used to determine the effectiveness of the proposed model are then specified. Then, data collection issues including various factors affecting the suitability of the proposed model and the experimental design are discussed. Lastly, a description of the computer modeling is followed by development of statistical hypotheses that operationalize the research questions.

4.1 Model Selection Criteria

The objective of this study is to compare existing methods used for micro-level forecasting with the proposed Transition Matrices method to determine if the Transition Matrices method offer any advantages over others. The following criteria are typical for model selection in forecasting [for revenue management].

- **Simplicity** – Numerous authors (Wickham, 1995; Zickus, 1998; and Zeni, 2001) have demonstrated the need for a simple model as an important factor in short-term forecasting. The simplicity of the model is determined by the degree of computational difficulty in producing the forecast. For example, even within

time-series forecasting, one can choose models with levels of complexity ranging from the extrapolation of simple means to auto-regressive integrated moving averages (ARIMA) models. However, it should be cautioned here that care must be taken not to choose from a subset of the simplest methods, as this certainly would not be representative of industry practices.

- **Ease of Implementation** – The trade-off here of course, is between an overly simplified mathematical model that is easy to implement and may not correctly address all of the factors versus a more accurate but complex mathematical model that is more difficult to implement. This criterion is obviously dependent on the previous one; the risks associated with a complex mathematical model could be magnified, and any resulting conclusions may not be valid. It should be noted that computing power might not be a factor here, as the airline industry has access to some of the best and latest technology.
- **Representative of Current Industry Practices** – Since the objective of this study is not only to propose a new method but also to evaluate its implementation potential; it is compared to methods that are in use. If the results of this study prove favorable, then the implementation step could be immediately considered.
- **Accuracy** – A forecast could never be 100% accurate due to randomness. However, the relative accuracy or the amount of inaccuracy that is willing to be tolerated is what determines the strength of the particular forecasting tool used. To measure the accuracy of the forecasts, various mathematical metrics are used (see Section 4.6). An important part of forecasting is that the forecasted results

should be constantly monitored. If the accuracy (as defined by the metrics used) of those results are stable then there should not be any reason for concern. Alternately, if forecasts develop a bias, or the accuracy is deteriorating, then the model chosen should be carefully re-evaluated or even discarded. As discussed in earlier sections, the impact of accurate forecasting is the most important underlying success factor in revenue management.

4.2 Forecasting Models used in Experiment Design

The following sections outlines the three current models chosen for comparison and the gives details of the proposed Markov Transition Matrix method.

4.2.1 Models used for Comparison

The first method used for comparison is the Classic Pickup model analysis performed by Duncanson in 1974, the second method is the Advanced Pickup model proposed by L'Heureux in 1986 and the third method is the Multiplicative Pickup model. The details of each of these methods were discussed in Section 2.4.2.2.

4.2.2 Numerical Examples of the Methods Chosen for Comparison

This section gives a systematic outline of the workings of each model chosen for comparison along with a worked example of the proposed model. Its purpose is to demonstrate the computational degree of complexities involved.

a) **A Classical Pickup Example**

The example below represents the bookings for a set of 8 flights of capacity 90 seats departing between March 5th and April 23rd given that the current date is between Apr. 23rd and Apr. 30th. The assumption here is that the flights were not censored so we do not have to deal with constraining issues.

Days Prior \ Departure	DP0	DP7	DP14	DP21	DP28	DP35	DP42	DP49	DP56
Mar 05	15	14	10	10	12	11	10	10	1
Mar 12	9	7	9	9	12	12	8	6	1
Mar 19	9	11	10	9	11	12	6	4	4
Mar 26	6	6	6	6	3	4	2	1	1
Apr 02	13	14	15	15	16	15	8	5	4
Apr 09	5	4	3	2	3	4	2	2	1
Apr 16	8	8	8	9	8	8	8	8	7
Apr 23	9	10	16	16	11	11	17	5	2
Apr 30		8	9	9	9	4	4	2	0
May 07			9	8	6	5	5	6	5
May 14	FC			22	23	21	16	15	10
May 21					12	9	8	7	4
May 28						4	3	3	2
Jun 04							18	14	4
Jun 11								6	3
Jun 18									2

Table 3. Sample of bookings to date information

Given all these bookings from one observation period to the next, we now use this information to forecast the final bookings for the May 14th flight for example, as follows. First, we need to compute the pickup from one period to the next by simply subtracting the bookings in one time period from the previous time period's booking. We then get the following table of pickup values:

Pickup	7 to 0	14 to 7	21 to 14	28 to 21	35 to 28	42 to 35	49 to 42	56 to 49	Open to 56
Mar 05	1	4	0	-2	1	1	0	9	1
Mar 12	2	-2	0	-3	0	4	2	5	1
Mar 19	-2	1	1	-2	-1	6	2	0	4
Mar 26	0	0	0	3	-1	2	1	0	1
Apr 02	-1	-1	0	-1	1	7	3	1	4
Apr 09	2	1	1	-1	-1	2	0	1	1
Apr 16	0	0	-1	1	0	0	0	1	7
Apr 23	-1	-6	0	5	0	-6	12	3	2
Apr 30		-1	0	0	5	0	2	2	0
May 07			1	2	1	0	-1	1	5
May 14				-1	2	5	1	5	10
May 21					3	1	1	3	4
May 28						1	0	1	2
Jun 04							4	10	4
Jun 11								3	3
Jun 18									2

Table 4. Calculated Pickup values from Table 3

Next, we calculate the average pickup from day 21 to day 0 for the May 14th flight (for which there are already bookings on hand). So, for example here, the pickups are averaged over 8 observations and we get the following table:

Pickup	7 to 0	14 to 7	21 to 14
Mar 05	1	4	0
Mar 12	2	-2	0
Mar 19	-2	1	1
Mar 26	0	0	0
Apr 02	-1	-1	0
Apr 09	2	1	1
Apr 16	0	0	-1
Apr 23	-1	-6	0
Apr 30		-1*	0*
May 07			1*
May 14			
May 21			
May 28			
Jun 04			
Jun 11			
Jun 18			
Ave. Pickup	0.00	-0.375	0.125

Table 5. Averaged Pickup Values

(* - for the Classical PU method these values are not used in the average as the data is incomplete)

We then simply add the average pickups to the current number of bookings for the May 14th flight (there are 22 bookings on hand at 21 days before departure). Thus, the overall forecast for the May 14th flight is $FC = [0.00 + (-0.375) + 0.125 + 22] = 21.75$.

b) An Advanced Pickup Example

The key to the Advanced Pickup model is to estimate the aggregate pickup by summing estimates of the pickup over smaller disaggregated intervals and using information from flights yet to depart. As a result, using the same previous pickup matrix - where the starred values are now included, we have the new averages as follows:

For $PU_{7,0}$: the average = 0.00; for $PU_{14,7}$: the new average is now = -0.444; and for $PU_{21,14}$: the new average is now = 0.200. Thus, the overall forecast for the May 14th flight is $FC = [0.00 + (-0.444) + 0.200 + 22] = 21.76$.

It should be noted that some authors (Wickham, 1995; Zeni, 2001) use only the most recent 8 data points for any observation point prior to departure. The reason for doing so was if there was complete data for 8 departed flights then an eight period moving average would be suitable. In this case the revised forecast would now be $0.00 + (-1.000) + 0.250 + 22 = 21.25$.

c) A Multiplicative Pickup Example

The multiplicative pickup method gives the following results: the average bookings to date at $DP_0 = 9.25$ and the average bookings to date at DP_{21} using only departed flights is 9.50 and using all available data is 10.455. If we use only departed flights then the pickup ratio is $9.25/9.50 = 0.9739$ which when multiplied by the BTD at DP_{21} gives a forecast of 21.42. If we use all available data points, then the pickup ratio is $9.25/10.455 = 0.8848$ which when multiplied by the BTD at DP_{21} gives a forecast of 19.47.

d) The Proposed Markovian Transition Matrix Example

A Transition Matrix for pickup between two consecutive states uses the probability of a single reservation either remaining booked or becoming available. This depends on the average net effects of the reservations picked up or lost between any two consecutive observation points (OP's). The probability is calculated as a ratio of the average net pickup between any two consecutive OP's to the aircraft capacity.

It should be noted that this is different from the Multiplicative Pickup (MP) method in that the MP method simply uses the ratio of the average bookings picked up at DPO to the average bookings picked up at any other OP. The probability calculations for the transition matrices can be done in either of two ways similar to the methods used for the Advanced Pickup method or by using the values from the Classical Pickup calculations. That is, we can either use a simple average of all flights or we can use a moving average which considers only a specific number of most recent flights.

Consequently, in determining the probabilities for the transition matrices, a similar approach is utilized. As a result there will be three types of transition matrices forecasts produced: The first is based on a similar approach used by the Classical Pickup method whereby only the departed flights data are analyzed; the second type of transition matrices forecasting uses an approach similar to the Advanced Pickup method whereby all of the known flights data are analyzed; and the third type of transition matrices forecasting method uses an approach whereby only a particular number of the most recent flight data are analyzed.

It was observed in the previous literature that the most common number of complete flight departures analyzed was 8 flights. This number was therefore chosen as the most

recent number of flights to analyze amounting to an 8-period simple moving average value. Either method would include data of flights which have departed and flights which are yet to depart. The details of this method using the same pickup table calculated above are as follows:

Pickup	7 to 0	14 to 7	21 to 14	28 to 21	35 to 28	42 to 35	49 to 42	56 to 49	Open to 56
Mar 05	1	4	0	-2	1	1	0	9	1
Mar 12	2	-2	0	-3	0	4	2	5	1
Mar 19	-2	1	1	-2	-1	6	2	0	4
Mar 26	0	0	0	3	-1	2	1	0	1
Apr 02	-1	-1	0	-1	1	7	3	1	4
Apr 09	2	1	1	-1	-1	2	0	1	1
Apr 16	0	0	-1	1	0	0	0	1	7
Apr 23	-1	-6	0	5	0	-6	12	3	2
Apr 30		-1	0	0	5	0	2	2	0
May 07			1	2	1	0	-1	1	5
May 14				-1	2	5	1	5	10
May 21					3	1	1	3	4
May 28						1	0	1	2
Jun 04							4	10	4
Jun 11								3	3
Jun 18									2
CP avg.	0.000	-0.375	0.125	0.000	-0.125	2.000	2.500	2.500	2.625
Rate	0.000	-0.004	0.001	0.000	-0.001	0.022	0.028	0.028	0.029
AP avg.	0.000	-0.444	0.200	0.091	0.833	1.769	1.929	3.000	3.188
Rate	0.000	-0.005	0.002	0.001	0.009	0.020	0.021	0.033	0.035
AP(8) avg.	0.000	-1.000	0.250	1.000	1.375	0.375	2.375	3.500	3.750
Rate	0.000	-0.011	0.003	0.011	0.015	0.004	0.026	0.039	0.042

Table 6. Transition Probabilities (rates) calculated from Pickup values

Using the CP averages we have for example, the pickup rate for DP14 to DP7 (the highlighted value) as:

$$\begin{aligned}
 \text{Rate (PU}_{14,7}) &= \frac{CP_{14,7}}{\text{AircraftCapacity}(C)} \\
 &= \frac{-0.38}{90} \\
 &= -0.004
 \end{aligned}$$

Since this rate is negative it implies that there was a negative pickup in that corresponding period and thus it would be placed in the lower left hand box of the

transition matrix as the probability of losing bookings for this particular interval. The calculations for the AP rates, whether we are using all available flights or the most recent flights are done in a similar manner. The transition matrices are now constructed for each interval as shown:

	A_{DP14}	B_{DP14}		A_{DP7}	B_{DP7}		A_{DP0}	B_{DP0}
A_{DP21}	0.9986	0.0014	A_{DP14}	1.00	0.00	A_{DP7}	1.00	0.00
B_{DP21}	0.00	1.00	B_{DP14}	0.0042	0.9958	B_{DP7}	0.00	1.00

These transition matrices are then pre-multiplied by the starting vector, in this case $A_{DP21} = 68$ and $B_{DP21} = 22$ to give the results at each step. The final forecast at DP0 yields $A_{DP0} = 67.9976$ and $B_{DP0} = 22.0024$. Similar calculations can be done using the rates determined by the Advanced Pickup methods. The table below gives a comparative summary of the results for a sample flight.

Flt. #	Class	Capac.	DP21	DP0	Pickup Methods					Markov Chains Method		
					BTD	Actual	CP (8)	AP (all)	AP (8)	MP (all)	MP (8)	CP (8)
14	Y	90	22	33	21.75	21.76	21.25	19.47	21.42	22.00	22.04	21.94

Table 7. Comparative Results between Pickup Methods used and MC methods

4.2.3 Forecasting Implementation Issues

This section describes forecasting implementation issues applicable to Phases I and II of the data analysis.

In the previous section of this research, various types of forecasting methods are illustrated by example. Since there is an intrinsic difference in the way forecasts are made for the Classical Pickup method and the Advanced Pickup methods as more data become available (that is, as time to departure nears), it is therefore necessary to illustrate how these methods differ during implementation.

Each booking profile will (not) be updated by new bookings_to_date (BTD) data as it becomes available depending on whether it is the Advanced (Classical) Pickup calculations that are being determined. That is, even though partial BTD data (and its equivalent pickup rates and transition probabilities) is available as the time to departure nears, that information would not be used in the calculations for the Classical Pickup method until after the flight has departed as illustrated in Table 5.

This updating is illustrated by another iteration of Tables 3-5 and shown below in Tables 3a-5a. If today is now Apr 30 and we have new BTD information (7 days has advanced) as shown below in Table 3a, the pickup calculations are updated as shown in Table 4a and the new average pickup rates are shown in Table 5a.

Days Prior Departure	DP0	DP7	DP14	DP21	DP28	DP35	DP42	DP49	DP56
Mar 05	15	14	10	10	12	11	10	10	1
Mar 12	9	7	9	9	12	12	8	6	1
Mar 19	9	11	10	9	11	12	6	4	4
Mar 26	6	6	6	6	3	4	2	1	1
Apr 02	13	14	15	15	16	15	8	5	4
Apr 09	5	4	3	2	3	4	2	2	1
Apr 16	8	8	8	9	8	8	8	8	7
Apr 23	9	10	16	16	11	11	17	5	2
Apr 30	10	8	9	9	9	4	4	2	0
May 07		12	9	8	6	5	5	6	5
May 14	FC		26	22	23	21	16	15	10
May 21				14	12	9	8	7	4
May 28					7	4	3	3	2
Jun 04						22	18	14	4
Jun 11							8	6	3
Jun 18								5	2

Table 3a. Sample of bookings to date including new information on Apr 30 flight

Pickup	7 to 0	14 to 7	21 to 14	28 to 21	35 to 28	42 to 35	49 to 42	56 to 49	Open to 56
Mar 05	1	4	0	-2	1	1	0	9	1
Mar 12	2	-2	0	-3	0	4	2	5	1
Mar 19	-2	1	1	-2	-1	6	2	0	4
Mar 26	0	0	0	3	-1	2	1	0	1
Apr 02	-1	-1	0	-1	1	7	3	1	4
Apr 09	2	1	1	-1	-1	2	0	1	1
Apr 16	0	0	-1	1	0	0	0	1	7
Apr 23	-1	-6	0	5	0	-6	12	3	2
Apr 30	2	-1	0	0	5	0	2	2	0
May 07		3	1	2	1	0	-1	1	5
May 14			4	-1	2	5	1	5	10
May 21				2	3	1	1	3	4
May 28					3	1	0	1	2
Jun 04						4	4	10	4
Jun 11							2	3	3
Jun 18								3	2

Table 4a. Calculated Pickup values from Table 3a

Pickup	7 to 0	14 to 7	21 to 14
Mar 05	1	4	0
Mar 12	2	-2	0
Mar 19	-2	1	1
Mar 26	0	0	0
Apr 02	-1	-1	0
Apr 09	2	1	1
Apr 16	0	0	-1
Apr 23	-1	-6	0
Apr 30	2	-1	0
May 07		3*	1*
May 14			-1*
May 21			
May 28			
Jun 04			
Jun 11			
Jun 18			
Ave. Pickup	0.333	0.444	0.111

Table 5a. New averaged pickup values

(* - for the Classical PU method these values are not used in the average as the classical methods do not incorporate data from flights that are yet to depart)

These new values can be compared to the previous averages of 0.000, -0.375 and 0.125 respectively. Similarly for the Advanced Pickup averages (where the starred values are now included), we now have for $PU_{7,0}$: the average = 0.333; for $PU_{14,7}$: the new average is now = -0.100; and for $PU_{21,14}$: the new average is now = 0.091.

4.3 The Type of Demand Distribution

Since we cannot conclusively fit a distribution pattern to the data, it becomes necessary to model it using different probability distributions. It is known that the Normal distribution is an excellent approximation for censored Poisson when the mean of the Poisson is greater than 1000, and it is a good approximation for the Poisson when the mean is greater than 10. But bookings can only be non-negative whole numbers, forcing any fitted distribution to be a truncated Normal, the Censored Poisson, or the Adjusted Gamma. The Weibull distribution has also been found to be applicable.

However, it should be noted that it is theoretically and practically impossible to fit a single distribution to every sample of booking data.

An alternative approach is to use bootstrapping, which is a statistical technique based on resampling with replacement from the observed data; it does not require distribution assumptions and can be effectively used with smaller sample sizes. Even though some of the sample sizes for the various flight numbers were relatively large, bootstrapping was applied uniformly across all flight samples regardless of the actual sample sizes of the observed flight data. Thus, it was not necessary to model the demand distributions.

4.4 The Capacity of the Aircraft

In the initial phase of this study, two levels of capacity C are studied: High and Low. High capacity flights are defined as those whereby the total bookings at departure were greater than or equal to 96. Low capacity flights are defined as those whereby the total bookings at departure were lesser than or equal to 80. These cutoff values are conveniently chosen to match the data available by eliminating approximately the middle third of the data. In essence, the capacity of 80 or less seats represents the bottom 35% of the data and the capacity of 96 or greater represents the top 33% of the data. The percentiles are chosen to minimize any conflicts with overlapping capacities and to offer distinct boundaries. This selection method could be easily adapted to fit any other relative levels of capacity.

4.5 Decision Factors

The decision factors relating to forecasting for revenue management are: the number of observation points (days prior) used, the level of fare class aggregation, and the type of data used.

4.5.1 Aggregation by Time Fences

Flights are normally open for booking from at most one year or 360 days prior to departure. Data snapshots are then reported at various unequal time intervals decrementing until departure time. The time intervals in this study were decremented by an appropriately chosen number such that there was no loss of data integrity and that there was no burden on the data reporting or storage components of the system.

The number of time-frames available for this study was 16 occurring at semi-regularly spaced intervals, starting from departure until 180 days out. This allows us to select a subset of these points to determine if the data storage and reporting requirements could be reduced. It also allows us to determine if the choice (spacing) of time intervals affects the accuracy of the forecasting methods. A base-information set will be used in all cases and this will then be augmented by the other intervals. This base-information set will consist of 9 observation points up to 18 days prior to departure. These are taken at 18, 13, 10, 7, 5, 3, 2, 1, and 0 days prior to departure.

Up to 18 days prior to departure was chosen to be common to all scenarios since this interval would contain essential bookings_to_date (BTD) data. It should be noted that the number of observation points used should be large enough to enable relatively accurate forecasts that are not overly sensitive to the omitted intervals. Simultaneously,

the objective should also be to omit time intervals that may be too old or too adjacent to provide useful information.

The following seven (7) observation point (OP) scenarios will be tested:

- (1) Base points plus all other available OP's (BP-AA) = Base Case + DP23 + DP30 + DP40 + DP60 + DP90 + DP120 + DP180. This will be referred to as the Full Information model. It contains the full complement of 16 OP's.
- (2) Base points plus alternate OP's (BP-ALT1) = Base Case + DP30 + DP60 + DP120. This will be referred to as the Alternate Information model I. It contains 3 additional OP's geometrically spaced by doubling the reported days prior.
- (3) Base points plus alternate OP's (BP-ALT2) = Base Case + DP23 + DP40 + DP90 + DP180. This will be referred to as the Alternate Information model II. It contains 4 additional OP's unevenly spaced.
- (4) Base points plus selected subset I (BP-1) = Base Case + DP23 + DP30 + DP40 + DP60. This will be referred to as the Information Subset model I. This also contains 4 additional OP's but it's those that are the most recent.
- (5) Base points plus selected subset II (BP-2) = Base Case + DP30 + DP60 + DP90 + DP120. This will be referred to as the Information Subset model II. This model contains 4 additional OP's that are evenly spaced. The objective is to determine if this type of spacing is advantageous.
- (6) Base points plus selected subset III (BP-3) = Base Case + DP23 + DP30 + DP40 + DP60 + DP90. This will be referred to as the Information Subset

model III. This model contains 5 additional OP's and it omits the last two OP's. The objective is to determine if these last OP's are necessary.

(7) Base points, and selected subset IV (BP-4) = Base Case + DP23 + DP30 + DP60 + DP90 + DP120. This will be referred to as the Information Subset model IV. This model contains five additional OP's and it omits two OP's. The objective is to determine if these OP's are necessary.

These settings are depicted in Table 8 below.

Scenario	DP															
	180	120	90	60	40	30	23	18	13	10	7	5	3	2	1	0
1 (BP-AA)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2 (BP-ALT1)		x		x		x		x	x	x	x	x	x	x	x	x
3 (BP-ALT2)	x		x		x		x	x	x	x	x	x	x	x	x	x
4 (BP-1)				x	x	x	x	x	x	x	x	x	x	x	x	x
5 (BP-2)		x	x	x		x		x	x	x	x	x	x	x	x	x
6 (BP-3)			x	x	x	x	x	x	x	x	x	x	x	x	x	x
7 (BP-4)		x	x	x		x	x	x	x	x	x	x	x	x	x	x

Table 8. Observation Point Scenarios to be tested

For any other OP scheme not utilizing all available data collections points, those columns of data are first deleted before any pickup calculations are done. For example, in Table 3: DP28, DP42, and DP56 could be deleted to create another OP scheme. The pickup rates and transition probabilities are then calculated as shown before in Section 4.2.2 (d). That is, for any of the alternative OP schemes, other than the scheme utilizing all available data, the pickup values between successive OPs are computed by aggregating the pickups between the skipped OPs. For example, in the OP scheme utilizing all available data, the pickup between DP56 and DP49 for the March 12th flight would be 5 bookings and from DP49 to DP42 the pickup would be 2 bookings. However, utilizing an alternative OP scheme in which the DP49 data would be deleted, the pickup from DP56 to DP42 would then be equal to 7 bookings which is the aggregated amount of the skipped pickup information.

4.5.2 Aggregation by Fare Classes

Airlines use single-letter class codes to distinguish between different fare products. For example, Y might be used for full-fare coach, M and Q for discounts, V for deeper discounts, and other classes, which vary by airline (Brumelle and McGill, 1993). Modern revenue management systems forecast demand for each one of these fare classes by using historical booking data from the same fare class of similar flight departures. This data is usually aggregated by departure time, day of the week and time of day (Lee, 1990).

The number of fare classes available for this study was six, all belonging to the same cabin, with two of them having relatively higher fares than the other four classes. The only difference within the classes was the type of amenities offered in the two higher fare-paying classes compared to the remaining four lower paying fare classes. This allows us to select a subset of these classes to determine if the fare structure could be reduced or condensed, possibly allowing for an increase in revenue. It also allows us to determine if the number and type of fare classes to be forecasted affects the accuracy of the forecasting methods.

The following three (3) subsets of fare classes will be tested:

- (1) All available fare classes (FC-AA) = Y + B + M + H + K + V. This will be referred to as the All Fare Class model. It contains the full complement of 6 fare classes. This could be considered as the status quo (or base case) scenario.

(2) Collapsed subset I (FC-1) = (Y+B) + (M+H) + (K+V). This will be referred to as the 3-fare class model. The fares used would now be calculated as an average of the two highest fare classes followed by an average of the next two highest fare classes, then followed by an average of the last two lowest fare classes. The objective is to determine if the number of fare classes could be reduced while possibly increasing revenues.

(3) Collapsed subset II (FC-2) = (Y+B) + (M+H+K+V). This will be referred to as the 2-fare class model. This fare subset calculates the resulting fares as an average of the two higher fare classes followed by an average of the next four lower fare classes. The objective is to determine if the number of fare classes could be reduced to a bare minimum (for example, business v. leisure) while possibly increasing revenues.

These settings are depicted in Table 9 below.

<i>Scenario</i>	<i>Classes</i>					
1 (FC-AA)	Y	B	M	H	K	V
2 (FC-1)	Y + B		M + H		K + V	
3 (FC-2)	Y + B		M + H + K + V			

Table 9. Fare Class Scenarios to be tested

4.5.3 Censored and Uncensored Data

The number of seats an airline can sell on a flight is determined by the booking limits set by the revenue management system. An airline continues to accept reservations in a fare class until the booking limit is reached. At that point, the airline stops selling seats in that fare class. It also stops collecting valuable data. Demand for travel in that fare class may exceed the booking limit, but the data does not reflect this. In such cases, the data is said to be censored or “constrained” (Zeni, 2001).

For this study data was available for flights which possessed no constraining (they were never sold out).

4.6 Measures of Performance

The output from each of the various models consists of the forecasted and the observed or actual final reservations for each scenario as well as the errors or residuals (defined as the difference between the observed and forecasted values). The amount of forecasting error using any of the statistical measures of accuracy is calculated only at time of departure (DP0). The error is defined as the difference between the actual number booked at DP0 and the number of forecasted bookings for DP0. The standard numerical measures of performance for forecasting will be used as well as one non-standard measure. These are as follows:

- **The Mean Absolute Deviation (MAD)** – is the average of the absolute values of the forecast errors. It is by far the easiest statistical measure of forecast errors. It is defined mathematically as:

$$MAD = \frac{1}{n} \sum_{i=1}^n |Actual_i - Forecast_i|$$

This type of error calculation is very useful when the cost of forecasting errors is proportional to the absolute value of the error.

- **The Mean Square Error (MSE)** – is the average of the squared differences between the observed and forecasted values. Because of this type of calculation, large deviations tend to be highlighted due to the squared term. It is defined mathematically as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Actual_i - Forecast_i)^2$$

Using MSE as a primary error measurement does indicate that the user prefers to have several smaller deviations compared to one large deviation.

- **The Mean Absolute Percentage Error (MAPE)** – is the average of the absolute values of the percentage errors. The main advantage the MAPE offers is that it is a magnitude-free calculation as compared to the MAD and MSE. That is, if the quantity being forecasted is measured in the thousands or millions, then the MAD and MSE values are correspondingly large. The MAPE is defined mathematically as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Actual_i - Forecast_i}{Actual_i} \right| \times 100\%$$

The MAPE as a result, is one of the easiest measurements to interpret since it is not dependent on the magnitude of the data, that is, it is dimensionless. A significant disadvantage with the MAPE however, is that it is undefined when the actual number of bookings is equal to zero.

- **The Theil's Inequality Coefficient (U)** – is the square root of the ratio of the MSE to the sum of the mean square values of the actual and predicted reservations. It is defined mathematically as:

$$U = \sqrt{\frac{MSE}{\frac{1}{n} \sum_{i=1}^n (Actual_i)^2 + \frac{1}{n} \sum_{i=1}^n (predicted_i)^2}}$$

The usefulness of this measurement is that the numerator represents the actual forecast error and the denominator represents a comparative statistic, which

normalizes the overall result so that it becomes dimensionless like the MAPE. However, much like the MAPE, the Theil's coefficient is undefined when the denominator is zero. The range of this coefficient is such that (perfect prediction) $0 \leq U \leq 1$ (poor prediction). This coefficient would be the best determinant of the relative performances of each model. Furthermore, the proportions of inequality, which decomposes the Theil's Inequality into bias, variance and co-variance proportions, can be used to determine the source of the forecast error.

4.7 Data Properties

Data was obtained from Zeni's (2001) research on unconstraining the demand estimates from censored data. The author started with a large data set obtained from a major airline and samples of flights with no data censoring throughout their entire booking history were selected. As a result this data set represented the actual demand for the flights in the sample.

The data represented 1727 flights departing during 1997 and 1998. These were produced from a mixture of high demand and low demand flights. High demand flights are classified as those flights whereby the total bookings at departure were greater than or equal to 96. Low demand flights were defined as those with total bookings at departure of lesser than or equal to 80. There are at least 20 observations for each flight, which are then artificially censored with 7 degrees of censoring for each observation. The observations were censored by reducing the capacity in decrements of 10% starting from 100%.

There are 6 fare classes, all within the economy cabin for each flight. Within each of the fare classes there are 2 rows of data. The first row represents the uncensored demand and also has the fare value. The second row represents the censored demand for the given reduced capacity. For example, Y_A is actual uncensored demand for the Y class. Y_C is the censored demand for the Y class. Demand is listed by the number of days prior to departure (DP) with DP0 being demand at departure.

Flight numbers were disguised (to protect the airline), but are recorded so that the data could be grouped appropriately. The complete history of a flight could therefore be determined by grouping at the flight number. For each flight there are 16 observation points corresponding to various days prior to departure ranging from zero to 180 days out. There were 6 fare classes (Y B M H K V) of data available for each flight and various combinations will be used each time since in this study all of the seats are physically indistinguishable. That is, there is no separate business or leisure class, just different fare levels.

Other properties of the data were that: 1) all of the flights were weekday departures and holiday periods were excluded. This sample selection excluded the possibility of data spikes or highly seasonal effects, 2) each flight also had a minimum of 20 observations up to a maximum of 43 observations. The variation in the number of observations was due to schedule changes, flight cancellations and irregular operations (Zeni, 2001).

4.8 Experimental Design

The experiment is initially conducted in two phases: Phase I involves utilizing the actual historical data to evaluate the forecasting methods and Phase II involves making comparisons under the various simulated environmental conditions. The details of each phase are follows:

4.8.1 Phase I

- Determines the parameters necessary for using Markov Chain Transition Matrices as a forecasting tool according to the proposed methodology (See Sections 3.2.2)
- Compares the three forecasting methodologies currently in use with the proposed methods by using the statistical measures of accuracy
- Selects the best parameter values for use in the proposed method as input to the simulation model for Phase II experimentation
- Reduces the experimental design settings based on initial results

This phase utilizes the base level of each factor which also represents the conditions under which actual data was collected. The demand data will be the uncensored values. The base case scenario for the capacity of the aircraft will be set at actual levels. The demand used is the observed one. These settings are illustrated in Table 10 below. Each setting will be evaluated using the previously outlined accuracy criteria.

Decision Factor	Settings
Observation Points (OP's)	OP16 – OP 1 (that is, DP0 to DP180)
Fare Classes (FC)	Y, B, M, H, K, V (actual)
Data Type (DT)	Uncensored
Environmental Factors	Settings
Capacity of the aircraft (C)	Observed: High and Low Levels
Probability Distribution	Observed

Table 10. Base Case Scenario

The Base Case Scenario is used to compare the chosen pickup forecasting methods (CP, AP, and MP) with the proposed Markov Chains (MC) transition matrices method using the accuracy criteria previously outlined.

4.8.2 Phase II

- Utilizing the best values of the parameters for the proposed model as determined in Phase I, simulates its performance under the various environmental conditions and decision factors as outlined below
- Compares the three forecasting methodologies with the proposed method under the simulated environmental conditions and decision factors, using statistical measures of accuracy

A full factorial analysis will be conducted on the remaining scenarios. Here, a combination of the factors will determine which forecasting method works best under the various scenarios. These will be simulated for a total of 42 scenarios (see Table 11). For this study data was available for flights which possessed no constraining (they were never sold out).

This data was then artificially constrained to obtain actual and constrained data at various levels of censoring. These levels started from 100% (no censoring) and were decremented by 10% each time up to a censored level of 30%. As a result we have eight levels of censoring available. However, since no differences may be observed at the 90% level of censoring or too many differences may be observed at the 30% level of censoring, a compromised level of 70% was chosen for the simulation runs.

Decision Factor	Levels	Settings
Observation Points (OP's)	7	BP-AA,....,BP-4
Fare Classes (FC)	3	FC-AA, FC-1, FC-2
Data Type (DT)	2	Uncensored, Censored @70%
Environmental Factors	Levels	Settings
Capacity of the aircraft (C)	1	Observed
Probability Distribution	1	Observed
Demand Variability (σ)	1	Observed
Total number of cases	42	

Table 11. Summary of the number of cases to be simulated

4.9 Simulation and Computer Modeling

The simulated forecasting environments represent possible scenarios that can occur given the factors under investigation. The basic assumptions for the remaining factors which are not investigated are that the bookings are independent, there were no extraneous events (acts of nature) influencing demand, and the sample of data was randomly selected which would not influence the outcome of our experiment.

This study delves into more detail than previous studies where the demand distribution was typically assumed to be of a Normal pattern for sake of simplicity. This simulation model does not assume that the bookings follow any predictable distribution except that the lower fares are booked in advance of higher fares mainly because of the restrictions imposed by the airlines on the different fare types.

A simulated computer modeling of this system is done using a combination of Microsoft Access programming and Microsoft Excel. Data analysis is conducted using Excel and SAS. Data validation and test runs are done using the sample size determined at the 95% confidence level. Various statistical measurements of forecast accuracy such as MAD, MSE, MAPE, and the U statistic are used to assess the accuracy of the

forecasting. ANOVA is used for comparative analysis of the performance of the proposed system versus the other three chosen systems under the various environmental conditions.

For the simulation aspect, common random numbers (CRN's) are used for each replication to reduce variation and to provide a greater level of precision (Law and Kelton, 2007). The empirical data collected from the real reservation system are used to validate the simulation model by comparing it to the output from the computer-generated model.

From the database of historical reservations booking data, demand is forecasted at the various fare levels offered. These forecasts could then serve as input to optimization models, which determines yield for the different fare levels. However, since we do not consider the no-show rates we are forced to assume an overbooking level of zero. Yield is measured at time of departure (DP0) for all scenarios. This causes all scenarios to be equally biased since they are all measured at the same end-point.

4.9.1 The Number of Replications

Each iteration of the simulation begins with the same set of initial conditions and demand is generated according to the various probability distributions. One disadvantage of the fixed-sample-size procedure based on n replications is that the analyst has no control over the confidence-interval half-length [or the precision of $\bar{X}(n)$]; for fixed n , the half-length will depend on $\text{Var}(X)$ - the population variance of the X_j 's (Law and Kelton, 2007). Therefore, we need a procedure, which provides the basis for determining the number of replications.

Let \bar{e} be the average error from the forecasted demand. Then if we want to estimate \bar{e} , with a confidence level of 95% and a relative error of 0.05 the number of replications needed is given by:

$$n_r^*(\gamma) = \min \left\{ i \geq n : \frac{t_{i-1, 1-\alpha/2} \sqrt{S^2(n)/i}}{|\bar{e}(n)|} \leq \gamma' \right\} \text{ where}$$

$\gamma' = \frac{\gamma}{1+\gamma}$ is the “adjusted” relative error needed to get the *actual* relative error of γ .

$n_r^*(\gamma)$ is solved iteratively by increasing i by 1 until both inequalities are satisfied (Law and Kelton, 2007).

In our case, $\gamma = 0.05$, which implies $n_r^*(\gamma) = 150$ replications. Therefore, for each set of decision factors as well as environmental conditions, we will replicate it 150 times. We can then be sure that the estimate of \bar{e} has a relative error of at most 0.05 with 95% certainty.

4.9.2 Limitations of the Model

Previous studies have all concurred that the performance of a particular forecasting methodology is highly dependent upon numerous factors such as the amount of historical data, the length of the time-period being forecasted as well as the method itself. In the airline industry, the amount of historical data is certainly the least troublesome.

In today’s computing environment with storage capacity getting cheaper by the month and data compression techniques being perfected, the amount of data stored by the

airlines is unmatched. However, because of the proprietary nature of this data and its value to each airline, obtaining it is indeed the limitation. Since travel offerings vary according to markets, fares offered, and aircraft capacity, this study is limited in scope by the kind of data available (see Section 4.7).

One of the major assumptions made when modeling airline demand is the independence of fare classes. This study, and others like it, has assumed that the fare classes are independent and that the lower fare classes book before the higher fare classes. In practice, this is not entirely true as some business class fares are booked relatively early in advance because of the perks offered by business class travel. Also, since in practice some fare classes are deemed to be sold out (because of booking limits), passengers wishing to travel on a specific itinerary may be forced to “buy-up” to a higher fare class. From an airline perspective of course, this is a sell-up situation.

This research does not model sell-ups explicitly because to do so would require an entirely different set of data about consumer buying behavior and that is probably the one area in which the airline does not maintain data. Most airlines, including the one from which the data set was obtained, do handle sell-ups. The notable exceptions are those airlines, which are now introducing or have been using a standardized pricing option.

Chapter 5

Results

This chapter analyzes the results in two phases, parallel to the experimental design described in Chapter 4. In Phase I, the accuracy of the proposed forecasting method using Markov Chains transition matrices and the currently used models in the industry are compared on an actual data set. This base case scenario comparison represents the simplest level of the experimental and decision factors which were summarized in Table 10. The statistical measures of performance used to determine the effectiveness of the proposed model are then calculated.

Phase II of the analysis compares the performance of the proposed method to others under the various simulated environmental conditions and decision factors as outlined in the experimental design. Each experimental scenario is assessed via the measures of performance to determine which forecasting method works best under which environmental scenario(s).

5.1 Phase I Analysis

This phase compares the proposed transition matrices method with existing methods used for micro-level forecasting to determine if they compare favorably. As previously illustrated in a worked example, there are two slightly different ways of calculating the Advanced Pickup (AP) and the Multiplicative Pickup (MP) forecasts. The difference in the calculation of the transition probabilities arises in the choice of including all of the known data points or utilizing a particular subset of most recent departures. The second method is in effect a simple moving average and could be influenced by any recent trends. Because of

these two differing options, there are two Advanced Pickup methods and two Multiplicative Pickup methods operationalized.

5.1.1 Capacity Analysis

In preliminary analyses, two types of aircraft capacities were considered: High and Low. Based on the data available with capacities ranging from 76 to 156 seats, it was determined that capacities in the range of 96 and above (the upper third of the capacity range) were designated as High capacity; and capacities below 80 (the lowest third) were designated as Low capacity.

As Figure 3 shows, there is no discernible pattern of forecasting accuracy across all capacities for any forecasting method.

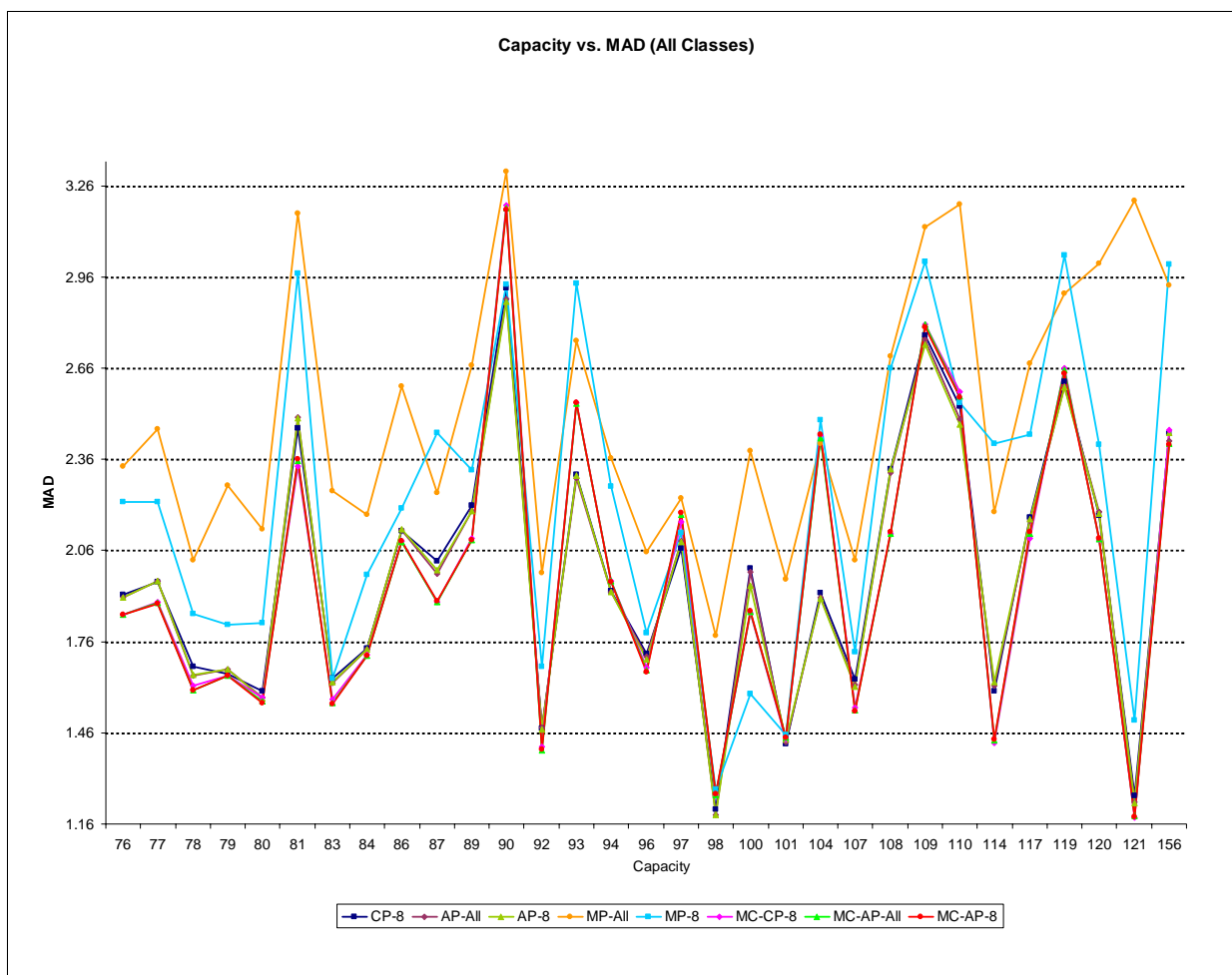


Figure 3. Capacity vs. MAD for all classes

When the data were analyzed across all capacities, ANOVA and Tukey's tests on all of the other performance measurements (MSE, MAPE, and U) also confirm this finding. These findings are illustrated by Figures I.1 – I.3 and Tables I.1 – I.4 in Appendix I.

As a result, this initial discretionary segregation of the data set will be dropped and the entire data set will be treated as homogenous with analysis being performed on capacities ranging from 76 to 156 seats. Consequently, capacity will be eliminated from future consideration as an environmental factor in further experimentation and analysis.

5.1.2 Fare Class Analysis - Overall

In this study, there were six fare classes available for analysis: V, K, H, M, B, and Y, arranged in ascending order of prices. Based on the data available, it was determined that the two distinctly higher paying fare classes (B and Y) could be considered as equivalent to “business” or “first-class” seats and the remaining four fare classes (V, K, H, and M) are to be considered part of a “leisure” or “economy” cabin even though there was no physical cabin partitioning. In the base case scenario, all of these fare class data are analyzed.

Table 12 which lists the methods in descending order for each performance measure shows that the MC-AP-All method performed best for the lower (V, K, H, and M) classes. The MC-AP-All method was tied for best performance with the MC-AP-8 method in the B class which represented the second highest fare class. The AP-All method was however the best performer overall in the Y class. For all of these classes, both of the multiplicative pickup methods (MP-xx) were the worst performers. (Note: In Table 12, bold horizontal lines separate statistically significant methods as per Tukey tests with Alpha = 0.05).

Class	MAD	MSE	MAPE	U value
All	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8	AP-All AP-8 CP-8 MC-AP-All MC-AP-8 MC-CP-8	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8
	<i>MP-8</i> <i>MP-All</i>	<i>MP-8</i> <i>MP-All</i>	<i>MP-8</i> <i>MP-All</i>	<i>MP-8</i> <i>MP-All</i>
V	MC-AP-8 MC-AP-All MC-CP-8 AP-8 AP-All CP-8 MP-8	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8 MP-8	MC-AP-All MC-AP-8 MC-CP-8 MP-8 AP-All AP-8 CP-8	MC-AP-All MC-AP-8 MC-CP-8 MP-8 AP-All AP-8 CP-8
	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>
K	MC-AP-All MC-AP-8 MC-CP-8 AP-8 AP-All CP-8 MP-8	AP-8 AP-All CP-8 MP-8 MC-AP-All MC-AP-8 MC-CP-8	MC-CP-8 MC-AP-All MC-AP-8 MP-8 AP-8 AP-All CP-8	MC-AP-8 MC-AP-All MC-CP-8 MP-8 AP-8 AP-All CP-8
	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>
H	MC-AP-All MC-AP-8 MC-CP-8 CP-8 AP-All AP-8 MP-8	CP-8 AP-8 AP-All MC-CP-8 MC-AP-All MC-AP-8 MP-8	MC-AP-All MC-CP-8 MC-AP-8 CP-8 AP-All AP-8 MP-8	MC-CP-8 MC-AP-All MC-AP-8 CP-8 AP-All AP-8 MP-8
	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>	<i>MP-All</i>
M	MC-AP-All MC-AP-8 MC-CP-8 AP-All AP-8 CP-8	MC-AP-8 MC-AP-All MC-CP-8 AP-8 AP-All CP-8	MC-AP-All MC-AP-8 MC-CP-8 AP-All AP-8 CP-8	AP-All AP-8 CP-8 MC-AP-All MC-CP-8 MC-AP-8
	<i>MP-All</i> <i>MP-8</i>	<i>MP-All</i> <i>MP-8</i>	<i>MP-All</i> <i>MP-8</i>	<i>MP-All</i> <i>MP-8</i>
B	MC-AP-All MC-AP-8 MC-CP-8 AP-8 AP-All CP-8 MP-All MP-8	MC-CP-8 MC-AP-All MC-AP-8 CP-8 AP-All AP-8 MP-8 MP-All	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8 MP-All MP-8	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8
				<i>MP-8</i> <i>MP-All</i>
Y	AP-All AP-8 CP-8 MC-AP-8 MC-CP-8 MC-AP-All MP-All	AP-All AP-8 CP-8 MC-AP-All MC-AP-8 MC-CP-8 MP-All	MC-AP-8 MC-AP-All MC-CP-8 AP-All AP-8 CP-8 MP-All	AP-All CP-8 AP-8 MC-AP-8 MC-AP-All MC-CP-8
	<i>MP-8</i>	<i>MP-8</i>	<i>MP-8</i>	<i>MP-All</i> <i>MP-8</i>

Table 12. Rankings of the forecasting methods overall and by class

5.1.3 Fare Class Analysis - Individual

a) V Class

For the V class, which is the lowest “economy” fare class, studied in this data, the MC-AP-All and the MC-AP-8 methods were tied for best on average according to the performance measures. For this fare class, the MC-AP-8 was best according to the MAD and MSE and the MC-AP-All was best according to the MAPE and U values. ANOVA analysis followed by Tukey’s testing on all of the performance measurements indicated significant p -values and significant differences between means for the above fare class across all capacities.

b) K Class

For the K class, which is the second lowest “economy” fare class, studied in this data, the MC-AP-All proved best on average, but the MSE values indicated superiority for the AP-8 method. This was one of the few instances where any of the performance measures proved better for the pickup methods compared to the transition matrices methods. For this class, the CP-8 method proved to be the second worst. A breakdown by the MAD, MSE, MAPE performance measures and the Theil’s co-efficient reinforces these findings as evidenced by Table 12. Again, ANOVA analysis followed by Tukey’s testing of all of the performance measurements indicated significant p -values and significant differences between means for the above fare class across all capacities as shown in Table 12.

c) H Class

For the H class the transitions matrices methods MC-AP-All and MC-CP-8 proved best on average, but the MSE values indicated superiority for the Classical Pickup method

CP-8. This was another instance where any of the performance measures proved superior to transition matrix based methods. ANOVA analysis followed by Tukey's testing on all of the performance measurements indicated significant p -values and significant differences between means for the above fare class across all capacities. Whether it is an anomaly or not will be further investigated with the larger simulated data set in Phase II.

d) M Class

For the M class, which is the highest "economy" fare class, studied in this data, the MC-AP-All proved best on average, but the U values indicated superiority for the AP-All method. This was another instance where any of the performance measures for the pickup methods proved better than the performance measures for any of the transition matrices methods. For this fare class, the best Pickup method (AP-All) was equal in performance to the worst performing of the Markov Chain methods (MC-CP-8). Again, ANOVA analysis followed by Tukey's testing on all of the performance measurements indicated significant p -values and significant differences between means for the above fare class across all capacities.

e) B Class

For the B class, the MC-AP-8 method proved best according to the MAPE and U performance measurements, second best according to the MAD and third best according to the MSE measurement. For all capacities, there was remarkable consistency amongst the three transition matrix based methods. In comparison, across the other 5 pickup methods, there were significant fluctuations from one method to another.

Analysis of the MSE, MAPE, and U values reveal similar disparities among the methods. Since the B class is the second most expensive fare class where passengers tend to book much closer to the time of departure than earlier in the booking profile, this may tend to skew the pickup ratios used for the multiplicative pickup methods and as a result those methods would be the least accurate of all of the methods. Performing ANOVA analysis and Tukey's test on the B class performance measurements show no significance at the 0.05 p -value level and no differences between means for any of the performance measurements except for the U values.

f) Y Class

Whereas the MC-AP-8 was only best in 2 of the 4 performance measurements for the B class, it was only best in the MAPE category for the Y class, which was the highest fare class data available. Analysis of the MSE, MAPE and U values showed that the AP-All method was the best performer for this fare class.

When ANOVA and Tukey's test were performed on the Y class data, only the MAPE values showed no significance at the 0.05 p -value level and no differences between means. For all of the other performance measurements in the Y class there is statistical significance at the 0.05 p -value level and a difference between means according to the Tukey's test.

Since the Y fare-class is the most expensive fare class and closest in designation to a business or first class, it is intuitively understandable that the pickup methods perform the best in this class. As the booking-profile tends to be more uniform earlier and fares are booked much later, this leads to rapid changes in demand as the time to departure nears. As

a result, the pickup ratios and averages tend to be much greater which in turn can lead to fewer fluctuations in the forecasts.

g) Conclusions

Based on the previous analysis, the following generalizations are possible:

- The Markov Chains transition matrices methods performed best for the lower-fare classes. This could possibly be attributed to the fact that the booking profiles for these classes demonstrate a gradual increase as compared to the higher-fare “business type” classes where there are sudden drastic increases in the booking profiles at particular observation points.
- Forecasts for the lower-fare classes proved to be the most accurate for all methods. Again, this could be attributed to the natural rule of thumb, which indicates that lower fare classes book before higher fare classes and are thus easier and more accurate to forecasts.

Even though nested booking limits are used (whereby the higher fare is always available), the very nature of consumers’ behavior dictates that the lower fares will be bought first and well in advance. Consequently, reducing the number of fare classes may force customers to purchase up to a higher fare level, and could enable better forecast accuracy.

5.1.4 Summary of Phase I Analysis

Phase I analysis supports the following conclusions:

- *Capacity is irrelevant to forecast accuracy*

Preliminary analysis showed that there was no pattern in forecasting accuracy for any method as capacity varied. This was verified by using a statistical Runs test for randomness (i.e. Wald-Wolfowitz test).

- *Multiplicative Pickup (MP) methods can be dropped from further experimental consideration*

ANOVA results showed statistically significant differences between the MP methods and all other methods. The MP methods were shown to be the worst performing and as a result, both of the MP methods can be dropped from further experimentation in Phase II.

- *Markov Chains (MC) transition matrices approaches promises superior accuracy*

Overall, across all capacities and all fare classes, the MC methods demonstrated superiority when compared to the pickup methods. Additionally, of the 24 fare-class/performance measurement combinations possible, the MC methods were top contenders in 18 instances and were ranked only inferior than the pickup methods (AP-All, AP-8 and CP-8) for the Y class. However, the differences were statistically insignificant. The MC methods were never the worst performing for any class or at any capacity according to any of the performance measurements, a distinction only held by the multiplicative pickup methods.

Three of the six instances (H, K, and Y classes) where the MC methods were outperformed were in the MSE category. Based on these results, it is apparent that the MC approaches promise superior accuracy.

5.2 Phase II Analysis

Phase II involves simulation experiments to generalize and further analyze the findings from Phase I results.

Simulation of the demand data involved bootstrapping from the actual data used in Phase I analysis. Bootstrapping from the original observations was done as follows:

1. From the original data set, at each flight and fare class level, calculate the average pickup values. It should be noted that there were an unequal number of observations for each flight number. The number of observations ranged from a minimum of 20 observations to a maximum of 43 observations. As a result, the number of averaged observations varies.
2. At each flight and fare class level, calculate the average of all of the DP180 values to be used as the new starting value for the simulated data set.
3. Since the simulated data set is to consist of a minimum of 150 replications at each flight/fare class combination, populate a new data table with 150 entries of the averaged DP180 value.
4. Selects one row at random from the pickup table created in Step 1 and adds these values to the averaged DP180 value, to obtain the DP120 data.
5. Repeat for all possible flight and fare class combinations.

Because of bootstrapping, the simulated data set could be considered “distribution-free.” (For complete details on the simulation of a new dataset, see Appendix II) This in effect, allows us to ignore probability distributions and resultant demand variations from consideration as environmental factors. It should be noted that even though bootstrapping was utilized to make the simulated data distribution-free, Mann-Whitney tests were performed to ensure that the simulated data resembled the actual data set. Distribution fitting techniques performed on the simulated data set also indicated similarities between the actual and simulated data sets.

As (i) the multiplicative pickup methods have proven inferior; and (ii) the capacity of the aircraft has been deemed irrelevant to the forecasting methodologies, we now focus on the decision factors as illustrated in Table 13.

Decision Factor	Levels	Settings
Forecasting methods	6	CP, AP-All, AP-8 (EP), MC-CP, MC-AP-All, MC-AP-8
Observation Points (OP's)	7	BP-AA, ..., BP-4
Fare Classes (FC)	3	FC-AA, FC-1, FC-2
Data Type (DT)	2	Uncensored, Censored @70%
Total number of cases	252	

Table 13. Revised summary of the number of cases simulated

5.2.1 ANOVA RESULTS

As the ANOVA results show in Table 14, the type of forecasting method used is statistically significant according to all of the statistical models except the MAPE measure. The ANOVA results also demonstrate that the number of Fare Classes used indeed have an effect on all of the models. This decision factor is statistically significant at the $\alpha = 0.05$ level for all measures. Further, the ANOVA results illustrate that the number of

OP's and the Type of Data utilized have an effect on all of the performance measures. These decision factors are statistically significant at the $\alpha = 0.05$ level for all measures. Results also show that the MAD measure had the strongest R^2 value compared to all of the other measures.

MAD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	94072.16	1190.79	491.32	<.0001
Error	15040	36451.92	2.42		
Corrected Total	15119	130524.08			
R-Square	Coeff Var	Root MSE	MAD Mean		
0.72	20.48	1.56	7.60		
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Fare_Class	2	68163.68	34081.84	14062.1	<.0001
OP	6	1316.43	219.40	90.53	<.0001
Fare_Class*OP	12	89.69	7.47	3.08	0.0002
Type	1	11933.11	11933.11	4923.58	<.0001
Type*Fare_Class	2	2425.48	1212.74	500.37	<.0001
Type*OP	6	32.61	5.43	2.24	0.0364
Method	5	5841.82	1168.36	482.07	<.0001
Fare_Class*Method	10	3099.29	309.93	127.88	<.0001
OP*Method	30	289.23	9.64	3.98	<.0001
Type*Method	5	880.82	176.16	72.69	<.0001

MSE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	106906722.8	1353249.7	276.96	<.0001
Error	15040	73487727.8	4886.2		
Corrected Total	15119	180394450.6			
R-Square	Coeff Var	Root MSE	MSE Mean		
0.592628	56.81244	69.90102	123.0382		
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Fare_Class	2	59278499.25	29639249.62	6065.97	<.0001
OP	6	534849.42	89141.57	18.24	<.0001
Fare_Class*OP	12	193138.65	16094.89	3.29	<.0001
Type	1	18249317.04	18249317.04	3734.91	<.0001
Type*Fare_Class	2	8932097.03	4466048.51	914.02	<.0001
Type*OP	6	49232.30	8205.38	1.68	0.1216
Method	5	8755102.20	1751020.44	358.36	<.0001
Fare_Class*Method	10	7672782.61	767278.26	157.03	<.0001
OP*Method	30	51027.27	1700.91	0.35	0.9996
Type*Method	5	3190677.02	638135.40	130.60	<.0001

MAPE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	22057.71	279.21	59.34	<.0001
Error	15040	70767.52	4.71		
Corrected Total	15119	92825.23			
R-Square	Coeff Var	Root MSE	MAPE Mean		
0.24	141.41	2.17	1.53		
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Fare_Class	2	9443.73	4721.86	1003.52	<.0001
OP	6	444.34	74.06	15.74	<.0001
Fare_Class*OP	12	108.16	9.01	1.92	0.0279
Type	1	8323.44	8323.44	1768.95	<.0001
Type*Fare_Class	2	3355.37	1677.69	356.55	<.0001
Type*OP	6	130.41	21.74	4.62	0.0001
Method	5	114.48	22.90	4.87	0.0002
Fare_Class*Method	10	9.71	0.97	0.21	0.9958
OP*Method	30	126.32	4.21	0.89	0.6312
Type*Method	5	1.75	0.35	0.07	0.9961

U

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	79	20.38	0.26	148.24	<.0001
Error	15040	26.18	0.002		
Corrected Total	15119	46.56			
R-Square	Coeff Var	Root MSE	U Mean		
0.44	20.68	0.042	0.203		
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Fare_Class	2	11.478	5.739	3297.02	<.0001
OP	6	0.767	0.128	73.48	<.0001
Fare_Class*OP	12	0.036	0.003	1.74	0.0524
Type	1	4.867	4.867	2796.27	<.0001
Type*Fare_Class	2	1.522	0.761	437.31	<.0001
Type*OP	6	0.028	0.005	2.65	0.0142
Method	5	1.114	0.223	128.04	<.0001
Fare_Class*Method	10	0.264	0.026	15.15	<.0001
OP*Method	30	0.271	0.009	5.19	<.0001
Type*Method	5	0.036	0.007	4.15	0.0009

Table 14. Overall ANOVA results for Decision Factors

5.2.2 Main Effects

In examining the conflicting results provided by the four performance measures in detail (See Table 14), the following observation was made: as the MAPE and U statistics are scalar measurements, they tend to exaggerate errors when applied to actual observations of small magnitude. In view of this fact, the primary emphasis is placed on the MAD and MSE statistics in further analysis.

a) Forecasting Method Used

ANOVA results indicate that the Advanced Pickup method is best for the forecasting models compared to all other methods according to the MAD and MSE statistical measures. These findings are illustrated in Table 15. Tukey's tests also indicate that the Pickup Methods are statistically significantly different from the MC methods according to both the MAD and MSE measures.

	FC Method	Fare Classes	OP's	Data Type
MAD	4 = 6 = 5 > 1 > 3 = 2	FC-2 > FC-1 > FC-AA	3 = 2 > 5 = 1 = 7 = 6 > 4	U > C
MSE	4 = 6 = 5 > 1 = 3 = 2	FC-2 > FC-1 > FC-AA	1 = 3 = 2 = 5 = 7 = 6 > 4	U > C
MAPE	4 = 5 = 6 = 1 = 2 = 3	FC-AA > FC-1 > FC-2	3 = 2 = 5 = 7 = 1 = 6 = 4	C > U
U	4 = 6 = 5 > 1 = 3 = 2	FC-AA > FC-1 > FC-2	3 = 2 > 5 = 1 = 7 > 6 > 4	C > U
<i>Key</i>	1 = Classical Pickup [CP] 2 = Advanced Pickup [AP-All] 3 = Advanced Pickup [AP-8] 4 = MC- CP 5 = MC-AP-All 6 = MC-AP-8	FC-AA = All available fare classes FC-1 = 3 fare class system FC-2 = 2 fare class system	1 = BP-AA 2 = BP-ALT1 3 = BP-ALT2 4 = BP-1 5 = BP-2 6 = BP-3 7 = BP-4	U = Uncensored C = Censored

Table 15. Rankings of Decision Factor Schemes According to Performance Measures

(Note: Statistically significant differences (i.e. using Tukey's test at $\alpha > 0.05$ level) are indicated by '>', insignificant differences are indicated by '='). Individual analysis results are included in Appendix I.

b) **Observation Points Utilized**

The results indicate that the BP-1 scheme, which includes the base case (from DP0 to DP18) and a selected subset of the most recent additional OP's (DP23, DP30, DP40, and DP60), is the best overall. The OP schemes BP-ALT1 (Base Case + DP30 + DP60 + DP120) and BP-ALT2 (Base Case + DP23 + DP40 + DP90 + DP180) proved to have the

worst effect on the forecasts as illustrated by the rankings according to the performance measures in Table 15.

Tukey's tests results indicate that OP scheme 4 (BP-1) is significantly different from all other OP schemes and is the best. Consequently, the focus for further analysis will be shifted to comparisons between OP scheme 4 (BP-1) – a proposed scenario; and the industry standard of OP scheme 1 (BP-AA).

c) **Number of Fare Classes Used**

The results indicate that it is the fare class scheme FC-AA, which includes all of the available fare classes individually (disaggregated) that is the best overall according to the MAD and MSE performance measures. The FC scheme FC-2 (which aggregates the two highest paying fare classes and then aggregates the remaining four fare classes) proved to have the worst effect on the forecasts according to the MAD and MSE measurements as illustrated by the rankings in Table 15.

Tukey's tests results indicate that all of the fare class schemes are significantly different from each other with a disaggregated system (FC-AA) being the best. Thus, the focus for further analysis will be shifted to the fare class structure of FC-AA, as the two other proposed aggregation scenarios, FC-1 and FC-2, have proven to be inferior.

This finding, that a disaggregated forecasting system is the most effective while being counterintuitive to statistical principles, confirms previous findings. As discussed earlier, in the hotel industry, Weatherford et al. (2001) presented a model for testing aggregation versus disaggregation when used to forecast revenues. Their results indicated

that full disaggregation by rate category and length of stay produces smaller forecast errors.

d) **Type of Data used**

As shown in Table 15, forecasting methods perform better with censored (or constrained) data compared to uncensored data according to both the MAD and MSE measures. However, this is merely a self-fulfilling prophecy, because if the data is censored or constrained, in effect, it prevents the error rates from ballooning up because the forecasts level out when the censoring limit is reached.

Even though Tukey's tests results indicate that Censored data (C) is significantly different from Uncensored (U) data according to all of the statistical models, further analysis will continue to focus on both Censored data (which is the prevailing industry standard) and Uncensored data which represents recent industry practices experimenting with single fare systems.

5.2.3 **Interactions between Main Effects**

a) **Forecasting Method and Data Type**

As shown by the ANOVA results in Table 14, the interactions between forecasting method and data type used is statistically significant according to both the MAD and MSE measures. Further analysis of the crosstabs results shown in Table 16 indicates that the interactions are indicative of more pronounced differences between methods for uncensored data, as compared to censored data. This is illustrated in Table

16 where the solid horizontal lines depict statistically significant differences according to the Tukey tests.

Level of Type	Level of method	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
C	2	1260	6.2866	2.0713	77.4329	46.7140
C	3	1260	6.2919	2.0739	77.6309	46.8854
C	1	1260	6.4312	2.1382	81.4010	49.4263
C	5	1260	7.0641	2.4044	97.0403	61.4936
C	6	1260	7.0662	2.4050	97.1276	61.5586
C	4	1260	7.1451	2.4235	99.1482	62.3990
		Avg.	6.7141		88.2968	
U	2	1260	7.5721	2.2456	117.2528	65.4506
U	3	1260	7.5808	2.2501	117.5162	65.6673
U	1	1260	7.7390	2.2943	123.0034	69.3472
U	5	1260	9.3234	3.7810	195.4173	170.5959
U	6	1260	9.3273	3.7803	195.4682	170.4063
U	4	1260	9.4029	3.7812	198.0199	171.6575
		Avg.	8.4909		157.7796	

Table 16. Crosstabs Table showing 2nd order means and standard deviations for method and data type interactions

b) Forecasting Method and the Number of Observation Points utilized

The interaction between the forecasting method used and the number of observation points is only statistically significant according to the MAD measure as illustrated in [Table 14](#). It was hypothesized that because of the dynamism involved in the pickup rates, the number of observation points utilized would indeed have an effect on the accuracy of the forecasting methods. However, this hypothesis is consistently disproved by the second order interactions, which indicate that regardless of which system of OP's are used, the rankings of the forecasting methods are the same.

Results indicate that the Advanced Pickup method utilizing all available data is best for the forecasting models compared to all other methods. This result can be attributed to a number of factors as follows:

- The Markov transition matrices methods are not as accurate in long-horizon predictions;
- The rate of change of bookings (pickup rates) are more accurately captured by a deterministic system (e.g. Advanced Pickup), rather than a probabilistic or stochastic type system (e.g. transition matrices) initially; and
- When the transition matrices methods begin to experience forecasts inaccuracies, those trends are irreversible.

When Tukey Tests (illustrated in Table 17) are done across individual forecasting methods for each OP scheme, OP scheme 4 (BP-1) or OP scheme 6 (BP-3) consistently performs the best, while OP scheme 3 (BP-ALT2) performs the worst according to the MAD measure. As previously mentioned, OP scheme 2 (BP-ALT1) comes in as the second worst method overall when assessing the effect of the number of OP's used on the forecasting accuracy. These findings are shown in Table 18 where the solid horizontal lines depict statistically significant differences.

AP (2)	N	Subset for alpha = .05			MC_AP (5)	N	Subset for alpha = .05		
		1	2				1	2	3
OP4	360	6.4382			OP4	360	7.5542		
OP6	360	6.8285	6.8285		OP6	360	7.9154		
OP7	360		7.0077		OP7	360	7.9990	7.9990	
OP5	360		7.0100		OP1	360	8.1355	8.1355	8.1355
OP2	360		7.0103		OP5	360	8.2596	8.2596	8.2596
OP1	360		7.1040		OP2	360		8.7035	8.7035
OP3	360		7.1067		OP3	360			8.7890
Sig.		.229	.642		Sig.		.070	.071	.120
EP (3)	N	Subset for alpha = .05			MC_EP (6)	N	Subset for alpha = .05		
		1	2				1	2	3
OP4	360	6.4428			OP4	360	7.5557		
OP6	360	6.8362	6.8362		OP6	360	7.9182		
OP7	360		7.0157		OP7	360	8.0024	8.0024	
OP2	360		7.0180		OP1	360	8.1383	8.1383	8.1383
OP5	360		7.0180		OP5	360	8.2637	8.2637	8.2637
OP1	360		7.1113		OP2	360		8.7072	8.7072
OP3	360		7.1126		OP3	360			8.7918
Sig.		.222	.650		Sig.		.068	.070	.119
CP (1)	N	Subset for alpha = .05			MC_CP (4)	N	Subset for alpha = .05		
		1	2				1	2	3
OP4	360	6.5740			OP4	360	7.6305		
OP6	360	6.9823	6.9823		OP6	360	7.9973		
OP2	360		7.1684		OP7	360	8.0831	8.0831	
OP5	360		7.1684		OP1	360	8.2237	8.2237	8.2237
OP7	360		7.1684		OP5	360	8.3387	8.3387	8.3387
OP1	360		7.2671		OP2	360		8.7798	8.7798
OP3	360		7.2671		OP3	360			8.8648
Sig.		.208	.644		Sig.		.069	.078	.136

Table 17. Tukey Test results for comparisons of FC methods and OPs

Level of OP	Level of method	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
4	2	360	6.4382	2.1076	83.8013	51.2019
4	3	360	6.4428	2.1086	83.9396	51.1925
4	1	360	6.5740	2.1620	87.9674	54.5137
4	5	360	7.5542	3.4405	134.8179	137.4721
4	6	360	7.5557	3.4398	134.8512	137.3016
4	4	360	7.6305	3.4471	136.8308	138.5993
		Avg.	7.0325		110.3680	
6	2	360	6.8285	2.2084	94.1454	57.6768
6	3	360	6.8362	2.2119	94.3620	57.7816
6	1	360	6.9823	2.2721	98.8010	61.3467
6	5	360	7.9154	3.4698	146.7008	145.6118
6	6	360	7.9182	3.4696	146.7579	145.4953
6	4	360	7.9973	3.4803	148.9765	146.7698
		Avg.	7.4130		121.6239	
7	2	360	7.0077	2.2659	99.5299	61.1622
7	3	360	7.0157	2.2699	99.7870	61.3783
7	1	360	7.1684	2.3246	104.4834	64.5686
7	5	360	7.9990	3.2939	142.5632	137.0562
7	6	360	8.0024	3.2941	142.6509	136.9575
7	4	360	8.0831	3.3020	145.0224	137.9852
		Avg.	7.5461		122.3395	
1	2	360	7.1040	2.2973	102.3343	63.2551
1	3	360	7.1113	2.3021	102.5949	63.5143
1	1	360	7.2671	2.3543	107.5984	66.8521
1	5	360	8.1355	3.4831	154.3340	149.8476
1	6	360	8.1383	3.4825	154.3820	149.7065
1	4	360	8.2237	3.4923	156.8751	150.9109
		Avg.	7.6633		129.6865	
5	2	360	7.0100	2.2677	99.5749	61.2255
5	3	360	7.0180	2.2720	99.8322	61.4626
5	1	360	7.1684	2.3246	104.4834	64.5686
5	5	360	8.2596	3.2277	143.0720	132.5867
5	6	360	8.2637	3.2282	143.1593	132.4931
5	4	360	8.3387	3.2333	145.4690	133.3580
		Avg.	7.6764		122.5985	
2	2	360	7.0103	2.2677	99.5927	61.2435
2	3	360	7.0180	2.2715	99.8370	61.4749
2	1	360	7.1684	2.3246	104.4834	64.5686
2	5	360	8.7035	3.2356	149.9177	128.8904
2	6	360	8.7072	3.2356	149.9965	128.7892
2	4	360	8.7798	3.2400	152.2609	129.5751
		Avg.	7.8978		126.0147	
3	2	360	7.1067	2.2983	102.4213	63.3172
3	3	360	7.1126	2.3034	102.6622	63.5832
3	1	360	7.2671	2.3543	107.5984	66.8521
3	5	360	8.7890	3.2362	152.1960	128.4783
3	6	360	8.7918	3.2376	152.2876	128.4362
3	4	360	8.8648	3.2398	154.6535	129.3338
		Avg.	7.9887		128.6365	

Table 18. Crosstabs Table showing 2nd order means and standard deviations for method and data type interactions.

c) **Forecasting Method and the Number of Fare Classes used**

Overall, the interaction between the forecasting method and the number of fare classes used is statistically significant according to both the MAD and MSE measures. When Tukey tests (shown in Table 19) are done across all forecasting methods for each fare class scheme, fare class scheme FC-AA consistently is the best according to both the MAD and MSE measures. The first order interactions also indicate that regardless of which fare class scheme is used, the rankings of the forecasting methods are the same throughout. These results indicate the Advanced Pickup forecasting method utilizing all available data is the best under any fare class structure for the simulated dataset. These findings are shown in Table 20 where the solid horizontal lines depict statistically significant differences between fare class scheme averages.

AP (2)	N	Subset for alpha = .05			MC_AP (5)	N	Subset for alpha = .05		
		1	2	3			1	2	3
FC-AA	840	4.7411			FC-AA	840	5.1360		
FC-1	840		7.1966		FC-1	840		8.0588	
FC-2	840			8.8503	FC-2	840			11.3864
Sig.		1.000	1.000	1.000	Sig.		1.000	1.000	1.000
<hr/>									
EP (3)	N	Subset for alpha = .05			MC_EP (6)	N	Subset for alpha = .05		
		1	2	3			1	2	3
FC-AA	840	4.7445			FC-AA	840	5.1380		
FC-1	840		7.2057		FC-1	840		8.0640	
FC-2	840			8.8589	FC-2	840			11.3882
Sig.		1.000	1.000	1.000	Sig.		1.000	1.000	1.000
<hr/>									
CP (1)	N	Subset for alpha = .05			MC_CP (4)	N	Subset for alpha = .05		
		1	2	3			1	2	3
FC-AA	840	4.8385			FC-AA	840	5.2110		
FC-1	840		7.3600		FC-1	840		8.1456	
FC-2	840			9.0567	FC-2	840			11.4653
Sig.		1.000	1.000	1.000	Sig.		1.000	1.000	1.000

Table 19. Tukey Test results for comparisons of FC methods by fare class

Level of Fare Class	Level of method	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
FC-AA	2	840	4.7411	0.9814	48.2901	19.9878
FC-AA	3	840	4.7445	0.9828	48.3831	20.0149
FC-AA	1	840	4.8385	1.0063	50.6790	21.1688
FC-AA	5	840	5.1360	1.0847	55.3929	23.5994
FC-AA	6	840	5.1380	1.0861	55.4556	23.6194
FC-AA	4	840	5.2110	1.0991	56.9883	24.1697
		Avg.	4.9681		52.5315	
FC-1	2	840	7.1966	1.4735	96.9441	40.6542
FC-1	3	840	7.2057	1.4777	97.1989	40.8188
FC-1	1	840	7.3600	1.5030	101.7970	43.0851
FC-1	5	840	8.0588	1.7119	124.3021	62.6472
FC-1	6	840	8.0640	1.7148	124.4330	62.7075
FC-1	4	840	8.1456	1.7230	126.7102	63.5294
		Avg.	7.6718		111.8975	
FC-2	2	840	8.8503	1.8846	146.7943	63.1159
FC-2	3	840	8.8589	1.8884	147.1386	63.3637
FC-2	1	840	9.0567	1.9367	154.1306	67.2290
FC-2	5	840	11.3864	3.2045	258.9914	175.0969
FC-2	6	840	11.3882	3.2032	259.0051	174.8544
FC-2	4	840	11.4653	3.2114	262.0536	176.0981
		Avg.	10.1677		204.6856	

Table 20. Crosstabs Table showing 2nd order means and standard deviations for method and fare class interactions.

d) The Number of Observation Points and the Data Type used

In general, the interaction between the number of observation points and data type is only statistically significant according to the MAD measure. Again, it was hypothesized that the data type would have no effect on the accuracy of the forecasts regardless of which OP scheme was used. This is because the data is not constrained (censored) only until the booking-limits have been reached for a particular class, and there is no distinction between censored and uncensored data up to the observation point at which censoring starts.

This is evident by the results of the second-order interactions indicated by Table 34, which show that for the two data types, BP-1 (OP4) is the best performer according to all performance measures and is statistically different from other methods according to both the MAD and MSE statistic. Additionally, OP scheme 3 (Base Case + DP23 + DP40 + DP90 + DP180) proved to be the worst performer according to all performance measures.

Tukey tests (enumerated in Table 21) across all OP schemes by data type indicate that the forecasting methods perform better with censored data. These findings are shown in Table 22 where the solid horizontal lines indicate statistically significant differences. The focus of future analyses will be confined to OP schemes 1 and 4 (BP-AA and BP-1 respectively).

CENSORED DATA (MAD)							UNCENSORED DATA (MAD)				
OP	N	Subset for alpha = .05					OP	N	Subset for alpha = .05		
		1	2	3	4	5			1	2	3
4	1080	6.0864					4	1080	7.9787		
6	1080		6.4692				6	1080	8.3568	8.3568	
7	1080		6.6499	6.6499			7	1080		8.4423	8.4423
1	1080		6.7517	6.7517			5	1080		8.5437	8.5437
5	1080			6.8091	6.8091		1	1080		8.5749	8.5749
2	1080				7.0656	7.0656	2	1080		8.7301	8.7301
3	1080					7.1673	3	1080			8.8100
Sig.		1.000	.058	.660	.117	.944	Sig.		.092	.100	.111

Table 21. Tukey Test results for comparisons of OPs by Data Type

Level of Type	Level of OP	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
C	4	1080	6.0864	2.1035	76.2315	49.0912
C	6	1080	6.4692	2.1878	84.7233	53.8184
C	7	1080	6.6499	2.2050	86.8078	53.8827
C	1	1080	6.7517	2.2849	92.2791	59.2892
C	5	1080	6.8091	2.2481	88.4743	54.2487
C	2	1080	7.0656	2.3932	93.4960	58.4777
C	3	1080	7.1673	2.4191	96.0656	60.2598
		Avg.	6.8188		90.3077	
U	4	1080	7.9787	3.2756	144.5046	134.9472
U	6	1080	8.3568	3.3180	158.5245	143.0254
U	7	1080	8.4423	3.1734	157.8711	134.8167
U	5	1080	8.5437	3.1356	156.7226	130.9143
U	1	1080	8.5749	3.3349	167.0938	147.1703
U	2	1080	8.7301	3.1566	158.5333	127.3446
U	3	1080	8.8100	3.1627	161.2074	127.1087
		Avg.	8.5763		159.9921	

Table 22. Crosstabs Table showing 2nd order means and standard deviations for number of OPs and data type interactions.

e) **The Number of Observation Points and the Number of Fare Classes used**

Overall, the interaction of the number of observation points and fare class is statistically significant according to the both the MAD and MSE measures. It was hypothesized that because of the booking profiles of certain fare classes, the number of observation points utilized would have an effect on the forecasts by fare class. For example, in the Y and B classes (“business” classes) there are generally no bookings prior to DP23; consequently utilizing OP scheme 4 (BP-1) should give the best results since it only uses the most recent OP’s up to DP60.

Table 23 which shows the Tukey tests results indicate that for all fare class schemes; OP scheme 4 (BP-1) or OP scheme 6 (BP-3) is the best performer according to

the MAD statistical measure. This finding is also replicated by the Tukey test results of the MSE measure. OP scheme 3 (BP-ALT2), which utilizes the base case plus four unevenly spaced OP's (Base Case + DP23 + DP40 + DP90 + DP180) proved to be the worst performer according to both statistical measurements. This is summarized in Table 24 where the solid horizontal lines depict statistically significant differences.

The results also show that the fare class scheme FC-AA under OP scheme 4 is the best performer. Under these settings, only infatuation effects are evident and not cross-effects.

FARE CLASSES DATA (MAD)

FC-AA	N	Subset for alpha = .05			
		1	2	3	4
OP4	720	4.4584			
OP6	720		4.7361		
OP7	720			4.9279	
OP1	720			4.9306	
OP5	720			5.0814	
OP2	720				5.2836
OP3	720				5.3591
Sig.		1.000	1.000	.065	.799
FC-1	N	Subset for alpha = .05			
		1	2	3	4
OP4	720	6.9892			
OP6	720		7.4164		
OP7	720		7.5998	7.5998	
OP1	720			7.6880	
OP5	720			7.7830	
OP2	720				8.0624
OP3	720				8.1638
Sig.		1.000	.323	.325	.898
FC-2	N	Subset for alpha = .05			
		1	2		
OP4	720	9.6500			
OP6	720	10.0863	10.0863		
OP7	720		10.1104		
OP1	720		10.1649		
OP5	720		10.3476		
OP2	720		10.3713		
OP3	720		10.4430		
Sig.		.067	.231		

Table 23. Tukey Test results for comparisons of OPs by Fare Classes

Level of Fare Class	Level of OP	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
FC-AA	4	720	4.4584	0.9244	43.9843	19.0616
FC-AA	6	720	4.7361	0.9504	49.0591	20.8326
FC-AA	7	720	4.9279	0.9783	51.7363	21.3529
FC-AA	1	720	4.9306	0.9956	53.0435	22.4696
FC-AA	5	720	5.0814	1.0160	53.7719	21.7315
FC-AA	2	720	5.2836	1.1194	57.3618	23.6325
FC-AA	3	720	5.3591	1.1364	58.7637	24.2515
		Avg.	5.0531		53.9560	
FC-1	4	720	6.9892	1.5086	96.9879	51.3416
FC-1	6	720	7.4164	1.5349	107.9577	55.1249
FC-1	7	720	7.5998	1.5462	110.6302	53.3049
FC-1	1	720	7.6880	1.5853	115.7326	57.7018
FC-1	5	720	7.7830	1.5975	112.8130	52.6969
FC-1	2	720	8.0624	1.7565	118.2922	55.0222
FC-1	3	720	8.1638	1.7786	120.8692	56.0573
		Avg.	7.7856		114.3825	
FC-2	4	720	9.6500	2.9670	190.1318	143.1207
FC-2	6	720	10.0863	2.9647	207.8549	150.8545
FC-2	7	720	10.1104	2.8324	204.6519	142.1780
FC-2	5	720	10.1649	2.8389	201.2105	139.1300
FC-2	2	720	10.3476	2.9000	202.3902	135.7871
FC-2	1	720	10.3713	2.9487	220.2833	154.4404
FC-2	3	720	10.4430	2.8946	206.2766	134.7821
		Avg.	10.2539		207.1112	

Table 24. Crosstabs Table showing 2nd order means and standard deviations for number of OPs and fare class interactions.

f) The Number of Fare Classes and Type of Data used

It was hypothesized that there would be no interaction between the type of data used and the fare class scheme applied. However, to the contrary, the ANOVA results indicate that the interaction between the fare class scheme applied and the data type used is statistically significant according to both of the measures.

This is indeed borne out by the Tukey analysis of the second-order interactions as shown in Table 25, which indicate that all of the fare class schemes exhibit statistically

significant differences. As the cross-tabs results in Table 26 indicate (where the solid horizontal lines depict statistically significant differences), the forecasts are more accurate for censored data, especially when applied to the FC-AA fare class scheme.

CENSORED DATA (MAD) Tukey HSD ^a				
FARECLASS	N	Subset for alpha = .05		
		1	2	3
FC-AA	2520	4.4887		
FC-1	2520		6.9183	
FC-2	2520			8.7354
Sig.		1.000	1.000	1.000

UNCENSORED DATA (MAD)				
FARECLASS	N	Subset for alpha = .05		
		1	2	3
FC-AA	2520	5.4476		
FC-1	2520		8.4253	
FC-2	2520			11.5999
Sig.		1.000	1.000	1.000

Table 25. Tukey Test results for comparisons of Fare Classes by Data Type

Level of Type	Level of Fare Class	N	MAD		MSE	
			Mean	Std Dev	Mean	Std Dev
C	FC-AA	2520	4.4887	0.9543	42.0585	18.2510
C	FC-1	2520	6.9183	1.4888	86.1029	36.9359
C	FC-2	2520	8.7354	1.8764	136.7291	56.8207
		Avg.	6.7142		88.2968	
U	FC-AA	2520	5.4476	0.9352	63.0045	21.3109
U	FC-1	2520	8.4253	1.4667	137.6922	57.8521
U	FC-2	2520	11.5999	3.0630	272.6421	169.0072
		Avg.	7.3351		120.7388	

Table 26. Crosstabs results showing 2nd order means and standard deviations for data type and fare class interactions.

5.3 Phase II Conclusions

Phase II analysis supports the following conclusions:

- *Of the proposed Observation Point (OP) schemes, scheme 4 (BP-1) is the best*

In evaluating the overall ANOVA results, Scheme 4, which utilizes data only up to sixty days prior to departure, results in the best forecast accuracy. Curtailing data collection and the forecasting effort to the most recent sixty days avoids compounding of errors made when forecasting into the distant future. All pickup methods continue their initial distortions at early stages. These results suggest that focusing on shorter-term forecasts only enhances final forecast accuracy. The Markov transition matrices based methods also perform the best under this observation point scheme.

- *An individual (disaggregated) fare class structure is the best*

None of the proposed aggregated fare class structures proved better than the current disaggregated system used in practice. Contrary to well-known statistical principles, an individual fare class system (FC-AA) gives better forecasting accuracy than any proposed aggregation of fare classes. There are also no statistically significant differences between any of the methods within this fare class structure with regard to forecasting accuracy.

- *Forecasting Methods are more accurate for censored data.*

As discussed earlier, truncating the bookings data increases the forecasting accuracy, especially when the booking limit is reached, and the surplus demand is truncated, resulting in a constant booking profile for a fare class.

Chapter 6

Conclusions and Directions for Future Research

6.1 Summary of Research Findings

This study investigates a new approach to forecasting airline seat demand and its consequent application to revenue management. It is different than previous studies which utilized traditional forecasting methods such as Pickup methods, Time Series algorithms and Econometric based approaches. Even though those methods had some underlying moving average component, which made them dynamic, they were not truly interactive models. The application and use of Markov Chains based transition matrices allowed the proposed forecasting model to be stochastic and as a result, captured a different picture of the reservation system.

Ideally, the following states: available, booked, cancelled, departed, no-showed, or walkups are applicable for Transition Matrix modeling. Of these states, departures and no-shows are absorbing states; while available, booked, cancellations and walkups are transient states. However, because of data limitations, only a simple 2-state model consisting of the Available and Booked states was utilized. Research findings support the effectiveness of Transition Matrices modeling by using only the two the above-mentioned states. It is reasonable to postulate that the availability of historical transitional probability information among other states is likely to improve forecasting accuracy.

Initially, based on the literature review, five pickup methods were selected for comparison to the proposed transition matrices based methods. These methods were the Classical Pickup, the Advanced Pickup using all known departures, the Advanced Pickup using the 8 most recent departures, the Multiplicative Pickup using all known departures

and the Multiplicative Pickup using the 8 most recent departures. The 8 most recent departures were chosen simply because it was the observed number from the literature and it could possibly be interpreted as the bare minimum number of departures in order to obtain any meaningful results.

A similar reasoning was applied in constructing the transition matrices based methods. The first transition matrix method calculated probabilities in a manner comparable to the Classical Pickup method, the second transition matrix method calculated probabilities analogous to the Advanced Pickup method (using all known departures), and the third transition matrix method calculated probabilities similar to the Advance Pickup method utilizing the 8 most recent departures.

Comparisons of the results were done by the use of four performance measures. Three traditional performance measures used were the mean absolute deviation (MAD), the mean square error (MSE) and the mean absolute percentage error (MAPE). The other non-traditional performance measure utilized was the Theil's Inequality Co-efficient (U), an economics based forecasting accuracy measure, which compares the relative performances of each model. As the statistical analysis progressed, and because of conflicting results obtained by some of the measures (the MAPE and U which are scalar and tended to exaggerate smaller differences), the final analysis whittled down the performance measures to the MAD and MSE.

Initial investigations were based on artificially segregating the data into two types of capacities: High and Low. Preliminary analysis showed no discernible patterns between capacity and forecasting accuracy. Because of the preliminary findings, that capacity was irrelevant, the entire data set was recombined to be analyzed in its entirety.

This early investigation also showed that the transition matrices methods were better at short-term forecasting (which is needed for the airline industry).

The experimental design consisted of two phases. In Phase I, the accuracy of the proposed forecasting methods using Markov Chains transition matrices and the currently used models in the industry were compared on an actual data set. This base case scenario comparison represented the simplest level of the experimental and decision factors.

Phase II of the analysis compared the performance of the proposed methods to others under the various simulated environmental conditions and decision factors as outlined in the experimental design. Each experimental scenario was assessed via the measures of performance to determine which forecasting method works best under which environmental scenario(s). This phase of the experimental design involved a thorough evaluation of a simulated dataset to generalize the findings from Phase I. Bootstrapping was utilized to generate a distribution free random dataset. This technique was particularly advantageous because in almost all of the previous analyses covered by the literature, there was an assumption of an underlying probability distribution. Consequently, utilizing bootstrapping enabled the removal of this assumption.

6.2 Conclusions

The first objective of this research was to establish the feasibility of Markov Chains based transition matrices as a forecasting tool for airline seat demand. As the research findings conclusively demonstrate, Markov Chains based transition matrices are indeed a viable tool for airline seat demand forecasting. Further, it has been shown that this tool even offers an improvement over existing methods under certain conditions.

Phase I of the analysis provided the following conclusions: capacity was irrelevant to forecasting accuracy, the multiplicative pickup methods could be dropped from further experimental consideration because of their poor performance, and the transition matrices based approaches promised superior accuracy. The results of the ANOVA analysis and Tukey testing showed the same statistical significance when analyzing High, Low or All capacities. This confirmed that forecasting accuracy for any method is not affected by capacity. Similar testing also demonstrated that the multiplicative pickup methods were the worst performers and thus could be removed from further consideration.

In Phase I, across all capacities and all fare classes, the Markov Chains (MC) methods demonstrated superiority when compared to the pickup methods. Additionally, of the 24 fare-class/performance measurement combinations possible, the MC methods were best performers in 18 instances and were equally effective as others in the rest of instances.

The results of the ANOVA analysis of the main effects of four decision factors indicated some contradictory results for the four performance measures. Since the MAPE and U statistics are scalar measurements, and they tend to exaggerate errors when applied to actual observations of small magnitude, it was decided to concentrate solely on MAD and MSE for further analysis. The findings suggest that the forecasting method, the data type and the fare class scheme used do have a significant impact on the forecasting accuracy. Given the fact that most of the previous literature did not focus on these factors or any combination thereof, these results reveal new dynamics of forecasting environments.

In considering the second-order interactions, all were significant for either both, or one of the performance measures. Overall, Phase II analysis supported the following conclusions:

1) the observation point scheme which utilizes observations up to 60 days prior to departure is the best as demonstrated by the ANOVA results for BP-4,

2) an individual (disaggregated) fare class structure is best – this is contrary to common statistical principles but supported by one other study, and

3) the forecasting methods are more accurate for censored data – which is a self-fulfilling prophecy since the data is curtailed and thus there is no change in the forecasts.

As this research conclusively demonstrates, Markov Chains transition matrices are effective forecasting tools for airline seat demand. Findings also suggest that aggregation levels of fare classes impacts forecast accuracy in a negative manner, aggregation levels by time periods impacts forecast accuracy in that only the most recent observation points should be considered, and the forecasting model performance does indeed depend on (i.e. interacts with) aggregation levels of (i) fare classes and (ii) observation points utilized.

Since most reservation systems are usually available for bookings from 360 days prior to departure, and forecasts have to be made from this time henceforth, it then becomes obvious in terms of managerial implications and from an applications perspective; that a bridge between forecasting from DP_{360} until DP_{60} must be constructed. A hybrid forecasting system could therefore be the most effective under these circumstances. For example, the Classical Pickup or any one of the Advanced Pickup methods could be used from DP_{360} until DP_{60} and then the Markov Chains Transition

Matrices approaches can be utilized from DP_{60} until DP_0 . This would ensure the most accurate forecasting of the number of bookings.

The results from this research offer a new perspective on what has been a dormant area in revenue management. While implementation may be sometime in the future, it is hoped that the methods prescribed here can serve as a stimulant to change in the forecasting approaches utilized by the airline industry and other industries, which utilize revenue systems in a constrained capacity environment. Even if the transition matrices methodology is not used in its entirety, a hybrid approach is certainly not out of the question.

6.3 Directions for Future Research

The literature on empirically validated research in yield and revenue management is indeed rich and even though several studies support and build upon each other's findings, there are some that are contradictory. Most of the studies are somewhat narrow in focus, but this is due to the particular nature of yield management being made up of different components. The initial review of revenue management literature and practice demonstrates several areas have not been adequately addressed:

- (1) Most studies are somewhat narrow in focus in terms of systems integration;
- (2) While the majority of yield management applications involve multi-stage decisions, research efforts and implementation foci have been mostly limited to single-leg applications;

(3) Demand modeling efforts have assumed the providers' current market segmentation practices (i.e. fare-class structures) to be consistent with customer behavior and discounted the potential effects of complexity in pricing strategies.

One of the main findings of the literature review and an area of concern, which was not covered in this research, is the accountability for no-shows and walk-ups. These environmental variables have proven to be difficult to model and forecast primarily because of the lack of data availability.

Forecasting models in revenue management are primarily based on the independent-demand model, which is seen as affected by price and timing. However, demand models that consider availability of competitive alternatives and conformance of offerings to customer expectations are largely missing from forecasting models in practice. There also seems to be a gap between practice and research in understanding the dynamics between the extent of premiums paid (or penalties charged) for travel flexibility afforded to the customers and resultant no-show behavior of customers.

One of the major shortcomings of forecasting applications in yield and revenue management is the failure to obtain information on lost customers. Lost customer data, which may be cancellations or no-shows, offer vital information for use as input to the inventory control and overbooking modules. At this stage, it is impossible to assess the potential impact of this shortcoming on total system effectiveness. Identifying the profiles and analyzing the various reasons for cancellations and no-shows would definitely enable more accurate forecasting and facilitate better overbooking controls.

A frustration shared by both revenue management researchers and practitioners surrounds the analysis of data. Researchers have defined alternative forecasting and

optimization approaches that they believe may lead to improved revenue management decisions, but often cannot access the appropriate data to test out their theories (Lieberman, 2004). Another noteworthy characteristic of forecasting research in the context of revenue management is the clear lack of significant research that assesses the suitability of various forecasting tools for revenue management applications.

This study has identified the following areas where research agendas have not fulfilled the practitioners' needs or provided novel guidelines for implementation:

A data-driven issue has been identified as the countdown forecasting systems, an area that has not been addressed either as a stand-alone component, or as an integral part of the broader forecasting function.

Although overbooking research is probably most voluminous within the context of revenue management, its dependency on forecast accuracy has not yielded generalized conclusions. The culprit is the complementary nature of class segments within the same cabin, coupled with the pricing schemes. As industry applications experiment with reduced fare classes and limited price changes, researchers also face the challenges and opportunities to address the impact of such practices.

Related to the issue mentioned immediately above, is the gap in understanding customer preferences and their price-elasticity for the flexibility afforded by higher fare classes. Providing more insight to this issue would facilitate both better pricing decisions, as well as improved ability to forecast fare-class specific no-show rates.

Multiple-leg flights require the correct assessment of trade-offs involved for pricing and offering individual service components separately as opposed to collectively.

Efforts in this arena involve a spectrum of pricing decisions, which treat the origin-destination as a single service or the sum of individual components.

Appendix 1

Data Analysis Results

This appendix provides in detail, the results obtained for Phase I of the data analysis. Initially, two types of capacities were analyzed: High and Low. This segmentation was artificially done based on the data available. Even though neither the primary or secondary research questions are to determine the effect of capacity on forecasting accuracy, it is a necessary environmental condition to consider and as such it was included in the preliminary data analysis. As the results indicate, capacity indeed turns out to be a non-factor and could be eliminated from future consideration.

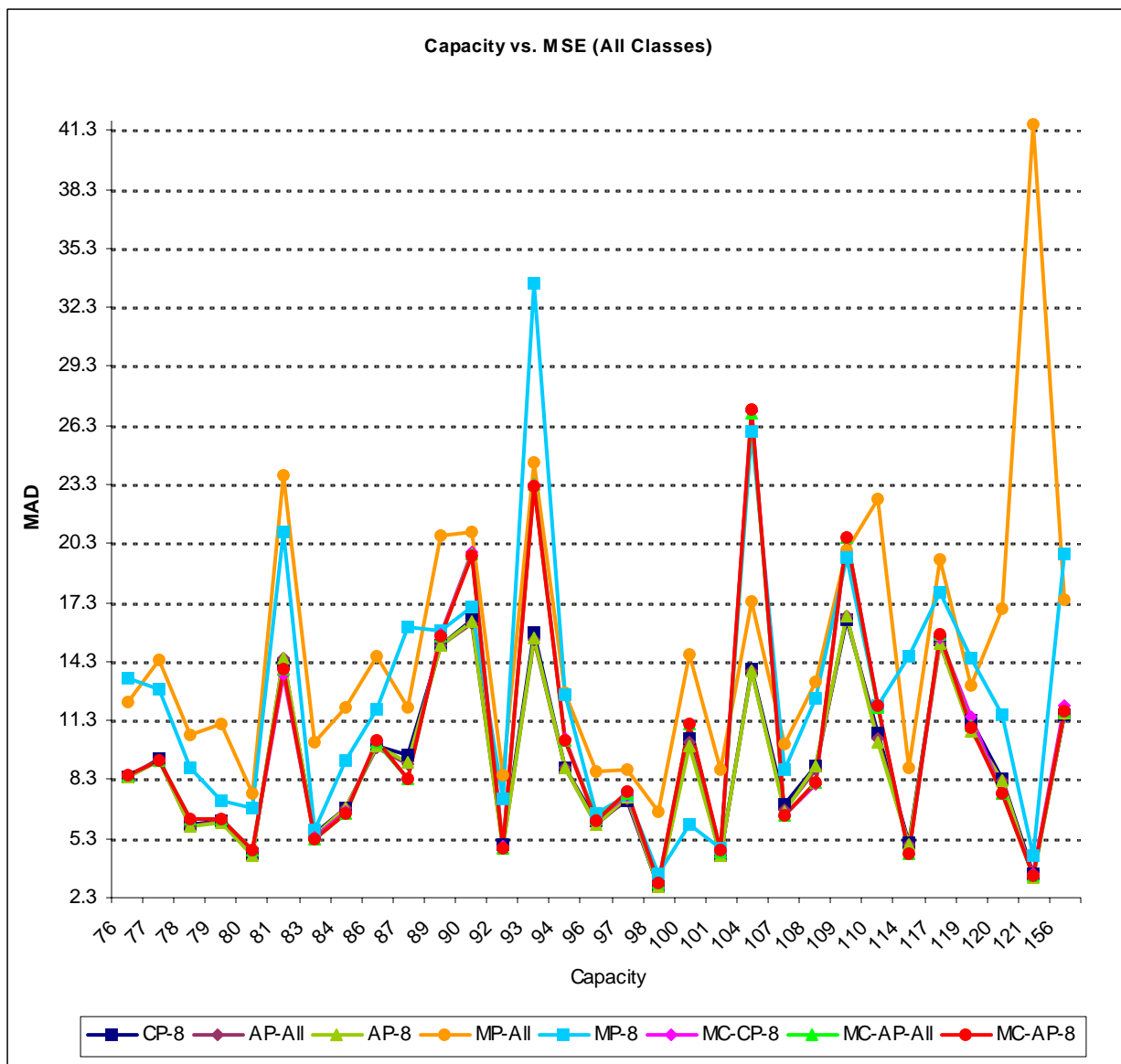


Figure I.1 Capacity vs. MSE for all classes

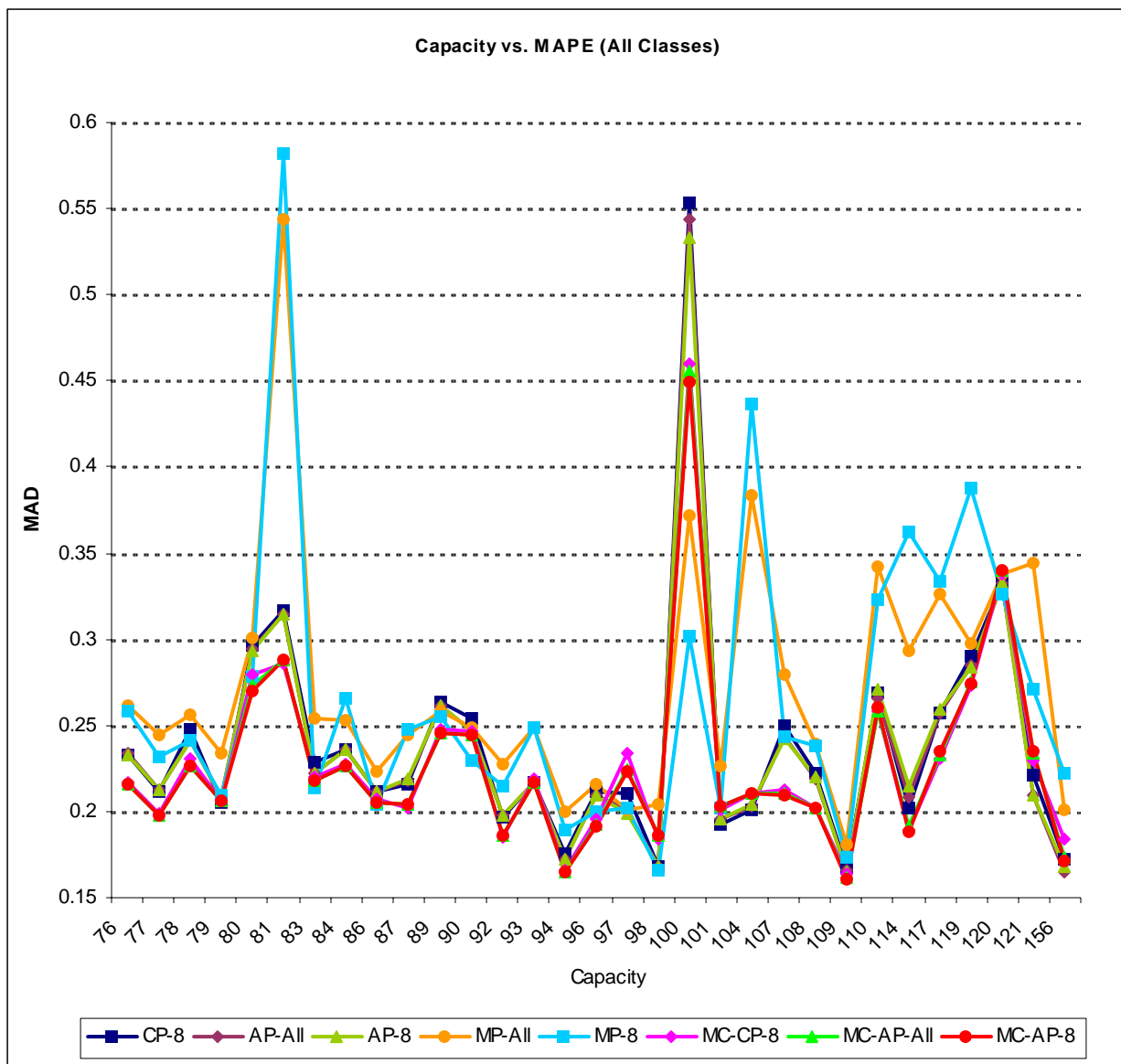


Figure I.2 Capacity vs. MAPE for all classes

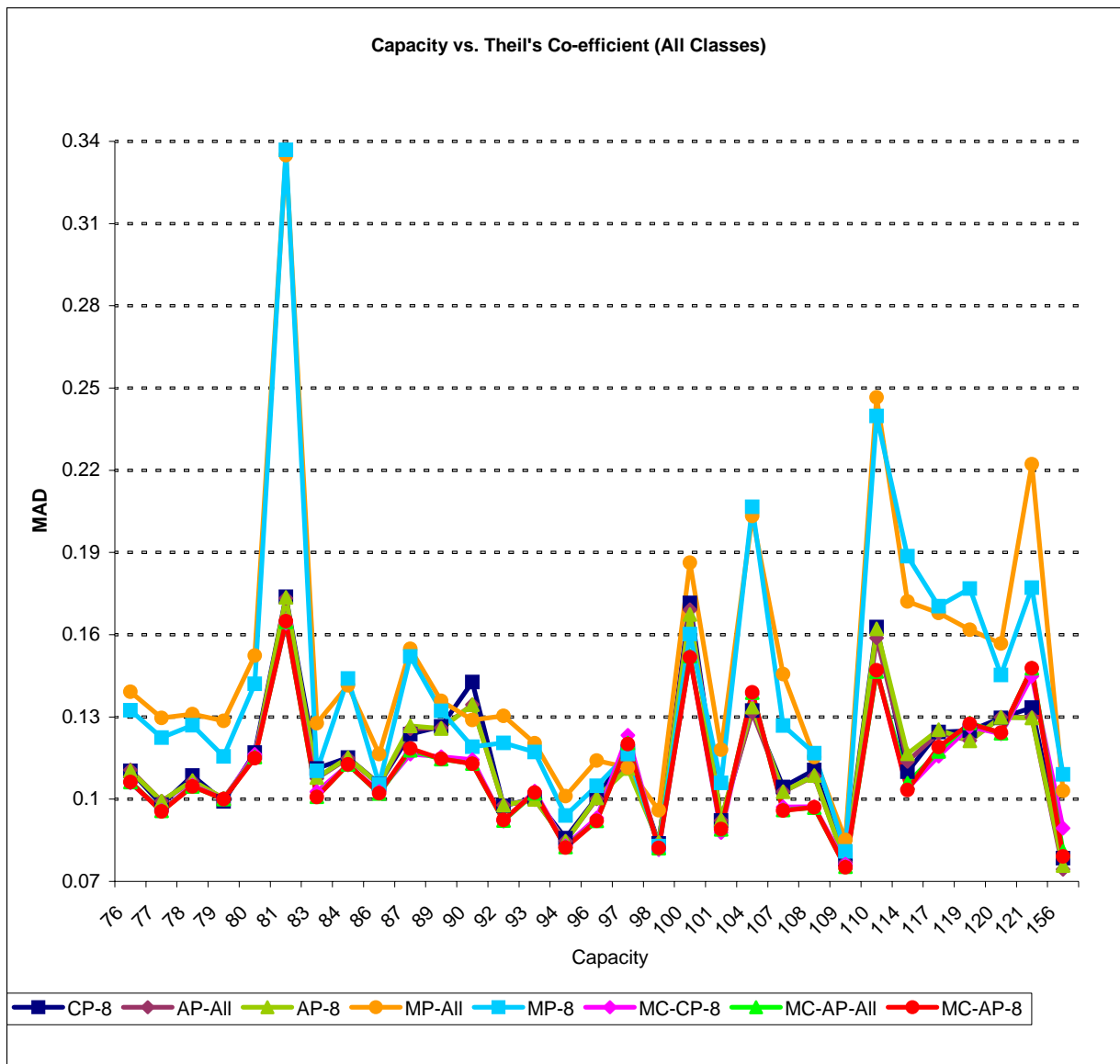


Figure I.4 Capacity vs. U for all classes

Anova: Single Factor **AllCapMAD**SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
CP-8	4980.0000	9609.0000	1.9295	5.3336
AP-All	4980.0000	9572.6222	1.9222	5.2791
AP-8	4980.0000	9574.2194	1.9225	5.2861
MP-All	4980.0000	11987.1865	2.4071	8.0643
MP-8	4980.0000	10825.0092	2.1737	7.9907
MC-CP-8	4980.0000	9481.6816	1.9040	6.1686
MC-AP-All	4980.0000	9467.2814	1.9011	6.1579
MC-AP-8	4980.0000	9466.1021	1.9008	6.1627

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	1201.5716	7.0000	171.6531	27.2232	0.0000	2.0098
Within Groups	251156.2252	39832.0000	6.3054			
Total	252357.7969	39839.0000				

		MC-AP-8	MC-AP-All	MC-CP-8	AP-All	AP-8	CP-8	MP-8	<u>MP-All</u>
		1.9008	1.9011	1.9040	1.9222	1.9225	1.9295	2.1737	<u>2.4071</u>
MC-AP-8	1.9008		No	No	No	No	No	Yes	Yes
MC-AP-All	1.9011			No	No	No	No	Yes	Yes
MC-CP-8	1.9040				No	No	No	Yes	Yes
AP-All	1.9222					No	No	Yes	Yes
AP-8	1.9225						No	Yes	Yes
CP-8	1.9295							Yes	Yes
MP-8	2.1737								Yes
<u>MP-All</u>	<u>2.4071</u>								

Table I.1. ANOVA and Tukey Tests results for All Capacities according to the MAD statistic

Anova: Single Factor **AllCapMSE**

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
CP-8	4980	45096.5017	9.0555	1025.1070
AP-All	4980	44685.4655	8.9730	1022.2638
AP-8	4980	44726.2660	8.9812	1026.2589
MP-All	4980	69006.0982	13.8566	1836.2853
MP-8	4980	63315.8928	12.7140	2304.3887
MC-CP-8	4980	48766.2688	9.7924	1615.3483
MC-AP-All	4980	48658.2986	9.7707	1603.8191
MC-AP-8	4980	48677.7195	9.7746	1606.8731

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	121046.1976	7	17292.3139	11.4896	0.0000	2.0098
Within Groups	59948873.8873	39832	1505.0430			
Total	60069920.0849	39839				

	AP-All	AP-8	CP-8	MC-AP-All	MC-AP-8	MC-CP-8	MP-8	<u>MP-All</u>
	8.9730	8.9812	9.0555	9.7707	9.7746	9.7924	12.7140	<u>13.8566</u>
AP-All	8.9730	No	No	No	No	No	Yes	Yes
AP-8	8.9812		No	No	No	No	Yes	Yes
CP-8	9.0555			No	No	No	Yes	Yes
MC-AP-All	9.7707				No	No	Yes	Yes
MC-AP-8	9.7746					No	Yes	Yes
MC-CP-8	9.7924						Yes	Yes
MP-8	12.7140							No
<u>MP-All</u>	<u>13.8566</u>							

Table I.2. ANOVA and Tukey Tests results for All Capacities according to the MSE statistic

AllCapMAPE

Anova: Single Factor

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
CP-8	4980	1141.07139	0.229130801	0.158007405
AP-All	4980	1137.042638	0.228321815	0.157658694
AP-8	4980	1137.282199	0.228369919	0.156992886
MP-All	4980	1291.047515	0.259246489	0.16978988
MP-8	4980	1254.073021	0.251821892	0.228591278
MC-CP-8	4980	1087.035802	0.218280282	0.139783944
MC-AP-All	4980	1082.054324	0.217279985	0.13930146
MC-AP-8	4980	1081.741695	0.217217208	0.138666394

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	8.907460541	7	1.272494363	7.898835013	1.32594E-09	2.009819866
Within Groups	6416.895071	39832	0.161098993			
Total	6425.802532	39839				

	MC-AP-8	MC-AP-All	MC-CP-8	AP-All	AP-8	CP-8	MP-8	MP-All
	0.217217	0.217280	0.218280	0.228322	0.228370	0.229131	0.251822	0.259246
MC-AP-8	0.217217	No	No	No	No	No	Yes	Yes
MC-AP-All	0.217280		No	No	No	No	Yes	Yes
MC-CP-8	0.218280			No	No	No	Yes	Yes
AP-All	0.228322				No	No	No	Yes
AP-8	0.228370					No	No	Yes
CP-8	0.229131						No	Yes
MP-8	0.251822							No
MP-All	0.259246							

Table I.3. ANOVA and Tukey Tests results for All Capacities according to the MAPE statistic

AllCapUvalue

Anova: Single Factor

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
CP-8	4980	546.7232914	0.109783793	0.023249374
AP-All	4980	544.7464923	0.109386846	0.023043527
AP-8	4980	545.3254445	0.109503101	0.023122082
MP-All	4980	690.4080121	0.138636147	0.039809202
MP-8	4980	657.3967909	0.132007388	0.041413988
MC-CP-8	4980	525.1207165	0.105445927	0.021560641
MC-AP-All	4980	524.2881751	0.10527875	0.021738889
MC-AP-8	4980	524.283089	0.105277729	0.021725175

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	6.047803067	7	0.863971867	32.04897847	1.06677E-44	2.009819866
Within Groups	1073.78547	39832	0.02695786			
Total	1079.833273	39839				

	MC-AP-8	MC-AP-All	MC-CP-8	AP-All	AP-8	CP-8	MP-8	MP-All
	0.105278	0.105279	0.105446	0.109387	0.109503	0.109784	0.132007	0.138636
MC-AP-8	0.105278	No	No	No	No	No	Yes	Yes
MC-AP-All	0.105279		No	No	No	No	Yes	Yes
MC-CP-8	0.105446			No	No	No	Yes	Yes
AP-All	0.109387				No	No	Yes	Yes
AP-8	0.109503					No	Yes	Yes
CP-8	0.109784						Yes	Yes
MP-8	0.132007							No
MP-All	0.138636							

Table I.4. ANOVA and Tukey Tests results for All Capacities according to the U statistic

Appendix 2

Data Simulation Procedure

This appendix describes in detail, the data simulation procedure used for Phase II of the data analysis.

In the literature review section of this research, it was observed that there are usually 3 types of distribution assumptions made for the bookings data: the Truncated Normal, the Censored Poisson, and the Adjusted Gamma. Using a distribution fitting software (ExpertFit), it was indeed determined that these were the three most likely distributions which could be fitted to the data. While this research is not an exercise in distribution fitting, it is necessary to determine the approximate distribution and parameters to ensure accurate replication of the simulated data set.

It was therefore thought that bootstrapping, a modern statistical technique based on resampling with replacement from the original samples could allow us to reduce the number of scenario analyses needed for Phase II. It does not require distribution assumptions and can be effectively used with smaller sample sizes. Even though the sample sizes are not small for the current data set, the number of scenarios is large. Initially, a total of 756 scenarios were accounted for primarily because the distribution patterns of the bookings data and their variabilities which needed to be modeled. Even though bootstrapping was to be applied, it is still necessary to check the resulting simulated data set for resemblance to the original data.

Bootstrapping of the original data was done as follows:

5. From the original data set, at each flight and fare class level, calculate the pickup values. It should be noted that there were an unequal number of observations for each flight number. The number of observations ranged from a minimum of 20 observations to a maximum of 43 observations. As a result, the number of rows in each pickup table would vary from 20 to 43.
6. At each flight and fare class level, calculate the average of all of the DP180 values to be used as the new starting value for the simulated data set.
7. Since the simulated data set is to consist of a minimum of 150 replications at each flight/fare class combination, populate a new data table with 150 entries of the averaged DP180 value.
8. Select one row at random from the pickup table created in Step 1 and add these values to the averaged DP180 value to obtain a new row of simulated data.
9. Repeat for all possible flight and fare class combinations.

The new (simulated) data set now consists of 60 flights times 150 replicates = 9000 flights. This was built up from the original data set of a total of 1727 flights. It should also be noted that one major problem was encountered and solved when simulating the new data set. Since it is entirely possible for some pickup values to be negative (representing cancellations), and when combined with the averaged DP180 value resulting in negative bookings at DP120 and other observation points; it was necessary to find a solution to this problem. As a result, 175 replicates were used and then the resulting data set was cleansed of any negative bookings. 175 replicates were used since it was observed that at the worst case there were approximately 23 rows (out of the initially simulated 150 rows) for some flight numbers which contained negative values.

After generation of the simulated data set, distribution fitting techniques performed on the simulated data set also indicated similarities between the actual and simulated data sets both with respect to the fitted distributions and their underlying parameters. Statistical verification was done by using Mann-Whitney tests.

BIBLIOGRAPHY

- Alstrup, J., Boas, S., Madsen, O.B.G., Vidal, R., and Victor, V. "Booking policy for flights with two types of passengers," *European Journal of Operational Research* (27), 1986, pp 274-288.
- Andrew, W., Cranage, D., and Lee, C. "Forecasting hotel occupancy rates with time series models: an empirical analysis," *Hospitality Research Journal* (14:2), 1990, pp 173-181.
- Baker, T.K., and Collier, D.A. "A comparative revenue analysis of hotel yield management heuristics," *Decision Sciences* (30:1), Winter 1999, p 239.
- Baliga, J. "Yield management relies more on integrated data handling," *Semiconductor International*), May 1998, p 13.
- Beckman, J.M. "Decision and team problems in airline reservations," *Econometrica* (26), 1958, pp 134-145.
- Beckman, M.J., and Bobkowski, F. "Airline Demand: An analysis of some frequency distributions," *Naval Research Logistics Quarterly* (5), 1958, pp 43-51.
- Belobaba, P., and Weatherford, L.R. "Comparing decision rules that incorporate customer diversion in perishable asset revenue management situations," *Decision Sciences* (27:2), Spring 1996, p 343.
- Belobaba, P.P. "Airline Yield Management An overview of seat inventory control," *Transportation Science* (21:2), May 1987, pp 63-73.
- Belobaba, P.P. "Application of a probabilistic decision model to airline seat inventory control," *Operations Research* (37:2), Mar 1989, pp 183-197.

- Belobaba, P.P. "Air travel demand and airline seat inventory management," Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1987.
- Ben-Akiva, M., and Lerman, S. *Discrete Choice Analysis: Theory and Application to Travel Demand* MIT Press, Cambridge, MA, 1985.
- Bernar, A. "Firms find more use for neural nets," *PC Week* (5:6), Feb 9, 1988.
- Bhatia, A.V., and Parekh, S.C. "Optimal Allocation of Seats by Fare," AGIFORS Reservations Study Group, Trans World Airlines, 1973.
- Bitran, G.R., and Mondschein, S.V. "An application of yield management to the hotel industry considering multiple day stays," *Operations Research* (43:3), May/June 1995, p 427.
- Bodily, S.E., and Weatherford, L.R. "Perishable-asset revenue management: Generic and multiple-price yield management with diversion," *Omega* (23:2), Apr 1995, p 173.
- Bollapragada, S., Cheng, H., Phillips, M., and Garbiras, M. "NBC's optimization systems increase revenues and productivity," *Interfaces* (32:1), Jan-Feb 2002, pp 47-60.
- Buhr, J. "Optimal sales limits for two sector flights," AGIFORS Symposium Proceedings, Lagonissi, Greece, 1982.
- Caixeat-Filho, J.V., van Swaay-Neto, J.M., and Wagemaker, A.d.P. "Optimization of the Production Planning and Trade of Lily Flowers at Jan de Wit Company," *Interfaces* (13:1), Jan-Feb 2002, pp 35-46.
- Carroll, W.J., and Grimes, R.C. "Evolutionary change in product management:

- Experiences in the car rental industry," *Interfaces* (25:5), Sept 1995, pp 84-104.
- Chase Jr., C.W. "Revenue management: A review," *The Journal of Business Forecasting Methods & Systems* (18:1), Spring 1999, p 2.
- Chatwin, R.E. "Multiperiod airline overbooking with a single fare class," *Operations Research* (46:6), Nov-Dec 1998, pp 805-819.
- Ciancimino, A., Inzerillo, G., Lucidi, S., and Palagi, L. "A mathematical programming approach for the solution of the railway yield management problem," *Transportation Science* (33:2), May 1999, pp 168-181.
- Collier, D.A. *Service Management: Operating Decisions* Prentice-Hall, Englewood Cliffs, NJ, 1987.
- Colvin, G. "The seller's instant Net advantage," *Fortune* (142:2), Jul 10, 2000, p 74.
- Coulter, K.S. "The application of airline yield management techniques to a holiday retail shopping setting," *The Journal of Product and Brand Management* (8:1), 1999, p 61.
- Cross, R.G. "Launching the revenue rocket: How revenue management can work for your business," *Cornell Hotel and Restaurant Administration Quarterly* (38:2), Apr 1997, p 32.
- Cross, R.G. *An Introduction to Revenue Management* McGraw-Hill Companies, New York, NY, 1985, pp. 443-468.
- Cross, R.G. *Revenue Management* Broadway Books, New York, NY, 1998, p. 276.
- Cross, R.G., and Samuelson, R.D. "Using revenue management to improve the bottom line," *Electrical World* (211:11), Nov 1997, p 44.
- Dana Jr., J.D. "Using yield management to shift demand when the peak time in

- unknown," *The Rand Journal of Economics* (30:3), Autumn 1999, p 456.
- Dana Jr., J.D. "Advance-purchase discounts and price discrimination in competitive markets," *Journal of Political Economy* (106:2), Apr 1998, p 395.
- Davis, P. "Airline ties profitability to yield management," *SIAM News* (27:5), May 1994.
- Deetman, C. "Booking levels," AGIFORS Symposium Proceedings, New York, NY, 1964.
- Desiraju, R. "Strategic service pricing and yield management," *Journal of Marketing* (63:1), Jan 1999, pp 44-56.
- DiDio, L. "Airmax management package keeps vacant seats to minimum," *Computerworld* (32:37), Sep 14, 1998, p 29.
- Donaghy, K., McMahon-Beattie, U., and McDowell, D. "Implementing yield management: lessons from the hotel sector," *International Journal of Contemporary Hospitality Management* (9:2), 1997, p 50.
- Dyson, E. "What the traffic will bear." *Forbes* (143), May 1989, p 282.
- Elimam, A.A., and Dodin, B.M. "Incentives and yield management in improving productivity of manufacturing facilities," *IIE Transactions* (33:6), Jun 2001, p 449.
- Etschmaier, M.M., and Rothstein, M. "Operations Research in the Management of the Airlines," *Omega* (2), 1974, pp 160-175.
- Feng, Y., and Gallego, G. "Perishable asset revenue management with Markovian time dependent demand intensities," *Management Science* (46:7), Jul 2000, p 941.
- Feng, Y., and Gallego, G. "Optimal starting times for end of season sales and optimal stopping times for promotional fares," *Management Science* (41), 1995, pp 1371-

1391.

Feng, Y., and Xiao, B. "Optimal policies of yield management with multiple predetermined prices," *Operations Research* (48:2), Mar 2000, p 332.

Feng, Y., and Xiao, B. "A continuous-time yield management model with multiple prices and reversible price changes," *Management Science* (46:5), May 2000, pp 644-657.

Gallego, G., and Van Ryzin, G. "A multiproduct dynamic pricing problem and its applications to network yield management," *Operations Research* (45:1), Jan/Feb 1997, p 24.

Gallego, G., and Van Ryzin, G. "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science* (40:8), Aug 1994, pp 999-1020.

Geraghty, M.K., and Johnson, E. "Revenue management saves National Car Rental," *Interfaces* (27:1), Jan 1997, pp 107-127.

Gorin, T.O. "Airline Revenue Management: Sell-up and Forecasting Algorithms," Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1995.

Gourville, J., and Soman, D. "Pricing and the psychology of consumption," *Harvard Business Review* (80:9), Sept 2002, pp 90-96.

Greenberg, H.J. "Mathematical Programming Glossary,"
<http://www.cudenver.edu/~hgreenbe/glossary/>, 1996-2002.

Griesshaber, R., and Titze, B. "Realistic Passenger Booking Behaviour and the Simple Low-fare/High-fare seat allotment model," AGIFORS Symposium Proceedings,

1983, pp. 197-223.

Griffin, R.K. "Factors of successful lodging yield management systems," *Hospitality Research Journal* (19:4), 1996, pp 17-30.

Hadjinicola, G.C., and Panayi, C. "The overbooking problem in hotels with multiple tour operators," *International Journal of Operations & Production Management* (17:9), 1997, pp 874-885.

Hanks, R.B., Noland, R.P., and Cross, R.G. "Discounting in the hotel industry: A new approach," *Cornell Hotel and Restaurant Administration Quarterly* (33:3), Jun 1992, pp 40-45.

Harris, F.H., and Peacock, P. "Ten steps to yield management success," *Marketing Management* (4:2), Fall 1995, p 45.

Hersh, M., and Ladany, S.P. "Optimal Seat allocation for flights with intermediate stops," *Computational Operations Research* (5), 1978, pp 31-37.

Jauncey, S., Mitchell, I., and Slamet, P. "The meaning and management of yield in hotels," *International Journal of Contemporary Hospitality Management* (7:4), 1995, p 23.

Jones, P., and Hamilton, D. "Yield Management: Putting People in the Big Picture," *Cornell Hotel and Restaurant Administration Quarterly* (33:1), Feb 1992, p 89.

Kimes, S.E. "Perceived fairness of yield management," *Cornell Hotel and Restaurant Administration Quarterly* (43:1), Feb 2002, pp 21-30.

Kimes, S.E. "Revenue management on the links," *Cornell Hotel and Restaurant Administration Quarterly* (41:1), Feb 2000, p 120.

Kimes, S.E. "Developing a restaurant revenue-management strategy," *Cornell Hotel and*

- Restaurant Administration Quarterly* (40:5), Oct 1999, p 18.
- Kimes, S.E. "Implementing restaurant revenue management," *Cornell Hotel and Restaurant Administration Quarterly* (40:3), Jun 1999, p 16.
- Kimes, S.E. "The Basics of Yield Management," *Cornell Hotel and Restaurant Administration Quarterly* (30:3), Nov 1989, p 14.
- Kimes, S.E. "Yield Management: A Tool for Capacity-Constrained Service Firms," *Journal of Operations Management* (8:4), Oct 1989, p 348.
- Kimes, S.E., Chase, R.B., Choi, S., Lee, P., and Ngonzi, E. "Restaurant revenue management," *Cornell Hotel and Restaurant Administration Quarterly* (39:3), Jun 1998, p 32.
- Kimes, S.E., and Wagner, P. "Preserving your revenue-management system as a trade secret," *Cornell Hotel and Restaurant Administration Quarterly* (42:5), Oct 2001, pp 8-15.
- Kimes, S.E., and Wirtz, J. "Perceived Fairness of demand based pricing for restaurants," *Cornell Hotel and Restaurant Administration Quarterly* (43:1), Feb 2002, pp 31-37.
- Ladany, S.P., and Arabel, A. "Optimal cruise-liner passenger cabin pricing policy," *European Journal of Operational Research* (55), 1991, pp 136-147.
- Ladany, S.P., and Bedi, D.N. "Dynamic rules for flights with an intermediate stop," *Omega* (5), 1977, pp 721-730.
- Lautenbacher, C.J., and Stidham Jr., S. "The underlying Markov decision process in the single-leg airline yield management problem," *Transportation Science* (33:2), May 1999, pp 136-146.

- Law, A., and Kelton, W.D., *Simulation Modeling and Analysis*, McGraw-Hill, NY, 2007.
- Lee, A.O. "Airlines reservations forecasting: probabilistic and stochastic models of the booking process," in: *Flight Transportation Laboratory*, Massachusetts Institute of Technology, Cambridge, MA, 1990.
- Lee, T.C., and Hersh, M. "A model for dynamic airline seat inventory control with multiple bookings," *Transportation Science* (27), 1993, pp 252-265.
- Lee-Ross, D., and Johns, N. "Yield management in hospitality SMEs," *International Journal of Contemporary Hospitality Management* (9:2), 1997, p 66.
- Lennon, C.O. "Market Reactions to fare constraints and its effects on passenger revenue," AGIFORS Symposium Proceedings, Nathanya, Israel.
- Liang, Y. "Solution to the continuous time dynamic yield management model," *Transportation Science* (33:1), Feb 1999, pp 117-123.
- Lieberman, W.H. "Debunking the myths of yield management," *Cornell Hotel and Restaurant Administration Quarterly* (34:1), Feb 1993, p 34.
- Littlechild, S.C., and Shutler, M.F. *Operation Research in Management* Prentice Hall International (UK) Ltd, Hemel Hempstead, UK, 1991.
- Littlewood, K. "Forecasting and control of passenger bookings," AGIFORS Symposium Proceedings, Nathanya, Israel.
- Lyle, C. "A statistical analysis of the variability in aircraft occupancy," AGIFORS Symposium Proceedings, Terrigal, Australia, 1970.
- Mayer, M. "Seat Allocation, or a simple model of seat allocation via sophisticated ones," AGIFORS Symposium Proceedings, Key Biscayne, FL, 1976.
- McEvoy, B.J. "Integrating operational and financial perspectives using yield management

- techniques: an add-on matrix model," *International Journal of Contemporary Hospitality Management* (9:2), 1997, p 60.
- McGill, J. "Optimization and Estimation Problems in Airline Yield Management," in: *Faculty of Commerce and Business Administration*, University of British Columbia, Vancouver, BC, 1989.
- McGill, J., and Van Ryzin, G. "Revenue Management: Research Overview and Prospects," *Transportation Science* (33:2), May 1999, pp 233-256.
- Metters, R., and Vargas, V. "Yield management for the non-profit sector," *Journal of Service Research* (1:3), Feb 1999, pp 215-226.
- Murtagh, B.A., and Mitra, S. "Yield management with side conditions," *Asia - Pacific Journal of Operational Research* (19:1), May 2002, p 71.
- Noone, B., and Griffin, P. "Enhancing yield management with customer profitability analysis," *International Journal of Contemporary Hospitality Management* (9:2), 1997, p 75.
- Norman, E.D., and Mayer, K.J. "Yield management in Las Vegas casino hotels," *Cornell Hotel and Restaurant Administration Quarterly* (38:5), Oct 1997, p 28.
- O'Leary, M. "The CIO 100: Travel Services - Reservation Required," *CIO* (2:11), Aug 1989, pp 80-82.
- Orkin, E.B. "Boosting Your Bottom Line With Yield Management," *Cornell Hotel And Restaurant Administration Quarterly* (28:4), Feb 1988, p 52.
- Pachauri, M. "Consumer Behavior - A literature review," *Marketing Review* (2:3), Spring 2002, p 319.

- Peters, L. "Evaluating the benefits of yield improvement investments," *Semiconductor International* (21:3), Mar 1998, p 44.
- Pitt, L.F., Berthon, P., Watson, R.T., and Ewing, M. "Pricing strategy and the Net," *Business Horizons* (44:2), Mar-Apr 2001, pp 45-54.
- Quain, W.J. "Revenue enhancement, part 3: Picking low-hanging fruit--a simple approach to yield management," *Cornell Hotel and Restaurant Administration Quarterly* (40:2), Apr 1999, p 76.
- Quain, W.J., Sansbury, M., and Quinn, D. "Analyzing Sales-Mix Profitability," *Cornell Hotel and Restaurant Administration Quarterly* (33:2), Apr 1992, p 56.
- Ratliff, R., and Vinod, B. "Airline pricing and revenue management: a future outlook," *Journal of Revenue and Pricing Management* (4:3), May 2005, p 302.
- Reece, W.S., and Sobel, R.S. "Diagrammatic approach to capacity-constrained price discrimination," *Southern Economic Journal* (66:4), Apr 2000, p 1001.
- Relihan, W.J., III "The Yield-Management Approach to Hotel-Room Pricing," *Cornell Hotel and Restaurant Administration Quarterly* (30:1), May 1989, p 40.
- Richter, H. "Thirty years of Airline Operations Research," *Interfaces* (19:3), Jul-Aug 1989, pp 3-9.
- Richter, H. "The differential revenue method to determine optimal seat allotments by fare types," AGIFORS Symposium Proceedings, Lagonissi, Greece, 1982.
- Rothstein, M. "An airline overbooking model," *Transportation Science* (5:2), May 1971, pp 180-192.
- Rothstein, M. "Airline overbooking: Fresh approaches are needed," *Transportation*

- Science* (9:2), May 1975, pp 169-173.
- Rothstein, M. "OR and the Airline overbooking problem," *Operations Research* (33:2), Mar-Apr 1985, pp 237-248.
- Rothstein, M., and Stone, A.W. "Passenger Booking Levels," AGIFORS Symposium Proceedings, Noordwijk, The Netherlands, 1967.
- Sa, J. "Reservation Forecasting in Airline Yield Management," Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1987.
- Shlifer, R., and Vardi, Y. "An airline overbooking policy," *Transportation Science* (9), 1975, pp 101-114.
- Skwarek, D.K. "Competitive Impacts of Yield Management Components: Forecasting and Sell-up Methods," Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1997.
- Smith, B.C., Leimkuhler, J.F., and Darrow, R.M. "Yield Management at American Airlines," *Interfaces* (22:1), Jan/Feb 1992, p 8.
- Subramanian, J., Stidham Jr., S., and Lautenbacher, C.J. "Airline Yield Management with Overbooking, Cancellations, and No-Shows," *Transportation Science* (33:2), May 1999, pp 147-167.
- Sweetman, B. "Neural net yield answers; 'thinking' computers may offer a breakthrough in revenue management," *Air Transport World* (33:11), Nov 1996, p 77.
- Talluri, K., and Van Ryzin, G. "An analysis of bid-price controls for network revenue management," *Management Science* (44:11), Nov 1998, p 1577.
- Talluri, K., and Van Ryzin, G., *Theory and Practice of Revenue Management*, Springer,

NY, 2006.

Taneja, N.K. *Airline Traffic Forecasting: A regression analysis approach* Lexington Books, Lexington, MA, 1978.

Taylor, C.J. "The determination of passenger booking levels," AGIFORS Symposium Proceedings, Fregene, Italy, 1962.

Thompson, H.R. "Statistical Problems in Airline Reservation Control," *Operations Research Quarterly* (12), 1961, pp 167-185.

Upchurch, R., Ellis, T., and Seo, J. "Revenue Management underpinnings: an exploratory review," *International Journal of Hospitality Management* (21), 2002, pp 67-83.

Van Ryzin, G., and McGill, J. "Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels," *Management Science* (46:6), Jun 2000, p 760.

Wang, K. "Modelling the interaction between payload restriction, passenger demand, and reservation booking levels," AGIFORS, Lagonissi, Greece, 1982.

Wang, S. "A hybrid threshold curve model for optimal yield management: Neural networks and dynamic programming," *Computers & Industrial Engineering* (40:1, 2), Jun 2001, p 161.

Weatherford, L.R. "Optimization of perishable asset revenue management problems that allow prices as decision variables," *International Journal of Services Technology and Management* (2:1, 2), 2001, p 71.

Weatherford, L.R. "Length of stay heuristics: Do they really make a difference?" *Cornell Hotel and Restaurant Administration Quarterly* (36:6), Dec 1995, p 70.

Weatherford, L.R. "Perishable Asset Revenue Management in General Business

- Situations," in: *Darden Graduate School of Business Administration*, University of Virginia, Charlottesville, VA, 1991.
- Weatherford, L.R., and Belobaba, P.P. "Revenue impacts of fare input and demand forecast accuracy in airline yield management," *The Journal of the Operational Research Society* (53:8), Aug 2002, p 811.
- Weatherford, L.R., and Bodily, S.E. "A Taxonomy and Research Overview of Perishable-Asset Revenue," *Operations Research* (40:5), Sep/Oct 1992, p 831.
- Weatherford, L.R., Kimes, S.E., and Scott, D.A. "Forecasting for hotel revenue management: Testing aggregation against disaggregation," *Cornell Hotel and Restaurant Administration Quarterly* (42:4), Aug/Sep 2001, p 53.
- Weatherford, L.R., and Kimes, S.E. "A comparison of forecasting methods for hotel revenue management," *International Journal of Forecasting* (19), 2003, p 401.
- Weatherford, L.R., and Polt, Stefan. "Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues," *Journal of Revenue and Pricing Management* (1:3), May 2002, p 234.
- Weigand, R.E. "Yield Management: Filling buckets, papering the house," *Business Horizons* (42:5), Sept 1999, p 55.
- Wickham, R.R. "Evaluation of Forecasting Techniques for Short-Term Demand of Air Transportation," Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1995.
- Wollmer, R.D. "An airline seat management model for a single route when lower fare classes book first," *Operations Research* (40), 1992, pp 26-37.
- Yeoman, I., and Watson, S. "Yield management: a human activity system," *International*

Journal of Contemporary Hospitality Management (9:2), 1997, p 80.

You, P.S. "Dynamic Pricing in Airline Seat Management for flights with multiple legs,"

Transportation Science (33), 1999, pp 192-206.

Zaki, H. "Forecasting for airline revenue management," *The Journal of Business*

Forecasting Methods & Systems (19:1), Spring 2000, p 2.

Zeni, Richard H. "Improved Forecast Accuracy in Airline Revenue Management by

Unconstraining Demand Estimates from Censored Data," Graduate School,

Rutgers – The State University of New Jersey, Newark, NJ, 2001.

Zeni, Richard H. "The value of analyst interaction with revenue management systems,"

Journal of Revenue and Pricing Management (2:1), Jan. 2003, p 37.

Zhao, W., and Zheng, Y.S. "General Dynamic Models for Airline Seat Allocation," in:

The Wharton School, University of Pennsylvania, Philadelphia, PA, 1998.

Zickus, Jeffrey S. "Forecasting for Airline Network Revenue Management: Revenue and

Competitive Impacts," Department of Civil and Environmental Engineering,

Massachusetts Institute of Technology, Cambridge, MA, 1988.