

Competition and Congestion in the National Aviation System: Multi-agent, Multi-stakeholder Approaches for Evaluation and Mitigation

by

Vikrant Suhas Vaze

B.Tech., Civil Engineering,
Indian Institute of Technology Bombay (2005)

M.S. in Transportation,
Massachusetts Institute of Technology (2007)

M.S. in Operations Research,
Massachusetts Institute of Technology (2010)

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

Doctor of Philosophy in the field of Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

© Massachusetts Institute of Technology 2011. All rights reserved.

Author

Department of Civil and Environmental Engineering

July 31, 2011

Certified by

Cynthia Barnhart

Professor of Civil and Environmental Engineering and Engineering Systems

Associate Dean of Engineering

Thesis Supervisor

Accepted by

Heidi M. Nepf

Chair, Departmental Committee for Graduate Students

Competition and Congestion in the National Aviation System: Multi-agent, Multi-stakeholder Approaches for Evaluation and Mitigation

by

Vikrant Suhas Vaze

Submitted to the Department of Civil and Environmental Engineering
on July 31, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the field of Transportation

Abstract

The US National Aviation System (NAS) is a complex system with multiple, interacting agents including airlines, passengers, and system operators, each with somewhat different objectives and incentives. These interactions determine the state of the system. NAS congestion and delays result in additional operating costs and reduced profitability for the airlines, a decrease in the level-of-service to passengers, and a decrease in the efficiency of NAS resource utilization. We evaluate the congestion impacts on the NAS stakeholders while explicitly accounting for their interactions and propose congestion mitigation mechanisms that are beneficial to these different stakeholders.

We measure the extent to which the NAS capacity is being inefficiently utilized. We show that at the current level of passenger demand, delays are avoidable to a large extent if we control the negative effects of competitive airline scheduling practices, thus providing critical insights into the nature and causes of delays.

We develop a detailed framework using data fusion and discrete choice modeling for generating disaggregate passenger travel data. We characterize the impacts of airline network structures, schedules and operational decisions on passenger delays.

We propose a parametric game-theoretic model consistent with the most popular characterization of frequency competition. We prove that the level of congestion in a system of competing airlines is an increasing function of 1) the number of competing airlines, 2) a measure of the gross profit margin, and 3) the frequency sensitivity of passenger demand.

We propose a game-theoretic model of frequency competition under slot constraints and provide empirical and algorithmic justifications of the suitability of the Nash equilibrium solution concept for modeling these games. We devise and assess new administrative strategies for congestion mitigation. We show that a small reduction in the total number of allocated slots translates into a substantial reduction in delays, and also a considerable improvement in airlines' profits.

We develop an equilibrium model of frequency competition in the presence of delay costs and congestion prices. We find that the success of congestion pricing critically depends on the characteristics of frequency competition in individual markets. We also identify critical differences between flat pricing and marginal cost pricing.

Key words: Airline Scheduling, Airline Frequency Competition, National Aviation System, Stakeholders, Multi-agent Models, Nash Equilibrium, Game Theory, Price of Anarchy, Passenger Delays, Cancellations, Missed Connections, Cost of Passenger Disruptions, Administrative Slot Controls, Slot Reduction, Congestion Pricing.

Thesis Supervisor: Cynthia Barnhart

Title: Professor of Civil and Environmental Engineering and Engineering Systems
Associate Dean of Engineering

Acknowledgments

First and foremost, I would like to thank my research advisor, Prof. Cynthia Barnhart. Cindy has been a constant source of knowledge, guidance and encouragement. This thesis has greatly benefited from her immense knowledge of both operations research and aviation, her insightful comments and her unique approach balancing mathematical rigor and real-world applicability of research in such interdisciplinary field. Apart from her guidance on technical aspects of research, her refreshingly positive approach, openness, decisiveness, strong work ethics, kindness and humane attitude towards students makes her one of the best advisors and human beings I have ever come across. Not once in the three years I have worked with her, have I felt pressured to work hard and yet these three years have easily been my most productive years in terms of the work done. Cindy is the single most important reason I feel excited about research and academia. Just like for many of her other students, she has been a great role model for me too.

I would like to thank my truly amazing thesis committee members, Prof. Amedeo Odoni, Dr. Peter Belobaba and Prof. Asuman Ozdaglar. I am greatly indebted to Amedeo for his guidance and suggestions related to research and teaching throughout my time as a PhD student. He has been an invaluable source of guidance as a thesis committee member. Apart from that I also feel very fortunate to have worked with Amedeo, first as a teaching assistant, then as a co-instructor, for the 1.200 course. I have tried to learn as much as possible from his tremendous expertise and experience as a great researcher and a great teacher. He is also one of the kindest and most helpful people I have come across. A special thanks to Peter for investing a lot of time and energy in improving various aspects of my research. I strongly feel that his insightful and forthright comments and industry knowledge have greatly contributed to enhancing the contents of this thesis. I would like to thank Asu for helping me with the game-theoretic aspects in this research. I have greatly benefited from her technical knowledge, experience and advice.

I am fortunate to have worked with Antony Evans, Douglas Fearing and Pavithra

Harsha during different phases of my PhD life. They have all been great friends, colleagues and collaborators, with whom I hope to continue working in the future too. I feel that my long and fruitful research collaboration with Doug has been particularly helpful towards increasing my productivity and I have learnt a lot in the process. It has been a pleasure to have Lavanya Marla and Virod (Ta) Chiraphadhanakul as friends and labmates, with whom I have shared so many interesting discussions on research, course-work, and innumerable other topics. A special thanks to Ta for bearing with me through the thick and thin of graduate student life and being a true friend. I cannot forget my friends from MIT-Portugal Program, including Alda Mendes, Joao Pita, Nikolas Pyrgiotis and Miguel Santos, who have made the conferences and seminars so much of fun. I would also like to thank Carolina Osorio, Maya Abou-Zeid, Varun Ramanujam and the rest of my friends from MIT's transportation group who have made it a vibrant and active community.

I am very grateful to Maria Marangiello for being so helpful with scheduling of meetings, travel arrangements and innumerable other logistical issues in her extremely efficient and ever-smiling manner. I would also like to thank Jeanette Marchocki, Kris Kipp and Patricia Glidden at the CEE Headquarters, who have often gone out of their way to help students and have always done so with a smile on their faces. A quick trip down to the CEE Headquarters followed by a quick chat has often been a refreshing little break in my day's work.

Sarvee Diwan and Kushal Kedia are the two individuals without whom I would not have been able to get even half of the work done, for which I am eternally thankful to them. I am very fortunate to have shared with them those amazing coffee and lunch breaks and those relaxing discussions on any and every topic on earth. Without them, I don't know what I would have done. I must certainly acknowledge and thank the rest of the awesome *DuDo* gang including Abishek Kashinath, Mukul Kabir, Sahil Sahni and Vivek Raghunathan. Our regular and frequent visits together to *DuDo*, *Muddy*, *Sunny* and *Bengal* became an integral part of my PhD life at MIT.

I would like to thank my roommates at the amazing *88C*, including Parikshit Shah, Vaibhav Rathi, Chaitanya Misal and Kushal Kedia. I would like to espe-

cially thank Pari and Kushal for those interesting and relaxing conversations late into the night, after a long and tiring day of work. A big thanks to Harshad Kasture, Raghavendra Hosur, Abhinav Akhoury, Ahmedali Lokhandwala, Ahmed Kirmani, Andrea Colaço, Aniruddh Sarkar, Harish Sundaresh, Manas Shah, Muralidharan Vijayaraghavan, Srikanth Jagabathula, Srikanth Patala, Vibhu Prakash Saxena and the rest of the MIT *desi junta* for making the social life at MIT so much fun.

Most importantly, I would like to thank my family. Nothing would have been possible without the constant love, support and encouragement from my parents, my sister Varada, and my fiancée Neha. Finally, I would like to thank god for giving me this opportunity.

Executive Summary

The US National Aviation System (NAS) is a large and complex system consisting of multiple, interacting agents including airlines, passengers, and the system operators, each with a somewhat different set of objectives and incentives. Interactions between the actions and decisions of these autonomous agents determine the state of the system. Over the past several years, congestion and delays in the NAS are imposing a tremendous cost on the US economy. Delays have resulted in additional operating costs and a decrease in profitability for the airlines, a decrease in quality of service to the passengers through passenger delays and disruptions, and a decrease in the efficiency of resource utilization in the NAS.

In this thesis, we first evaluate the congestion impacts on the various stakeholders in the NAS while explicitly accounting for the interactions between the different decision-makers. Next, we propose congestion mitigation mechanisms that are beneficial from the perspectives of these different stakeholders.

- Demand management strategies are expected to improve the system efficiency by mitigating the congestion introduced by competitive airline scheduling practices. We measure the extent to which the capacity of the US domestic air transportation network is being inefficiently utilized. We formulate the problem as a large-scale, network-based, mixed-integer linear programming problem and solve it using a sequence of relaxations and greedy heuristics. The solution serves as a lower bound on the minimum level of delays that can be achieved given the existing levels of passenger demand and airport capacity. We show that at the current level of passenger demand, delays are avoidable to a large extent if we control the negative effects of competitive airline scheduling practices. These results provide critical insights into the nature and causes of aviation delays, allowing better planning and utilization of the aviation infrastructure.
- A lack of publicly available detailed data on passenger travel has thus far prevented extensive analyses of passenger delays. We develop a detailed framework using data fusion and discrete choice modeling for generating disaggregate passenger travel

data. We use the resulting data to gain critical insights into passenger travel, delays and disruption patterns in the US. Scheduling and operational policies and decisions by the airlines significantly alter the passenger delays and disruptions, which in turn affect the overall level-of-service experienced by the passengers. We present a sequence of data mining and statistical modeling analyses that characterize the impacts of airline network structures, schedules and operational decisions on passenger delays. Apart from the analyses and findings presented in this thesis, we foresee a large variety of further applications of this passenger delays framework for passenger-centric approaches in airline scheduling, air traffic flow management, and aviation policy-making.

- An airline is expected to attract more passengers by increasing its frequency share in a market. Frequency competition affects airlines' capacity allocation decisions, which in turn have a strong impact on airline profitability and on airport congestion. We propose a parametric game-theoretic model consistent with the most popular characterization of frequency competition. We prove the suitability of Nash equilibrium for modeling airline frequency competition and mathematically show the extent to which competition worsens the congestion problem. Our model is general enough to accommodate somewhat differing beliefs about the market share-frequency share relationship. We propose two alternative simple frequency adjustment rules and prove that under mild conditions, either of them converges to an equilibrium state, thus confirming the stability of the equilibrium state. We prove that the level of congestion in a system of competing airlines is an increasing function of the number of competing airlines, the ratio of average fare to operating cost per seat and the frequency sensitivity of passenger demand.
- Administrative slot controls have been in place at some of the most congested US airports for decades, even though large delays have often coexisted with such measures. We propose a game-theoretic model of airline frequency competition under slot constraints and devise and assess new administrative strategies for congestion mitigation. We develop a fast, dynamic programming-based algorithm to obtain

a Nash equilibrium. The model predictions are validated against actual frequency data, with the results indicating a close fit to reality. We use the model to evaluate different strategic slot allocation schemes. The most significant result of this research shows that, under the assumptions of our modeling framework, a small reduction in the total number of allocated slots translates into a substantial reduction in flight and passenger delays, and also a considerable improvement in airlines' operating profits. We also tested the sensitivity of our results to many of our assumptions and approximations. We found that the major conclusions were robust to individual assumptions and in many cases our original results were somewhat conservative.

- Airport congestion pricing has often been advocated in literature as a means of controlling demand for airport operations and for achieving social welfare maximization by making each airport user pay for the delay cost it imposes on the other users. Competition between airlines affects the extent to which an airline would be willing to pay for airport slots. We develop an equilibrium model of airline frequency competition in the presence of delay costs and congestion prices. Our work provides a computational framework for understanding the impacts of congestion pricing under competitive effects. Our results based on a small hypothetical network provide some critical insights. Most importantly, our results show that variation in the number of passengers per flight plays a vital role in determining the degree of attractiveness of congestion pricing to the airlines. A significant part of the impact of congestion pricing could not be accounted for in the previous studies because of the assumptions of constant load factors and constant aircraft sizes. The framework presented in this chapter captures some important characteristics of the competitive equilibrium solution under congestion prices, which have not been captured by the previous studies. We find that the effectiveness of a congestion pricing scheme critically depends on three essential characteristics of frequency competition in individual markets, the same ones that affect the level of congestion as mentioned above. We also identify some critical differences between the

flat pricing and marginal cost pricing of airport slots, and present some favorable characteristics of marginal cost pricing.

Contents

Abstract	4
Executive Summary	12
1 Introduction	27
1.1 Causes of Delay	27
1.2 Decision-makers and Stakeholders in the National Aviation System	29
1.3 Congestion Mitigation Strategies	30
1.3.1 Administrative Slot Controls	31
1.3.2 Congestion Pricing	32
1.3.3 Slot Auctions	32
1.4 Thesis Outline	33
1.4.1 Chapter 2: Minimization of System-wide Delays in the Absence of Competition	33
1.4.2 Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions	35
1.4.3 Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion	36
1.4.4 Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation	38
1.4.5 Chapter 6: Pricing Mechanisms for Airport Congestion Mitigation	39
1.5 Thesis Contributions	41

1.5.1	Chapter 2: Minimization of System-wide Delays in the Absence of Competition	41
1.5.2	Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions	42
1.5.3	Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion	43
1.5.4	Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation	43
1.5.5	Chapter 6: Pricing Mechanisms for Airport Congestion Mitigation	44
2	Minimization of System-wide Delays in the Absence of Competition	47
2.1	Introduction	47
2.2	Motivation for Single Airline Scheduling	48
2.3	Airline Schedule Development	50
2.4	Problem Statement	52
2.5	Modeling Framework	53
2.6	Solution Algorithm	59
2.7	Data Sources and Implementation Details	64
2.8	Results	66
2.9	Summary	70
3	Quantification and Analysis of Passenger Delays and Disruptions	73
3.1	Introduction	73
3.2	Passenger Delay Estimation Methodology	76
3.2.1	Passenger Itinerary Flow Estimation	77
3.2.2	Delay Calculation	79
3.3	Passenger Delay Results	81
3.4	Regression Model for Passenger Delay Estimation	89
3.5	Analysis of Flight Cancellations	94
3.5.1	Airports and Carriers	95

3.5.2	Flight Frequency and Load Factors	98
3.5.3	Carrier Effect	101
3.6	Analysis of Missed Connections	107
3.6.1	Airports and Carriers	108
3.6.2	Schedule Banking	110
3.6.3	Modeling Missed Connections	114
3.7	Summary	117
4	Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion	119
4.1	Introduction	119
4.2	Frequency Planning under Competition	121
4.3	Model	125
4.4	Best Response Curves	127
4.5	2-Player Game	135
4.5.1	Existence and Uniqueness	136
4.5.2	Realistic Parameter Ranges	141
4.5.3	Myopic Best Response Dynamic	143
4.5.4	Alternative Dynamic	153
4.6	N-Player Symmetric Game	155
4.7	Price of Anarchy	162
4.8	Summary	164
5	Administrative Mechanisms for Airport Congestion Mitigation	167
5.1	Introduction	167
5.1.1	Demand Management	168
5.1.2	Literature Review	171
5.1.3	Contributions	172
5.2	Model	174
5.2.1	Model Extension 1: Fare Differentiation	177
5.2.2	Model Extension 2: Market Entry Deterrence	179

5.3	Solution Algorithm	179
5.3.1	Myopic Best Response Algorithm	182
5.3.2	Dynamic Programming Formulation	183
5.4	Validity of Nash Equilibrium Outcome	184
5.4.1	Data Sources and Implementation Details	184
5.4.2	Empirical Validation	186
5.4.3	Game Dynamics	188
5.4.4	Robustness Verification	190
5.5	Evaluation of Simple Slot Reduction Strategies	192
5.5.1	Numerical Results	193
5.5.2	Sensitivity to Model Parameters	201
5.5.3	Effect of Aircraft Upgauges	202
5.5.4	Effects of Demand Uncertainty and Revenue Management Practices	205
5.5.5	Effects of Passenger Connections	210
5.6	Summary	213
6	Pricing Mechanisms for Airport Congestion Mitigation	217
6.1	Introduction	217
6.2	Literature Review	218
6.3	Model	222
6.4	Flight Delay Model	226
6.5	Solution Algorithm	227
6.5.1	Solution Algorithm for <i>Type I Experiments</i>	227
6.5.2	Solution Algorithm for <i>Type II Experiments</i>	228
6.6	Experimental Setup and Data Sources	230
6.7	Delay Function Fitting	231
6.8	Numerical Results	233
6.8.1	Flat Pricing Results	233
6.8.2	Marginal Cost Pricing Results	243

6.9	Summary	249
7	Conclusions and Recent Events	253
7.1	Conclusions and Contributions	253
7.1.1	Chapter 2: Minimization of System-wide Delays in the Absence of Competition	254
7.1.2	Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions	255
7.1.3	Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion	256
7.1.4	Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation	257
7.1.5	Chapter 6: Pricing Mechanisms for Airport Congestion Mitiga- tion	258
7.2	Recent Events	260
A	Abbreviations	271
A.1	Carrier Abbreviations	271
A.2	Airport Abbreviations	272

List of Figures

2-1	Modeling framework	55
2-2	Timetable development algorithm	61
2-3	Greedy heuristic for hub markets with small demand	62
2-4	Greedy heuristic for non-hub markets with small demand	63
2-5	Distribution of utilization ratio over a day at JFK airport	68
2-6	Number of operations by airport	71
3-1	Linear regression model to bypass the passenger allocation and re- accommodation process	89
3-2	Cancellation rates for the top 50 busiest airports	96
3-3	Cancellation rates by carrier and carrier type	98
3-4	Distribution of flight departures at the two airports with the highest cancellation rates (LGA and ORD)	99
3-5	Misconnection rates for the top 50 busiest airports	108
3-6	Misconnection rates by carrier and carrier type	110
3-7	Example of banked hub operations (NW at MEM)	111
3-8	Example of de-banked hub operations (AA at ORD)	112
4-1	A typical shape of profit function (Case A)	129
4-2	A typical shape of profit function (Case B)	129
4-3	A typical shape of profit function (Case C)	130
4-4	A typical best response curve	133
4-5	Histograms of parameter values	142
4-6	Best response curves in 2-player game	144

5-1	Typical shape of the segment profit function	181
5-2	Typical shape of the best response function	181
5-3	Empirical validation of frequency predictions	187
5-4	Total operating profit as a function of slot reductions under a proportionate allocation scheme assuming constant aircraft sizes	195
5-5	Total operating profit as a function of slot reductions under a reward-based allocation scheme assuming constant aircraft sizes	195
5-6	Total number of passengers carried as a function of slot reductions under a proportionate allocation scheme assuming constant aircraft sizes	196
5-7	Total number of passengers carried as a function of slot reductions under a reward-based allocation scheme assuming constant aircraft sizes	196
5-8	Effect of limited upgauging on total number of passengers under a 12.3% slot reduction	205
5-9	Expected spill per seat obtained using the spill modeling approach for a fare discount ratio of 0.7 (Source: Belobaba and Farkas, 1999) . . .	207
5-10	Average spill fare obtained using the spill modeling approach for a fare discount ratio of 0.7 (Source: Belobaba and Farkas, 1999)	208
6-1	Estimation of delay function parameters using linear regression	232
6-2	Demand for airport operations as a function of slot prices for different values of α	236
6-3	Passenger delay costs as a function of slot prices for different values of α	238
6-4	Total operating profits of the airlines as a function of slot prices for different values of α	238
6-5	Demand for airport operations as a function of slot prices for different fare levels	240
6-6	Passenger delay costs as a function of slot prices for different fare levels	241
6-7	Total operating profits of the airlines as a function of slot prices for different fare levels	241

6-8	Demand for airport operations as a function of slot prices for different number of competitors	243
6-9	Passenger delay costs as a function of slot prices for different number of competitors	244
6-10	Total operating profits of the airlines as a function of slot prices for different number of competitors	244
7-1	Trend in jet fuel prices	261
7-2	Trend in the number of passengers	262
7-3	Trend in number of flights	263
7-4	Trend in average arrival delays	265
7-5	Trend in number of operations for ASQP reporting airlines at LGA .	266
7-6	Trend in NAS delays per flight at LGA	267
7-7	Trend in average arrival delay at LGA	269

List of Tables

1.1	Trend in number of passengers, flights and delays	29
2.1	Comparison of airport utilization ratios	66
2.2	Network performance metrics	67
2.3	Aircraft size distributions	69
2.4	Total flight delay under various weather scenarios	69
3.1	Passenger delay estimates for 2008	81
3.2	Regression parameter estimates	91
3.3	Summary of error in passenger delay estimates at different aggregation levels	93
3.4	Delay estimates using the regression-based approach for 2007 and 2008	94
3.5	Correlation between average flight frequency and flight cancellation rates across carrier-segments	100
3.6	Variation in Southwest (WN) Airlines' flight cancellation rates based on daily flight frequency	100
3.7	Carrier effects	103
3.8	Effects of primary hub on cancellation rates	105
3.9	Cancellation rates and large delays for flights entering and exiting the primary hub	107
3.10	Worst 9 airports in terms of misconnection rate	113
3.11	Worst 9 airports in terms of misconnection rates with average connection times	113

3.12	Schedule banking coefficients for primary carriers at IAD, ORD, JFK, and CLE	114
3.13	Estimation results for misconnection rate model 1 (with flight delays)	115
3.14	Estimation results for misconnection rate model 2 (with flight delays and connection times)	116
3.15	Estimation results for misconnection rate model 3 (with flight delays, connection times and regional and low-cost dummies)	116
5.1	Stability of algorithm results to starting point perturbations	189
5.2	Model prediction error across different quarters	190
5.3	Stability of algorithm results to starting point perturbations across different quarters	191
5.4	Sensitivity of prediction accuracy (in MAPE) to model parameters	191
5.5	Effect of a 12.3% slot reduction on system-wide performance metrics	199
5.6	Increase in operating profits due to a 12.3% slot reduction	200
5.7	Sensitivity of slot reduction impacts to the maximum average segment load factor value under a 12.3% slot reduction	201
5.8	Sensitivity of increase in total profit due to slot reduction to model parameters (under 12.3% reduction with proportionate allocation)	202
5.9	Sensitivity of decrease in passengers carried due to slot reduction to model parameters (under 12.3% reduction with proportionate allocation)	203
5.10	Sensitivity of increase in total profit due to slot reduction to model parameters (under 12.3% reduction with reward-based allocation)	203
5.11	Sensitivity of decrease in passengers carried due to slot reduction to model parameters (under 12.3% reduction with reward-based allocation)	204
5.12	Effect of a 12.3% slot reduction under <i>spill modeling approach</i>	208
5.13	Effect of a 12.3% slot reduction under completely prorated fares approach	212
6.1	Model results for the <i>base case</i>	234
6.2	Marginal cost pricing equilibrium results (Experiment 6)	245

A.1	Carrier names and IATA codes	271
A.2	Airport names and IATA codes	272

Chapter 1

Introduction

Congestion in the National Aviation System (NAS) is imposing a tremendous cost on the US economy. In the recently concluded Total Delay Impact Study [7] commissioned by the Federal Aviation Administration (FAA), researchers estimated the total cost of domestic air traffic delays to be around \$31.2 billion for calendar year 2007, including \$8.3 billion in additional aircraft operating costs, \$16.7 billion in passenger delay costs, and an estimated \$6.2 billion in other indirect costs of delays to the economy. The magnitude of these delay costs can be properly grasped by noting that during the same period, the aggregate profits of US domestic airlines were \$5.0 billion [5]. Even though air travel demand and airport congestion has reduced over the last 2-3 years due to the recent economic recession, large delays are expected to return once the economic crisis subsides [96]. Following is a look at the most important causes of these delays.

1.1 Causes of Delay

For the year 2007, Bureau of Transportation Statistics [71] categorized delays to around 50% of the delayed flights as delays caused by the National Aviation System (NAS). Weather and volume were the top two causes of these NAS delays, together responsible for 84.51% of the NAS delays. Delays due to volume are those caused by scheduling more airport operations than the available capacity, while the delays due

to weather are those caused by airport capacity reductions under adverse weather conditions. Both these types of delays are due to airlines scheduling more operations than realized capacity. Such mismatches between demand and capacity are a primary cause of flight delays in the United States. From here onwards, we will refer to this phenomenon as the *demand-capacity mismatch*.

Before proceeding further, let us differentiate between two different types of demands. On the one hand is the demand for airport capacity in terms of the number of flight operations scheduled at an airport. This needs to be contrasted with passenger demand for air travel. It is the former that affects the airport congestion most directly. Table 1.1 shows the values of total number of passengers, total number of flights and total arrival delays to flights in the US. These values are obtained from the Bureau of Transportation Statistics (BTS) website [77, 72]. The data on passengers and flights corresponds to all the domestic operations of all the US carriers with an annual revenue of at least \$20 million as reported to the BTS [77]. The data on total flight arrival delays corresponds to certified U.S. air carriers that account for at least one percent of domestic scheduled passenger revenues as reported to the BTS [72]. All the values in Table 1.1 are normalized such that the values for the year 2000 are all equal to 100. Passenger demand dipped in the first two years of the first decade of this century, following the economic recession around the turn of century and the events on September 11, 2001. The period from 2002 to 2007 saw a sustained growth in passenger demand. By 2007, passenger demand was 13.28% higher compared to that in 2000. Interestingly, the number of scheduled flight operations was 24.46% higher and total arrival delays to flights were 38.58% higher.

The disproportionate rise in the number of flight operations compared to a relatively moderate increase in the number of passengers implies that the average number of passengers per flight reduced by around 9% from 2000 to 2007. This suggests that there is more to the demand-capacity mismatch than simply the rate of passenger growth outpacing the rate of airport capacity expansion. The disproportionate increase in the number of flights is a result of scheduling decisions by the airlines, reflecting, as will be detailed in this thesis, the effects of competition.

Table 1.1: Trend in number of passengers, flights and delays

Year	Number of Passengers [77]	Number of Flights [77]	Total Arrival Delays to Flights (Minutes) [72]
2000	100	100	100
2001	93.34	96.47	78.15
2002	92.06	102.32	59.75
2003	97.29	119.65	75.18
2004	105.04	126.09	103.58
2005	109.62	126.98	107.80
2006	109.81	122.86	120.99
2007	113.28	124.46	138.58
2008	108.70	118.60	119.11
2009	103.07	110.73	91.82
2010	105.00	110.03	88.30

Scheduling decisions by the airlines under competition are considered to be partially responsible for exacerbating the demand-capacity mismatch and therefore the congestion problem. However, these decisions cannot be analyzed in isolation. The decisions and actions of important stakeholders of the National Aviation System (NAS) of the United States, such as airlines, passengers and system operators (including airport authorities and the Federal Aviation Administration (FAA)) are highly inter-dependent.

1.2 Decision-makers and Stakeholders in the National Aviation System

Besides fares, a flight schedule is the most vital aspect of an airline’s ability to attract passengers and market share [17]. By providing more frequent flights, an airline attracts more passengers. Thus, airline scheduling decisions are directly affected by passenger preference for availability of air service at desirable times. The relative attractiveness of an airline to the passengers and hence, the market share of an airline, depends not only on the attractiveness of the airline’s own schedule but also the attractiveness of its competitors’ schedules. Therefore, an airline needs to take

into account the scheduling decisions of other competing airline(s) while making its own scheduling decisions. System operators (such as the FAA and airport authorities) are responsible for the safety and efficiency of the air transportation system. As a result, the system operators often impose rules and restrictions on the usage of aviation infrastructure; rules necessitated by safety, capacity, weather, emissions, noise, and other considerations. The airlines need to take these rules and restrictions into account while making their scheduling and operational decisions. Airline decisions about network structures, schedules and operations, in turn, have a significant effect on the level-of-service to the passengers. Finally, airline network structures and scheduling decisions under competition from other airlines affect the efficiency of resource utilization and systemwide congestion levels, which are of direct concern to the system operators.

Thus, airlines, passengers and system operators are the important decision-makers in the NAS and their decisions, which are highly interdependent, determine the overall system performance. Therefore, any congestion mitigation strategy needs to take into account the interactions between these various decision-makers. Furthermore, the impacts of congestion and any congestion mitigation strategies need to be assessed from the perspectives of these different stakeholders. In this thesis, we model the NAS as a multi-agent system of these interacting autonomous agents. First, we evaluate the impacts of congestion and delays from the perspectives of the various stakeholders. In Chapter 2, we look at the congestion problem from the perspective of system operators. In Chapter 3, we focus on the passenger perspective and in Chapter 4, we model the problem from the perspective of the airlines. Finally, in Chapters 5 and 6, we propose mechanisms for congestion mitigation and assess the impacts of these mechanisms on the NAS stakeholders.

1.3 Congestion Mitigation Strategies

Increasing capacity and decreasing demand are the two natural ways of bringing the demand-capacity mismatch into balance. Capacity enhancement measures such

as building new airports, construction of new runways, etc. are investment intensive, require long-time horizons, and might not be feasible in many cases due to geographic, environmental, socio-economic and political issues associated with such large projects. On the other hand, demand management strategies such as administrative slot controls, market-based mechanisms, or any combinations thereof, have the potential to restore the demand-capacity balance over a medium- to short-time horizon with comparatively little investment. Demand management strategies refer to any administrative or economic policies and regulations that restrict airport access to users.

The capacity of an airport is often measured in terms of the available number of *slots* for takeoffs and landings at that airport. According to the US code, a slot is a reservation for an instrument flight rule takeoff or landing of an aircraft by an air carrier in air transportation [36]. All the demand management strategies proposed in the literature and practiced in reality can be broadly categorized as administrative controls and market-based mechanisms, although various hybrid schemes have also been proposed. The demand management problem involves two types of decisions, namely, (1) slot determination, which involves deciding the total number of slots to be allocated, and (2) slot allocation, which involves the decision on distribution of these slots among the different users. These decisions can be taken either sequentially, such as in an auction or administrative mechanism, or simultaneously, such as in a congestion pricing mechanism.

1.3.1 Administrative Slot Controls

Currently, four major airports in the United States, namely, LaGuardia (LGA), John F. Kennedy (JFK), and Newark (EWR) airports in the New York region, and Reagan (DCA) airport at Washington D.C., have administrative controls limiting the number of flight operations. Outside of the US, administrative controls are commonplace at busy airports. Several major airports in Europe and Asia are *schedule-coordinated*, where a central coordinator allocates the airport slots to airlines based on a set of pre-determined rules. Under the current practices, both in and outside of the US,

the criteria governing the slot allocation process are typically based on historical precedents and use-it-or-lose-it rules. Under these rules, an airline is entitled to retain a slot that was allocated to it in the previous year (sometimes called as the *grandfathering rights*), contingent on the fact that the slot was utilized for at least a certain minimum fraction of time over the previous year. An airline failing to utilize a slot frequently enough, however, is in danger of losing it (sometimes called as the *use-it-or-lose-it* rules).

1.3.2 Congestion Pricing

Researchers have shown that market-based mechanisms, if implemented properly, result in efficient allocation of airport resources. Congestion pricing and slot auction are two of the most popular market mechanisms proposed in the literature. Classical studies such as those of Vickrey [97], Levine [62], and Carlin and Park [30] proposed congestion pricing based on the marginal cost of delays. Such pricing schemes, in theory, maximize social welfare through optimal allocation of public resources. Under congestion pricing, the total cost to the airlines includes the delay cost as well as the congestion price.

1.3.3 Slot Auctions

The idea of airport slot auctions was first proposed by Grether et al. [48]. Rassenti, Smith, and Bulfin [82] showed how combinatorial auction design is suitable for airport slot auctions and highlighted the associated efficiency gains through experiments. Since then, several researchers (Cramton et al. [39], Ball, Donohue and Hoffman [9], Dot Econ Limited [63] and Harsha [51], to mention a few) have shown the advantages of slot auctions. The reader is referred to Ball, Donohue, and Hoffman [9] and Harsha [51] for detailed accounts of various commonly raised concerns regarding slot auctions and ways of addressing them. In spite of the many attractive properties of the auctioning mechanisms, an auction by itself does not alleviate airport congestion, but rather allocates a fixed set of resources in a more efficient way. So, to that extent,

auctions are similar to administrative controls, as they too pose an implicit need to make a tradeoff between delays and resource utilization.

In this thesis, we focus on these demand management strategies with primary focus on administrative slot controls (in Chapter 5) and congestion pricing (in Chapter 6).

1.4 Thesis Outline

In the following five sub-sections, we briefly outline the structure of this thesis.

1.4.1 Chapter 2: Minimization of System-wide Delays in the Absence of Competition

Besides safety, efficiency in utilization of the NAS capacity is of utmost importance from the system operators' perspective. The administrative and market-based demand management strategies, described in Section 1.3, are expected to bring demand and supply into balance by removing inefficiencies in the NAS. However, the extent to which the system-wide delays can be reduced by these mechanisms is still unclear. On the one hand, restricting airport utilization to a very low level can practically ensure the absence of congestion related delays, but this could mean that the airport is highly underutilized and all passenger demand might not be satisfied. On the other hand, scheduling a very large number of operations can satisfy all passenger demand but the delays could reach unacceptable levels. An important question is what minimum level of airport utilization and delays needs to be permitted in order to satisfy all passenger demand.

In this research, we measure the extent to which airport capacity in the US domestic air transportation network is being inefficiently utilized. The aim is to build a schedule that minimizes delays in the absence of frequency competition. In order to obviate the effects of competition, we assume a single airline that satisfies all passenger demand without compromising the level-of-service for passengers. The problem is modeled as a large-scale, network-based, mixed-integer programming (MIP) problem

and solved using a sequence of linear programming relaxations and greedy heuristics. A network delay simulator [70] is used to estimate the delays for the resulting network. The delay values for the single airline network are compared with those for the existing network under various realistic scenarios. These delay estimates serve as theoretical lower bounds on system-wide delays when airport capacity is allocated most efficiently.

All the analysis is based on extensive amounts of publicly available data on airline schedules, passenger flows and airport capacities. Detailed flight and passenger flow information was obtained from the Bureau of Transportation Statistics website [71, 73]. Actual realized airport capacity values for one entire year were used for calculation of expected flight delays.

The value of the maximum possible delay reduction provides valuable information to the system operators. It indicates the maximum potential impact of implementing efficient demand management strategies. If insignificant, then passenger demand has already reached a level where large delays are inevitable and capacity enhancement is the only realistic means of delay reduction. On the other hand, if the results suggest substantial delay reduction under the single airline case, then the existing level of passenger demand can be efficiently served using the existing infrastructure with much lower delays and there is ample opportunity for congestion mitigation using demand management strategies.

Our results in Chapter 2 show that there is a significant room for improvement in the level of congestion even with the existing airport infrastructure. Passenger demand is currently at a level where delays are avoidable to a large extent. Given the available capacity, efficient administrative controls and/or market-based mechanisms can potentially lead to substantial reductions in airport congestion and delays. These results provide critical insights into the nature and causes of aviation delays, allowing better planning and utilization of aviation infrastructure. In particular, the results help differentiate between the delays caused by insufficient capacity and delays caused by inefficient utilization of capacity. The models and solution methods presented here can also be used for analyzing the best-case delay levels under different future

scenarios, with different levels of demand and capacity. These results also emphasize the need to devise intelligent mechanisms and incentives that will result in airlines gradually migrating their schedules from those in place today towards the delay-minimizing schedules presented in this chapter, in the presence of market competition. We tackle this problem of devising such intelligent mechanisms in Chapters 5 and 6.

1.4.2 Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions

Airline passengers are an important stakeholder group in the NAS. As discussed earlier, over recent years, passengers too have suffered enormously from congestion effects in the NAS. Therefore, a good understanding of the nature, causes and magnitude of passenger delays is essential. Previous literature has shown that flight delays are not a good proxy measure for passenger delays [23]. In particular, passengers suffer large amounts of delay due to itinerary disruptions such as flight cancellations and missed connections. Unfortunately, a lack of publicly available detailed data on passenger travel has thus far prevented extensive analyses of passenger delays. In Chapter 3, we develop a detailed framework for generating disaggregate passenger travel data and use the resulting data to gain critical insights into passenger travel, and delay and disruption patterns in the US.

In the first part of Chapter 3 (Section 3.2), we briefly present a methodological framework based on a discrete-choice multinomial Logit model for generating disaggregate passenger itinerary flow data. The framework uses various data fusion and statistical modeling techniques. Statistical estimation of the model is performed using passenger booking data from one large network carrier in the US for one quarter in 2007. The resulting parameter estimates are used to estimate the number of passengers that traveled on each individual itinerary for the entire year 2007 for all the major carriers in the US. Using this rich database on passenger travel, we are able to perform a large variety of passenger-centric analyses.

Airlines' strategic decisions regarding their network structures, hub locations, con-

necting bank structures, flight frequencies and flight departure schedules impact an airline's attractiveness to the passengers. At the same time, these very factors also affect various fixed and operating costs to the airlines. Therefore, in making such strategic decisions, airlines need to balance schedule attractiveness and cost implications, while accounting for competitors' decisions. Similarly, on the day of operations, airlines need to take various decisions such as flight cancellations, aircraft and crew reassignments, passenger re-booking etc. to address various kinds of irregularities and disruptions. While taking these operational decisions, an airline needs to balance often-conflicting objectives of minimizing the passengers' inconvenience and minimizing the cost of disruption to the airline. Scheduling and operational policies and decisions by the airlines significantly alter the passenger delays and disruptions, which in turn affect the overall level-of-service experienced by the passengers. In the subsequent Sections of Chapter 3 (Sections 3.3 through 3.6), we present a sequence of analyses on the impacts of airline network structures, schedules and operational decisions on passenger delays and travel disruptions.

1.4.3 Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion

Profitability is one of the most important, if not the most important, objectives of US airlines, most of which are owned by private shareholders. Since deregulation of the US domestic airline industry in 1978, airline profits have been highly volatile. Several major US carriers have incurred substantial losses over the last decade with some of them filing for Chapter 11 bankruptcy and some others narrowly escaping bankruptcy. Provision of excess seating capacity is one of the reasons often cited for the poor economic health of airlines [56, 69, 29]. Due to the so called S-curve relationship [17] between market share and frequency share, an airline is expected to attract disproportionately more passengers by increasing its frequency share in a market. To increase their market share, airlines engage in frequency competition by providing more flights per day on competitive routes. As a result, they prefer

operating many flights with small aircraft rather than operating fewer flights with larger aircraft. The average aircraft sizes in domestic US markets have been falling continuously over the last couple of decades (until the recent economic crisis) in spite of increasing passenger demand [20]. Similarly, the average load factors, i.e., the ratio of the number of passengers to the number of seats, on some of the most competitive and high demand markets have been found to be lower than the industry average.

Besides direct connections to airline profitability, airline frequency competition is intricately connected to the worsening congestion and delays at the major US airports. As shown in Table 1.1 earlier, increases in passenger demand coupled with decreases in the average number of passengers per flight have led to a great increase in the number of flights being operated, especially between the major airports, leading to congestion. Thus, frequency competition affects airlines' capacity allocation decisions, which in turn have a strong impact on airline profitability and on airport congestion.

In Chapter 4, we propose a parametric game-theoretic model, which is consistent with a popular characterization of frequency competition. Our model is general enough to accommodate somewhat differing beliefs about the market share-frequency share relationship. First, we characterize the curves representing the optimal frequency of an airline as a function of its competitors' frequencies, otherwise known as the *best-response curves*. Focusing on a 2-airline competition case, we state and prove the conditions for the existence and uniqueness of all the possible types of equilibrium states (*pure strategy Nash equilibria*, to be precise).

As the first of the two major results in this chapter, we propose 2 alternative simple frequency adjustment rules (otherwise known as *myopic learning dynamics*) for the 2-airline case and prove that under mild conditions, either of them converges to an equilibrium state. This means that the equilibrium is highly stable and that even if each airline simply optimizes its own profit in response to the competitors' actions, they iteratively converge to equilibrium frequencies. This also substantiates the predictive power of the Nash equilibrium concept for modeling airline frequency competition. The author is not aware of any previous study which proves these stability properties of airline competition games. We found this result to be highly

beneficial 1) for modeling airline response to congestion mitigation mechanisms, and 2) for the computation of a Nash equilibrium for an airline frequency competition game, in Chapters 5 and 6.

Moving to an N -airline (for any integer $N \geq 2$), symmetric case, we characterize the entire set of possible equilibria and also identify the worst-case equilibrium, which corresponds to the highest congestion and lowest profitability for the airlines as a group. As the second major result of this chapter, we prove that the congestion level and the degree of inefficiency in the system is an increasing function of the number of competing airlines, the ratio of average fare to operating cost per seat and the exponent in the S-curve relationship (reflecting the degree of competition). This is the first study, to the best of the author's knowledge, which actually proves that the S-curve relationship between market share and frequency share has direct and negative implications to airline profitability and airport congestion, as has been speculated in multiple previous studies. Furthermore, these results provide the intuition behind our analysis of effectiveness of congestion pricing mechanisms in Chapter 6. Our results on the important factors affecting the effectiveness of congestion pricing are found to be consistent with the results in this chapter.

1.4.4 Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation

As discussed in Chapters 2, 3, and 4, competitive airline scheduling and airport congestion are intricately related and this relationship has adverse implications, 1) to NAS operators through systemwide delays and inefficiency, 2) to passengers through passenger delays and disruptions, and 3) to airlines through reduced profitability and flight delays. In Chapters 5 and 6, we propose simple mechanisms for congestion mitigation which address these issues.

We focus on LaGuardia Airport (LGA) at New York, where administrative mechanisms are currently used to control the slot allocation and thus to manage demand. However, it is one of the most congested and delay-prone airports in the United States.

We propose modifications to the existing slot controls at this airport for congestion mitigation. If such modifications are implemented, then the airlines in turn need to modify their scheduling decisions while ensuring that their schedules remain sufficiently attractive from the passengers' perspectives. We model the airline response to changes in slot controls through models of airline frequency competition. Models of airline competition used in Chapters 5 and 6 are extensions of the model presented in Chapter 4.

To the best of the author's knowledge, no previous study has incorporated slot controls into airline competition models. We provide a solution algorithm with strong computational performance for obtaining the equilibrium outcome under our proposed slot allocation. We provide further justification of the credibility of the Nash equilibrium solution concept in two different ways, 1) through empirical validation of the model outcome, and 2) through a computational demonstration of the convergence properties of the myopic learning dynamics for non-equilibrium situations. Finally, under our proposed slot allocation mechanisms, we evaluate system performance from the perspectives of the various stakeholders.

1.4.5 Chapter 6: Pricing Mechanisms for Airport Congestion Mitigation

Congestion pricing and slot auctions are the two most prevalent market-based mechanisms mentioned in the micro-economic literature. These mechanisms have often been claimed to alleviate the demand-capacity mismatch by placing monetary prices on airport slots, which reflect the true economic cost of using the slots. These market-based mechanisms rely on the ability of the airlines to assess the value of airport slots, while bidding for slots in the case of auctions and for determining the demand for slots at a given level of prices in the case of congestion pricing. However, while ascertaining its own valuation of an airport slot, an airline needs to account for competition from other airlines operating at that airport. In this chapter, we model the airline frequency decisions under congestion pricing mechanisms through explicit modeling of

competition and assess the dependence of the effectiveness, or lack thereof, of airport congestion pricing mechanisms on the characteristics of the competition in airline markets.

The incremental profitability of having an extra flight in a particular market largely depends on the number of additional passengers that the airline will be able to carry because of the additional flight, which in turn depends on the schedule of flights offered by the competitor airlines in the same market. So total demand for airport slots under congestion pricing should reflect these competitive interactions. Many prior studies have accounted for airline competition under pricing using conventional micro-economic models of firm competition [25, 26, 80, 81]. By assuming constant load factors and constant aircraft seating capacities, they fail to recognize the important distinguishing features of the airline industry where the quantity produced is captured by three different entities: number of flights, number of seats and number of passengers carried. By not capturing the variation in number of passengers per flight, earlier studies tend to underestimate benefits of congestion pricing to the airlines. In this chapter, we model airline frequency competition under congestion pricing using a popular market-share model of frequency competition, which accounts for these peculiar characteristics of the airline industry and generate insights that were not possible with the previous models.

Our model is similar to the models of frequency competition presented in Chapters 4 and 5, but it additionally accounts for the impacts of flight delays costs and the congestion prices to the airlines' profitability. We model average flight delay as a function of airport utilization ratio (the ratio of the number of scheduled operations to capacity) and estimate the model parameters using data on actual airport capacities, demands and delays. We develop an iterative algorithm with good computational properties to solve the congestion pricing problem to an equilibrium. Our experimental setup consists of a small hypothetical network of three airports, with one of the three airports being subjected to congestion pricing. We run computational experiments under flat- as well as marginal cost-pricing scenarios. We vary important characteristics of our markets and test their impacts on the effectiveness

of the congestion pricing mechanism.

We find that the effectiveness of a congestion pricing scheme critically depends on three essential characteristics of frequency competition in individual markets: 1) the number of competing airlines, 2) a measure of the gross profit margin for airlines (defined as the ratio of average fare to operating cost per seat), and 3) frequency sensitivity of passengers, which is nothing but the exponent in the S-curve relationship. These are the same three parameters that affect the level of congestion introduced by competition as described in Chapter 4. Moreover, our results indicate the important differences between flat pricing and marginal cost pricing mechanisms. We show that a marginal cost pricing mechanism is able to deter the airlines from scheduling very frequent flights without penalizing them with very high congestion toll payments. Most importantly we prove that, in addition to delay reduction benefits, a significant part of congestion pricing benefits to the airlines are in the form of reduction in operating costs due to increased number of passengers per flight. Our models of competition are able to capture this important effects which could not be captured by previous studies.

We conclude the thesis in Chapter 7.

1.5 Thesis Contributions

In the following, we briefly describe the major contributions of this thesis to the existing body of research. We detail the contributions of each chapter sequentially.

1.5.1 Chapter 2: Minimization of System-wide Delays in the Absence of Competition

The main contributions of the research presented in this chapter are threefold. First, we propose a novel, optimization-based approach for attributing the congestion-related delays in the NAS to two different causes, namely, delays due to insufficient airport capacity and delays due to inefficient utilization of available capacity due to

airline competition. Second, we develop an aggregated, integrated airline scheduling model with a proxy objective function for delay minimization and an elaborate heuristic-based approach for an approximate solution of this large-scale (non-binary) MIP. Finally, and most importantly, this is the first study which proves that there is a significant room for reducing the level of congestion even with the existing airport infrastructure without compromising the passenger level-of-service, if we can control the negative impacts of airline competition through efficient demand management strategies. Thus, in this chapter, we make a strong case for the need for the subsequent research (presented in the Chapters 5 and 6) on demand management-based mechanisms for congestion mitigation.

1.5.2 Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions

The major contributions of the research in this chapter fall into three broad categories. First, we develop a detailed approach for disaggregating publicly available aggregate passenger flow data which, among other applications, facilitates the usage of a pre-existing passenger delay calculation heuristic to a much wider dataset (viz. to aggregate data corresponding to any major US airline for any of the last 18 years). Second, we analyze the spatio-temporal patterns in passenger delays using these estimated disaggregate passenger flows and present numerous insights into the factors affecting passenger delays. Such insights could not be generated in any of the prior studies due to the lack of comprehensive passenger itinerary flow data. Third, we investigate the causes of passenger travel disruptions by applying data analysis and statistical modeling to historical flight and passenger data. Apart from the analyses and findings presented in Chapter 3, our methodology and the resulting passenger itinerary flow data has already been used to estimate the overall costs of passenger delays as one component of the Total Delay Impact Study commissioned by the FAA [7]. We foresee a large variety of further applications of this passenger delays framework for passenger-centric approaches in airline scheduling, air traffic flow management,

and aviation policy-making.

1.5.3 Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion

The major contributions of the research in this chapter are threefold. First of all, ours is the first study that models the S-curve-based airline frequency competition using game-theoretic tools. The S-curve has been mentioned in many empirical studies and has also been an important part of the airline industry lore. Second, we provide credibility to the idea of using Nash equilibrium as a means of modeling airline frequency competition by proving the convergence of two alternative simple frequency adjustment rules (otherwise known as *myopic learning dynamics*) to a Nash equilibrium. Finally, using the idea of Nash equilibrium, we prove that the S-curve relationship between market share and frequency share has direct and negative implications to airline profitability and airport congestion, as has been speculated in multiple previous studies. These results make a strong case for careful incorporation of airline frequency competition into any assessment of the impacts of demand management mechanisms (as presented in Chapters 5 and 6).

1.5.4 Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation

The results presented in this chapter are the most significant practical contributions of this thesis. The main contributions of this chapter fall into four categories. First, we propose a game-theoretic model of frequency competition under slot constraints as an evaluation methodology for slot allocation schemes. Second, we provide a solution algorithm with good computational performance for solving the problem to a Nash equilibrium. Third, we provide justification of the credibility of the Nash equilibrium solution concept in two different ways, through empirical validation of the model outcome and through a computational proof of the convergence properties of the learning dynamics for non-equilibrium situations. Finally, under simple slot allocation

schemes, we evaluate system performance from the perspectives of the passengers and the competing airlines, and provide insights to guide the demand management policy decisions. The most significant result of the research in this chapter shows that a small reduction in the total number of allocated slots translates into a substantial reduction in airport congestion and passenger delays, as well as a considerable improvement in airlines' profits, under the assumptions of our modeling framework. We also tested the sensitivity of these results to many of our underlying assumptions and found that our conclusions are robust against small changes in underlying assumptions and in many cases our original estimates of the benefits of slot reduction were, in fact, somewhat conservative. Thus our administrative slot allocation-based strategies are shown to be beneficial to all the major NAS stakeholders at the same time.

1.5.5 Chapter 6: Pricing Mechanisms for Airport Congestion Mitigation

As mentioned earlier, Chapter 5 presents the most significant practical contributions of this thesis. In Chapter 6, we extend the game-theoretic model of frequency competition, presented in Chapter 5, to incorporate slot prices and provide computational results using a small hypothetical network. These results serve as a proof-of-concept for assessing the effectiveness of congestion pricing mechanisms under frequency competition.

The major contributions of the research in this chapter are threefold. First, we develop a model of airline frequency competition that explicitly accounts for the relationship between the number of flights operated, number of seats flown and the number of passengers carried by an airline under slot pricing. To the best of author's knowledge, this is the first computational study on congestion pricing that accounts for this relationship. Second, using a small hypothetical network, we evaluate the impacts of congestion prices on the various stakeholders and investigate the dependence of effectiveness of congestion pricing mechanisms on the different characteristics of frequency competition in individual markets. Third, we provide computational re-

sults under flat prices, as well as under a marginal cost pricing equilibrium. Our results show that variation in the number of passengers per flight plays a vital role in determining the degree of attractiveness of congestion pricing to the airlines. A significant part of the impact of congestion pricing could not be accounted for using the earlier models which were based on the assumptions of constant load factors and constant aircraft sizes. The framework presented in this chapter captures some important characteristics of the competitive equilibrium solution under congestion prices, which have not been captured by the previous studies. The congestion pricing results in this chapter serve as a proof-of-concept and provide several interesting insights, which need more detailed verification through computational experiments with larger data sets.

Our results in Chapters 2 through 5 are based on year 2007-2008 when congestion and delays were rampant in the United States NAS. In Chapter 7 we conclude the thesis and discuss the applicability of our conclusions in the context of some recent changes in the NAS. This section provides a nice further validation of some of our results based on the real-world events in NAS over the last three years or so.

Chapter 2

Minimization of System-wide Delays in the Absence of Competition

2.1 Introduction

Efficiency in utilization of the National Aviation System's capacity is of utmost importance from the system operators' perspective. The administrative and market-based demand management strategies are expected to bring demand and supply into balance by removing inefficiencies in the NAS. However, the extent to which the system-wide delays can be reduced by these mechanisms is still unclear. On the one hand, restricting airport utilization to a very low level can practically ensure the absence of congestion related delays, but this could mean that the airport is highly underutilized and all passenger demand might not be satisfied. On the other hand, scheduling a very large number of operations can satisfy all passenger demand but the delays could reach unacceptable levels. An important question is what minimum level of airport utilization and delays needs to be permitted in order to satisfy all passenger demand. In this chapter, we assess the maximum possible impact of these demand management strategies.

Before proceeding further, let us define some terminology that will be used frequently in this chapter. In all our models, a market is defined as a passenger origin and destination pair. A segment is defined as an origin and destination pair for non-stop flights. A path is defined as a sequence of segments along which a passenger is transported from origin to destination. A flight leg is defined as a combination of origin, destination, departure time and arrival time of a non-stop flight. An itinerary is a sequence of flight legs along which a passenger is transported from origin to destination. We will refer to the actual network of flights operated by multiple airlines in the US domestic markets in 2007 as the *existing network*. Also we will refer to our delay minimizing network as the *single airline (or SA) network*.

The rest of this chapter is organized as follows. Section 2.2 explains the motivation behind solving the single airline scheduling problem. Section 2.3 briefly reviews the literature on airline scheduling and highlights the important differences between the previous research and the problem at hand. Section 2.4 provides a detailed problem statement for this study. Section 2.5 describes the modeling framework. Section 2.6 outlines the set of algorithms used to solve the problem. Section 2.7 provides details of data sources and implementation. A summary of results is provided in section 2.8. Finally, we conclude with a discussion of the main contributions and the directions for future research in section 2.9.

2.2 Motivation for Single Airline Scheduling

As described in Chapter 1, demand management strategies refer to any administrative or economic regulation that restricts airport access to users. It should be noted that these strategies refer to managing the demand for flight arrival and departure slots at an airport to meet a given level of passenger demand, and not to managing passenger demand itself. Few of the most congested US airports, such as Kennedy (JFK), Newark (EWR) and Laguardia (LGA) Airports in New York City area, O'Hare (ORD) Airport at Chicago and Reagan (DCA) Airport at Washington DC, have been slot controlled in one way or the other for a long time. Current slot allocation strategies,

based on administrative controls, are inefficient because of large barriers to market entry [63] and use-it-or-lose-it rules that encourage over-scheduling practices [51]. In response to these shortcomings, market-based mechanisms such as slot auctions and congestion pricing have been proposed as efficient means of reducing congestion. There is extensive literature [39, 9, 40, 63, 45, 48, 47] suggesting that market-based approaches, if designed properly, allocate scarce resources efficiently and promote fair competition. Harsha shows that market-based mechanisms can lead to airline schedule changes that reduce the demand for runway capacity, without reducing the number of passengers being transported [51]. This is achieved by better utilization of capacity in off-peak hours and by greater usage of larger aircraft.

In theory, these pricing and auction mechanisms should bring the demand and supply into balance by removing the inefficiencies in the system. However, the extent to which the system-wide delays can be reduced by these mechanisms is still unclear. On the one hand, restricting airport utilization to a very low level can practically ensure the absence of congestion related delays, but this could mean that the airport is highly underutilized and all the passenger demand might not be satisfied. On the other hand, scheduling a very large number of operations can satisfy all the passenger demand but the delays could reach unacceptable levels. An important question is what minimum level of airport utilization and delays needs to be permitted in order to satisfy all the passenger demand.

In this research, we measure the extent to which airport capacity in the US domestic air transportation network is being inefficiently utilized. The aim is to build a schedule that minimizes delays in the absence of frequency competition. In order to obviate the effects of competition, we assume a single airline that satisfies all passenger demand without compromising the level-of-service for passengers. A network delay simulator is used to estimate the delays for the resulting network [70]. The delay values for the single airline network are compared with those for the existing network under various realistic scenarios. All the days in an entire year are divided into 5 categories based on the total duration of capacity reduction on that day across all the busy airports. One representative day from each category is chosen for delay

calculations. These delay estimates serve as theoretical lower bounds on system-wide delays when airport capacity is allocated most efficiently. The value of maximum possible delay reduction will indicate the maximum potential impact of implementing efficient demand management techniques. If insignificant, then passenger demand has already reached a level where large delays are inevitable and capacity enhancement is the only realistic means of delay reduction. On the other hand, if the results suggest substantial delay reduction under the single airline case, then the existing level of passenger demand can be efficiently served using the existing infrastructure with much lower delays and there is ample opportunity for congestion mitigation using demand management strategies.

2.3 Airline Schedule Development

The airline schedule development process includes decisions regarding daily frequency, departure times, aircraft sizes and crew schedules. Due to the enormous size and complexity of the airline schedule development process, the problem is typically broken down into four stages: 1) timetable development; 2) fleet assignment; 3) maintenance routing; and 4) crew scheduling [10]. The task of deciding the set of flight legs to be operated along with the corresponding origin, destination and departure time for each leg is called timetable development. Although the entire timetable development problem can be modeled as an optimization problem, practitioners typically focus on incremental changes to existing schedules. The fleet assignment problem involves a profit maximizing assignment of fleet types to flight legs. Hane et al. [49] proposed a leg-based fleet assignment model which assumes independent leg demand and average fares. Jacobs, Johnson, and Smith [54] and Barnhart, Knicker, and Lohatepanont [14] proposed itinerary-based fleet assignment models that produce significant profit improvement over the leg-based models. Maintenance routing is the assignment of specific aircraft to individual flight legs while satisfying the periodic aircraft maintenance requirements. The maintenance routing problem is typically solved as a feasibility problem or as a through-revenue maximization problem [10]. The problem

of assigning a cost minimizing combination of pilot and cabin crews to each flight leg is called crew scheduling. Because of the complicated duty rules and pay structure for airline crews, crew pairing has long been regarded as a challenging problem. Crew pairing is modeled as a set partitioning problem and solved using techniques such as column generation and branch-and-price [13, 42, 57].

The aforementioned sequential solution process may result in suboptimal solutions. Many researchers have tried to integrate some of the stages into simultaneous optimization problems. Rexing et al. [83] proposed joint models for flight re-timing within time windows and fleet assignment. Lohatepanont and Barnhart [64] present an integrated model for incremental schedule development and itinerary-based fleet assignment. Clarke et al. [31] and Barnhart et al. [11] have proposed models to incorporate the effect of maintenance routing while making the fleet assignment decisions. There is a large body of literature on the integration of maintenance routing and crew scheduling problems including Cohn and Barnhart [33], Cordeau et al. [37] and Klabjan et al. [58].

All the models mentioned above aim to produce precise schedules that maximize planned profit. Models developed for the purpose of this study are different in several important ways. Rather than producing an operable schedule, the main purpose is to obtain a bound on delays. Because of the complex and stochastic relationship between schedule and delays, any such bound will have to be approximate. Therefore there is no point in developing a very precise schedule. Moreover, the problem of single airline schedule development is even larger in size than the schedule development problem for any existing airline, which itself is solved sequentially due to tractability issues. Therefore in this study, we use aggregate models that are sufficient for our purposes while maintaining tractability. Instead of profit maximization, the objective is delay minimization subject to satisfaction of demand and level-of-service requirements. Therefore, only the relevant decisions such as timetable development and fleet assignment are included in the problem. The output of our models is a flight schedule with departure times and fleet types corresponding to each flight. We do not solve the maintenance routing and crew scheduling problems in this research. Harsha [51]

has proposed an aggregated, integrated airline scheduling and fleet assignment model (AIASFAM) to help airlines place a bid in a slot auction. The itinerary-based version of this model is an extension of the Barnhart, Knicker, and Lohatepanont [14] model, with more aggregate time-line discretization for computational tractability. The models presented in this study share some characteristics with the AIASFAM model.

2.4 Problem Statement

In order to obtain a lower bound on airport congestion, we assume the existence of a single monopolistic airline. The problem at hand is to design a schedule for this single airline with the objective of minimizing airport congestion, while satisfying the entire passenger demand and maintaining a comparable level-of-service. An important modeling consideration is how to capture the passenger demand satisfaction requirements for every market and every time period of the day. The single carrier must be able to transport all passengers who are currently transported by existing airlines, from their respective origins to their respective destinations. To model this, we divide the day into four time intervals and ensure that all the passengers who are currently transported during a particular interval continue to be transported during the same interval in the new schedule. We define the level-of-service (as perceived by air passengers) as the number of stops in an itinerary. Almost 97.6% of all US domestic passengers traveled on non-stop and one-stop itineraries in 2007 [73]. Hence, in the single airline scheduling model, we assume that all the passengers must be transported on itineraries with at most one stop.

Service frequency is another important criterion of level-of-service in the current competitive environment. Therefore, in our model, we require that the single airline provide at least the same daily frequency on each non-stop segment as the *effective frequency* provided by the existing carriers. Cohas, Belobaba, and Simpson [32] propose a model of effective frequency available to air passengers faced with a choice between multiple competing carriers. When more than one airline operates in a mar-

ket, the effective frequency depends on how closely the schedules are matched. For example, consider two competing carriers, each offering n flights per day. If one airline schedules flights at a time when the other airline does not offer service, then the effective frequency increases. However, if one airline schedules all its flights close to the departure times of flights by the other airlines, then the number of different options to the passengers does not increase above n . Thus, the important criterion in deciding the effective frequency is the closeness of competing airline schedules. We calculate the effective non-stop frequency for a segment as the total number of non-stop flights offered by all carriers as long as the flight departure times are not within less than one hour of each other. If the departure times of two flights are separated by less than one hour, then we assume the two flights to be equivalent to a single flight. The minimum frequency to be provided on each non-stop segment by the single airline must be greater than or equal to the effective frequency currently provided by all the existing carriers on that segment. This constraint ensures that the passengers experience the same or higher effective frequency in each market. This constraint combined with the time-of-the-day demand satisfaction criteria also ensures that there is negligible shift in the passengers' arrival and departure times in comparison to the desired values of the same.

Before designing a schedule, decisions must be taken regarding the network structure. Network design involves decisions about network type i.e. hub-and-spoke or point-to-point, choice of hubs, choice of non-stop segments, choice of allowable airports for passenger connections. One possible approach would be to include all these decisions into our single-airline optimization problem. For the problem size under consideration, that would lead to an integer optimization problem involving over one-hundred million variables. Instead, we solve the problem sequentially in three stages.

2.5 Modeling Framework

Figure 2-1 provides a schematic description of the overall modeling framework. The first stage is the Network Design (ND) stage, which involves decisions about the

number and location of hubs, candidates for non-stop routes and allowable airports for passenger connections. The network structures of existing airlines were used as a guideline for our network design stage. Many of the major airlines in US domestic market today have a set of 4 or 5 major hubs. The direct flights are allowed to bypass the hub for a few important markets with large demand. Our selection of hubs was made based on qualitative criteria including the number of operations in the existing network, available capacity, geographic location and weather. Atlanta (ATL), Denver (DEN), Dallas/Fort Worth (DFW) and Chicago O’Hare (ORD) were chosen because they are in the top five US airports in terms of both existing capacity as well as the number of operations in the existing network. None of the airports in the New York area were chosen because of their low capacities. Los Angeles (LAX) was not chosen because of its geographically extreme location in the continental US. Phoenix (PHX) was chosen because of large number of operations in the existing network and the maximum capacity being available for a large fraction of the time due to good weather conditions. Our choices of non-stop segments bypassing the hub were made based on the market demand corresponding to the non-stop segments. Any market with a daily demand of at least 250 passengers was included as a candidate for non-stop flights. We allow passengers to connect only at the hubs.

The second stage involves the daily Frequency Planning and Fleet Assignment (FPFA) problem. A delay minimizing schedule should have fewer flights per day and better distribution of flight timings to avoid clustering of demand near peak hours. Obtaining a good feasible solution is the main aim of our FPFA stage. A good solution will keep the number of flights to a minimum, so that airport usage is minimized. We tried using a variety of formulations with different objective functions for this stage. Our initial modeling efforts for this stage showed that there are multiple optimal solutions that minimize the total number of flights but differ in terms of the amount of slack in the seating capacity. In order to produce an efficient schedule it is important to choose the most appropriate aircraft size for each segment so as to avoid excessive seating capacity. We achieved this by choosing cost coefficients (denoted by $c_{s,k}$ for segment s and fleet k) such that the overall cost increases with increasing seating

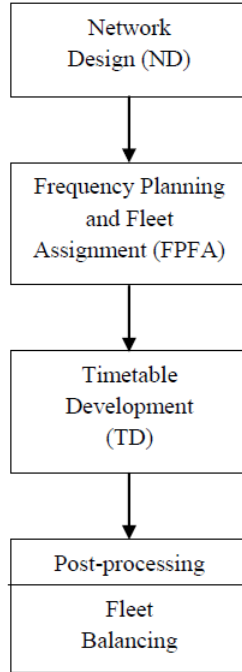


Figure 2-1: Modeling framework

capacity and cost per seat decreases with increasing seating capacity. These costs are consistent with those that the airlines report through Form 41 financial reports [74]. Satisfaction of the daily demand and the minimum daily frequency requirement are the two main constraints. Output of this second stage includes the daily frequency of service on each segment and fleet types assigned to each segment. Constraints 2.1 through 2.5 provide integer programming formulation for the FPFA problem.

FPFA Formulation

Notation:

- K = Set of fleet types
- S = Set of segments
- P = Set of paths
- M = Set of markets
- $c_{s,k}$ = Operating cost of fleet type k on segment s , $s \in S$ and $k \in K$

- C_k = Seating capacity of fleet type $k, k \in K$
- D_m = Daily demand in market $m, m \in M$
- f_s = Minimum daily frequency to be provided on segment $s, s \in S$
- $P(m)$ = Set of paths associated with market $m, m \in M$
- $\delta_s^p = \begin{cases} 1 & \text{if path } p \text{ contains segment } s \\ 0 & \text{otherwise} \end{cases} \quad p \in P \text{ and } s \in S$

Decision variables:

- $x_{s,k}$ = Number of flights of fleet type k on segment s per day, $s \in S$ and $k \in K$
- y_p = Number of passengers on path p per day, $p \in P$

Formulation:

$$\text{Minimize} \quad \sum_{s \in S} \sum_{k \in K} C_{s,k} x_{s,k}$$

$$\text{Subject to:} \quad \sum_{p \in P(m)} y_p = D_m \quad \forall m \in M \quad (2.1)$$

$$\sum_{k \in K} x_{c,k} C_k \geq \sum_{p \in P} \delta_s^p y_p \quad s \in S \quad (2.2)$$

$$\sum_{k \in K} x_{c,k} \geq f_s \quad \forall s \in S \quad (2.3)$$

$$x_{s,k} \in \mathbb{Z}^+ \quad \forall s \in S \text{ and } k \in K \quad (2.4)$$

$$y_p \in \mathbb{Z}^+ \quad p \in P \quad (2.5)$$

Constraint 2.1 ensures that the total daily demand for each market is satisfied. Constraint 2.2 ensures that the total number of seats on each segment is sufficient for carrying all the passengers whose paths contain that segment. Constraint 2.3 enforces that the daily frequency of service in each market is at least equal to the effective frequency currently provided by the existing carriers in that market. Constraints 2.4 and 2.5 restrict the allowable values for the number of passengers and number

of flights to non-negative integers. Alternatively, the number of passengers could be modeled as continuous variables without significant impact on solution quality. The integrality of variables corresponding to the number of flights, however, is critical to obtaining meaningful solutions.

The third stage involves actual Timetable Development (TD). Similar to the approach adopted by Harsha [51], the departure and arrival times are aggregated to the nearest hour to keep the number of decision variables low. Given the daily frequencies and fleet assignment for each segment, output of this stage produces the scheduled set of flight legs. Constraints 2.6 through 2.11 provide an integer programming formulation of the TD model.

The utilization ratio is defined as the ratio of demand to capacity of a server, which in this case is an airport. Queuing theory suggests that the average flight delay is an increasing and convex function of the utilization ratio [60]. Considering the tremendous size of the problem at hand, using a non-linear objective function would make the problem intractable. Total delay is a nonlinear and stochastic function of the number of scheduled flights. Therefore, we aim to minimize the maximum utilization ratio as a surrogate objective function for the scheduling problem. Due to the convex relationship between the utilization ratio and delays, the effect of the maximum utilization ratio on total delay in a queuing network is disproportionately high. The objective function in the TD formulation is to minimize the maximum utilization ratio across all busy airports across all airport-time period (ATP) pairs. The duration of each ATP is 1 hour in this case. The hourly utilization ratio is the ratio of the sum of all flight frequencies corresponding to that ATP to the hourly capacity of the airport. Thus, the maximum utilization ratio is a deterministic and linear function of the flight frequencies. Constraint 2.6 enforces the satisfaction of demand for each market-time period (MTP) pair. Constraint 2.7 ensures that the total number of seats for each flight leg is at least equal to the total number of passengers whose itineraries contain that flight leg. Constraint 2.8 ensures that the minimum daily frequency requirement is satisfied. Constraint 2.9 relates the maximum utilization ratio to the operations in each ATP. Constraints 2.10 and 2.11 restrict the possible

values for the number of passengers and the frequencies to non-negative integers.

TD Formulation

Notation:

- A = Set of airports
- I = Set of itineraries
- F = Set of flight legs
- $L(m)$ = Set of MTPs associated with market $m, m \in M$
- D_l = Demand in MTP $l, l \in L(m)$ and $m \in M$
- C_f = Seating capacity for fleet type assigned to flight leg $f, f \in F$
- f_s = Minimum daily frequency to be provided for segment $s, s \in S$
- HC_t = Hourly capacity (i.e. maximum total number of operations) for ATP $t, t \in T(a)$ and $a \in A$
- $T(a)$ = Set of ATPs associated with airport $a, a \in A$
- $I(l)$ = Set of itineraries associated with MTP $l, l \in L(m)$ and $m \in M$
- $F(s)$ = Set of flight legs associated with segment $s, s \in S$
- $\delta_f^i = \begin{cases} 1 & \text{if itinerary } i \text{ contains flight leg } f \\ 0 & \text{otherwise} \end{cases} \quad i \in I \text{ and } f \in F.$
- $\gamma_f^t = \begin{cases} 1 & \text{if ATP } t \text{ is utilized by flight leg } f \\ 0 & \text{otherwise} \end{cases} \quad t \in T \text{ and } f \in F.$

Decision variables:

- x_f = Frequency of flight leg $f, f \in F$
- y_i = Number of passengers on itinerary $i, i \in I$
- r_{max} = Maximum utilization ratio for airport hourly capacities

Formulation:

Minimize r_{max}

$$\text{Subject to: } \sum_{i \in I(l)} y_i = D_l \quad \forall l \in L(m), m \in M \quad (2.6)$$

$$x_f c_f \geq \sum_{i \in I} \delta_f^i y_i \quad f \in F \quad (2.7)$$

$$\sum_{f \in F(s)} x_f \geq f_s \quad \forall s \in S \quad (2.8)$$

$$\sum_{f \in F} \gamma_f^t x_f \leq r_{max} HC_t \quad \forall t \in T(a), a \in A \quad (2.9)$$

$$x_f \in \mathbb{Z}^+ \quad \forall f \in F \quad (2.10)$$

$$y_i \in \mathbb{Z}^+ \quad \forall i \in I \quad (2.11)$$

2.6 Solution Algorithm

As mentioned earlier, due to large problem size, obtaining an exact solution is difficult. Additionally, because of the aggregate nature of our analysis, approximate solution methods are sufficient. We solve the FPFA linear programming (LP) relaxation and round up the resulting solution to the nearest integer values greater than or equal to the LP optimal solution. Due to the nature of the constraints in FPFA formulation, none of the constraints is violated if segment frequencies are increased.

Solution to the FPFA problem involves determining daily frequency values, which are relatively large integers. The impact of rounding up is comparatively small. But for the TD problem, the solutions are highly fractional because the hourly frequencies are much smaller than daily frequency values. Therefore, the rounding up procedure worsens the objective function dramatically. Much of the solution's non-integrality stems from the markets in which demand is extremely low per day. Therefore, the LP solution has very small fractions of flight legs serving small markets, and the solution is not of sufficient quality for our purposes. Therefore, we solved the TD problem in two steps. The TD solution procedure is described schematically in Figure 2-2. In the

first step, the TD LP relaxation was solved for a smaller sub-problem involving all the markets with a daily demand of at least 250 passengers. These constituted over 60% of the total demand. These markets include all the candidates for non-stop service bypassing the hub. This LP solution was rounded upward to the nearest integers. Due to the nature of constraints in the TD formulation, an increase in value of any x variable in a feasible solution does not affect feasibility. Moreover, most of flights in these important markets serve as connecting flights for smaller markets. Therefore the additional seating capacity made available due to flight rounding is very likely to be utilized to carry passengers in remaining smaller markets. The remaining problem was solved using a greedy heuristic, as depicted in Figures 2-3 and 2-4.

In the first step of the heuristic, as described in Figure 2-3, additional non-stop flights are scheduled to satisfy the demand in markets with a daily demand of less than 250 passengers and with at least one endpoint at a hub (the *Hub Markets*). The markets are processed one after the other in decreasing order of demand. Additional flights are scheduled such that the maximum utilization ratio across all the affected ATPs is minimized at each step. Scheduling an additional non-stop flight increases the utilization ratio for the origin airport during the departure hour and also increases the utilization ratio of the destination airport during the arrival hour. Among all the departure time choices for a MTP, the one which minimizes this maximum utilization ratio is chosen.

In the second step of the heuristic, as described in Figure 2-4, additional flight legs are scheduled to satisfy the demand for markets with a daily demand of less than 250 passengers, where neither endpoint is a hub (the *Non-hub Markets*). The demand for these small, non-hub markets has to be satisfied by one-stop itineraries. The algorithm processes the markets in decreasing order of demand and schedules additional flights on first or second or both legs of an itinerary so as to minimize the maximum utilization ratio among all the affected ATPs.

The optimization algorithm ignores the aircraft flow balance constraints, which is an important component of the airline scheduling procedure. The purpose of this study is not to come up with a schedule that can be operated using actual aircraft

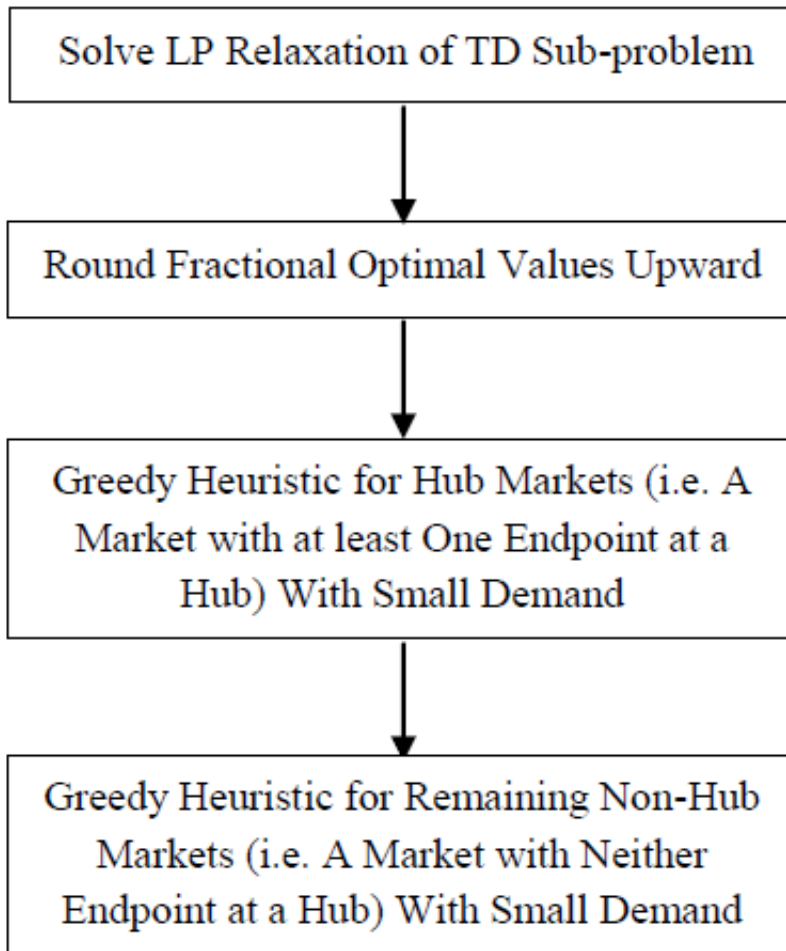


Figure 2-2: Timetable development algorithm

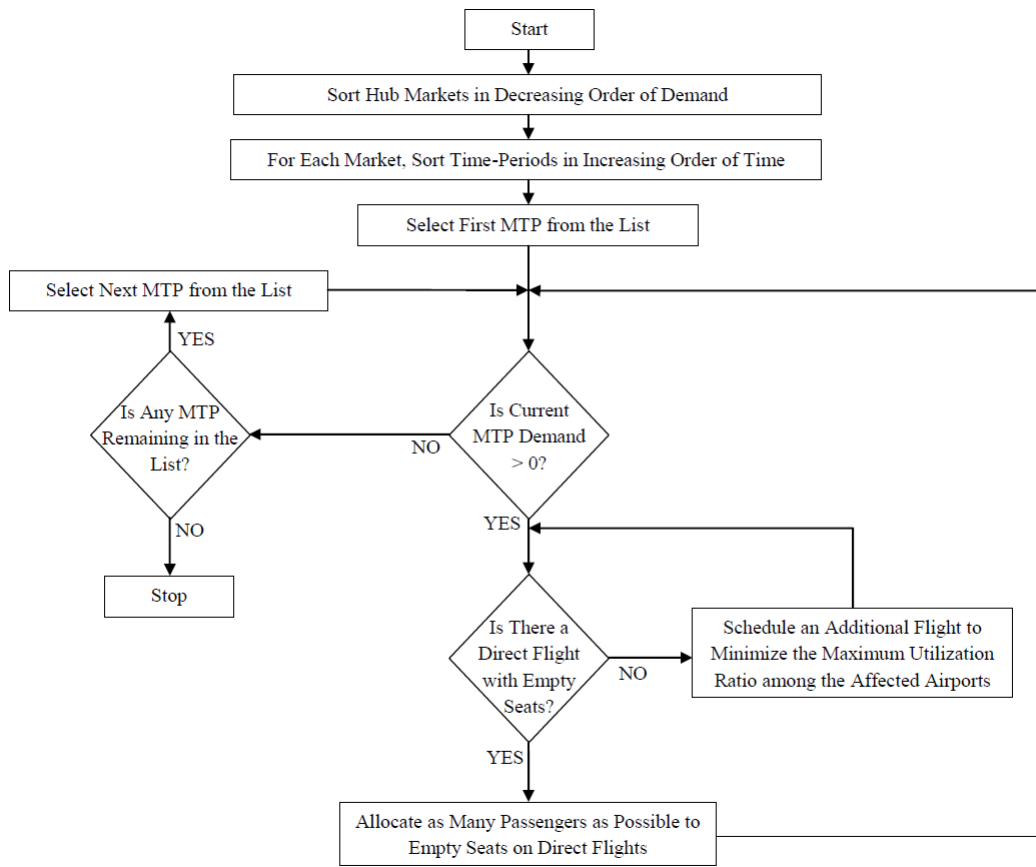


Figure 2-3: Greedy heuristic for hub markets with small demand

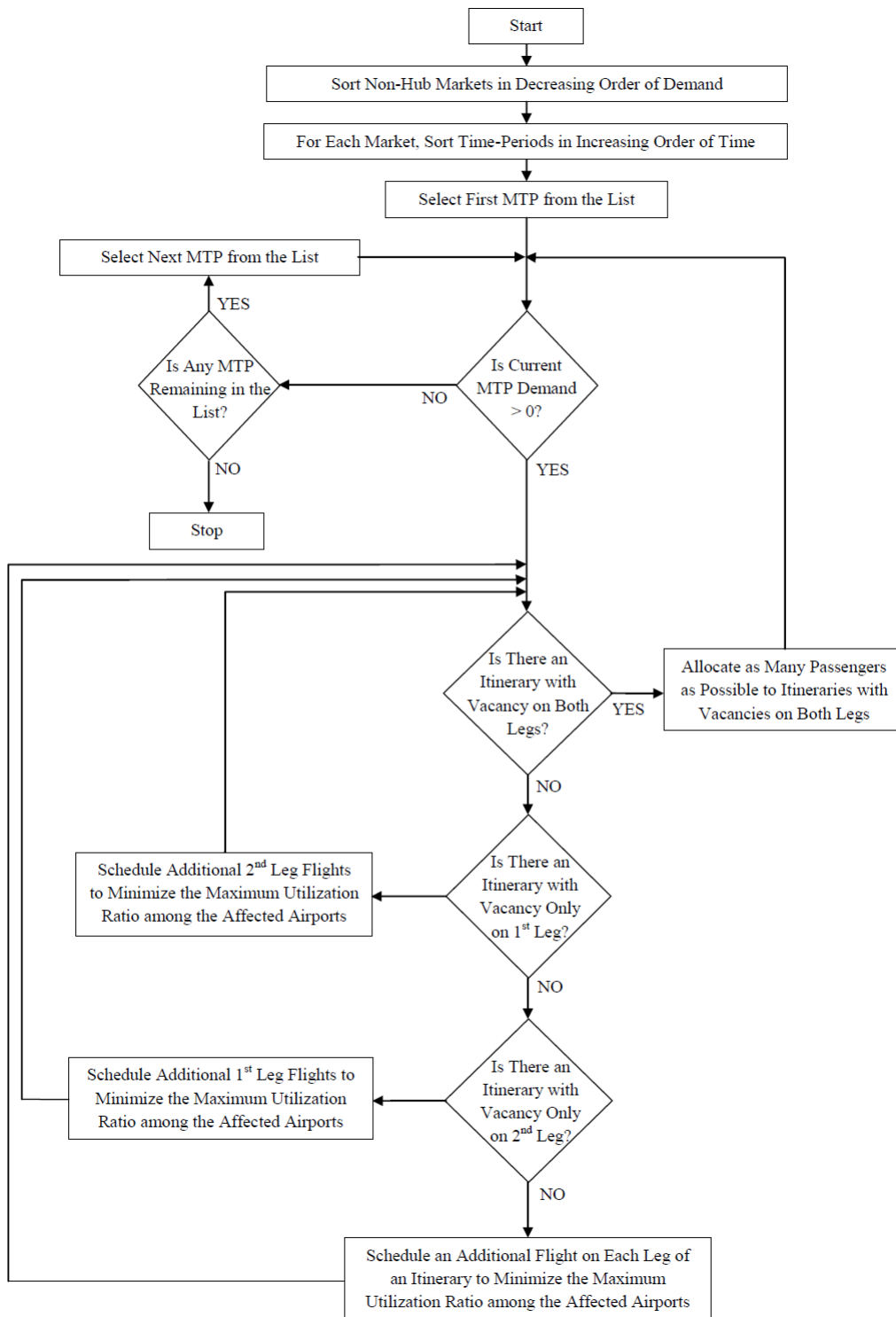


Figure 2-4: Greedy heuristic for non-hub markets with small demand

fleets- our goal is not to design a monopolistic airline schedule for the US- but rather to find a lower bound on the levels of delays. However, to assess the potential impact of aircraft balance requirements, after completing the optimization process, we performed a post-processing step wherein additional flights were added to the optimal unbalanced schedule in order to balance it. Fortunately, because the passenger demands typically are balanced in both directions of a market, the resulting schedule is not too far from a balanced schedule. In our algorithm, we add one flight at a time from an airport with a surplus of a particular aircraft type to an airport with a deficit of the same aircraft type. The departure time of each additional flight is chosen to greedily minimize the maximum utilization ratio of the affected ATPs at each step. The results presented in the next section provide the statistics on both the balanced and unbalanced single airline (SA) networks, and compare these statistics with the existing network.

Flight delays are estimated using the Approximate Network Delay (AND) model, described by Odoni, Pyrgiotis and Malone [70] as follows: AND is a stochastic and dynamic queuing model. It has two main components, a queuing engine (QE) and a delay propagation algorithm (DPA). AND treats a network of airports as a set of interconnected queuing systems where each airport is modeled as an $M(t) | E_k(t) | 1$ queuing system. The queuing system is characterized by a non-stationary Poisson arrival process, time-dependent k^{th} -order Erlang service-time distribution and a single server with infinite queuing capacity. AND iterates between QE and DPA by tracking each flight in the network and updating the airport demand profiles based on revised flight arrival and departure times.

2.7 Data Sources and Implementation Details

Schedules of major US domestic airlines were obtained from the Airline On-Time Performance Database provided by the Bureau of Transportation Statistics [71]. A 10% sample of the passengers carried by each airline per quarter is provided in the Airline Origin and Destination Survey (DB1B) available on the same website. An

estimate of the total passenger demand per OD market is obtained by multiplying the DB1B passenger number by 10. The top 20 airlines in the US, which constitute over 95% of the total traffic, are considered for the analysis. Demand for each market per day is divided into 4 time periods that roughly correspond to morning peak period, off-peak period, evening peak period and another off-peak period. Given the daily demands, the Decision Window Model by the Boeing Airplane Company [6] is used to calculate the demand in each MTP. Apart from these 20 airlines, there are several other flights such as cargo, general aviation and international flights which are not included in our analysis. We assume that these remaining flights continue to be operated as they currently are. We simply add those operations to the total operations at each airport for calculating the utilization ratios.

The Federal Aviation Administration has published benchmark capacity values for the 35 most congested US airports [44]. For each one of those airports, the report provides 3 values of capacity along with the corresponding probabilities of their realization, based on weather, wind and other conditions. The capacity is measured as the maximum number of operations (takeoffs and landings) possible at an airport per hour. Based on the capacity distributions for these airports, we use capacity values for each airport that correspond to the maximum capacity that is available for at least 95% of the time at that airport. We use these capacities only for calculating the utilization ratios. The delay calculation is based on the actual realized capacity values under each weather scenario.

The analysis is carried out for Tuesday, 16th October 2007. Because this analysis is aggregate in nature, we use only three generic fleet types. We call them Wide Body (WB), Narrow Body (NB) and Regional Jets (RJ). The seating capacity values for these three fleet types are chosen to be the average seating capacities of the corresponding aircraft in the 2007 fleet of the seven largest US airlines.

Realized hourly capacity data for the 35 busiest US airports was compiled for the entire year from April 2007 to March 2008 in order to generate realistic capacity reduction scenarios. Based on the total duration of capacity reduction across the 35 airports, we divided the 366 days in the year into 5 categories, namely, very good,

good, normal, bad and very bad, each containing approximately the same number of days. A typical (median) day belonging to each category was chosen and delays were computed using the corresponding realized capacity values.

2.8 Results

For our single airline network, the maximum utilization ratio over all airports over all ATPs was found to be 90.77%. In comparison, the maximum utilization ratio in the existing network was 160%. Table 2.1 provides the cumulative distribution of utilization ratios across the 35 busiest airports, across 24 hours of the day. The first column contains the percent utilization ratios and the next three columns show the number of times that value was exceeded in the existing network, unbalanced single airline (SA) network and balanced single airline (SA) network respectively. Each combination of airport and hour is counted as one observation. In case of the single airline networks, compared to the existing network, there is a substantial reduction in the hourly utilization ratios at the busy airports, which means that there would be substantially lower delays when bad weather reduces airport capacity.

Table 2.1: Comparison of airport utilization ratios

Utilization Ratio	Existing Network	Unbalanced Single Airline Network	Balanced Single Airline Network
> 150%	1	0	0
> 140%	3	0	0
> 130%	8	0	0
> 120%	19	0	0
> 110%	28	0	0
> 100%	55	0	0
> 90%	76	1	1
> 80%	133	60	83
> 70%	196	131	151
> 60%	275	196	210
> 50%	350	268	292

Table 2.2 compares some important metrics for the unbalanced and balanced single

airline networks to those for the existing network. US domestic flights in October 2007 had an average load factor of 78.45% [75]. The load factor for the single airline network is lower, reflecting that the single airline network design problem had less flexibility in selecting the most appropriate aircraft for each flight leg due to aggregation of aircraft sizes into 3 broad categories. We define block hours as the difference between the scheduled arrival time and scheduled departure time of a flight. The unbalanced single airline network requires 26% fewer flights and 8% lower total block hours than the existing network. For the year 2007, the DB1B Market database [73] shows that approximately 31.98% of the passengers traveling on major US airlines were connecting passengers, which is very close to that for the single airline network.

Table 2.2: Network performance metrics

Metric	Existing Network	Unbalanced Single Airline Network	Balanced Single Airline Network
Number of Flights	20,539	15,231	16,872
Connecting Passenger Percentage	31.98%	33.77%	33.77%
Load Factor	78.45%	71.77%	64.51%
Total Block Hours	44,103	40,756	45,739

In the single airline network, operations are more evenly spread over the day at the congested airports. We compared the means and standard deviations of hourly utilization ratios from 6:00 am to midnight at the 35 busiest airports in the US. For the existing network, the average values of the mean and the coefficient of variation (i.e., the ratio of standard deviation to mean) are 36.7% and 56.4% respectively, while for the balanced single airline network they equal 31.5% and 47.8% respectively. So the single airline schedule has not only lower average utilization but also lower variation in the number of operations scheduled. This effect is especially strong for the more congested airports. There are 9 airports, Atlanta (ATL), Washington Reagan (DCA), Newark (EWR), New York Kennedy (JFK), Los Angeles (LAX), New York Laguardia (LGA), Chicago O'Hare (ORD), Seattle-Tacoma (SEA) and San Francisco (SFO), which have a mean utilization of at least 70% between 6 am and midnight in the existing network. For these most congested airports, the mean utilization ratio

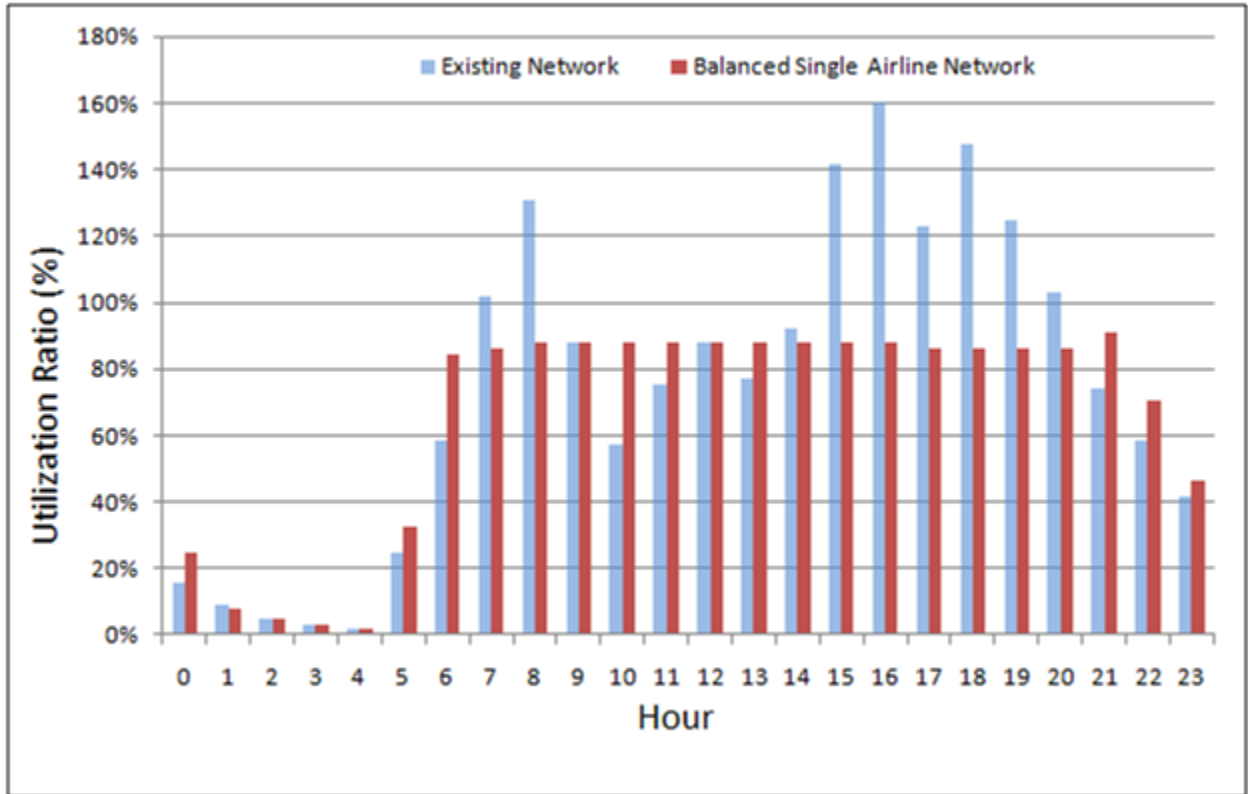


Figure 2-5: Distribution of utilization ratio over a day at JFK airport

decreases from 86.5% for the existing network to 68.7% for the balanced single airline network and the coefficient of variation averaged across these airports decreases from 29.6% for the existing network to 16.5% for the balanced single airline network. To illustrate, Figure 2-5 shows the distribution of the utilization ratio over a day at John F. Kennedy International Airport, New York (JFK), one of the busiest airports in US. This efficient utilization of off-peak hour capacity is one of the factors that contribute to lowering the congestion levels in the single airline network.

The other important factor that contributes to lower congestion is an increase in average aircraft size. Table 2.3 illustrates the number of flights being operated by each fleet type. The numbers in the parentheses indicate the percentage of all flights corresponding to that particular fleet type. The strong shift towards wide body aircraft can be observed in both the unbalanced and balanced networks. Most of the increase in the number of wide bodies comes from a decrease in usage of narrow bodies.

Table 2.3: Aircraft size distributions

Fleet Type	Wide Body	Narrow Body	Regional Jets
Existing Network	363 (1.77%)	12788 (62.26%)	7388 (35.97%)
Unbalanced Single Airline Network	3285 (21.57%)	6058 (39.77%)	5888 (38.66%)
Balanced Single Airline Network	3561 (21.11%)	6775 (40.16%)	6536 (38.74%)

Table 2.4 compares the total aircraft delay under 5 different capacity scenarios. The delay values are computed excluding any propagated delay due to late arriving aircraft. In each capacity scenario, the single airline network produces substantially lower delays, with the delay reduction ranging from 53% to 88%. As the congestion worsens, the absolute, as well as percentage, delay reduction increases. Given the way the scenarios are chosen, each chosen scenario can be considered equally likely. Thus, across different scenarios an average delay reduction of 81.72% can be achieved in the absence of competition. This provides an estimate of the inefficiencies due to the competitive scheduling practices of carriers using the existing airport infrastructure. The results suggest that congestion related delays could be reduced to less than one fifth of the existing level if there was no competition. In other words, competition is responsible for more than a 400% worsening of congestion related delays. Most of the improvement is due to efficient utilization of airport capacity during off-peak hours and a strong shift towards larger aircraft.

Table 2.4: Total flight delay under various weather scenarios

Scenario	Existing Network Delay (aircraft-min)	Balanced Single Airline Network Delay (aircraft-min)	Percentage Reduction
Very Good	7495.05	3552.97	52.60%
Good	14682.3	4090.06	72.14%
Normal	27998.76	5940.4	78.78%
Bad	35081.44	6289.88	82.07%
Very Bad	64026.52	7421.76	88.41%
Average	29856.82	5459.02	81.72%

In Figure 2-6, we show the total number of operations at the 35 busiest airports under the existing network and under the balanced single airline (SA) network. The

airports are arranged from left to right in decreasing order of daily utilization ratios in the existing network. The number of operations in the SA network is lower than in the existing network in all but 7 out of these 35 airports, with the total number of operations being 13.9% lower. This effect is especially strong in the cases of the leftmost four airports, ORD, EWR, JFK and LGA, which are the top four airports in terms of the daily utilization ratio, as well as average flight delays. The total number of operations at these 4 airports is 29.6% lower in the SA network than in the existing network. The DFW, DEN and PHX airports, which have comparatively lower utilization ratios and a lot of excess capacity in the existing network, are the only airports with significant increases in the number of operations. These were also chosen as 3 of the 5 hubs in the single airline network because of these characteristics. The overall increase in operations at these three airports is 62.9%. Finally, operations are substantially reduced at airports such as Salt Lake City (SLC), Cleveland (CLE), Cincinnati (CVG) and Memphis (MEM), which already have low utilization ratios in the existing network. This apparent anomaly stems from the fact that these are the hub airports for at least one major existing network carrier, but none of them has a strong local market with high passenger demand. These existing hubs tend to rely on connecting passenger flow, in the absence of which the number of operations is expected to be lower.

2.9 Summary

In this chapter, we calculated a lower bound on system-wide delays that can be achieved in the absence of competition while satisfying all passenger demand and maintaining a comparable level-of-service to that achieved with the existing networks of multiple competing airline. Aggregate integrated integer programming models of timetable development and fleet assignment were developed and solved using a heuristic-based approach. Obviously, getting rid of (or even reducing the extent of) competition between airlines is a highly unrealistic strategy in the real world. It is also not the point of the research presented in this chapter. However, what

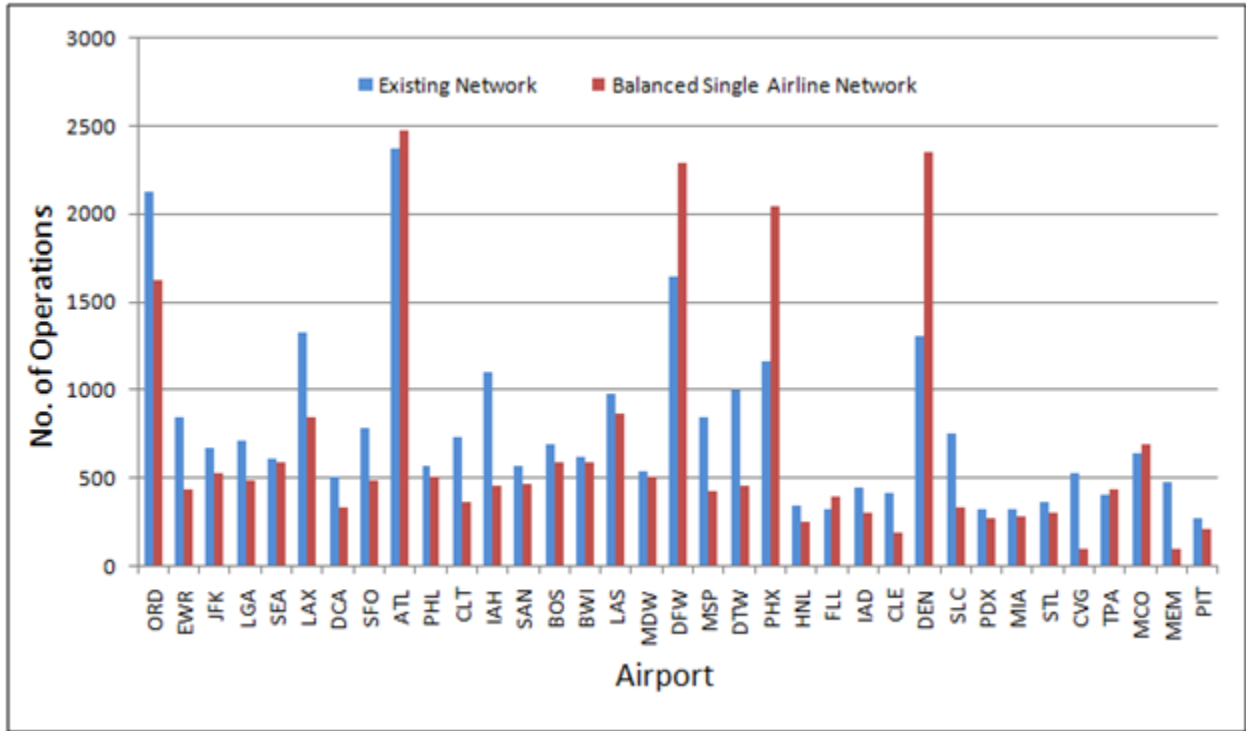


Figure 2-6: Number of operations by airport

our results show is that there is significant room for improvement in the level of congestion with the existing airport infrastructure. Passenger demand is currently at a level where delays are avoidable to a large extent. Given the available capacity, efficient administrative controls and/or market-based mechanisms can potentially lead to substantial reductions in airport congestion and delays.

These results provide critical insights into the nature and causes of aviation delays, allowing better planning and utilization of aviation infrastructure. In particular, the results help differentiate between the delays caused by insufficient capacity and delays caused by inefficient utilization of capacity. The models and solution methods presented here can also be used for analyzing the best-case delay levels under different future scenarios, with different levels of demand and capacity.

The critical next step is to devise intelligent mechanisms and incentives that will result in airlines gradually migrating their schedules from those in place today towards the delay-minimizing schedules presented in this chapter, in the presence of market competition. In Chapter 5 of this thesis we demonstrate that, under frequency com-

petition between carriers, simple changes to existing administrative slot controls at a congested airport can lead to improved schedule reliability and delay reduction while maintaining a comparable level-of-service through the provision of adequate frequency and seating capacity. Moreover, they also result in a considerable increase in airline profits. This is achieved through reductions in total allocated airport capacity and aircraft upgauges. Market-based slot pricing and auctioning mechanisms, mentioned earlier, are expected to achieve similar improvements in schedule reliability and delays. However, due to the associated monetary payments, the overall impact on airline profits needs to be evaluated carefully while accounting for the effects of airline schedule competition. We provide detailed analysis of the impacts of congestion pricing mechanisms under airline competition in Chapter 6, while the analysis of slot auctions is left as future work.

Chapter 3

Quantification and Analysis of Passenger Delays and Disruptions

3.1 Introduction

Passenger delays and disruptions significantly degrade the travel experience of passengers. Accurate estimation of passenger delays is necessary to evaluate the performance of the National Aviation System (NAS) adequately from the passengers' perspectives. Furthermore, a detailed analysis of the spatio-temporal patterns in passenger delays is important also for enabling passenger-centric decision making on part of the airlines as well as the policy-makers. With 2007 being one of the worst years in terms of delays in the NAS, several studies have tried to quantify the total costs of delays to the passengers in that year. As per the U.S. Congress Joint Economic Committee report, the total cost of passenger delays was estimated at \$12 billion for the year 2007 [87]. For the same year, the analysis by Air Transport Association estimated the passenger delay costs to be approximately \$5 billion [5]. The Center for Air Transportation Systems Research (CATSR) at the George Mason University estimated the value to be \$8.5 billion [88]. Such large differences in these estimates point to a need for a more transparent and rigorous approach to passenger delays estimation problem.

An important issue with each of these three passenger delay estimation studies is that they do not account properly for the passenger delays arising from flight cancel-

lations and/or missed connections. In particular, the Congress Economic Committee report ignores flight cancellations as well as missed connections [87]. The CATSR report takes flight cancellations into account, but ignores all passenger connections resulting in double counting of passengers on connecting flights [88]. Therefore, the CATSR methodology cannot account for missed connections. The Air Transport Association’s report does not account for flight cancellations and missed connections either [5].

Flight cancellations and missed connections are known to be the reasons behind a significant part of the delays to passengers [23]. Bratu and Barnhart used one month of proprietary passenger booking data from one large legacy carrier to perform an analysis of passenger delays, which showed that the delays due to itinerary disruptions such as cancellations and missed connections represent a significant component of the overall passenger delays [23]. The difficulty in extending this analysis system-wide is that the publicly available data sources do not contain passenger flows disaggregated by individual itineraries. For example, the public data sources do not provide information on how many passengers planned to take the 7:05am Delta Airlines flight from Boston Logan (BOS) to Atlanta (ATL) on a given day followed by the 12:25pm flight from Atlanta (ATL) to Miami (MIA), or even the number of non-stop passengers on either of these flights. The publicly available data sources contain either monthly or quarterly aggregates of passenger flows, reported based only on the origin, connection, and destination airports. In this research, we have developed methodologies to address these very limitations. We estimate passenger itinerary flows using a novel discrete choice-based approach. We use the resulting disaggregate data for accurate estimation of passenger delays and provide valuable insights into the important factors affecting passenger delays and disruptions.

A comprehensive estimation of passenger delays using the Bratu and Barnhart approach requires passenger booking data across different carriers for the entire period under consideration. Such data is not available publicly and is usually very difficult to acquire from the individual airlines. Zhu tried to address the problem of estimating passenger itinerary flows using an allocation approach based on linear

programming [101]. This approach does not allow for incorporating secondary factors, such as connection time, which play an important role in passenger itinerary choice. Also due to the extreme point optimal solution to the linear programming model, a large number of flights end up with 0% or 100% load factors as per the estimated passenger flows. The discrete choice-based methodology presented in this chapter allows us to overcome both these difficulties. Coldren, Koppelman and others have applied discrete choice models to estimate airline itinerary shares from passenger booking data [35, 34]. They use this approach to forecast the share of passenger demand for a market (i.e., all air travel from an origin to destination) that will use each of the available itineraries. Thus, theirs is a more general problem where all the routes between the origin and destination are considered simultaneously for the estimation process. In our problem, due to the manner in which publicly available passenger flow data is aggregated, we already have information at a somewhat more disaggregate level compared to these two studies. We know the number of passengers on each combination of carrier and route. Our problem is to estimate the share of passenger demand for a carrier-route combination that used each of the individual itineraries corresponding to that carrier-route combination. Our discrete choice-based methodology is similar to that employed by Coldren and Koppelman [35, 34].

Most existing studies analyzing passenger delays make a number of assumptions in order to use aggregate data because of the unavailability of disaggregate data. In order to circumvent the problem of unavailability of passenger itinerary flow data, the CATSR study ignores all passenger connections, thus treating each connecting passenger as equivalent of multiple non-stop passengers, one for each flight in the itinerary [88]. This study also assumes that the load factors for all flights on an origin-destination segment remain constant at the average monthly value. Tien, Ball and Subramanian develop a structural model of passenger delays but, due to a lack of disaggregate data, have to rely on various assumptions for the values of key parameters which are difficult to verify [95]. A measure drawback of these studies is that the results cannot be validated due to unavailability of passenger itinerary flows data. The extensive database of passenger itinerary flows and delays generated as a result of our

research will be highly beneficial for detailed validation of these earlier approaches.

The major contributions of our research fall into three broad categories: 1) an approach for disaggregating publicly available aggregate passenger flows data, 2) an analysis of the spatio-temporal patterns in passenger delays using these estimated disaggregate passenger flows, and 3) an investigation of the causes and costs of passenger travel disruptions by applying data analysis and statistical modeling to historical flight and passenger data. Section 3.2 provides a brief overview of our passenger delay estimation methodology. The reader is referred to Barnhart, Fearing and Vaze [12] for a much more detailed discussion of the methodology and the discrete choice model estimation results. Section 3.3 summarizes the passenger delay results and discusses several key findings. These findings enhance our understanding of the complex performance characteristics of the National Aviation System and demonstrate the breadth of analytical possibilities based on the methodologies that we have developed. Section 3.4 presents a simplified, linear regression-based approach for passenger delay estimation, bypassing the passenger allocation and re-accommodation process. Sections 3.5 and 3.6 develop insights into the disruption performance of the U.S. National Aviation System through data analysis and statistical modeling to analyze flight cancellations and missed connections, respectively. Section 3.7 concludes the chapter with a summary of our findings.

Several parts of this chapter make extensive use of the 2-letter carrier abbreviation codes and the 3-letter airport abbreviation codes. For ease of reference, each of the carrier and airport abbreviations used is listed in the Appendix A at the end of this thesis.

3.2 Passenger Delay Estimation Methodology

In this section, we describe the methodology used for estimating passenger delays. The passenger delay estimation process can be divided into two sequential steps: 1) Passenger Itinerary Flow Estimation, and 2) Delay Calculation, which are described in Sub-sections 3.2.1 and 3.2.2 respectively.

3.2.1 Passenger Itinerary Flow Estimation

In this sub-section, we describe the passenger itinerary flow estimation problem, data sources, pre-processing of data, and the discrete choice model for flow estimation.

Problem Statement: The problem can be stated in terms of itineraries, carrier-segments, and carrier-routes. We defined these three terms as follows. An *itinerary* is defined to be a sequence of connecting flights that represents a one-way trip, including scheduled times and airports for departure, connection (if any), and arrival. For an operating carrier providing non-stop flight service between the origin and destination airports, a *carrier-segment* is the combination of an operating carrier, origin, and destination. Finally, a *carrier-route* is a sequence of carrier-segments that represents the flight path a passenger could travel from the origin of the first carrier-segment to the destination of the last carrier-segment. Then a passenger itinerary flow estimation problem is that of combining the monthly aggregated passenger flow data by carrier-segments with the quarterly aggregated passenger flow data by carrier-route to allocate passengers to plausible itineraries.

Data Sources: The monthly aggregates of numbers of passengers and seats flown on each carrier-segment and aircraft type are obtained from the T-100 Domestic Segment (T-100) database [75]. A 10% sample of the quarterly aggregates of domestic passenger flows is obtained from the Airline Origin and Destination Survey (DB1B) [73]. Proprietary passenger booking data from one large carrier in the United States for the 4th quarter of 2007 is used for estimating the parameters of the discrete choice model.

Preprocessing of Data: Data pre-processing step involves, 1) generation of the set of potential itineraries, and 2) estimation of the number of passengers traveling on each carrier-route for each month. We generate one non-stop itinerary corresponding to each flight. For each one-stop carrier-route in DB1B ticket data, we create potential itineraries that have reasonable connection times (between 30 minutes and 5 hrs). Estimation of the number of passengers traveling on each carrier-route for each month is performed by scaling the quarterly carrier-route-wise DB1B 10% ticket sam-

ple flows using the monthly carrier-segment-wise T-100 flows for each carrier-segment associated with that carrier-route. Different scaling factors are used for each combination of carrier-segment and month. Such differential scaling is necessitated because of the absence of information on domestic component of international itineraries and non-uniformity of the sampling of domestic itineraries in DB1B ticket sample. Please refer to Barnhart, Fearing and Vaze [12] for more details on these issues and the differential scaling procedure.

Discrete Choice Model for Flow Estimation: Given the total passenger flow and a set of potential itineraries for each carrier-route for each month, the passenger itineraries flows are obtained by allocating portion of the passenger flow to each individual itinerary in the choice set. For each passenger, this is performed by randomly selecting an itinerary from the set of potential itineraries for each carrier-route for each month. The selection probability for each itinerary is given by the multinomial Logit function given by,

$$P(i) = \frac{e^{u(x_i)}}{\sum e^{u(x_j)}} \forall \text{ itineraries } i \quad (3.1)$$

The utility $u(x_i)$ of itinerary i is a function of various characteristics of the itinerary. It is well known that passengers prefer certain times of a day over others and certain days of the week over others. Therefore, we use dummy variables corresponding to combinations of time-of-the-day and day-of-the-week characteristics of each itinerary, in its utility function. Various studies on air transportation passenger choice have helped us determine additional characteristics to be included in the utility function, including connection times, aircraft size and cancellation dummy. Theis, Adler, Clarke, and Ben-Akiva demonstrate that passengers traveling on one-stop itineraries are sensitive to connection times, specifically exhibiting a disutility associated with both short and long connection times [93]. The study by Coldren and Koppelman suggests that passengers prefer traveling on larger aircraft [34]. Finally, recent work has shown that flight cancellation decisions are affected by flight load

factors - the fraction of seats filled on each flight [94]. This suggests flight cancellations are an important factor to consider, because we would expect fewer passengers to have been booked on canceled flights.

The model is estimated using proprietary booking data from one large network carrier in the United States for the 4th quarter of 2007. The model was found to be statistically significant and all the parameter estimates were found to be intuitively reasonable. For more details on the statistical estimation and validation results, please refer to Barnhart, Fearing and Vaze [12].

3.2.2 Delay Calculation

The passenger delay calculator developed by Bratu and Barnhart [23] calculates the passenger delays through re-accommodation of passengers using information on a single carrier. In reality, the passengers booked on one carrier are sometimes re-accommodated on other carriers. For the purpose of our study involving 20 major domestic carriers, we extend the algorithm to estimate the delays for passengers re-booked on a carrier different than planned. In this sub-section we describe important steps in the extended passenger delay calculator.

Identification of Disrupted Passengers: We assume that a non-stop passenger is disrupted if the corresponding flight is cancelled or diverted, while a one-stop passenger is disrupted if one or both of the flights in the planned itinerary are cancelled or if the passenger is unable to make his/her connection. We assume that a passenger misses a connection if the available connection time is less than 15 minutes. For a non-disrupted passenger, the delay equals the delay to the final flight in his/her itinerary, while for a disrupted passenger, the delay depends on the re-accommodation process.

Identification of the Disruption Time and Airport: For passengers on cancelled or diverted flights, we use the planned departure time and the origin airport of the first cancelled/diverted flight in their itinerary as the disruption time and airport respectively. For the passengers missing their connections, we use the actual arrival time of the first flight as the disruption time and the planned connection airport as the disruption airport.

Identification of Potential Re-accommodation Itineraries: In order to be conservative in our estimates, we put a limit on the re-accommodation delay for each disrupted passenger based on the time of disruption. For passengers disrupted during daytime hours, between 5:00am and 5:00pm, we limit the re-accommodation delay to 8 hours. For passenger disrupted during evening or pre-dawn hours, between 5:00pm and 5:00am, we set the limit to 16 hours to allow for re-accommodation the following day. A potential itinerary for re-accommodation has to be scheduled to depart from the passenger’s disruption airport at least 45 minutes after the disruption time, scheduled to end at the passenger’s planned destination airport no later than the passenger’s re-accommodation delay limit. We allow for the possibility that the recovery itinerary to which a passenger is assigned may in turn get disrupted and the passenger may be required to be re-booked again. However, the total delay to a passenger cannot exceed the re-accommodation delay limit.

Re-accommodation Heuristic: Disrupted passengers are re-accommodated, from disruption airport to their final destination, in the order of their disruption times. Similar to the assumption by Bratu and Barnhart each passenger is re-accommodated on an itinerary that is scheduled to arrive the earliest at the passenger’s destination [23]. The passenger delay for these passengers is the time they reach their final destination minus the planned arrival time, ignoring negative values. We first search for potential itineraries that use airlines matching the original itinerary (e.g., the two carriers on a multi-carrier one-stop itinerary), along with any sub-contracted or parent airlines. If such itinerary is found, then the passenger is re-booked on a potential re-accommodation itinerary that is scheduled to reach the passenger’s destination at the earliest time. If no such potential re-accommodation itinerary is found, then we attempt to re-accommodate the passengers using any potential re-accommodation itinerary across all carriers. If such itinerary is found, then the passenger is re-booked on a potential re-accommodation itinerary that is scheduled to reach the passenger’s destination at the earliest time. If no such itinerary is found, then the passenger’s delay is set to be equal to the re-accommodation delay limit.

3.3 Passenger Delay Results

Using the approach described in Section 3.2 above, we estimated passenger itinerary flows and passenger delays for US domestic passengers across the entire 2007 calendar year. Table 3.1 summarizes the number of passengers and total passenger delays for the 2007 US domestic air passengers by cause of delay. As shown in the Table 3.1, only around 52% of the delays to passengers are directly caused by flight delays. Approximately 30% of the passenger delays are caused by flight cancellations and approximately 18% are caused by missed connections.

Table 3.1: Passenger delay estimates for 2008

Cause	Number of Passengers	Delay (Hours)
Flight Delays	470,601,247	125,348,408
Flight Cancellations	10,190,837	74,229,945
Missed Connections	5,740,563	44,890,612
Total	486,532,647	244,468,965

Using the aggregated results in Table 3.1 combined with the disaggregated results derived from our approach, we highlight nine key findings regarding the breakdown and causes of passenger delays. In each case, we begin by stating the finding, and then providing further details, including any definitions or assumptions, as well as further discussion of the result.

Finding 1. *The ratio of average passenger delay to average flight delay is maximum for regional carriers and minimum for low-cost carriers, owing primarily to the cancellation rates and the connecting passenger percentages.*

As above, a passenger is identified based on the carrier operating the first flight in the itinerary. We categorize American Airlines (AA), Continental Airlines (CO), Delta Airlines (DL), Northwest Airlines (NW), United Airlines (UA), and US Airways (US) as the *legacy network carriers*; JetBlue Airways (B6), Frontier Airlines (F9), AirTran Airways (FL), and Southwest Airlines (WN) as the *low cost carriers*; and Pinnacle Airlines (9E), Atlantic Southeast Airlines (EV), American Eagle Airlines

(MQ), Comair (OH), Skywest Airlines (OO), Expressjet Airlines (XE), and Mesa Airlines (YV) as the *regional carriers*.

Across all carriers in 2007, the ratio of average passenger delay to average flight delay is 1.97. For individual carriers, it ranges between 1.49 for Southwest Airlines (WN) and 2.99 for Pinnacle Airlines (9E). For the legacy network carriers, this ratio ranges from 1.65 to 2.23, with an average value of 2.03. For regional carriers, it ranges from 2.27 to 2.99 with an average value of 2.61. Last, for low cost carriers, it ranges from 1.49 to 1.89 with an average value of 1.61.

The reasons for such disparity become clear when we look at the cancellation percentages and the percentages of connecting passengers. In the year 2007, the overall percentage of canceled flights was 2.4% and the percentage of connecting passengers was 27.2%. The regional carriers had both the greatest percentage of cancellations (3.4%) as well as the greatest percentage of connecting passengers (39.6%). Low-cost carriers had the lowest percentage of cancellations (1.2%) and the lowest fraction of connecting passengers (17.0%). Legacy network carriers fell between these two extremes, both for the percentage of cancellations (2.2%) and the percentage of connecting passengers (31.0%). As we show later in Section 3.4, the percentage of canceled flights and the percentage of connecting passengers are highly correlated with the ratio of average passenger delay to average flight delay.

Finding 2. *Passengers scheduled to transfer in one of 6 airports: Newark (EWR), Chicago O'Hare (ORD), New York La Guardia (LGA), Washington Dulles (IAD), New York Kennedy (JFK) or Philadelphia (PHL), were exposed to the longest average connecting passenger delays. For each of these airports, over 10% of scheduled connecting passengers had their itineraries disrupted. These 6 airports were also among the worst airports with respect to both average delays for departing flights and departure cancellations.*

We restrict this analysis to only the connecting passengers and consider data from the top 50 transfer airports in the U.S. These airports account for nearly 98.7% of all domestic connecting passengers in the U.S. On average, 12.2% of the passengers

scheduled to connect through the 6 airports listed had their itineraries disrupted compared to just 6.9% at the remaining 44 airports. These airports were the worst transfer airports in terms of average connecting passenger delay. Across these 6 airports, the average delay per connecting passenger of 78.5 minutes was 32.9 minutes more than that at the remaining 44 airports (45.6 minutes). These 6 airports are among the 9 worst transfer airports in terms of departure cancellation rates and the 7 worst transfer airports in terms of average delays for departing flights. The worst transfer airports based on departure cancellation rates also includes Reagan (DCA), Boston (BOS), and Dallas / Fort Worth (DFW). DFW is also on the list of transfer airports with the worst average delays for departing flights, rounding out that list.

Finding 3. *Domestic passenger connections are highly concentrated at the top three transfer airports: Atlanta (ATL), Chicago O’Hare (ORD), and Dallas / Fort Worth (DFW), representing approximately 43.2% of planned passenger connections. As such, ATL, ORD, and DFW are responsible for more than 40% of domestic missed connections, and contribute to more than 43% of all delays to connecting passengers.*

As above, we restrict this analysis to only connecting passengers and consider data from the top 50 transfer airports in the U.S. Approximately 43.2% of these connecting passengers connect either at ATL, ORD, or DFW. Consequently, the largest numbers of connecting passengers either miss their connections or are alternatively disrupted at one of these three airports, representing 44.5% of all disrupted connecting passengers and 40.5% of all misconnecting passengers. In comparison, only 15.3% of all connecting passengers connect at the next three largest transfer airports: Denver (DEN), Phoenix (PHX) and Houston (IAH), which contribute to 16.8% of the missed connections and 15.3% of the delays to connecting passengers. Among all transfer airports, ATL contributes the most to total connecting passenger delays (15.8%), because it has the highest number of connecting passengers (17.9% of all connecting passengers), even though its average connecting passenger delay is below average (43.9 minutes vs. 49.9 minutes on average). DFW contributes 13.0% of the total connecting passenger delays while servicing just 11.0% of scheduled connecting pas-

sengers due to a higher than average connecting passenger delay (59.07 minutes). For ORD, the discrepancy is even more extreme, as it services just 9.1% of all connecting passengers, but corresponds to 14.4% of the total connecting passenger delays. This substantial discrepancy is due to ORD having the second highest average connecting passenger delays (78.4 minutes behind EWR at 93.1 minutes).

Finding 4. *Average evening passenger delay is 86.8% greater than the average morning passenger delay. One important reason for this difference is the 89.4% greater average evening flight delay compared to the average morning flight delay. The other important reason is the greater ease of re-booking for the passengers disrupted in the morning compared to those disrupted in the evening, as reflected by 66.3% higher average disrupted passenger delay to evening passengers compared to that for the morning passengers.*

For this analysis, all passengers and flights are categorized as morning or evening depending on the planned departure time from their origin airport. Any passenger (or flight) with planned local departure time between midnight and 11:59 am is denoted as a morning passenger (or morning flight) while any passenger (or flight) with planned local departure time between noon and 11:59 pm is categorized as an evening passenger (or evening flight). Note that one-stop passengers are categorized depending on the planned departure time of the first flight in the itinerary. According to this definition, 41.0% of the flights were classified as morning flights and 43.8% of the passengers were classified as morning passengers.

The contribution of non-disrupted passenger delay to the total passenger delay depends mainly on the flight delays, while the contribution from the disrupted passengers depends on the percentage of disrupted passengers and average delay to disrupted passengers. For calendar year 2007, average delay for morning passengers was 20.3 minutes compared to 37.8 minutes for evening passengers. A large part of this difference can be attributed to the higher average delay to evening flights (18.5 minutes), compared to morning flights (9.8 minutes). In fact, 73.2% of overall flight delays are due to delays to evening flights.

The greater ease of re-booking is suggested by the fact that the average disrupted passenger delay to morning passengers is 320.3 minutes while that for the evening passengers is 532.6 minutes. This difference is, in part, due to the different maximum delay values used for morning and evening disruptions, though it is also heavily influenced by the near-term availability of re-booking alternatives. Evening passengers are much more likely to be disrupted at times where the next available re-booking alternative requires an overnight stay-over. Though the delay to disrupted passengers differs dramatically, the percentage of disrupted passengers does not differ much between morning (2.96%) and evening passengers (3.52%), which is due in part to the smaller difference between the percentage of canceled (or diverted) flights in the morning (2.1%) and evening (2.6%). As a result, the relative disparity between delays to morning and evening flights is greater than the disparity between delays to the morning and evening passengers. In other words, the ratio of average passenger delay to average flight delay in the morning (2.07) is slightly higher than that in the evening (2.04).

Finding 5. *The average passenger delay for the three months of summer and the three months of winter was 56% higher than for the remaining six months, with June being the worst month for air travel in terms of both total as well as average passenger delays.*

For this analysis we consider June, July, and August as the summer months; and December, January, and February as the winter months. Average passenger delay in the summer months was 37.4 minutes while that in winter months was 36.0 minutes. In the remaining 6 months, however, the average passenger delay was only 23.5 minutes. June and February were the only two months with average passenger delays greater than 40 minutes. On the other end of the spectrum, September and November were the only two months with average passenger delays of less than 20 minutes. In terms of total passengers, the summer and winter months fall on opposite extremes, with average passengers per month being 9.9% above the annual average in the summer and 10.3% below the annual average in the winter. The end result is

that total monthly passenger delays are 36.5% higher during the summer as compared to an average month, whereas total monthly passenger delays are only 7.0% higher during the winter. If load factors were to increase during the winter, it is possible that the winter months would become the worst months for travel based on passenger delays.

Finding 6. *Delay to the non-stop disrupted passengers depends on the ease of re-booking and therefore is lower for origin-destination pairs with higher daily frequency. Average delay to disrupted non-stop passengers on routes with at least 10 daily flights per carrier is 31.4% lower than the overall average for disrupted passengers, and on routes with at most 3 daily flights per carrier, it is 15.3% higher than the overall average.*

Disruptions to non-stop passenger itineraries occur due only to flight cancellations. The average delays for these disrupted passengers are dependent on the ease of re-booking which, in part, depends on the number of direct flights offered by the carrier for the corresponding origin-destination pair. The overall average delay to disrupted non-stop passengers is 443.6 minutes. When the carrier has a daily frequency of at least 10 flights for the origin-destination pair, this average decreases to 304.1 minutes. On the other hand, when the carrier has at most 3 flights per day, it is more difficult to obtain a suitable recovery itinerary, increasing the average delay of disrupted non-stop passengers to 511.5 minutes.

Finding 7. *The relative benefits of flight frequency in terms of the ease of re-booking depend significantly on load factors. On carrier-segments that have less than a 75% average load factor, average delays to disrupted non-stop passengers are 385.9 minutes as compared to 216.5 minutes when considering only those carrier-segments with 10 or more flights per day, representing a 43.9% improvement due to increased frequency. On carrier-segments with at least a 75% average load factor, a frequency of 10 or more flights per day leads to a relative reduction in disrupted non-stop passenger delays of only 16.9% (455.5 minutes vs. 378.5 minutes).*

For this analysis, we consider only those combinations of carriers and segments

which have at least 2 flights per day. Average delays to disrupted nonstop passengers on carrier-segments with at least 75% average load factor (455.5 minutes) is 18% higher than on carrier-segments with less than 75% average load factor (385.9 minutes). On average, nonstop passengers disrupted on low load factor (less than 75% full), high frequency (10+ flights per day) carrier-segments experience 55.0% less delay than their low load factor, low frequency (2 - 6 flights per day) counterparts, whereas for high load factor (at least 75% full) carrier-segments, the increasing flight frequency only reduces average delays by 23.6%. That is, though increasing flight frequency is beneficial for all disrupted nonstop passengers, the impact is largest when there are ample seats available for re-booking.

Finding 8. *Monday and Saturday have, by far, the lowest ratio of average passenger delay to average flight delay and these are the only two days when the ratio is lower than the overall average value for the week. One part of the reason is the lower percentage of canceled flights and another is the significantly higher percentage of morning passengers on these two days.*

The ratio of average passenger delay to average flight delay on Monday is 1.75 and on Saturday it is 1.88. For the remaining five days of the week, this ratio ranges between 2.00 and 2.03 compared to an overall average of 1.97 for the week. One reason for this difference is the lower percentage of canceled (and diverted) flights on these two days; 2.2% for Mondays and 1.9% for Saturday, compared to the average of 2.5% for the remaining 5 days of the week. Another important reason for this difference is the higher percentage of morning travelers on these two days. On Monday and Saturday, 47.6% and 48.3% of passengers respectively are morning passengers while only 42.3% are morning passengers for the remaining 5 days. As discussed in Finding 4, average delays for morning passengers are significantly lower than for evening passengers, due to shorter flight delays and easier re-booking earlier in the day. It is interesting to note that on Monday, average flight delays (10.3 minutes) are almost equivalent to the average flight delays throughout the week (10.2 minutes), which implies that the difference in ratio is entirely due to reduced passenger delays.

Finding 9. *Southwest Airlines has the lowest average passenger delay, nearly 55% lower than its competitors, even though its average flight delay is only 36.3% lower than other airlines. The primary reason for the smaller magnitude of passenger delays is the relative infrequency of disruptions to passenger itineraries; both in terms of a much lower number of flight cancellations and much lower percentage of missed connections.*

Over the last 5 calendar years (2005 - 2009), Southwest Airline has been ranked as the airline with highest overall on-time flight arrival performance (BTS, 2010), among all the airlines that predominantly serve the continental United States. This excludes Aloha Airlines (AQ) and Hawaiian Airlines (HA), among all the ASQP reporting carriers. Southwest has also had the overall best on-time arrival performance for 4 out of these 5 years, including 2007, which is the year of our analysis. The following analysis is performed for all the ASQP reporting carriers, excluding Aloha Airline and Hawaiian Airlines.

Southwest Airlines has an average passenger delay of 15.6 minutes, less than half of the 33.7 minute average value for the other airlines. The primary driver of this difference is not the difference between the average flight delays of Southwest Airlines (10.5 minutes) and that of the other airlines (16.2 minutes), nor the difference in average delay to non-disrupted passengers for Southwest (11.1 minutes) and that of other airlines (17.2 minutes). The major driver of Southwest Airline's passenger on-time performance is the relative infrequency of itinerary disruptions. For instance, only 1.0% of Southwest flights are canceled as compared to 2.8% of flights for other carriers. Consequently, the percentage of passengers on canceled flights for Southwest (0.9%) is nearly a third of that for other airlines (2.4%). In addition, the percentage of passengers missing a connection (out of all passengers) on Southwest (0.4%) is nearly one fourth of that for the other carriers (1.4%). This is due to the smaller percentage of one-stop passengers (15.5% for Southwest compared to 30.0% for the other carriers) and to the propensity for longer connection times (41.9% of passenger connections are longer than 1.5 hours for Southwest, compared to 36.1% for the other carriers). Thus, for Southwest airlines, only 29.6% of all passenger delay is due to

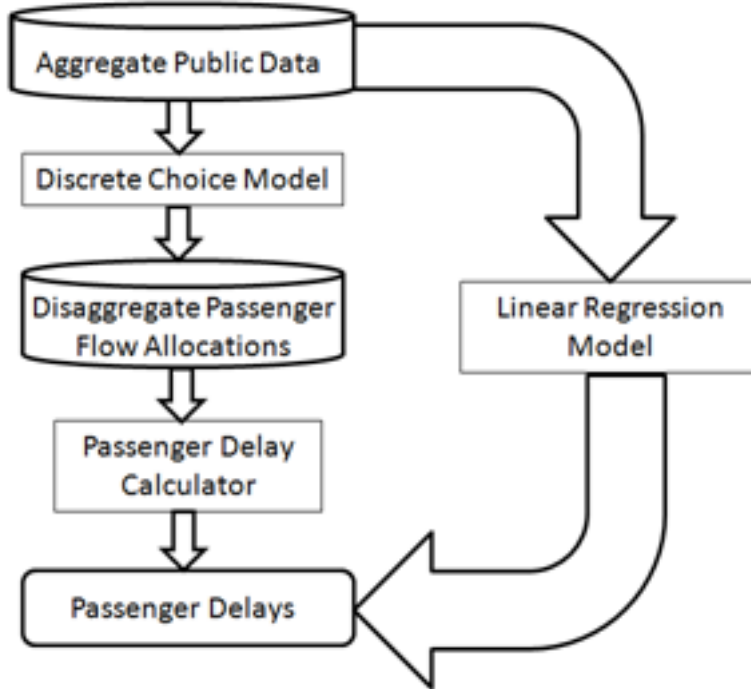


Figure 3-1: Linear regression model to bypass the passenger allocation and re-accommodation process

itinerary disruptions, while for other airlines, delays due to itinerary disruptions are responsible for 50.9% of all passenger delays.

3.4 Regression Model for Passenger Delay Estimation

In this section, we develop a linear regression model to 1) identify critical characteristics of airline networks, schedules, and passenger itineraries that affect passenger delays; and 2) estimate passenger delays directly given public data, thus bypassing the process of passenger allocation and re-accommodation. This simplified process is schematically depicted along the right path in Figure 3-1.

Flight delays influence passenger delays in the most direct way. In the absence of cancellations and missed connections, and assuming equal numbers of passengers per flight, average passenger delay will equal average flight delay. Other factors such as

cancellations and missed connections tend to increase average passenger delays beyond the average flight delay. In our regression models, we use the average passenger delay as the dependent variable, y_i .

We restrict the independent variables in our models to those which are available in public data sets such as ASQP, DB1B, etc. The purpose of this model is to see how well we can predict passenger delays based on the publicly available data without utilizing the complex itinerary flow estimation and delay calculation processes described in Section 3.2. For model estimation, we use the results of our passenger delay calculations. Each observation corresponds to a single day and airline combination. Thus, the 20 airlines in our data and the 365 days in 2007 correspond to the availability of 7300 observations for estimating the model. To describe the model, we utilize the following notation, where each observation i corresponds to a carrier-day combination:

- d_i^p = average passenger delay corresponding to observation i
- d_i^f = average flight delay corresponding to observation i
- LF_i = load factor for the carrier and month corresponding to observation i
- f_i^{canc} = fraction of canceled flights corresponding to observation i
- f_i^{conn} = fraction of connecting passengers for the carrier and quarter corresponding to observation i
- f_i^{60} = fraction of flights with at least 60 minutes of delay corresponding to observation i
- $I(\cdot)$ = the indicator function for the expression argument
- β_0 = intercept; and
- β_j = coefficient of independent variable x_j .

Average daily flight delays, the daily fraction of canceled flights, and the daily fraction of flights with at least 60 minutes of delay can be obtained from the Airline

Service Quality Performance (ASQP) data for each carrier for each day. Average monthly load factors can be obtained from T-100 segment data for each carrier for each month. Average connecting passenger percentages can be obtained from DB1B data for each carrier for each quarter. Various model specifications were tested and we present here the model specification that was found to be the most suitable. The independent variables for this model are calculated as shown in Equations 3.2 through 3.6.

$$x_{i1} = d_i^f \quad \forall i \quad (3.2)$$

$$x_{i2} = f_i^{canc} \quad \forall i \quad (3.3)$$

$$x_{i3} = f_i^{canc} xI(LF_i > 0.8) \quad \forall i \quad (3.4)$$

$$x_{i4} = f_i^{conn} \quad \forall i \quad (3.5)$$

$$x_{i5} = f_i^{60} x f_i^{conn} \quad \forall i \quad (3.6)$$

Using these definitions, the linear regression model is given by Equation 3.7.

$$y_i = \beta_0 + \sum_{j=1}^5 \beta_j x_{ij} \quad \forall i \quad (3.7)$$

Table 3.2: Regression parameter estimates

Parameter Description	Parameter	Estimate	Std. Error	p-value
Intercept	β_0	-1.34	0.24	0.00
Average flight delay	β_1	1.00	0.01	0.00
Fraction of canceled flights	β_2	458.77	2.92	0.00
Fraction of canceled flights * High load factor dummy	β_3	96.79	4.62	0.00
Fraction of connecting passengers	β_4	10.14	0.50	0.00
Fraction of connecting passengers * Fraction of flights with at least 60 minutes of delay	β_5	139.14	4.53	0.00

The estimate of β_1 is 1.00, suggesting that all else being equal, change in average flight delay results in equal change in average passenger delays. The magnitude of the estimate for β_2 demonstrates that the fraction of canceled flights has a very strong impact on passenger delays. The greater the fraction of canceled flights, the greater the average passenger delay, because passengers on canceled flights must be re-accommodated on later itineraries. Once a passenger is disrupted, re-booking requires seat availability on alternate itineraries, which means that higher average load factors reduce the probability that the passengers will be quickly re-accommodated. Thus, we would expect cancellations to have a greater impact on passenger delays when load factors are high. This effect is demonstrated in our model results through the significant positive estimated value of β_3 , which parameterizes the interaction of cancellations and load factors. In the publicly available data, there is no information on connection times, thus the best proxy for missed connections is the percentage of passengers with connections. In this context, the positive estimated value of β_4 is reasonable because the higher the ratio of connecting passengers to total passengers, the higher the proportion of missed connections, and hence the higher the average passenger delay. Out of all the connecting passengers, the fraction missing their connection depends on the fraction of flights that have large delay. We would expect the fraction of connecting passengers to have a greater impact on passenger delays when the fraction of flights with large delays is high. This effect is demonstrated through the significant positive value of estimate for β_5 .

Next, we assess the error in passenger delay estimates obtained from this simple regression model using publicly available data. Using the estimated parameter values reported in Table 3.2, we calculate the average passenger delay for each carrier-day combination. Average passenger delays for each carrier for each month are then calculated using simple averaging of the daily values. T-100 segments database provides the total segment passengers for each month for each carrier, while the DB1B database provides the fraction of connecting passengers for each carrier for each quarter. Combining the two, we estimate the number of monthly passengers for each carrier. Multiplying the average passenger delays by the number of passengers for each

carrier-month combination provides an estimate of the total passenger delay for each carrier-month. Aggregating across all carriers for the entire year, the estimate comes out to be 247,602,145 hours compared to the 244,468,965 hours estimated through the passenger allocation and delay calculation processes. That is, using our simplified regression approach, we are able to estimate annual delays within 1.28% of the totals listed in Table 3.1.

Table 3.3: Summary of error in passenger delay estimates at different aggregation levels

Aggregation Level	Passenger Allocation and Delay Calculation	Regression-based Delay Estimation
By Carrier-Day	11.1%	15.1%
Daily	10.3%	12.4%
Monthly	3.3%	8.0%
Quarterly	2.7%	8.0%

To further validate our regression approach, we compare the passenger delays obtained by the regression analysis with those obtained by applying the passenger delay calculator directly to the proprietary booking data. The percentage errors in estimates at different levels of aggregation are presented in Table 3.3. The second column lists the error in passenger delay estimates obtained by applying passenger allocation followed by delay calculation using our estimated passenger itinerary flows. The third column lists the error in passenger delays obtained from our simplified regression-based delay estimation approach. In both cases, the error is with respect to delay estimates based on the proprietary booking data. Table 3.3 demonstrates that the error decreases with increased level of aggregation for both approaches. At all aggregation levels, the errors are higher for the regression-based approach as compared to the passenger allocation and delay calculation approach. For the entire quarter, the regression model estimates are within 8.0% of the estimates based on the proprietary booking data. This suggests that the simplified regression approach provides a good alternative for estimating total delays or the total cost of delays if a more thorough analysis is not required.

As an example of the potential applications of this simplified approach to passenger delay estimation, we applied the model to estimate the passenger delays for the year 2008. The model inputs such as flight delays, flight cancellations rates, connecting passenger percentages and load factors were obtained from public data for 2008. Passenger delays for the entire 2008 year were estimated using the regression parameter estimates listed in Table 3.2. Table 3.4 compares aggregate statistics on flight schedules and passenger itineraries for the years 2007 and 2008 and summarizes total passenger delays. For 2008, the estimated average passenger delays were 6.7% less than those for 2007, mainly due to 8.8% lower average flight delays and a 7.6% lower cancellation rate. However, because of a 6.0% reduction in the number of passengers, the total passenger delay for 2008 was estimated to be 12.2% lower than that for 2007.

Table 3.4: Delay estimates using the regression-based approach for 2007 and 2008

	2007	2008	Difference
Number of Flights	7,455,458	7,009,726	-6.0%
Avg. Flight Delay (min)	15.3	14.0	-8.8%
% of Flights Cancelled	2.4%	2.2%	-7.6%
Avg. Load Factor	76.6%	76.1%	-0.6%
% of Connecting Passengers	31.7%	32.5%	2.4%
% of Flights with ≥ 60 Minute Delay	7.2%	6.6%	-8.7%
Passengers	474,003,389	445,704,815	-6.0%
Total Passenger Delay (hours)	247,602,145	217,310,671	-12.2%
Avg. Passenger Delay (min)	31.3	29.3	-6.7%
Avg. Passenger Delay/Avg. Flight Delay	2.05	2.09	2.3%

3.5 Analysis of Flight Cancellations

Flight cancellations are the second largest source of passenger delays. For calendar year 2007, only 2.1% of passengers were disrupted due to flight cancellations, and yet, these passengers accumulated 30.4% of the total passenger delays experienced for the year. Thus, it is important to understand the factors that influence flight cancellations. In this section, we attempt to identify these factors and present models

to distinguish their impacts. The majority of our analysis of flight cancellations is based on flight performance information provided in the ASQP database. In our discussion, we will often use the term *flight cancellation rate* (or simply, *cancellation rate*), defined as the ratio of the number of canceled flights to the number of scheduled flights, which we express as a percentage.

3.5.1 Airports and Carriers

Flight cancellation rates vary dramatically across airports and carriers. However, these effects are strongly related, because each airport has a different distribution of operations (arrivals and departures) across carriers. In this section, we demonstrate the dependence of flight cancellation rates on airports and carriers. In Sub-section 3.5.3, we will present models to separate these effects.

For the analysis of cancellation rates across airports, we consider the top 50 busiest airports in the U.S. in terms of number of flight operations per day. These airports constitute 77.9% of all flight operations, with 99.2% of ASQP flights departing from and/or arriving at one of these 50 airports. In our analysis, we categorize flights based on their departure airport.

For 2007, the overall cancellation rate was 2.2%, and across the top 50 airports, the cancellation rate was 2.1%. Figure 3-2 shows the cancellation rates by airport arranged in decreasing order of cancellation rate. Only 15 out of the top 50 airports have a cancellation rate greater than the overall average and 16 greater than average across the top 50. Among the top 50 airports, there is a substantial variation in cancellation rates, with the average cancellation rate across the worst 8 airports (3.8%) being more than 2.5 times the average cancellation rate across the remaining 42 airports (1.5%). These 8 airports, corresponding to only 18.6% of flight departures, contribute 31.4% of all flight cancellations. LGA (5.2%) and ORD (4.4%) are the airports with highest cancellation rates, the only two airports with cancellation rates more than twice the overall average. Each of the next six airports, EWR, DCA, BOS, JFK, IAD and DFW, has a cancellation rate between 3.2% and 3.8%. After DFW, there is a significant drop-off, with no other airport in the top 50 having a

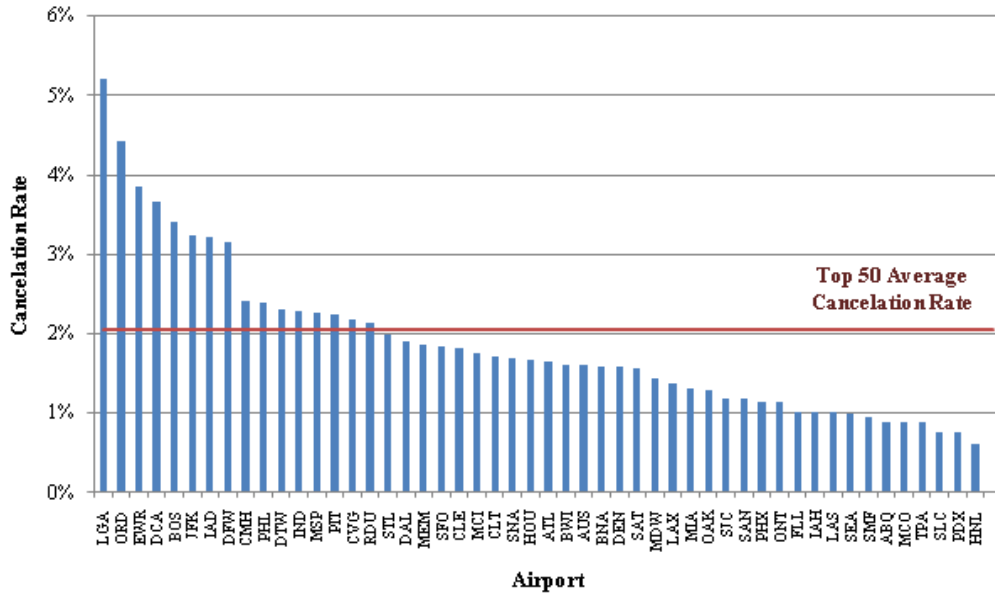


Figure 3-2: Cancellation rates for the top 50 busiest airports

cancellation rate of more than 2.4%.

In terms of the number of canceled flights, ATL, DFW and ORD top the list among U.S. airports, which is hardly surprising given that they are also the three busiest airports in terms of number of scheduled departures. These three airports correspond to 14.6% of flight departures and 19.7% of flight cancellations. The cancellation rate at ATL is well below the overall average, but the total number of cancellations is high because ATL is the busiest domestic airport, responsible for 5.6% of flight departures. ORD is the second busiest domestic airport with 5.0% of departures, but has the largest number of flight cancellations. It is interesting that the next three busiest airports in terms of total number of departures (DEN, LAX and PHX) correspond to 15.3% of all departures and yet only 9.6% of all cancellations. This result is due to the fact that the average cancellation rate for ATL, DFW and ORD (3.0%) is more than twice of that of DEN, LAX and PHX (1.3%).

For carrier-specific analysis, we classify carriers that have less than 80% of their operations in the continental U.S. as *non-continental carriers*. We further categorize the remaining 17 *continental carriers* as *legacy network carriers*, *low-cost carriers* and *regional carriers*. We categorize American Airlines (AA), Continental Airlines (CO),

Delta Airlines (DL), Northwest Airlines (NW), United Airlines (UA), and US Airways (US) as *legacy network carriers*. JetBlue Airways (B6), Frontier Airlines (F9), AirTran Airways (FL), and Southwest Airlines (WN) are classified as *low-cost carriers*; and Pinnacle Airlines (9E), Atlantic Southeast Airlines (EV), American Eagle Airlines (MQ), Comair (OH), Skywest Airlines (OO), Expressjet Airlines (XE), and Mesa Airlines (YV) as *regional carriers*. Aloha Airlines (AQ), Hawaiian Airlines (HA) and Alaska Airlines (AS) are the three non-continental carriers. For this analysis, a passenger scheduled to travel on a one-stop itinerary which includes flights operated by two different carriers is categorized based on the carrier for the first flight in the itinerary.

Among the four categories of carriers, cancellation rates are highest for the regional carriers, and lowest for the low-cost carriers, followed closely by the non-continental carriers. Legacy network carriers fall between these two extremes. The average cancellation rate for regional carriers (3.2%) is more than three times the average cancellation rate for low-cost carriers (1.0%). As a result, 39.0% of the passenger delays for regional carriers are caused by flight cancellations, as compared to 23.2% for low-cost carriers. The average cancellation rate for legacy network carriers is 2.0% and for non-continental carriers it is 1.2%. Figure 3-3 plots the cancellation rate for each airline arranged in decreasing order. In the plot, regional carriers are highlighted in blue, legacy network carriers in green, regional carriers in orange, and non-continental carriers in grey. The worst 5 carriers in terms of cancellation rates are all regional carriers, and no regional carrier has a cancellation rate below the overall average. On the other hand, every one of the low-cost and non-continental carriers has a cancellation rate below this average.

It can be difficult to separate out carrier performance from the impacts of airports. Among the legacy carriers, AA (2.8%) and UA (2.4%) have the two highest cancellation rates and are the only two legacy carriers with a cancellation rates higher than the overall average. Similarly, the cancellation rate of MQ (4.2%) is higher than all other regional carriers. In Figure 3-4, we chart the distribution of flight departures for the two worst airports in terms of flight cancellation rates (LGA and ORD). MQ,

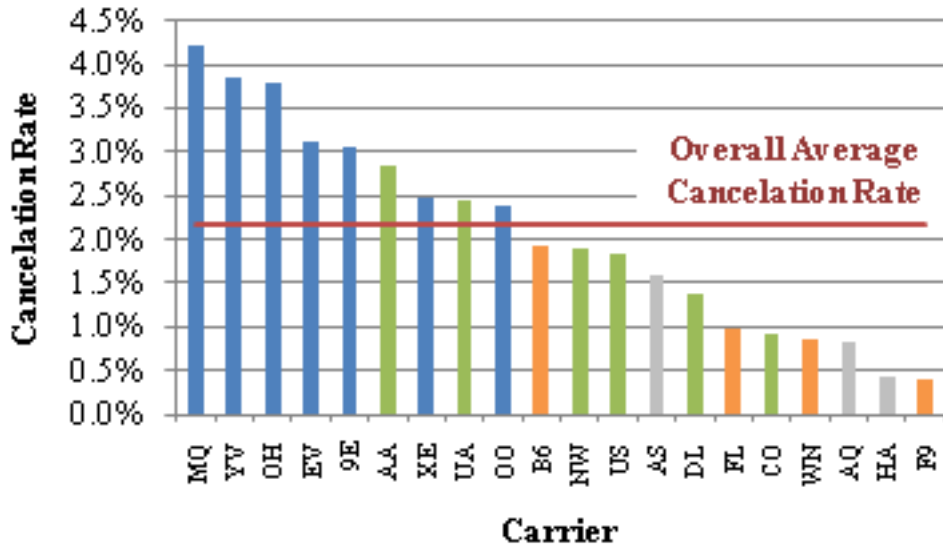


Figure 3-3: Cancellation rates by carrier and carrier type

UA and AA are the top three carriers in terms of number of departures at LGA and ORD. Of the flights departing from either LGA or ORD, 22.3% are operated by MQ, 20.5% by UA and 19.4% by AA. In addition, approximately 18.7% of the flights operated by MQ, UA, or AA depart from either LGA or ORD. This interdependence between the carrier-specific and airport-specific factors is explored in further detail in Sub-section 3.5.3.

3.5.2 Flight Frequency and Load Factors

Flight frequency and load factors play an important role in airline decisions about whether or not to cancel a flight [85, 94]. Higher frequency and lower load factors decrease the delays to disrupted passengers. In this section, we focus on how these factors impact the cancellation decision as opposed to the re-accommodation process. For our analysis, we compute average daily flight frequencies, average cancellation rates and average load factors for each carrier-segment (as defined in Section 3.2) over the course of the year. To perform these calculations, we combine the flight performance data in ASQP with the aggregate passenger demand data in T-100.

All else being equal, our results suggest that airlines prefer canceling flights on

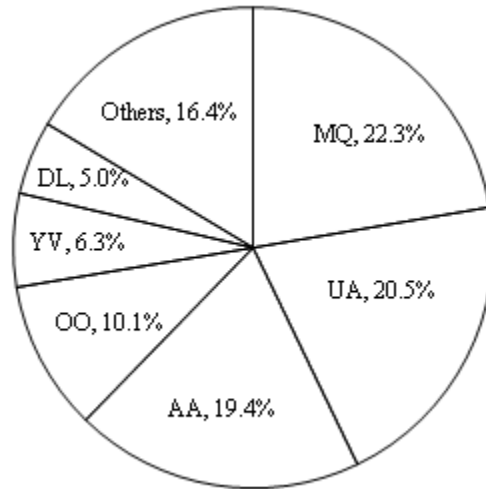


Figure 3-4: Distribution of flight departures at the two airports with the highest cancellation rates (LGA and ORD)

segments with higher daily frequency, most likely because higher frequency facilitates an easier recovery of passenger itineraries. In the ASQP database, there is a positive correlation of +7.3% between average daily frequency and cancellation rate, which is statistically significant at more than the 99% confidence level. The correlation between flight frequency and cancellation rates is especially strong for non-regional carriers in the continental U.S. For legacy network carriers, the correlation coefficient is +32.0%, and for low-cost carriers, it is +34.5%. The correlation is weaker for regional (+6.5%) and non-continental (+3.9%) carriers. Table 3.5 shows the correlation coefficient along with its statistical significance for each of the 10 carriers in continental United States, excluding the regional carriers. The correlation coefficient is positive for all 10 carriers and is statistically significant with at least a 98% confidence level for all carriers except F9.

The correlation between flight frequency and cancellation rate is highest for Southwest Airline (WN), so we conduct further analysis of Southwest’s cancellation rates. For Table 3.6, we categorize segments based on average daily flight frequency, and display the average cancellation rates for each group. The table shows dramatic variation in cancellation rates across the three categories: at least 10 flights per day, between 4 and 9 flights per day, and at most 3 flights per day. The 1.7% cancellation

Table 3.5: Correlation between average flight frequency and flight cancellation rates across carrier-segments

Carrier	Correlation	p-value
AA	+51.6%	0.00
B6	+33.6%	0.00
CO	+16.0%	0.02
DL	+22.8%	0.00
F9	+3.3%	0.78
FL	+20.5%	0.00
NW	+21.3%	0.00
UA	+40.3%	0.00
US	+33.5%	0.00
WN	+71.0%	0.00

rate for segments with at least 10 flights per day is more than double the cancellation rate for segments with 4 to 9 flights per day. The segments with the highest frequency correspond to 43.0% of Southwest’s cancellations but only 22.2% of its flights. On the other extreme, for Southwest segments with 3 or fewer flights per day, the average cancellation rate is only 0.4%, representing 12.3% of Southwest cancellations.

Table 3.6: Variation in Southwest (WN) Airlines’ flight cancellation rates based on daily flight frequency

Daily Frequency	% of WN Cancellations	% of WN Flights	Cancellation Rate
At least 10	43.0%	22.2%	1.7%
4 to 9	44.7%	51.7%	0.7%
At most 3	12.3%	26.1%	0.4%

Load factors represent another important consideration in flight cancellation decisions, because they directly impact the ease of passenger recovery. In this regard, high load factors are a problem for two reasons; they indicate that more passengers will need to be re-accommodated and that there will be fewer seats available on later flights. Therefore, all else being equal, airlines should prefer canceling flights on segments with lower load factors rather than higher load factors. To test this hypothesis, we divided all carrier-segment combinations into two categories by comparing the load factor with the median load factor value. High load factor category consists of all

carrier-segment combinations with load factors greater than the median load factor and low load factor category consists of all the carrier-segment combinations with load factors less than or equal to the median load factor. Note that there was less than a 2% difference between the average daily frequencies for the high load factor category (4.83) and the low load factor category (4.89). Nonetheless, the average cancellation rate for the low load factor category of carrier-segments (2.4%) was found to be approximately 25% greater than that for high load factor category (1.9%), confirming that load factors are a critical part of the cancellation decision.

3.5.3 Carrier Effect

Scheduling, operational, and philosophical differences between different carriers clearly impact cancellation rates. At the same time, congestion and weather patterns at an airport impact the cancellation rates for all flights at the airport, across carriers. Because the distribution of airport operations varies significantly across carriers, we would expect some carriers to have worse cancellation rates than others. For example, DL which has a primary hub in ATL is likely not forced to cancel as many flights as AA, which has a primary hub in ORD (due to persistent weather/capacity issues). Therefore, it is not clear how much of the difference between DL’s 1.4% cancellation rate and AA’s 2.8% cancellation rate is due to network differences (i.e., where the airlines operate their flights). In an effort to separate the carrier-specific impacts from the airport-specific ones, we develop a metric called carrier effect. The goal of carrier effect is to measure the relative impact of each carrier’s cancellation decision-making.

First, for each airport, a , we set the *baseline cancellation rate*, $\hat{\rho}_a$, equal to the historical cancellation rate for scheduled departures by non-hub carriers at the airport. We say that a carrier is a non-hub carrier if its operations at the airport constitute less than 10% of its total operations. We choose to eliminate hub carriers from the baseline because of the additional flexibility these carriers have based on the large number of gates, aircraft, and crews at their disposal. In Equation 3.8, we define $\hat{\rho}_a$, letting N_c^a and C_c^a represent the number of departures and cancellations respectively for carrier c at airport a , and \mathcal{H}_a^- represent the set of non-hub carriers at airport a .

$$\hat{\rho}_a = \frac{\sum_{c \in \mathcal{H}_a^-} C_c^a}{\sum_{c \in \mathcal{H}_a^-} N_c^a} \quad (3.8)$$

Next, we calculate the carrier effect, E_c , for carrier c as the historical number of cancellations divided by the baseline number of cancellations. The baseline number of cancellations for each carrier, c , and airport, a , is calculated by multiplying the number of scheduled departures, N_c^a , by the baseline cancellation rate, $\hat{\rho}_a$. In Equation 2, we formally define the carrier effect, E_c .

$$E_c = \frac{\sum_a C_c^a}{\sum_a N_c^a \hat{\rho}_a} \quad (3.9)$$

A smaller value of carrier effect is more desirable, because it indicates fewer cancellations than the baseline based on the distribution of flight departure airports. Table 3.7 lists the historical and baseline cancellation rates, the carrier effect, and the rank based on historical cancellation rate and carrier effect for each carrier. The rows in the table are sorted in increasing order based on carrier effect.

Many of the differences between the rankings according to historical cancellation rate and carrier effect are small. Out of the 20 carriers, 11 have a difference in rank of 2 or less (4 zeros, 2 ones, and 5 twos). The largest rank improvement is with B6, which has a rank of 11 based on historical cancellation rates and 7 based on carrier effect. At its two busiest departure airports, JFK and BOS, the B6 cancellation counts are well below the baseline totals. CO is ranked 5th in terms of historical cancellation rates. It is the legacy carrier with the lowest cancellation rate in spite of the fact that it has one of its hubs at EWR, where other carriers have much higher cancellation rates. When this effect is accounted for, CO becomes the second best carrier in terms of the carrier effect. Excluding the non-continental carriers, WN has the 2nd lowest historical cancellation rate, because it operates predominantly at airports with low cancellation rates such as LAS, PHX and MDW. In terms of carrier effect, WN stays in 4th place overall, but moves below both CO and FL. HA has lowest cancellation rates,

Table 3.7: Carrier effects

Carrier	Historical Cancellation Rate	Baseline Cancellation Rate	Carrier Effect	Historical Cancellation Rate Rank	Carrier Effect Rank
F9	0.41%	1.81%	22%	1	1
CO	0.91%	2.39%	38%	5	2
FL	0.99%	2.20%	45%	6	3
WN	0.85%	1.44%	59%	4	4
DL	1.37%	2.23%	61%	7	5
HA	0.42%	0.68%	62%	2	6
B6	1.94%	2.69%	72%	11	7
NW	1.89%	2.44%	77%	10	8
US	1.84%	2.05%	90%	9	9
UA	2.43%	2.49%	98%	13	10
XE	2.48%	2.46%	101%	14	11
AS	1.60%	1.50%	107%	8	12
9E	3.07%	2.87%	107%	16	13
OO	2.37%	2.09%	114%	12	14
EV	3.12%	2.72%	114%	17	15
AQ	0.84%	0.71%	117%	3	16
AA	2.83%	2.36%	120%	15	17
OH	3.78%	3.12%	121%	18	18
MQ	4.22%	3.18%	133%	20	19
YV	3.83%	2.51%	153%	19	20

and nearly 90% of its operations are at airports in the Hawaii region. These airports have very low cancellation rates in general. Therefore, in terms of carrier effect HA drops a few slots into 6th, although it still performs quite well historically canceling only 62% of the baseline. The other Hawaii based carrier, AQ, which also has about 89% of its operations in the airports in the Hawaii region, has the largest absolute change in rank, dropping from 3rd place to 16th place, when the carrier effect, rather than absolute cancellation rate, is considered. Though the differences are in most cases minor, in context, each of the changes is easy to understand. Thus, we believe carrier effect represents a better metric for evaluating the cancellation-performance of domestic air carriers as compared to the historical cancellation rate.

Many major U.S. carriers operate hub-and-spoke networks and many others have focus airports where the bulk of their activity is concentrated. Large proportions of the one-stop passengers traveling on these carriers usually connect at these hubs or focus airports. Such concentration of activity has important implications for the flight cancellation rates. More operational flexibility at the hub airport enables better recovery processes, which should be reflected in lower cancellation rates for the hubbing carrier as compared to other carriers at the airport. To measure the impact of this effect, we extend the carrier effect developed above to measure the *hub-carrier effect*, E_c^{hub} . The hub-carrier effect for a given carrier is defined in Equation 3.10 as the ratio of its cancellation rate at its primary airport of operations, "hub" , to the cancellation rate of non-hub carriers at that airport.

$$E_c^{hub} = \frac{C_c^{hub}}{N_c^{hub} \hat{\rho}_{hub}} \quad (3.10)$$

In Equation 3.11, we define the carrier's *Coefficient of Hubbing* α_c^{hub} , as the ratio of hub-carrier effect to carrier effect. We use the carrier's coefficient of hubbing to determine how much additional flexibility each carrier has at its primary hub of operations.

$$\alpha_c^{hub} = \frac{E_c^{hub}}{E_c} \quad (3.11)$$

In Table 3.8, for each of the legacy network and low-cost carriers, we list the values of E_c^{hub} and α_c^{hub} , along with the carrier effect, E_c . With the exceptions of AA and WN, the coefficient of hub effect is lower than 1 for each of these carriers. WN has, by far, the most distributed operations across different airports. Only 7.1% of the WN operations are concentrated at LAS, which contains the largest number of WN operations. No other airline in Table 3.8 has less than 15% of its operations at its main airport. Therefore, any operational flexibility afforded by having a hub is likely not as high for WN as all the other carriers, resulting in WN losing out on any incremental advantage. AA is the other carrier with a coefficient of hubbing effect greater than 1.0. AA operates at a disadvantage relative to other carriers at DFW, because flight delays and cancellations are often propagated from its secondary hub at ORD. If we were to instead treat ORD as AA's primary hub, the hub-carrier effect would be 0.81 and the coefficient of hubbing would be 0.67, which is in line with the coefficient of hubbing for other carriers.

Table 3.8: Effects of primary hub on cancellation rates

Carrier	Main Hub	% of Operations at Main Hub	Carrier Effect (E_c)	Hub-Carrier Effect (E_c^{hub})	Coefficient of Hubbing (E_h)
AA	DFW	26.00%	1.22	1.39	1.14
B6	JFK	30.60%	0.77	0.62	0.79
CO	IAH	28.50%	0.47	0.26	0.54
DL	ATL	32.10%	0.64	0.41	0.64
F9	DEN	48.70%	0.26	0.24	0.91
FL	ATL	33.30%	0.51	0.42	0.82
NW	MSP	22.50%	0.78	0.57	0.74
UA	ORD	19.30%	0.97	0.68	0.7
US	CLT	15.30%	0.88	0.5	0.57
WN	LAS	7.10%	0.61	0.71	1.15

These results suggest that there is a substantial operational advantage for flights

departing from a carrier's primary hub. This makes sense, because at a primary hub, carriers typically have numerous aircraft and crew available, providing operational flexibility that can be exploited if there are any issues with aircraft availability or crew work requirements. To further confirm this intuition, we can compare flights arriving into the primary hub with those departing from it. The operational advantages associated with the primary hub should not be afforded to flights departing from other airports, even those arriving at the primary hub. On the other hand, the impact of the airport-specific issues such as bad weather, congestion, etc. on arriving and departing flights should be comparable. Thus, the difference in cancellation rates between flights entering and exiting each carrier's primary hub provides another measure of the operational flexibility afforded by the hub.

Across the 20 carriers, the average cancellation rate for flights arriving at their respective primary hub airports (1.7%) is 9.2% higher than the cancellation rate for flights departing from their respective primary hub airports (1.6%). Table 3.9 shows the cancellation rates for flights entering and exiting the primary hub for each carrier in the continental U.S. excluding the regional carriers. It can be observed from Table 3.9 that the cancellation rate for flights entering the primary hub is higher than that for the flights exiting the primary hub in the case of all carriers except WN. The cancellation rates are calculated for all of 2007, suggesting that this effect is both significant and persistent. For WN, the cancellation rate for flights entering and exiting the main hub is almost equivalent. Thus, WN's distributed operation appears to once again deprive it of the operational flexibility afforded to other carriers at their respective primary hub airports.

Table 3.9 also lists the percentage of flights entering and exiting the main hub that suffer large delays, where large is defined as any delay greater than or equal to 30 minutes. The overall percentage of flights with large delays arriving into a primary hub (13.2%) is 19.9% lower than that for the flights departing from the primary hub. The same effect that is observed in aggregate is also observed at the individual carrier level for all carriers except B6. The flight delay results are consistent with the cancellation rates in that they suggest that carriers are able to absorb more delay

Table 3.9: Cancellation rates and large delays for flights entering and exiting the primary hub

Carrier	Primary Hub	Cancellation Rate			% with ≥ 30 Minute Delay		
		Exiting Main Hub	Entering Main Hub	% Increase	Exiting Main Hub	Entering Main Hub	% Increase
AA	DFW	3.00%	3.10%	2.50%	19.30%	15.10%	-21.50%
B6	JFK	2.40%	2.40%	2.30%	20.40%	21.10%	3.50%
CO	IAH	0.40%	0.50%	22.80%	13.90%	10.70%	-23.40%
DL	ATL	0.90%	1.10%	21.00%	12.70%	10.50%	-17.20%
F9	DEN	0.40%	0.50%	27.80%	12.60%	9.30%	-26.30%
FL	ATL	0.80%	1.00%	19.60%	14.30%	11.80%	-17.50%
NW	MSP	1.50%	1.60%	12.20%	18.10%	13.50%	-25.50%
UA	ORD	3.20%	3.50%	8.50%	22.00%	17.20%	-21.50%
US	CLT	1.20%	1.50%	25.30%	18.80%	13.80%	-27.00%
WN	LAS	0.70%	0.70%	-0.90%	11.90%	9.40%	-20.70%
Total		1.60%	1.70%	9.20%	16.40%	13.20%	-19.90%

and still operate the departing flight out of a primary hub.

3.6 Analysis of Missed Connections

Missed connections are the most significant cause of travel disruptions for one-stop passengers. For these passengers, missed connections are responsible for 57.2% of all disruptions and 40.9% of all the delays. In this section, we analyze the most important factors affecting missed connections. In our discussion, we will often use the term, *misconnection rate* which is defined as the ratio between the number of one-stop passengers who missed their connections (due to delays on the first flight in their itinerary) and the total number of one-stop passengers, which, as with cancellation rate, we will express as a percentage. Note that one-stop passengers who have at least one canceled flight in their planned itineraries are excluded from both the numerator and the denominator of the expression for misconnection rates. The analysis in this section incorporates both the flight performance data in ASQP and the estimated passenger travel and delay data described in Section 3.2.

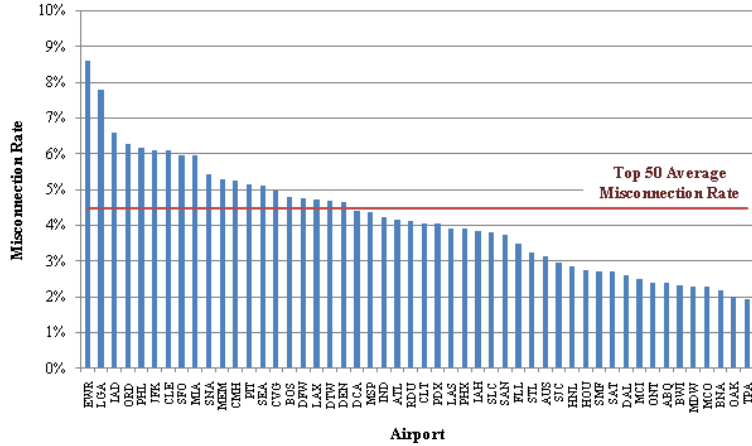


Figure 3-5: Misconnection rates for the top 50 busiest airports

3.6.1 Airports and Carriers

Just as we did for the case of the cancellation rates, for the airport-specific analysis of misconnection rates, we consider the top 50 airports in the U.S. in terms of number of flight operations per day. These top 50 airports correspond to 99.1% of planned one-stop passenger connections and 99.4% of missed passenger connections. For the following analysis, all passengers are categorized based on their connection airports.

For 2007, the average misconnection rate in the U.S. was 4.5%. For the top 50 airports, the misconnection rate ranges from 8.6% at EWR to 1.9% at TPA. In Figure 3-5, we plot misconnection rates at these airports arranged in decreasing order of misconnection rate. EWR and LGA (7.8%) are the two airports with, by far, the highest misconnection rates. At each of the next seven airports: IAD, ORD, PHL, JFK, CLE, SFO and MIA, the misconnection rate is in the range of 6.0% to 6.6%. After MIA, there is another significant drop-off, with the 41 remaining airports having misconnection rates of at most 5.4%. The average misconnection rate at the 9 worst connecting airports (6.4%) is greater than 1.5 times the misconnection rate (4.1%) at the remaining 41.

Obviously large delays to the first flight in an itinerary are primarily responsible for misconnections. Therefore, it is not a surprise that out of the 9 worst airports in terms of cancellation rates, EWR, JFK, LGA, ORD, PHL and SFO are the 6 worst airports

in terms of average arrival delays. But, clearly average arrival delays do not explain the whole story. For example, IAD, has a much lower average arrival delay than either ORD, PHL or JFK, but lies above these three in terms of misconnection rate. Another example is CLE, which is the 7th worst airport in terms of misconnection rates. CLE has a lower average arrival delay (15.0 minutes) than the overall US average (15.3 minutes), but ranks in this list above several other airports with much higher flight delays. We will address this apparent anomaly in Section 3.6.2 when we discuss schedule banking.

Much like our analysis of cancellation rates by carrier, here we categorize one-stop passengers based on the carrier of the first flight in the itinerary. Among the three categories of carriers in the continental United States, regional carriers are most severely impacted by missed connections. For regional carriers, 23.8% of all passenger delays (including both non-stop and one-stop passengers) are caused by missed connections. Low-cost carriers, on the other hand, are the least impacted by missed connections, with only 11.6% of all delays caused by misconnections. For legacy network carriers, 19.1% of all passenger delays are due to missed connections. The two drivers of this disparity are the percentage of connecting passengers and the misconnection rate, both of which are highest for regional carriers (39.6% and 6% respectively), intermediate for legacy network carriers (31.0% and 4.5%), and lowest for the low-cost carriers (17.0% and 2.8%). Average misconnection rate for the non-continental carriers is 3.5%, while the percentage of connecting passengers (11.3%) is even lower than that of the low-cost carriers. In Figure 3-6, we plot the misconnection rate by carrier in decreasing order of misconnection rates. As in Figure 3-3, regional carriers are highlighted in blue, legacy network carriers in green, regional carriers in orange, and non-continental carriers in grey.

The 5 worst carriers in terms on misconnection rates are regional carriers, and all 7 regional carriers have misconnection rates worse than the overall average (4.5%). AA and UA are the two worst legacy network carriers in terms of misconnection rates. In addition to these two, US and NW also have misconnection rates higher than the overall average. Some of these patterns in misconnection rates can be explained

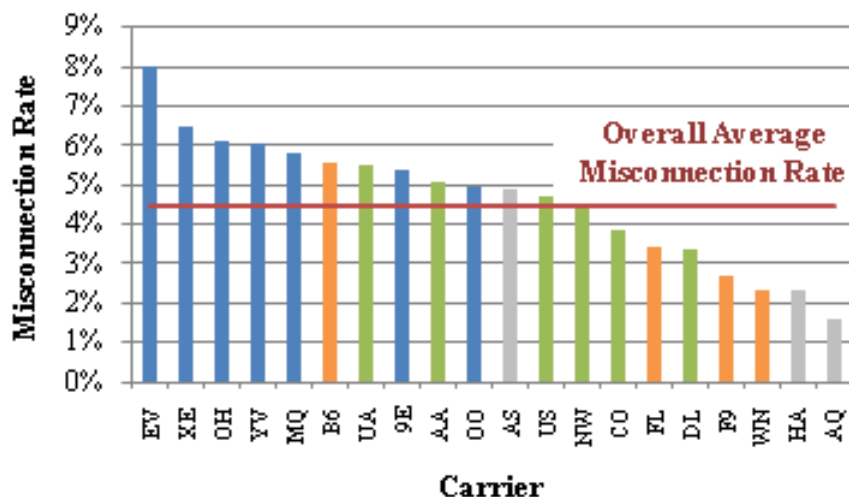


Figure 3-6: Misconnection rates by carrier and carrier type

based on average flight delays. For instance, EV is the worst carrier in terms of misconnection rates, and also in terms of average flight delays. B6 has the second highest average flight delays, and therefore performs much worse than other low-cost carriers in terms of missed connections. On the other end of the spectrum, both of the Hawaiian carriers perform exceptionally well in terms of misconnection rates because the Hawaiian airports experience very few delays, especially when compared to the continental U.S. We model the relationship between misconnection rates and various explanatory variables, including average flight delays, in Section 3.6.3.

3.6.2 Schedule Banking

Prior to the turn of century, most major hub-and-spoke carriers in the US operated one or more banked hubs. A banked hub for a carrier is a hub airport where a wave of flight arrivals (called an *arrival bank*) is followed soon by a wave of departing flights (called a *departure bank*), allowing passengers to connect between a flight in an arrival bank and a flight in the subsequent departure bank. The schedule of the hub operator carrier at a typical banked hub airport contains several such banks often separated by periods of limited activity. An example of a banked hub is provided in Figure 3-7, which shows the number of flight arrivals and departures for each hour of the day

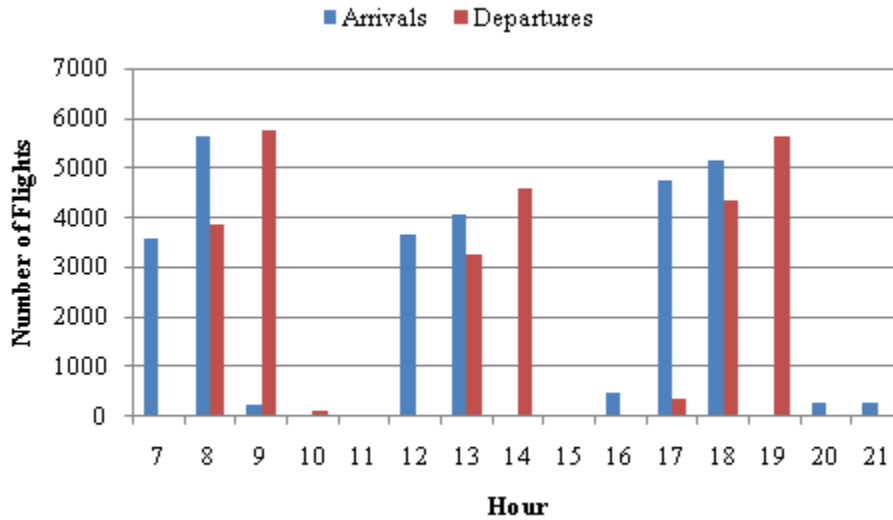


Figure 3-7: Example of banked hub operations (NW at MEM)

(from 7:00am to 10:00pm) for NW at MEM for the year 2007. Visually, it is easy to identify the three distinct banks operated by NW at MEM.

In the early 2000s, several major U.S. carriers as well as some European carriers started de-banking their schedules. De-banking allows carriers to balance resource utilization over the course of the day, reducing costs and increasing operational efficiency. An important effect of hub de-banking was an increase in average passenger connection times [55]. The trend was led by AA, who de-banked its hubs at ORD, DFW and MIA. Subsequently, UA de-banked its hubs at ORD and LAX, DL de-banked ATL and CO de-banked EWR. An example of a de-banked hub is provided in 3-8, which shows the flight arrivals and departures per hour of the day for AA at the ORD airport for the year 2007. It can be observed that the distribution of arrivals (as well as departures) per hour is much flatter than that shown in Figure 3-7. To measure the extent of banked operations by a carrier at an airport, we develop a metric called the *schedule banking coefficient*. The schedule banking coefficient for a carrier at an airport is defined as the coefficient of variation (i.e., the ratio of standard deviation to mean) of the number of arrivals per hour for that carrier at that airport, which we express as a percentage. Note that if the number of departures per hour were constant, the schedule banking coefficient would equal 0%. Larger schedule

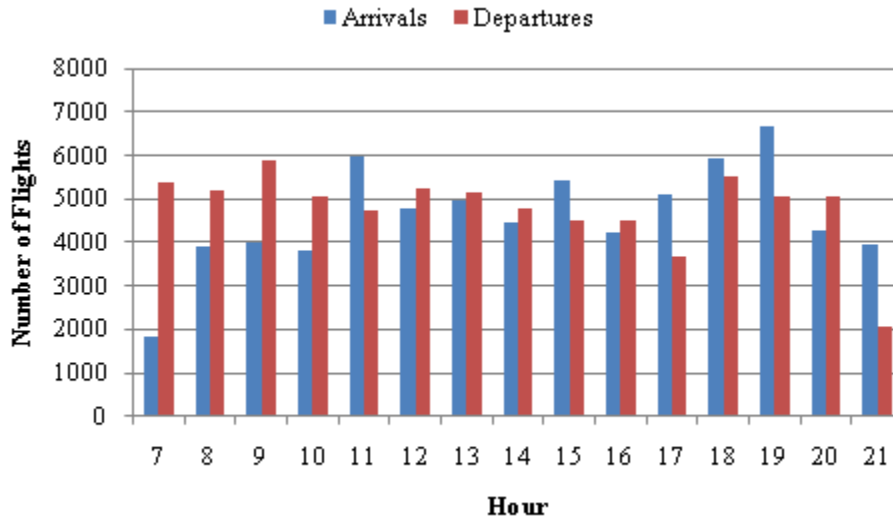


Figure 3-8: Example of de-banked hub operations (AA at ORD)

banking coefficients represent a greater extent of banked operations. For example, the schedule banking coefficient for NW at MEM is 120.9% while that for AA at ORD is 25.2%. This difference is also reflected in average connection times, with the average connection time at MEM for NW (78.9 minutes) being 21.1% lower than that for AA at ORD (100.0 minutes).

Using the schedule banking coefficient, we will now investigate how the extent of banking affects the misconnection rates at different airports. In Table 3.10, we provide another look at the list of the worst 9 airports in terms of misconnection rates, along with the corresponding average flight arrival delays.

As noted previously, the misconnection rate at IAD is higher than at ORD despite ORD having significantly higher average flight delays, which seems counterintuitive. Similarly, the misconnection rates at JFK and CLE are almost equal despite the former being significantly worse in terms of average flight delays. In Table 3.11, we add a column listing the average connection time for each of the airports included in Table 3.10, which helps to explain these apparent anomalies. For example, although the average flight delay at IAD is 6.0 minutes lower than at ORD, the average connection time at IAD is 12.8 minutes lower on average, resulting in a higher misconnection rate at IAD than at ORD. Similarly, although the average flight delay at CLE is

Table 3.10: Worst 9 airports in terms of misconnection rate

Airport	Misconnection Rate	Average Arrival Delay (min.)
EWR	8.60%	29
LGA	7.80%	23.8
IAD	6.60%	16.6
ORD	6.30%	22.6
PHL	6.20%	20.1
JFK	6.10%	23.8
CLE	6.10%	15
SFO	6.00%	17.7
MIA	6.00%	17.2

8.8 minutes less than at JFK, the average connection time is 22.1 minutes lower on average, resulting in nearly identical misconnection rates at JFK and CLE.

Table 3.11: Worst 9 airports in terms of misconnection rates with average connection times

Airport	Misconnection Rate	Average Arrival Delay (min.)	Average Connection Time (min)
EWR	8.60%	29	100.2
LGA	7.80%	23.8	90.1
IAD	6.60%	16.6	86.1
ORD	6.30%	22.6	98.9
PHL	6.20%	20.1	96.3
JFK	6.10%	23.8	103.9
CLE	6.10%	15	81.8
SFO	6.00%	17.7	102
MIA	6.00%	17.2	112.7

Given that the results presented in this research are based on the passenger itinerary flows obtained from discrete choice model estimation, we need to address the question of whether the differences in connection times are simply a construct of the passenger itinerary flow estimates or if they indicate something more fundamental about the schedule structure at these airports. In order to answer this question, we look at the schedule banking coefficients for the major carriers IAD, ORD, JFK and CLE. For each of these airports and each carrier that serves at least 10% of the

airport’s connecting passengers, Table 3.12 lists the schedule banking coefficients and the average connection times. The schedule banking coefficient for each major carrier at IAD is at least 3 times that for each major carrier at ORD, which results in much shorter average connection times at IAD than at ORD. Similarly, the schedule banking coefficient for each major carrier at CLE is at least 3 times that of B6 (the only major carrier at JFK) resulting in much shorter average connection times at CLE than at JFK. These results suggest that the lower average connection time values at IAD and CLE (and the resulting high misconnection rates) are due to the banked nature of the carrier operations at the airport rather than due to any artifacts of the passenger itinerary flow estimation procedure.

Table 3.12: Schedule banking coefficients for primary carriers at IAD, ORD, JFK, and CLE

Airport	Carrier	% of Airport’s Connecting Passengers	Schedule Banking Coefficient	Average Connection Time (min)
IAD	UA	54.60%	98.70%	88.7
IAD	YV	39.50%	99.30%	80.7
ORD	UA	36.50%	23.70%	99.6
ORD	AA	27.50%	25.20%	100
ORD	MQ	19.20%	25.30%	96.9
JFK	B6	81.20%	28.70%	105.1
CLE	XE	57.00%	64.20%	78.4
CLE	CO	40.60%	73.60%	85.6

3.6.3 Modeling Missed Connections

In this section, we present regression models to explain the variability in misconnection rates based on the insights gleaned above. As above, we categorize one-stop passengers based on the carrier that operates the first flight in the itinerary. In order to predict the misconnection rate using a linear regression approach, we aggregate individual passenger itineraries. For our model, each combination of carrier, connection airport and day corresponds to a single observation. In order to eliminate issues relating to sample size, we consider only those carrier-airport-day combinations which

include at least 100 connecting passengers. This approach results in 41,491 observations that cover approximately 98% of all one-stop passengers.

The dependent variable for our models is the average misconnection rate across the passengers corresponding to each observation. In our results, we present three regression models, each one building on the last. The incremental nature of these models allows us to determine the relative impact of each of the explanatory variables. As with the dependent variable, each of the explanatory variables is calculated by averaging the appropriate value across the passengers corresponding to the observation. Each of the regressions models is estimated by weighting the observations based on the number of connecting passengers corresponding to each carrier-airport-day combination.

As discussed at the end of Section 3.2, there is strong relationship between average flight delays at an airport and the corresponding misconnection rate. Thus, our first regression model attempts to predict the misconnection rate using average flight delays as the only explanatory variable, along with an intercept. Table 3.13 provides the estimation results for this first model.

Table 3.13: Estimation results for misconnection rate model 1 (with flight delays)

Parameter Description	Estimate	Std Error	p-value
Intercept	4.11E-03	2.49E-04	0
Average Flight Delay (min)	2.83E-03	1.16E-05	0

As expected, the coefficient of average flight delay is positive, meaning that the greater the average flight delay, the higher the misconnection rate. Also, both coefficient estimates are statistically highly significant with at least 99% confidence level. The adjusted R2 value is 0.5915, suggesting that average flight delays explain 59% of the variation in misconnection rates across our observations.

As mentioned in 3.6.2, in addition to flight delays, schedule banking and connection times impact the misconnection rates, because longer connections imply reduced risks of missing a connection. Therefore, in model 2, we add average connection time as another explanatory variable to the model. Table 3.14 shows the estimation results

for this second model.

Table 3.14: Estimation results for misconnection rate model 2 (with flight delays and connection times)

Parameter Description	Estimate	Std Error	p-value
Intercept	6.69E-02	1.57E-03	0
Average Flight Delay (min)	2.80E-03	1.14E-05	0
Average Connection Time (min)	-6.17E-04	1.53E-05	0

The coefficient estimate for average connection times is negative, implying that the higher the average connection time, the lower the misconnection rate. Also, all three coefficient estimates are statistically highly significant with at least a 99% confidence level. The adjusted R^2 value is 0.6070, suggesting that average connection times help explain another 1% of the variation in misconnection rates.

As seen in Figure 3-6 and discussed in Sub-section 3.6.1, among different carrier types, low-cost carriers have the lowest misconnection rates while regional carriers have the highest misconnection rates. To understand the magnitude of this effect, we add a 0-1 dummy variable each for the low-cost carriers and for the regional carriers. That is, any observation corresponding to the first flight being operated by a low-cost carrier will have value 1 for the low-cost carrier dummy and all other observations will have a value 0. Similarly, any observation corresponding to the first flight being operated by a regional carrier will have value 1 for the regional carrier dummy and all other observations will have a value 0. Table 3.15 shows the estimation results for this third and final model.

Table 3.15: Estimation results for misconnection rate model 3 (with flight delays, connection times and regional and low-cost dummies)

Parameter Description	Estimate	Std Error	p-value
Intercept	4.15E-02	1.78E-03	0
Average Flight Delay (min)	2.76E-03	1.14E-05	0
Average Connection Time (min)	-3.60E-04	1.75E-05	0
Low-cost Carrier Dummy	-7.61E-03	4.46E-04	0
Regional Carrier Dummy	8.26E-03	4.44E-04	0

In the third model, the coefficients for average flight delay and average connection time retained the appropriate signs (positive and negative, respectively). As anticipated, the coefficient estimate for the low-cost carrier dummy is negative and that for the regional carrier dummy is positive. Also, all the coefficient estimates are statistically highly significant with at least 99% confidence level. The adjusted R^2 value is 0.6149, suggesting that including carrier type helps to explain another 1% of the variation in misconnection rates.

3.7 Summary

Passengers are an important stakeholder group of the NAS. Quantification of passenger delays and understanding the causes of passenger delays is critical for evaluating the system performance and for motivating NAS policy and investment decisions. Flight delays are a poor proxy for passenger delays because a large proportion of the passenger delays are caused not directly by flight delays but due to passenger itinerary disruptions such as flight cancellations and missed connections. We developed a discrete choice framework for estimating disaggregate passenger flows and extended a pre-existing single-carrier passenger delay calculator to a multi-carrier setting. We found that on average each passenger suffered twice the delay suffered on average by each flight in the year 2007. Consequently, nearly half of the passenger delays in 2007 were attributed to passenger itinerary disruptions.

Airlines' strategic and operational decisions regarding network structures, hub locations, connecting bank structures, flight frequencies, flight departure schedules, flight cancellations and disrupted passenger re-accommodation significantly affect the passenger delays and disruptions. We analyzed these effects using a sequence of data analyses and statistical modeling tools and presented various insights into the variations and trends in passenger delays.

Apart from the analyses and findings presented in this chapter, our methodology and the resulting passenger itinerary flow data has already been used to estimate the overall costs of passenger delays as one component of the Total Delay Impact Study

commissioned by the FAA [7]. We foresee a large variety of further applications of this passenger delays framework for passenger-centric approaches in airline scheduling, air traffic flow management, and aviation policy-making. We hope that our research presented in this chapter serves as the beginning of several future passenger centric research studies that exploit the disaggregate passenger data generated through this research.

Chapter 4

Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion

4.1 Introduction

Since deregulation of the US domestic airline business in 1978, airlines have used fare and service frequency as the two most important instruments of competition. Passengers have greatly benefited from fare competition, which has resulted in a substantial decrease in real (inflation adjusted) airfares over the years. On the other hand, frequency competition has resulted in the availability of more options for air travel. The benefits of increased competition to the airlines themselves are not as obvious. Throughout the post-deregulation period, airline profits have been highly volatile. Several major US carriers have incurred substantial losses over the last decade with some of them filing for Chapter 11 bankruptcy and some others narrowly escaping bankruptcy. Provision of excess seating capacity is one of the reasons often cited for the poor economic health of airlines. Due to the so called S-curve relationship between

market share and frequency share, an airline is expected to attract disproportionately more passengers by increasing its frequency share in a market [17]. To increase their market shares, airlines engage in frequency competition by providing more flights per day on competitive routes. As a result, they prefer operating many flights with small aircraft rather than operating fewer flights with larger aircraft. The average aircraft sizes in domestic US markets have been falling continuously over the last couple of decades (until the recent economic crisis) in spite of increasing passenger demand [20]. Similarly, the average load factors, i.e., the ratio of the number of passengers to the number of seats, on some of the most competitive and high demand markets have been found to be lower than the industry average.

Apart from the chronic worries about the industry's financial health, worsening congestion and delays at the major US airports have become another cause of serious concern. Increases in passenger demand, coupled with decreases in average aircraft size have led to a great increase in the number of flights being operated, especially between the major airports, leading to congestion. The US Congress Joint Economic Committee has estimated that in calendar year 2007, delays cost around \$18 billion to the airlines and another \$12 billion to passengers [87].

Thus, frequency competition affects airlines' capacity allocation decisions, which in turn have a strong impact on airline profitability, as well as on airport congestion. In this chapter, we propose a game-theoretic framework, which is consistent with the most prevalent model of frequency competition. Section 4.2 provides background on airline schedule planning and reviews the literature on frequency competition. Section 4.3 presents the N-player game model. Best response curves are characterized in section 4.4. In section 4.5, we focus on the 2-player game. We provide the conditions for existence and uniqueness of a Nash equilibrium and discuss realistic parameter ranges. We then provide two different myopic learning models for the 2-player game and provide proof of their convergence to the Nash equilibrium. In section 4.6, we identify all possible equilibria in a N-player game with identical players and find the worst-case equilibrium. In section 4.7, we evaluate the price of anarchy and establish the dependence of airline profitability and airport congestion on airline frequency

competition. We conclude with a summary of main results in section 4.8.

4.2 Frequency Planning under Competition

The airline planning process involves decisions ranging from long-term strategic decisions such as fleet planning and route planning, to medium-term decisions about schedule development [16]. Fleet planning is the process of determining the composition of a fleet of aircraft, and involves decisions about acquiring new aircraft and retiring existing aircraft in the fleet. Given a fleet, the second step in the airline planning process involves the choice of routes to be flown, and is known as the route planning process. A route is a combination of origin and destination airports (occasionally with intermediate stops) between which flights are to be operated. Route planning decisions take into account the expected profitability of a route based on demand and revenue projections as well as the overall structure of the airline's network. Given a set of selected routes, the next step in the planning process is airline schedule development, which in itself is a combination of decisions about frequency, departure times and aircraft sizes for each route, and aircraft rotations over the network.

Frequency planning is the part of the airline schedule development process that involves decisions about the number of flights to be operated on each route. By providing more frequency on a route, an airline can attract more passengers. Given an estimate of total demand on a route, the market share of each airline depends on its own frequency as well as on competitor frequency. The S-curve or sigmoidal relationship between the market share and frequency share is a widely accepted notion in the airline industry [69, 17]. However, it is difficult to trace the origins and evolution of this S-shaped relationship in the airline literature [29]. Empirical evidence of the relationship was documented in some early studies and regression analysis was used to estimate the model parameters [92, 91, 89]. Over the years, there have been several references to the S-curve including Kahn [56] and Baseler [15]. In this chapter, we use a more general model that is compatible with the linear, as well as the S-curve assumptions. The mathematical expression for the S-curve relationship [89, 17] is

given by:

$$MS_i = \frac{FS_i^\alpha}{\sum_{j=1}^n FS_j^\alpha} \quad (4.1)$$

for parameter α such that $\alpha \geq 1$, where MS_i = market share of airline i , FS_i = frequency share of airline i , and n = number of competing airlines.

Some of the more recent empirical and econometric literature has focused on investigating the validity of the S-curve as the structure of airline business has evolved over the last few decades. The conclusions are mostly mixed. Wei and Hansen have provided statistical support for the S-curve, based on a nested Logit model for non-stop duopoly markets [98]. They conclude that by increasing the service frequency, an airline can get a disproportionately high share of the market and hence there is an incentive for operating more frequent flights with smaller aircraft. In another recent study, Button and Drexler observed limited evidence of the S-curve phenomenon in the 1990s [29]. But in the early 2000s, they found that the relationship between market share and frequency share is not S-shaped but rather is along a 45° straight line. This can be characterized by setting $\alpha = 1$ in equation 4.1. They, however, caution that the absence of empirical evidence for the S-curve does not necessarily mean that it does not affect airline behavior in a significant way. In an industry study, Binggeli and Pompeo concluded that the S-curve still very much exists in markets dominated by legacy carriers [19]. However, there is very little measurable evidence of the S-curve in markets where low cost carriers (LCCs) compete with each other and a straight line relationship is a better approximation for such markets. They call for a rethinking of the S-Curve principle that has been "hard-wired" in the heads of many network planners over the years.

In summary, recent evidence confirms that the market share is an increasing function of the frequency share and hence competition considerations affect the frequency decisions in an important way. However, the evidence is mixed about the exact shape of the relationships, in particular the exact value of the parameter α for different

types of markets.

Many of these studies go on to discuss the financial implications of the S-curve. Button and Drexler [29] associate it with provision of "excess capacity" and an "ever-expanding number of flights", while O'Connor [69] associates it with "an inherent tendency to overschedule". Kahn goes even further and raises the question of whether it is possible at all to have a financially strong and yet highly competitive airline industry at the same time [69].

Despite continuing interest in frequency competition based on the S-curve phenomenon, literature on game theoretic aspects of such competition is limited. Hansen [50] analyzed frequency competition in a hub-dominated environment using a strategic form game model. Dobson and Lederer [43] modeled schedule and fare competition as a strategic form game. Adler [1] used an extensive form game model to analyze airlines competing on fare, frequency and aircraft sizes. Each of these three studies adopted a successive optimizations approach to solve for a Nash equilibrium. Only Hansen [50] mentions some of the issues regarding convergence through discussion of different possible cases. But none of these three studies provides any conditions for convergence properties of the algorithm. Wei and Hansen [99] analyze three different models of airline competition and solve for equilibrium through explicit enumeration of the entire strategy space. Brander and Zhang [22] and Aguirregabiria and Ho [4] model airline competition as a dynamic game and estimate the model parameters using empirical data. Norman and Strandenes [67] also calibrate model parameters using empirical data but for a strategic form game. None of the studies mentioned so far provides any guarantee or conditions for existence or uniqueness of a pure strategy equilibrium. Brueckner and Flores-Fillol [28] and Brueckner [27] obtain closed form expressions for equilibrium decisions analytically. They focus on symmetric equilibria while ignoring the possibility of any asymmetric equilibria. Most of the previous studies involving game theoretic analysis of frequency competition, such as Adler [1], Pels et. al [79], Hansen [50], Wei and Hansen [99], Aguirregabiria and Ho [4], Dobson and Lederer [43], Hong and Harker [53], model market share using Logit or nested Logit type models, with utility typically being an affine function of the inverse of

frequency. Such relationships can be substantially different from the S-shaped relationship between market share and frequency share, depending on the exact values of utility parameters.

All of these studies involve finding a Nash equilibrium or some refinement of it. But there isn't sufficient justification of the predictive power of the equilibrium concept. Hansen [50] provides some discussion of the shapes of best response curves and stability of equilibrium points. But none of the studies has focused on any learning dynamics through which less than perfectly rational players may eventually reach the equilibrium state.

In this chapter, we use the most popular characterization of the S-curve model, as given by equation 4.1. The $\alpha = 1$ case is well suited for modeling markets dominated by low cost carriers, whereas markets dominated by legacy carriers can be suitably modeled using higher values of α . Thus, despite the mixed recent evidence about the exact shape of the market share-frequency share relationship, the model specified by equation 4.1 captures airline scheduling decisions well. We analyze a strategic form game among airlines with frequency of service being the only decision variable. We will only consider pure strategies of the players, i.e. we will assume that the frequency decisions made by the airlines are deterministic. We use the Nash equilibrium solution concept under pure strategy assumption. The research contributions of this chapter are threefold. First, we make use of the S-curve relationship between market share and frequency share and analyze its impact on the existence and uniqueness of pure strategy Nash equilibria. Second, we provide reasonable learning dynamics and provide theoretical proof for their convergence to the unique Nash equilibrium for the 2-player game. Third, we provide a measure of inefficiency, similar to the price of anarchy, of a system of competing profit-maximizing airlines in comparison to a system with centralized control. This measure can be used as a proxy to understand the effects of frequency competition on airline profitability and airport congestion.

4.3 Model

Let M be the total market size i.e. the number of passengers wishing to travel from a particular origin to a particular destination on a non-stop flight. In general, an airline passenger may have more than one flight in his itinerary. Conversely, two passengers on the same flight may have different origins and/or destinations. But for our analysis, we will ignore these network effects and assume the origin and destination pair of airports to be isolated from the rest of the network. Let $I = \{1, 2, \dots, n\}$ be the set of airlines competing in a particular non-stop market. Although most of the major airlines today follow the practices of differential pricing and revenue management, we will assume that the airfare charged by each airline remains constant across all passengers. Let p_i be the fare charged by each airline i . Further, we will assume that the type and seating capacity of aircraft to be operated on this non-stop route are known. Let S_i be the seating capacities for airline i and C_i be the operating cost per flight for airline i . Let α be the parameter in the S-curve relationship. A typical value suggested by literature is around 1.5. To keep our analysis general, we assume that $1 < \alpha < 2$. Our results are applicable even in the case of a linear relationship between market share and frequency share by taking the limit as $\alpha \rightarrow 1^+$.

Assumption 1. $1 < \alpha < 2$

Let x_i be the frequency of airline i . As per the S-curve relationship between market share and frequency share, the i^{th} airline's share of the market (MS_i) is given by:

$$MS_i = \frac{x_i^\alpha}{\sum_{j=1}^n x_j^\alpha}.$$

This is obtained by multiplying the numerator and denominator of the right hand side of equation (4.1) by $\left(\sum_{j=1}^n x_j\right)^\alpha$. The number of passengers (PAX_i) traveling

on airline i is given by:

$$PAX_i = \min \left(M \frac{x_i^\alpha}{n}, S_i x_i, \sum_{j=1} x_j^\alpha \right).$$

Airline i 's profit (Π_i) is given by:

$$\Pi_i = p_i * \min \left(M \frac{x_i^\alpha}{n}, S_i x_i, \sum_{j=1} x_j^\alpha \right) - C_i x_i.$$

We will assume that for every i , $C_i < p_i S_i$. In other words, the total operating cost of a flight is lower than the total revenue generated when the flight is completely filled. This assumption is reasonable because if it is violated for some airline i , then there is a trivial optimal solution $x_i = 0$ for that airline.

Assumption 2. $C_i < p_i S_i \forall i \in I$

From here onwards, our game-theoretic analysis proceeds as follows. In the next section (Section 4.4), we characterize the shapes of best response correspondences, that is, sets of optimal responses of a player as a function of the frequencies of the other player(s). This analysis, which focuses on the general frequency competition game model as described in this section, facilitates the subsequent analysis of Nash equilibria in Sections 4.5 through 4.7. In our Nash equilibrium analysis, we first focus on the two-player case (in Section 4.5) and later extend the analysis to symmetric N-player case (in Sections 4.6 and 4.7). We restrict our N-player game analysis to only the symmetric player case primarily for tractability reasons. As shown in our analysis in Section 4.5, even in 2-player case, the number of Nash equilibria can be as high as six depending on the combination of parameter values. The number of equilibria in frequency competition games with more players can be very high. In real-life airline markets, the parameters of airline frequency competition, such as,

fares, seating capacities, and operating costs of competing airlines are often not too far from each other. Therefore focusing on the symmetric player case is not that unrealistic. Furthermore, as shown in Sections 4.6 and 4.7, a thorough analysis of the symmetric player case presents several valuable insights. In Sections 4.6 and 4.7, we analyze both symmetric and asymmetric equilibria for the symmetric N-player case.

4.4 Best Response Curves

Let us define the effective competitor frequency, $y_i = \left(\sum_{j \in I, j \neq i} x_j^\alpha \right)^{1/\alpha}$, and

$$\Pi_i = \min(\Pi'_i, \Pi''_i)$$

where, $\Pi'_i = Mp_i \frac{x_i^\alpha}{x_i^\alpha + y_i^\alpha} - C_i x_i$ and $\Pi''_i = p_i S_i x_i - C_i x_i$.

Π'_i is a twice continuously differentiable function of x_i .

$$\frac{\partial \Pi'_i}{\partial x_i} = \frac{Mp_i \alpha x_i^{\alpha-1} y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^2} - C_i$$

and $\frac{\partial^2 \Pi'_i}{\partial x_i^2} = \frac{Mp_i \alpha x_i^{\alpha-2} y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^3} ((\alpha - 1) y_i^\alpha - (\alpha + 1) x_i^\alpha)$.

Π'_i has a single point of contraflexure at $x_i = y_i \left(\frac{\alpha-1}{\alpha+1} \right)^{1/\alpha}$ such that the function is strictly convex for all lower values of x_i and strictly concave for all higher values of x_i . Π'_i can have at most two points of zero slope (stationary points). If two such points exist, then the one with lower x_i will be a local minima in the convex region and the one with higher x_i will be a local maxima in the concave region. Therefore, Π'_i has at most one local maximum and exactly one boundary point at $x_i = 0$. Therefore, global maxima of Π'_i will be at either of these two points. Π''_i is a linear function of x_i with a positive slope. For a given combination of parameters α , M , p_i , C_i , S_i and a given effective competitor frequency y_i , the global maximum of Π_i can satisfy any one of the following three cases. These three cases are also illustrated in figures 4-1,

4-2 and 4-3 respectively.

Case A: $\Pi'_i \leq 0$ for all $x_i > 0$. Under this case, either a local maximum does not exist for Π'_i or it exists but value of the function Π'_i at that point is negative. In this case, a global maximum of $\Pi_i(x_i)$ is at $x_i = 0$. This describes a situation where the effective competitor frequency is so large that airline i cannot earn a positive profit at any frequency. Therefore, the best response of airline i is to have a zero frequency, i.e. not to operate any flights in that market.

Case B: Local maximum of Π'_i exists and the value of the function Π'_i at that local maximum is positive and less than or equal to $\Pi''_i(x_i)$. In this case, the unique global maximum of $\Pi_i(x_i)$ exists at the local maximum of $\Pi'_i(x_i)$. In this case, the optimum frequency is positive and at this frequency, airline i earns the maximum profit that it could have earned had the aircraft seating capacity been infinite.

Case C: A local maximum of Π'_i exists in the concave part and the value of the function $\Pi'_i(x_i)$ at this local maximum is greater than $\Pi''_i(x_i)$. In this case, $\Pi'_i(x_i)$ and $\Pi''_i(x_i)$ intersect at two distinct points (apart from $x_i = 0$). The unique global maximum of $\Pi_i(x_i)$ exists at the point of intersection with highest x_i value. This describes the case where optimum frequency is positive and greater than the optimum frequency under the assumption of infinite aircraft seating capacity. At this frequency, airline i earns lower profit than the maximum profit it could have earned had the aircraft seating capacity been infinite.

$\Pi'_i(0) = 0$ and for very low positive values of x_i , $\frac{\partial \Pi'_i}{\partial x_i}$ is negative. Therefore, at the first stationary point (the one with lower x_i value), the Π'_i function value will be negative. Moreover, as y_i tends to infinity, Π'_i is negative for any finite value of x_i . Therefore, $\Pi'_i(x_i) > 0$ for some x_i if and only if $\Pi'_i(x'_i) > 0$ for some stationary point x'_i . For a given combination of parameters α , M , p_i , C_i and S_i , there exists a threshold value of effective competitor frequency y_i such that, for any y_i value above this threshold, $\Pi'_i(x_i) \leq 0$ for all $x_i > 0$ and therefore the best response of airline i is $x_i = 0$. Let us denote this threshold by y_{th} and the corresponding x_i value as x_{th} . At $x_i = x_{th}$ and $y_i = y_{th}$,

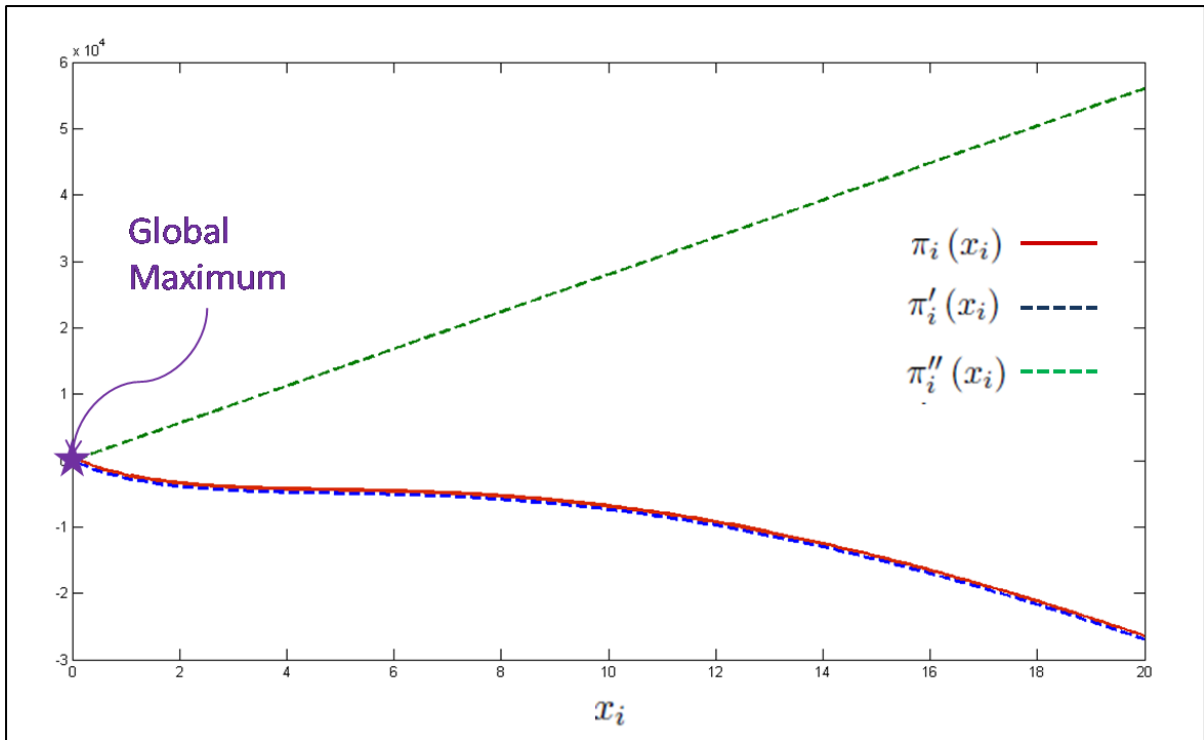


Figure 4-1: A typical shape of profit function (Case A)

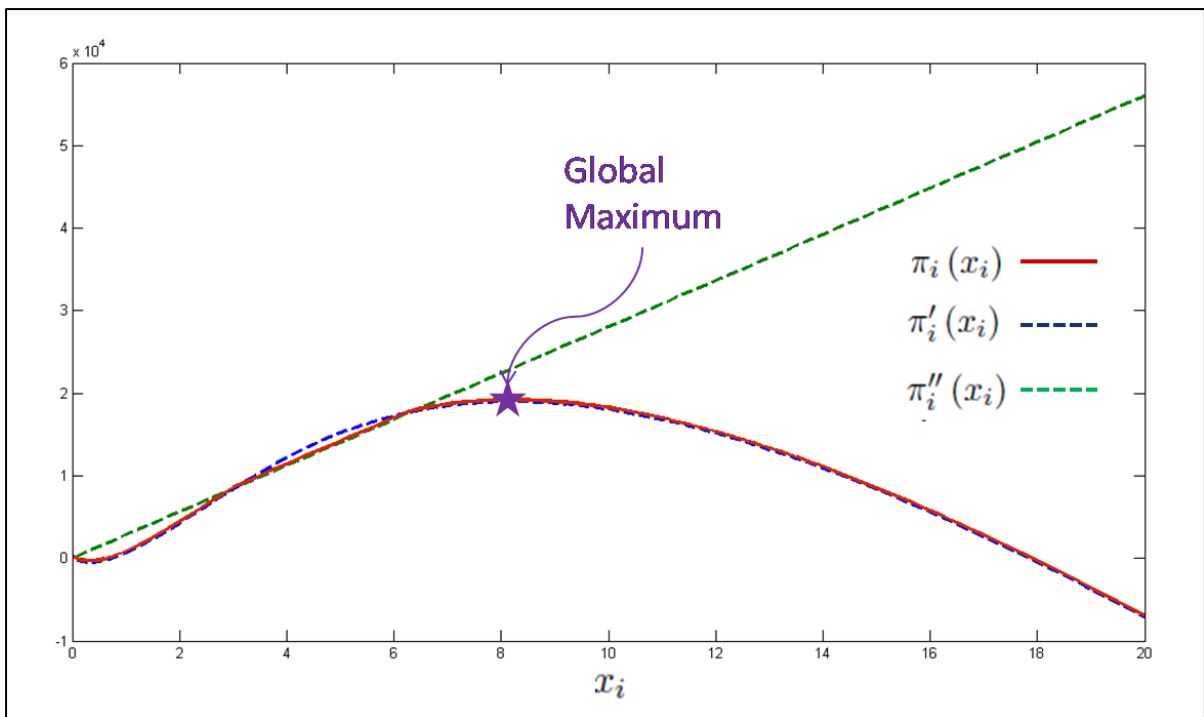


Figure 4-2: A typical shape of profit function (Case B)

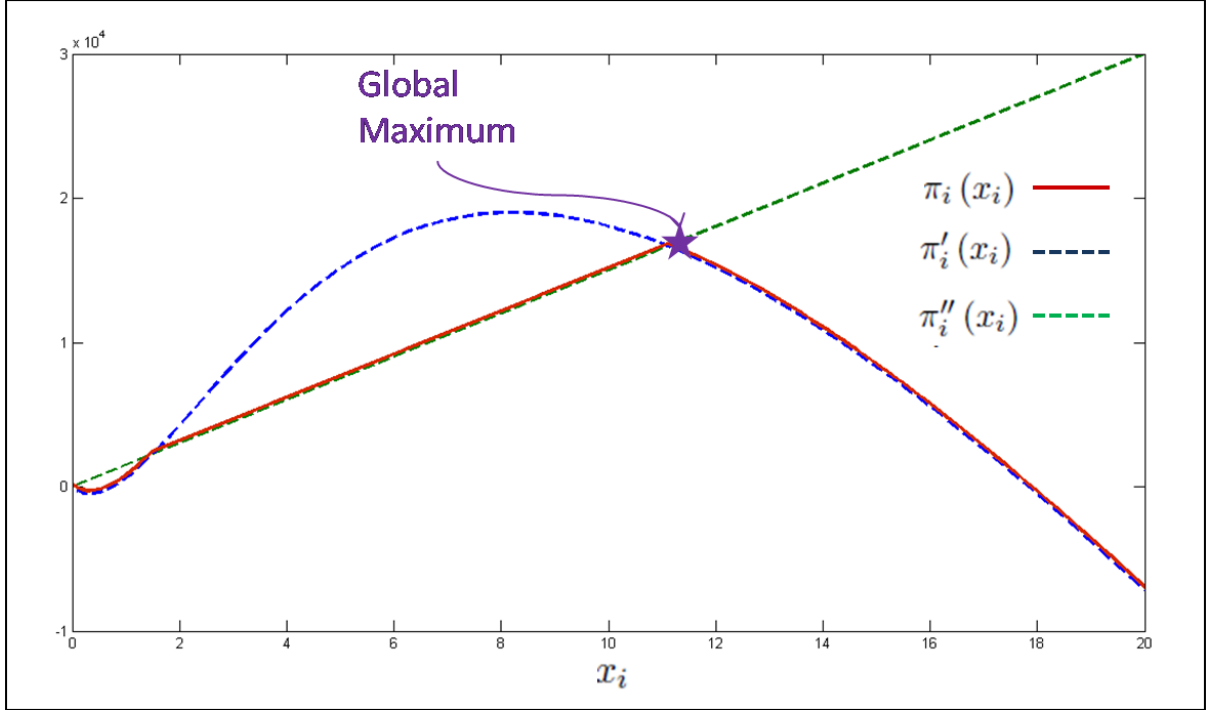


Figure 4-3: A typical shape of profit function (Case C)

$$\begin{aligned} \Pi'_i = 0, \quad \frac{\partial \Pi'_i}{\partial x_i} = 0, \quad \frac{\partial^2 \Pi'_i}{\partial x_i^2} \leq 0 \\ \Rightarrow x_{th} = (\alpha - 1) \frac{Mp_i}{\alpha C_i} \text{ and } y_{th} = (\alpha - 1)^{\frac{\alpha-1}{\alpha}} \frac{Mp_i}{\alpha C_i}. \end{aligned}$$

Of course, at $y_i = y_{th}$, $x_i = 0$ is also optimal. It turns out that it is the only y_i value at which there is more than one best response possible. This situation is unlikely to be observed in real world examples, because the parameters of the model are all real numbers with continuous distributions. So the probability of observing this exact idiosyncratic case is zero. If we arbitrarily assume that in the event of two optimal frequencies, an airline chooses the greater of the two values, then the best response correspondence reduces to a function which we will refer to as the best response function. The existence of two different maximum values at $y_i = y_{th}$ means that the best response correspondence is not always convex valued. Therefore, in general, a pure strategy Nash equilibrium may or may not exist for this game.

For y_i values slightly below y_{th} , the global maximum of Π_i corresponds to the stationary point of Π'_i in the concave part as described in case B above. Therefore, for y_i values slightly below y_{th} , at the stationary point of Π'_i in the concave part, $\Pi'_i < \Pi''_i$. However, as $y_i \rightarrow 0$, $\text{argmax}(\Pi'_i(x_i)) \rightarrow 0$. Therefore, the $\text{argmax}(\Pi_i(x_i))$ exists at the point of intersection of Π'_i and Π''_i curves, as explained in case C above. For y_i values slightly above 0, at the stationary point of Π'_i in the concave part, $\Pi'_i > \Pi''_i$. Therefore by continuity, for some y_i such that $0 \leq y_i \leq y_{th}$, there exists x_i such that, $\Pi'_i = \Pi''_i$, $\frac{\partial \Pi'_i}{\partial x_i} = 0$ and $\frac{\partial^2 \Pi'_i}{\partial x_i^2} \leq 0$. It turns out that there is only one such y_i value that satisfies these conditions. Let us denote this y_i value by y_{cr} , since this is critical value of effective competitor frequency such that case B prevails for higher y_i values (as long as $y_i \leq y_{th}$) and case C prevails for all lower y_i values. The value of y_{cr} and the corresponding x_i value, x_{cr} , is given by,

$$x_{cr} = \frac{M}{S_i} \left(1 - \frac{C_i}{\alpha p_i S_i} \right) \quad \text{and} \quad y_{cr} = \frac{\frac{M}{S_i} \left(1 - \frac{C_i}{\alpha p_i S_i} \right)}{\left(\frac{\alpha p_i S_i}{C_i} - 1 \right)^{\frac{1}{\alpha}}}.$$

For $y_i = 0$, as $x_i \rightarrow 0^+$, Π'_i keeps increasing and Π''_i keeps decreasing. However, $\Pi''_i < \Pi'_i$ for sufficiently low values of x_i . Therefore, Π_i is maximized when $\Pi'_i = \Pi''_i$. Let us denote this x_i value as x_0 . It is easy to see that $x_0 = \frac{M}{S_i}$. We will denote the range of y_i values with $y_i \geq y_{th}$ as region A, $y_{cr} \leq y_i < y_{th}$ as region B and $y_i < y_{cr}$ as range C.

In region C, Π_i is maximized for a unique x_i value such that $\Pi'_i = \Pi''_i$ and $\frac{\partial \Pi'_i}{\partial x_i} \leq 0$. The equality condition translates into,

$$\frac{M}{S_i} x_i^{\alpha-1} - x_i^\alpha = y_i^\alpha. \quad (4.2)$$

The left hand side (LHS) of equation (4.2) is strictly concave because $1 < \alpha < 2$. Further, the LHS is maximized at $x_i = \frac{\alpha-1}{\alpha} \frac{M}{S_i}$. The maximum value of LHS is at $y_i = (\alpha-1)^{\frac{\alpha-1}{\alpha}} \frac{M}{\alpha S_i}$. So for every y_i value, there are two corresponding x_i values satisfying equation (4.2) that correspond to the two points of intersection of the Π'_i

and Π_i'' curves. The one corresponding to the higher x_i value is of interest to us. That always corresponds to x_i values greater than $\frac{\alpha-1}{\alpha} \frac{M}{S_i}$. Differentiating both sides of equation (4.2) with respect to y_i ,

$$\frac{\partial x_i}{\partial y_i} = \alpha \frac{y_i^{\alpha-1}}{x_i^{\alpha-2}} \frac{1}{(\alpha-1) \frac{M}{S_i} - \alpha x_i} < 0.$$

So the best response of airline i in region C is strictly decreasing. Let us again differentiate with respect to y_i to obtain the second derivative of best response x_i ,

$$\frac{\partial^2 x_i}{\partial y_i^2} = \frac{\left(\frac{\partial x_i}{\partial y_i}\right)^2 (\alpha-1) x_i^{\alpha-3} \left(\alpha x_i + (2-\alpha) \frac{M}{S_i}\right) + \alpha (\alpha-1) y_i^{\alpha-2}}{\left((\alpha-1) \frac{M}{S_i} - \alpha x_i\right) x_i^{\alpha-2}} < 0. \quad (4.3)$$

Therefore, the best response curve is a strictly decreasing and concave function for all $0 \leq y_i < y_{cr}$.

In region B, Π_i is maximized for a unique x_i value such that $\frac{\partial \Pi_i'}{\partial x_i} = 0$ and $\frac{\partial^2 \Pi_i'}{\partial x_i^2} < 0$. The first order equality condition translates into,

$$Mp_i \alpha \frac{x_i^{\alpha-1} y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^2} = C_i. \quad (4.4)$$

Differentiating both sides of equation (4.4) with respect to y_i and again substituting equation (4.4) we get,

$$\frac{\partial x_i}{\partial y_i} = \frac{x_i}{y_i} \frac{x_i^\alpha - y_i^\alpha}{\left(1 + \frac{1}{\alpha}\right) x_i^\alpha - \left(1 - \frac{1}{\alpha}\right) y_i^\alpha}. \quad (4.5)$$

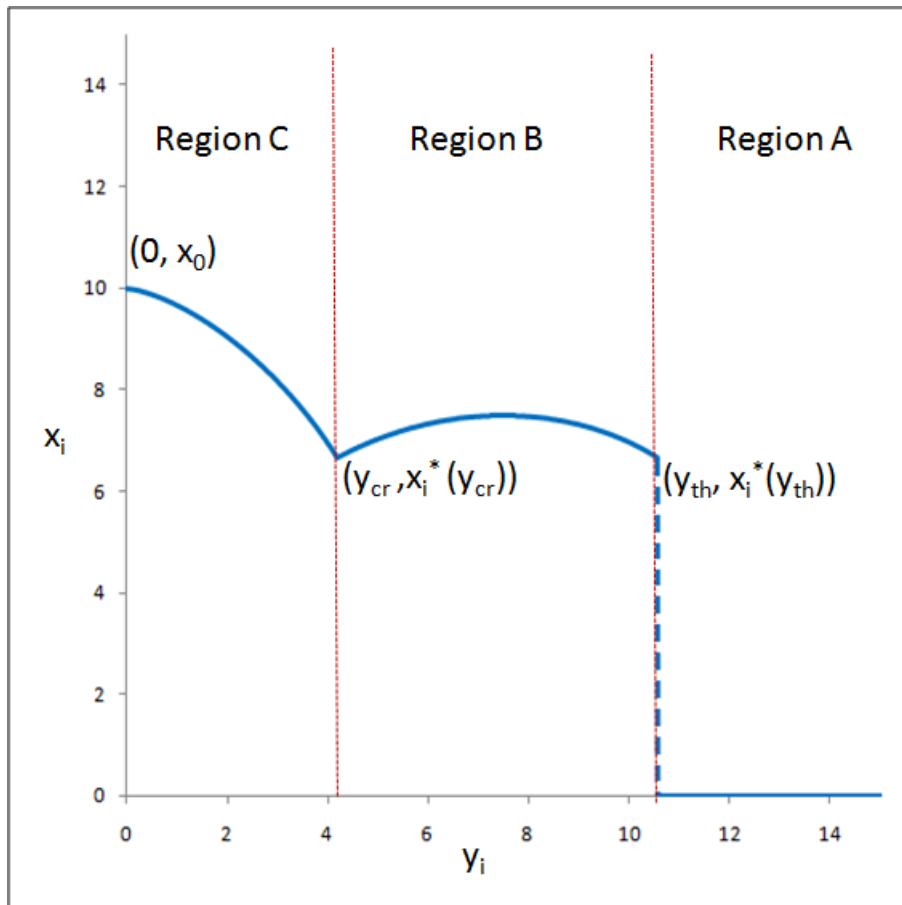


Figure 4-4: A typical best response curve

The second order inequality condition translates into,

$$\begin{aligned} & \frac{Mp_i \alpha x_i^{\alpha-2} y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^3} ((\alpha - 1) y_i^\alpha - (\alpha + 1) x_i^\alpha) < 0 \\ \Rightarrow & \left(1 + \frac{1}{\alpha}\right) x_i^\alpha - \left(1 - \frac{1}{\alpha}\right) y_i^\alpha > 0. \end{aligned} \quad (4.6)$$

So the denominator of the right hand side of equation (4.5) is positive. Therefore, $\frac{\partial x_i}{\partial y_i} = 0$ if and only if $x_i = y_i$, $\frac{\partial x_i}{\partial y_i} > 0$ if and only if $x_i > y_i$ and $\frac{\partial x_i}{\partial y_i} < 0$ if and only if $x_i < y_i$. Therefore, the best response curve $x_i(y_i)$ in region B has zero slope at $x_i = y_i$, is strictly increasing for $x_i > y_i$ and strictly decreasing for $x_i < y_i$. Substituting $x_i = y_i$ in equation (4.4) we get, $x_i = y_i = \frac{\alpha Mp_i}{4C_i}$.

Figure 4-4 describes a typical best response curve as a function of effective competitor frequency. In region A, the effective competitor frequency is so small that airline i attracts a large market share even with a small frequency. Therefore, the optimal frequency ignoring seating capacity constraints is so low that, the number of seats is exceeded by the number of passengers wishing to travel with airline i . As a result, the optimal frequency and the maximum profit that can be earned by airline i are decided by the aircraft seating capacity constraint. In this region, the optimal number of flights scheduled by airline i is just sufficient to carry all the passengers that wish to travel on airline i . In this region, airline i has 100% load factor at the optimal frequency. With increasing effective competitor frequency, the market share attracted by airline i reduces and hence fewer flights are required to carry those passengers. Therefore, the best response curve is strictly decreasing in this region. Once the effective competitor frequency exceeds a critical value y_{cr} , the seating capacity constraint ceases to affect the optimal frequency decision.

In region B, the effective competitor frequency is sufficiently large due to which the number of passengers attracted by airline i does not exceed the seating capacity. Therefore, the aircraft seating capacity constraint becomes redundant in this region. The optimal frequency is equal to the frequency at which the marginal revenue equals marginal cost, which is a constant C_i . As the effective competitor frequency increases, the market share of airline i at the optimal frequency decreases and the load factor

of airline i at optimal frequency also decreases. At a large value, y_{th} , of effective competitor frequency, the load factor of airline i at its optimal frequency reduces to a value $\frac{C_i}{p_i S_i}$ and the optimal profit drops to zero.

For all values of effective competitor frequency above y_{th} , i.e. in region C, there is no positive frequency for which the airline i can make positive profit. Therefore, the optimal frequency of airline i in region C is zero.

4.5 2-Player Game

Let x and y be the frequency of carrier 1 and 2 respectively. The effective competitor frequency for carrier 1 is y and that for carrier 2 is x . For any pure strategy Nash equilibrium (PSNE), the competitor frequency for each carrier can belong to any one of the three regions, A, B and C. So potentially there are 9 different combinations possible. We define the *type* of a PSNE as the combination of regions to which the competitor frequency belongs at equilibrium. We will denote each type by a pair of capital letters denoting the regions. For example, if carrier 1's effective competitor frequency, i.e. y , belongs to region B and carrier 2's effective competitor frequency, i.e. x , belongs to region C, then that PSNE is said to be of type BC. Accordingly, there are 9 different types of PSNE possible for this game, namely AA, AB, AC, BA, BB, BC, CA, CB and CC.

Frequency competition among carriers is the primary focus of this research. However, it is important to realize that frequency planning is just one part of the entire airline planning process. Frequency planning decisions are not taken in isolation, the route planning phase precedes the frequency planning phase. Once the set of routes to be operated is decided, the airline proceeds to the decision of the operating frequency on that route. This implicitly means that once a route is deemed profitable in the route planning phase, frequency planning is the phase that decides the number of flights per day, which is supposed to be a positive number. However, in AA, AB, BA, AC or CA type equilibria, the equilibrium frequency of at least one of the carriers is zero, which is inconsistent with the actual airline planning process. Moreover, for

ease of modeling, we have made a simplifying assumption that the seating capacity is constant. In reality, seating capacities are chosen considering the estimated demand in a market. If the demand for an airline in a market exceeds available seats on a regular basis, the airline would be inclined to use larger aircraft. Sustained presence of close to 100% load factors is a rarity. However type AC, BC, CA, CB and CC type equilibria involve one or both carriers having 100% load factors. Zero frequency and 100% load factors make all types of equilibria, apart from type BB equilibrium, suspect in terms of their portrayal of reality.

We will now investigate each of these possible types of pure strategy equilibria of this game and obtain the existence and uniqueness conditions for each of them.

4.5.1 Existence and Uniqueness

Proposition 1. *A type AA equilibrium cannot exist.*

Proof. If $x^* = 0$ then, $\Pi_2 = p_2 * \min(M, S_2 y) - C_2 y$, which is maximized at $y = \frac{M}{S_2}$ because $C_2 < S_2 p_2$. So $y^* > 0$ whenever $x^* = 0$. So this type of equilibrium cannot exist. \square

Proposition 2. *A type AB (and type BA) equilibrium cannot exist.*

Proof. Type AB equilibrium exists if and only if $x^* = 0$, $y^* > 0$ and $PAX_2 < S_2 y^*$. As shown before, if $x^* = 0$ then, Π_2 is maximized at $y = \frac{M}{S_2}$ as long as $C_2 < S_2 p_2$. So $PAX_2 = M = S_2 y^*$ whenever $x^* = 0$. So this type of equilibrium cannot exist. By symmetry, type BA equilibrium cannot exist either. \square

Proposition 3. *A type AC equilibrium exists if and only if $\frac{C_1}{S_1 p_1} \geq \frac{S_2}{S_1} \frac{1}{\alpha} (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$ and if it exists, then it is a unique type AC equilibrium.*

Proof. This type of equilibrium requires $x^* = 0$ and $y^* = \frac{M}{S_2}$. So if an equilibrium of this type exists, then it must be the unique type AC equilibrium. For this equilibrium to exist, the only condition we need to check is that $\frac{M}{S_2} = y \geq y_{th} = (\alpha - 1)^{\frac{\alpha-1}{\alpha}} \frac{M p_1}{\alpha C_1}$. For all $y^* = \frac{M}{S_2}$, $x^* = 0$ is true if and only if $\Pi_1 \leq 0$, for all $x \geq 0$. So type AC equilibrium will exist if and only if $\frac{C_1}{S_1 p_1} \geq \frac{S_2}{S_1} \frac{1}{\alpha} (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$. \square

By symmetry, a type CA equilibrium exists if and only if $\frac{C_2}{S_2 p_2} \geq \frac{S_1}{S_2} \frac{1}{\alpha} (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$ and if it exists, then it is the unique type CA equilibrium.

Proposition 4. *A type BB equilibrium exists if and only if $k \leq \left(\frac{1}{\alpha-1}\right)^{\frac{1}{\alpha}}$, $\frac{1}{k} \leq \left(\frac{1}{\alpha-1}\right)^{\frac{1}{\alpha}}$, $\frac{C_1}{S_1 p_1} < \alpha \frac{k^\alpha}{1+k^\alpha}$ and $\frac{C_2}{S_2 p_2} < \alpha \frac{1}{1+k^\alpha}$, where $k = \frac{C_1 p_2}{C_2 p_1}$, and if it exists, then it is a unique type BB equilibrium.*

Proof. In type BB equilibrium, $x^* > 0$, $y^* > 0$, $PAX_1 < S_1 x$ and $PAX_2 < S_2 y$. Therefore, $\Pi_1(x^*, y^*) = \Pi'_1(x^*, y^*)$ and $\Pi_2(x^*, y^*) = \Pi'_2(x^*, y^*)$. So Π_1 and Π_2 are both twice continuously differentiable at (x^*, y^*) . So type BB equilibrium exists if and only if there exist x and y such that $\frac{\partial \Pi'_1}{\partial x} = 0$, $\frac{\partial \Pi'_2}{\partial y} = 0$, $\frac{\partial^2 \Pi'_1}{\partial x^2} \leq 0$, $\frac{\partial^2 \Pi'_2}{\partial y^2} \leq 0$, $\Pi'_1 \geq 0$, $\Pi'_2 \geq 0$, $M \frac{x^\alpha}{x^\alpha + y^\alpha} < S_1 x$ and $M \frac{y^\alpha}{x^\alpha + y^\alpha} < S_2 y$. Solving the two First Order Conditions (FOCs) simultaneously, we get $x = \frac{\alpha M p_1}{C_1} \frac{k^\alpha}{(1+k^\alpha)^2}$ and $y = \frac{\alpha M p_1}{C_1} \frac{k^{\alpha+1}}{(1+k^\alpha)^2}$. So if this equilibrium exists, then it must be the unique type BB equilibrium.

The second order conditions (SOCs) can be simplified to $k \leq \left(\frac{\alpha+1}{\alpha-1}\right)^{\frac{1}{\alpha}}$ and $\frac{1}{k} \leq \left(\frac{\alpha+1}{\alpha-1}\right)^{\frac{1}{\alpha}}$. Also the $\Pi'_1 \geq 0$ and $\Pi'_2 \geq 0$ translate into,

$$k \leq \left(\frac{1}{\alpha-1}\right)^{\frac{1}{\alpha}} \quad (4.7)$$

$$\text{and } \frac{1}{k} \leq \left(\frac{1}{\alpha-1}\right)^{\frac{1}{\alpha}}. \quad (4.8)$$

Conditions (4.7) and (4.8) make the second order conditions redundant. Finally, the last two conditions translate into,

$$\frac{C_1}{S_1 p_1} < \alpha \frac{k^\alpha}{1+k^\alpha} \quad (4.9)$$

$$\text{and } \frac{C_2}{S_2 p_2} < \alpha \frac{1}{1+k^\alpha}. \quad (4.10)$$

Therefore, type BB equilibrium exists if and only if conditions (4.7), (4.8), (4.9) and (4.10) are satisfied. \square

Proposition 5. *A type BC equilibrium exists if and only if $\frac{C_1 S_1}{p_1 S_1 S_2} \leq (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$, $\frac{k^\alpha}{1+k^\alpha} \leq \frac{1}{\alpha}$, $\frac{1}{1+k^\alpha} \leq \frac{1}{\alpha} \frac{C_2}{p_2 S_2}$ and $\frac{1}{\alpha} \frac{C_1}{p_1 S_1} \geq \frac{1}{1 + \left(\frac{S_1}{S_2}\right)^{\frac{\alpha}{\alpha-1}}}$, where $k = \frac{C_1 p_2}{C_2 p_1}$, and if it exists, then it is a unique type BC equilibrium.*

Proof. In type BC equilibrium, $x^* > 0$, $y^* > 0$, $PAX_1 < S_1 x$ and $PAX_2 = S_2 y$. Therefore $\Pi_1(x^*, y^*) = \Pi'_1(x^*, y^*)$. So $\Pi_1(x)$ is twice continuously differentiable at (x^*, y^*) . For local maxima of Π_2 at (x^*, y^*) , we need $\Pi'_2 = \Pi''_2$ and $\frac{\partial \Pi'_2}{\partial y} \leq 0$.

A type BC equilibrium then exists if and only if there exists (x, y) such that $\frac{\partial \Pi'_1}{\partial x} = 0$, $\Pi'_2 = \Pi''_2$, $\frac{\partial^2 \Pi'_1}{\partial x^2} \leq 0$, $\frac{\partial \Pi'_2}{\partial y} \leq 0$, $\Pi'_1 \geq 0$ and $M \frac{x^\alpha}{x^\alpha + y^\alpha} < S_1 x$. The first two conditions translate into,

$$\frac{x^{\alpha-1} y^\alpha}{(x^\alpha + y^\alpha)^2} = \frac{C_1}{\alpha M p_1} \text{ and } \frac{y^\alpha}{x^\alpha + y^\alpha} = \frac{S_2}{M} y.$$

Solving these two equations simultaneously we get,

$$x = \left(\frac{M C_1}{\alpha p_1 S_2^2} \right)^{\frac{1}{\alpha-1}} y^{\frac{\alpha-2}{\alpha-1}} \quad (4.11)$$

$$\text{and } \left(\frac{y S_2}{M} \right)^{\frac{1}{\alpha-1}} - \left(\frac{y S_2}{M} \right)^{\frac{\alpha}{\alpha-1}} - \left(\frac{C_1}{\alpha p_1 S_2} \right)^{\frac{\alpha}{\alpha-1}} = 0. \quad (4.12)$$

The nonnegativity condition on airline 1's profit implies that $M p_1 \frac{x^\alpha}{x^\alpha + y^\alpha} \geq C_1 x$. Substituting equation (4.11) and (4.12) we get,

$$\frac{y S_2}{M} \leq \frac{1}{\alpha}. \quad (4.13)$$

The LHS of equation (4.12) is a strictly increasing function of y for $\frac{y S_2}{M} < \frac{1}{\alpha}$. Therefore, there exists a y that satisfies equation (4.12) and inequality (4.13) if and only if $\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - \left(\frac{C_1}{\alpha p_1 S_2}\right)^{\frac{\alpha}{\alpha-1}} \geq 0$, i.e. if and only if

$$\frac{C_1}{p_1 S_1} \frac{S_1}{S_2} \leq (\alpha - 1)^{\frac{\alpha-1}{\alpha}}, \quad (4.14)$$

and if it exists, then it is unique. Therefore, if a type BC equilibrium exists, then it must be a unique type BC equilibrium.

Simplifying the second order condition and substituting equation (4.11) and equation (4.12), we get $\frac{yS_2}{M} \leq \frac{\alpha+1}{2\alpha}$. Therefore, condition (4.13) makes the second order condition redundant.

First order condition (FOC) on $\Pi'_2(y)$ simplifies to $\frac{x}{y} \leq \frac{C_2 p_1}{C_1 p_2}$. Substituting equation (4.11) and equation (4.12) we get,

$$\frac{yS_2}{M} \geq \frac{k^\alpha}{1+k^\alpha}. \quad (4.15)$$

Therefore, there exists a y that satisfies equation (4.12), inequality (4.13) and inequality (4.15) if and only if

$$\begin{aligned} \frac{k^\alpha}{1+k^\alpha} \leq \frac{1}{\alpha} \text{ and } \left(\frac{C_1}{\alpha p_1 S_2} \right)^{\frac{\alpha}{\alpha-1}} &\geq \left(\frac{k^\alpha}{1+k^\alpha} \right)^{\frac{1}{\alpha-1}} - \left(\frac{k^\alpha}{1+k^\alpha} \right)^{\frac{\alpha}{\alpha-1}} \\ \iff \frac{k^\alpha}{1+k^\alpha} &\leq \frac{1}{\alpha} \end{aligned} \quad (4.16)$$

$$\text{and } \frac{1}{1+k^\alpha} \leq \frac{1}{\alpha} \frac{C_2}{p_2 S_2}. \quad (4.17)$$

Finally, the last condition, i.e. the condition that the seating capacity exceeds the number of passengers for airline 1, simplifies to $\frac{x^{\alpha-1}}{y^{\alpha-1}} \leq \frac{S_1}{S_2}$. Substituting equation (4.11) we get,

$$\frac{yS_2}{M} \geq \frac{C_1}{\alpha p_1 S_1}. \quad (4.18)$$

Combining with inequality (4.13) we get, $\frac{a}{\alpha} \geq \frac{yS_2}{M} \geq \frac{C_1}{\alpha p_1 S_1}$. Therefore, there exists

a y that satisfies equation (4.12), inequality (4.13) and inequality (4.18) if and only if,

$$\begin{aligned} \left(\frac{C_1}{\alpha p_1 S_2}\right)^{\frac{\alpha}{\alpha-1}} &\geq \left(\frac{C_1}{\alpha p_1 S_1}\right)^{\frac{1}{\alpha-1}} - \left(\frac{C_1}{\alpha p_1 S_1}\right)^{\frac{\alpha}{\alpha-1}} \\ \Leftrightarrow \frac{1}{\alpha} \frac{C_1}{p_1 S_1} &\geq \frac{1}{1 + \left(\frac{S_1}{S_2}\right)^{\frac{\alpha}{\alpha-1}}}. \end{aligned} \quad (4.19)$$

Therefore, type BC equilibrium exists if and only if inequality conditions (4.14), (4.16), (4.17) and (4.19) are satisfied. \square

By symmetry, a type CB equilibrium exists if and only if $\frac{C_2}{p_2 S_2} \frac{S_2}{S_1} \leq (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$, $\frac{1}{1+k\alpha} \leq \frac{1}{\alpha}$, $\frac{k^\alpha}{1+k\alpha} \leq \frac{1}{\alpha} \frac{C_1}{p_1 S_1}$, $\frac{1}{\alpha} \frac{C_2}{p_2 S_2} \geq \frac{1}{1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}}$, and if it exists, then it is a unique CB type equilibrium.

Proposition 6. *A type CC equilibrium exists if and only if $\frac{\left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}}{1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}} \leq \frac{1}{\alpha} \frac{C_1}{S_1 p_1}$ and $\frac{1}{1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}} \leq \frac{1}{\alpha} \frac{C_2}{S_2 p_2}$, and if it exists, then it is a unique type CC equilibrium.*

Proof. For type CC equilibrium, $x > 0$, $y > 0$, $PAX_1 = S_1 x$ and $PAX_2 = S_2 y$. Existence of local maxima of Π_1 at $x = x^*$ requires that $\frac{\partial \Pi_1'}{\partial x} \leq 0$. Similarly existence of local maxima of Π_2 at $y = y^*$ requires that $\frac{\partial \Pi_2'}{\partial y} \leq 0$. So for a type CC equilibrium to exist at (x, y) , the necessary and sufficient conditions to be satisfied are $\frac{x^\alpha}{x^\alpha + y^\alpha} M = S_1 x$, $\frac{y^\alpha}{x^\alpha + y^\alpha} M = S_2 y$, $\frac{\partial \Pi_1'}{\partial x} \leq 0$ and $\frac{\partial \Pi_2'}{\partial y} \leq 0$. Solving the two equalities simultaneously we get, $x = \frac{M}{S_1 \left(1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}\right)}$ and $y = \frac{M}{S_2 \left(1 + \left(\frac{S_1}{S_2}\right)^{\frac{\alpha}{\alpha-1}}\right)}$.

Therefore, if a type CC equilibrium exists, then it must be the unique type CC equilibrium. The two inequality conditions translate into,

$$\frac{\left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}}{1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}} \leq \frac{1}{\alpha} \frac{C_1}{S_1 p_1} \quad (4.20)$$

$$\text{and } \frac{1}{1 + \left(\frac{S_2}{S_1}\right)^{\frac{\alpha}{\alpha-1}}} \leq \frac{1}{\alpha} \frac{C_2}{S_2 p_2}. \quad (4.21)$$

Therefore, equation (4.20) and equation (4.21) together are necessary and sufficient conditions for type CC equilibrium to exist. \square

In any 2-player game, out of 9 possible types, 6 types of equilibria, namely AC, CA, BB, BC, CB and CC may exist depending on operating cost, fare and seating capacity values. Furthermore, all the necessary and sufficient conditions for the existence and uniqueness of each type of equilibrium can be expressed in terms of only 5 unitless parameters namely, $r_1 = \frac{C_1}{p_1 S_1}$, $r_2 = \frac{C_2}{p_2 S_2}$, $k = \frac{C_1 p_2}{C_2 p_1}$, $l = \frac{S_1}{S_2}$ and α , out of which l can be expressed as a function of the rest as $l = k \frac{r_2}{r_1}$. So there are only 4 independent parameters, which completely describe a 2-player frequency game. The total passenger demand M plays no part in any of the conditions.

4.5.2 Realistic Parameter Ranges

Up to 6 different pure strategy Nash equilibria may exist for a 2-player game depending on game parameters. Apart from α , the flight operating costs, seating capacities and fares are the only determinants of these parameters. In order to identify realistic ranges of these parameters, we looked at all the domestic segments in the United States with exactly 2 carriers providing non-stop service. There are 157 such segments that satisfied this criteria. Many of these markets cannot be classified as pure duopoly situations because passenger demand on many of these origin-destination pairs is served not only by the nonstop itineraries, but also by connecting itineraries offered by several carriers, often including the two carriers providing the nonstop service. Moreover, one or both endpoints for many of these non-stop segments are important hubs of one or both of these nonstop carriers, which means that connecting passengers traveling on this segment also play an important role in the profitability of this segment. Therefore, modeling these nonstop markets as pure duopoly cases can be a gross approximation. Our aim is not to capture all these effects into our frequency competition model but rather to identify realistic relative values of flight operating costs, seating capacities and fares. Despite these complications, these 157 segments are the real-world situations that come closest to the simplified frequency

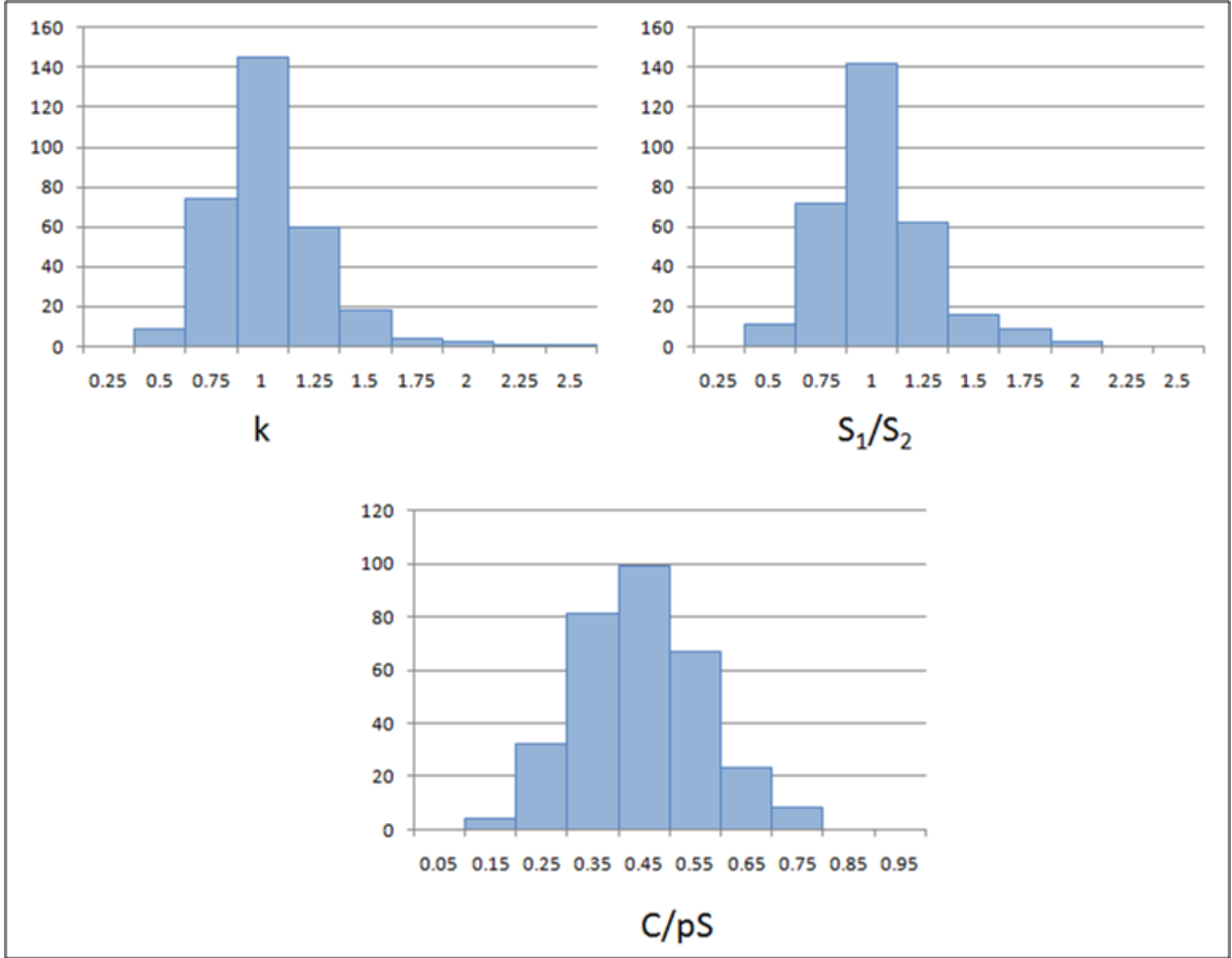


Figure 4-5: Histograms of parameter values

competition model that we have considered. Therefore, data from these markets were used to narrow down our modeling focus. Figure 4-5 shows the histograms of k , $\frac{S_1}{S_2}$ and $\frac{C}{pS}$. All k values were found to lie in the range 0.4 to 2.5, all $\frac{S_1}{S_2}$ were in the range 0.5 to 2 and all $\frac{C}{pS}$ values were found to lie in the range 0.18 to 0.8. We will restrict our further analysis to these ranges of values only. In particular, for later analysis, we will need only one of these assumptions, which is given by,

Assumption 3. $0.4 \leq k \leq 2.5$

For $\alpha = 1.5$, the conditions for type BB equilibrium were satisfied in 144 out of these 157 markets, i.e. almost 92% of the times. Conditions for type AC (or CA) equilibrium were satisfied in 71 markets, of which 8 were such that the conditions for both type AC and type CA equilibrium were satisfied together. Conditions for type

BC (or CB) equilibrium were satisfied in only 1 out of 157 markets and conditions for type CC equilibrium were never satisfied. In all the markets, the conditions for the existence of at least one pure strategy Nash equilibrium were satisfied. Out of 157 markets, almost 55% (86 markets) were such that type BB was the unique pure strategy Nash equilibrium.

We have already proved that AA, AB and BA type equilibria do not exist. Further, as discussed above, AC, CA, BC, CB and CC type equilibria are suspect in terms of portrayal of reality. Therefore, type BB equilibrium appears to be the most reasonable type of equilibrium. Indeed, the data analysis suggested that the existence conditions for type BB equilibrium were satisfied in most of the markets. So for the purpose of analyzing learning dynamics we will only consider the type BB equilibrium.

Now, we propose two alternative dynamics for the non-equilibrium situations.

4.5.3 Myopic Best Response Dynamic

Consider an adjustment process where the two players take turns to adjust their own frequency decision so that each time it is the best response to the frequency chosen by the competitor in the previous period. If x^i and y^i is the frequency decision by each carrier in period i , then x^i is the best response to y^{i-1} and y^{i-1} is the best response to x^{i-2} etc. We will prove the convergence of this dynamics for two representative values of α namely $\alpha = 1$ and $\alpha = 1.5$. We chose these two values because they correspond to two disparate beliefs about the market share-frequency share relationship. There is nothing specific about these two values that makes the algorithm converge. In fact given any value in between, we would probably be able to come up with a proof of convergence. But due to space constraints we will restrict our attention to these two specific values of α .

Let us define $\chi = x^\alpha$ and $\gamma = y^\alpha$. We will often use the $\chi - \gamma$ coordinate system in this section. Without any loss of generality, we assume that $k = \frac{C_1 p_2}{C_2 p_1} \leq 1$. We will denote the best response functions as $x_{BR}(y)$ and $y_{BR}(x)$ in the $x - y$ coordinate system and as $\chi_{BR}(\gamma)$ and $\gamma_{BR}(\chi)$ in the $\chi - \gamma$ coordinate system. Consider a two-dimensional interval I given by $x_{lb} \leq x \leq x_{ub}$, $y_{lb} \leq y \leq y_{ub}$ where,

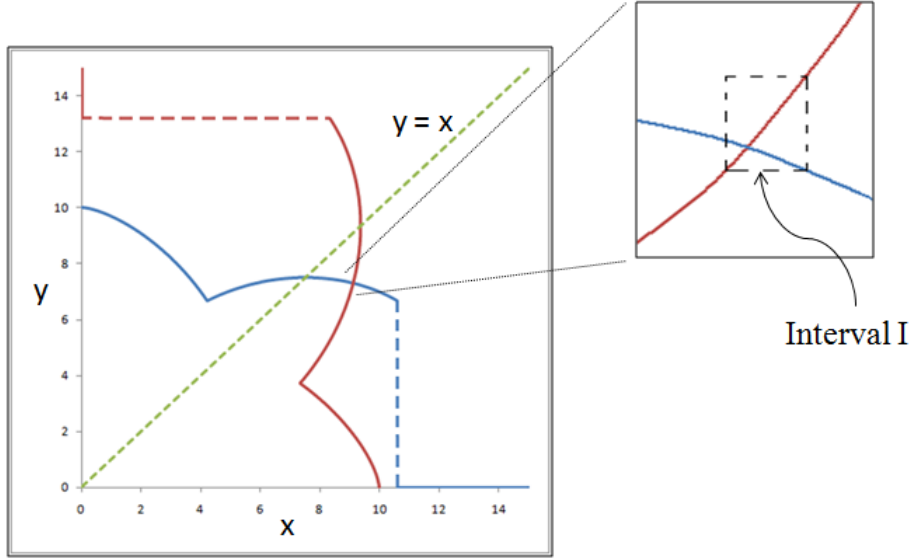


Figure 4-6: Best response curves in 2-player game

$$y_{ub} = \frac{\alpha M p_2}{4C_2}$$

$$x_{ub} = x_{BR}(y_{ub})$$

$$y_{lb} = y_{BR}(x_{ub})$$

$$x_{lb} = x_{BR}(y_{lb})$$

Figure 4-6 provides a pictorial depiction of interval I.

Proposition 7. *As long as the competitor frequency for each carrier remains in region B, regardless of the starting point: (a) the myopic best response algorithm will reach a point in interval I in a finite number of iterations, (b) once inside interval I, it will never leave the interval.*

Proof. Let us denote the frequency decisions of the two carriers after the i^{th} iteration by x^i and y^i respectively. At the beginning of the algorithm the frequency values are arbitrarily chosen to be x^0 and y^0 . If $i \geq 0$ is odd, then $x^i = x_{BR}(y^{i-1})$ and $y^i = y^{i-1}$. If $i \geq 0$ is even, then $y^i = y_{BR}(x^{i-1})$ and $x^i = x^{i-1}$.

Therefore for all $i \geq 2$, x_i is a best response to some y and y_i is a best response to

some x . Best response curve $x_{BR}(y)$ in region B has a unique maximum at $y = \frac{\alpha Mp_1}{4C_1}$ with $x_{BR}\left(\frac{\alpha Mp_1}{4C_1}\right) = \frac{\alpha Mp_1}{4C_1}$. By symmetry, the best response curve $y_{BR}(x)$ in region B has a unique maximum at $x = \frac{\alpha Mp_2}{4C_2}$ with $y_{BR}\left(\frac{\alpha Mp_2}{4C_2}\right) = \frac{\alpha Mp_2}{4C_2}$. $k \leq 1$ implies that $\frac{\alpha Mp_2}{4C_2} \leq \frac{\alpha Mp_1}{4C_1}$. Therefore, $y^i \leq \frac{\alpha Mp_2}{4C_2} = y_{ub}$ for all $i \geq 2$. $\frac{\partial x_{BR}}{\partial y} > 0$ for $y < \frac{\alpha Mp_1}{4C_1}$. Therefore, for all $i \geq 3$, $x^i = x_{BR}(y^{i-1}) \leq x_{BR}(y_{ub}) = x_{ub}$. So for all $i \geq 3$, $y^i \leq y_{ub}$ and $x^i \leq x_{ub}$.

Let us now prove that the type BB equilibrium point (x_{eq}, y_{eq}) is contained inside interval I . y_{eq} is a best response to x_{eq} . Therefore, $y_{eq} \leq \frac{\alpha Mp_2}{4C_2} = y_{ub}$. For $k \leq 1$, $x_{eq} = \frac{\alpha Mp_2}{4C_2} \frac{4k^{\alpha-1}}{(1+k^\alpha)^2} \geq \frac{\alpha Mp_2}{4C_2}$ and $y_{eq} = \frac{\alpha Mp_1}{4C_1} \frac{4k^{\alpha+1}}{(1+k^\alpha)^2} \leq \frac{\alpha Mp_1}{4C_1}$.

For all $y_{eq} \leq y \leq y_{ub}$, $\frac{\partial x_{BR}}{\partial y} \geq 0 \Rightarrow x_{eq} = x_{BR}(y_{eq}) \leq x_{BR}(y_{ub}) = x_{ub}$.

For all $x_{eq} \leq x \leq x_{ub}$, $\frac{\partial y_{BR}}{\partial x} \leq 0 \Rightarrow y_{eq} = y_{BR}(x_{eq}) \geq y_{BR}(x_{ub}) = y_{lb}$.

For all $y_{lb} \leq y \leq y_{eq}$, $\frac{\partial x_{BR}}{\partial y} \geq 0 \Rightarrow x_{eq} = x_{BR}(y_{eq}) \geq x_{BR}(y_{lb}) = x_{lb}$.

Thus, we have proved that $x_{lb} \leq x_{eq} \leq x_{ub}$, $y_{lb} \leq y_{eq} \leq y_{ub}$, that is, the type BB equilibrium is contained inside interval I .

Because of existence of a unique type BB equilibrium, the best response curves intersect each other at exactly one point denoted by (x_{eq}, y_{eq}) . Further, for all $x < x_{eq}$ and for all $y < y_{eq}$, the y_{BR} curve is above the x_{BR} curve and x_{BR} curve is to the right of y_{BR} curve. Also, for all $y < y_{eq}$, $x_{BR}(y) < x_{eq}$. Therefore, for all $x_i < x_{eq}$, if i is odd then $x_{i+1} = x_i$, $y_i < y_{i+1} \leq y_{ub}$ and if i is even then $x_i < x_{i+1} < x_{eq}$, $y_{i+1} = y_i$. So in each iteration, either x_i or y_i keeps strictly increasing until $y_i \geq y_{eq}$. In the very next iteration, $x_{i+1} = x_{BR}(y_i) \geq x_{eq}$ and $y_{i+1} = y_i \geq y_{eq}$. Thus, $x_{lb} \leq x_{eq} \leq x_{i+1} \leq x_{ub}$ and $y_{lb} \leq y_{eq} \leq y_{i+1} \leq y_{ub}$. We have proved part (a) of the proposition.

We have already proved that at the end of any iteration $i \geq 2$, $x_i \leq x_{ub}$ and $y_i \leq y_{ub}$. So for all i such that $x_{lb} \leq x_i \leq x_{ub}$ and $y_{lb} \leq y_i \leq y_{ub}$, all that remains to be proved is that $x_{lb} \leq x_{i+1}$ and $y_{lb} \leq y_{i+1}$. We first consider the case where i is even. $y_{i+1} = y_i$. As proved earlier, for all y such that $y_{lb} \leq y \leq y_{ub}$, $\frac{\partial x_{BR}}{\partial y} \geq 0$. Therefore, $y_{lb} \leq y_i \leq y_{ub} \Rightarrow x_{lb} = x_{BR}(y_{lb}) \leq x_{BR}(y_i) = x_{i+1} \leq x_{BR}(y_{ub}) = x_{ub}$. Therefore, $x_{lb} \leq x_{i+1} \leq x_{ub}$ and $y_{lb} \leq y_{i+1} \leq y_{ub}$. Now consider the case where i is odd. $x_{i+1} = x_i$. For all x_i such that $x_{eq} \leq x_i \leq x_{ub}$, $\frac{\partial y_{BR}}{\partial x} \leq 0$. Therefore, $y_{lb} = y_{BR}(x_{ub}) \leq y_{BR}(x_i) = y_{i+1}$. On the other hand, for all $x_i < x_{eq}$, $y_i < \frac{\alpha Mp_1}{4C_1}$,

$y_{i+1} = y_{BR}(x_i) > y_i \geq y_{lb}$. Therefore, if $x_{lb} \leq x_i \leq x_{ub}$, then $y_{lb} \leq y_{i+1}$. Thus we have proved that $x_{lb} \leq x_{i+1} \leq x_{ub}$ and $y_{lb} \leq y_{i+1} \leq y_{ub}$, if i is odd. Therefore, for any i such that (x_i, y_i) is in interval I , (x_{i+1}, y_{i+1}) is also in interval I . We have proved part (b) of the proposition. \square

Now we will prove that the absolute value of the slope of each of the best response curves inside interval I is less than 1 in the $\chi - \gamma$ coordinates. We will prove this for two representative values of α namely, $\alpha = 1.5$ and $\alpha = 1$.

Proposition 8. *For $\alpha = 1.5$, the absolute value of the slope of each of the best response curves inside interval I is less than 1 in the $\chi - \gamma$ coordinates.*

Proof. We will first prove that at $x = x_{ub}$, $|\frac{\partial \gamma_{BR}(\chi)}{\partial \chi}| < 1$.

$$\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = -\alpha \frac{\gamma}{\chi} \frac{1 - \frac{\gamma}{\chi}}{(\alpha + 1) \frac{\gamma}{\chi} - (\alpha - 1)}$$

The denominator of the right hand side (RHS) is always positive, due to the second order conditions. At $x = x_{ub}$, $x \geq y_{BR}(x)$, and hence, $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} \leq 0$. For $\alpha = 1.5$, solving for the point where $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = -1$, leads to a unique solution given by (x_{-1}, y_{-1}) , where,

$$y_{-1} = \frac{9}{32} \frac{Mp_2}{C_2} \text{ and } x_{-1} = 3^{\frac{2}{3}} \frac{9}{32} \frac{Mp_2}{C_2}.$$

Because $x_{ub} = x_{BR}\left(\frac{\alpha Mp_2}{4C_2}\right)$, we get,

$$\frac{4}{k} = \left(\frac{4C_2 x_{ub}}{1.5Mp_2}\right)^{2.5} + 2 \left(\frac{4C_2 x_{ub}}{1.5Mp_2}\right) + \left(\frac{4C_2 x_{ub}}{1.5Mp_2}\right)^{-0.5}.$$

Define $f(x) = \left(\frac{4C_2 x}{1.5Mp_2}\right)^{2.5} + 2 \left(\frac{4C_2 x}{1.5Mp_2}\right) + \left(\frac{4C_2 x}{1.5Mp_2}\right)^{-0.5}$. $f(x)$ is a strictly increasing function of x for $x \geq \frac{1.5Mp_2}{4C_2}$. $f(x_{ub}) = \frac{4}{k}$ and $f(x_{-1}) \approx 6.96$. $f(x_{ub}) < f(x_{-1})$ if and only if $k \geq 0.575$, which is always satisfied because one of the necessary conditions

for the existence of type BB equilibrium requires that $k \geq (\alpha - 1)^{\frac{1}{\alpha}} = 0.5^{\frac{2}{3}} > 0.575$. Therefore, $x_{ub} < x_{-1}$. Thus, we have proved that at $x = x_{ub}$, $-1 < \frac{\partial \gamma_{BR}(x)}{\partial x} < 0$.

Also for $x \geq \frac{\alpha M p_2}{4 C_2}$, $\frac{\partial y_{BR}}{\partial x} \leq 0$, therefore $y_{-1} = y_{BR}(x_{-1}) < y_{BR}(x_{ub}) = y_{lb}$. Next, we will obtain the coordinates of the point (which turns out to be unique) such that $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} = 1$ and prove that the y-coordinate at this point is less than y_{lb} . The condition,

$$\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} = 1.5 \frac{\chi}{\gamma} \frac{\frac{\chi}{\gamma} - 1}{(1.5 + 1) \frac{\chi}{\gamma} - (1.5 - 1)} = 1,$$

can be simplified to obtain,

$$x \approx 0.2029 \frac{1.5 M p_1}{C_1} \text{ and } y \approx 0.1091 \frac{1.5 M p_1}{C_1}.$$

Because $k \geq (\alpha - 1)^{\frac{1}{\alpha}} = 0.5^{\frac{2}{3}} > 0.589$, we get $y_{lb} > y_{-1} > 0.1091 \frac{1.5 M p_1}{C_1}$. So the y-coordinate of the point at which $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} = 1$ is less than y_{lb} . Because $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} \geq 0$ throughout interval I , $0 \leq \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} < 1$ for the $\chi_{BR}(\gamma)$ curve at $y = y_{lb}$.

Now, let us obtain the coordinates of the point (which turns out to be unique) such that $\frac{\partial \gamma_{BR}(x)}{\partial x} = 1$ and prove that the x-coordinate of this point is less than x_{lb} . Solving for $\frac{\partial \gamma_{BR}(x)}{\partial x} = 1$ we get,

$$y \approx 0.2029 \frac{1.5 M p_2}{C_2} \text{ and } x \approx 0.1091 \frac{1.5 M p_2}{C_2}.$$

In order to prove that $0.1091 \frac{1.5 M p_2}{C_2} < x_{lb} = x_{BR}(y_{lb})$, it is sufficient to prove that the y-coordinate of the point on the lower part of $x_{BR}(y)$ curve at which $x = 0.1091 \frac{1.5 M p_2}{C_2}$ is less than $y_{-1} = \frac{9}{32} \frac{M p_2}{C_2}$. This is easy to prove because for $y < \frac{\alpha M p_1}{4 C_1}$, the $x_{BR}(y)$ curve lies below $y = x$ line. Therefore, the y-coordinate corresponding to $x = 0.1091 \frac{1.5 M p_2}{C_2}$ is less than $0.1091 \frac{1.5 M p_2}{C_2}$ which is less than $\frac{9}{32} \frac{M p_2}{C_2}$. Therefore, at $x = x_{lb}$, $\frac{\partial \gamma_{BR}(x)}{\partial x} < 1$.

So far we have proved that $-1 < \frac{\partial \gamma_{BR}(x)}{\partial x} \leq 0$ at $x = x_{ub}$ and $\frac{\partial \gamma_{BR}(x)}{\partial x} < 1$ at $x = x_{lb}$. Therefore, $-1 < \frac{\partial \gamma_{BR}(x)}{\partial x} < 1$ for all x such that $x_{lb} \leq x \leq x_{ub}$. Also we

have proved that $0 \leq \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} < 1$ at $y = y_{lb}$ and $0 \leq \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma}$ at $y = y_{ub}$. Therefore, $-1 < \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} < 1$ for all y such that $y_{lb} \leq y \leq y_{ub}$.

Therefore for $\alpha = 1.5$, the absolute value of the slopes of each of the best response curves inside interval I is less than 1 in the $\chi - \gamma$ coordinates. \square

Proposition 9. *For $\alpha = 1$, the absolute value of the slope of each of the best response curves inside interval I is less than 1 in the $\chi - \gamma$ coordinates.*

Proof. For $\alpha = 1$, the $\chi - \gamma$ coordinate system is the same as the $x - y$ coordinate system. We will first prove that at $x = x_{ub}$, $|\frac{\partial \gamma_{BR}(\chi)}{\partial \chi}| < 1$.

For $\alpha = 1$,

$$\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = -\frac{1}{2} \left(1 - \frac{\gamma}{\chi} \right) > -\frac{1}{2}.$$

We know that at $x = x_{ub}$, $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} \leq 0$. Therefore, $x = x_{ub}$, $|\frac{\partial \gamma_{BR}(\chi)}{\partial \chi}| < 1$.

Next, we will obtain the coordinates of the point (which turns out to be unique) such that $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} = 1$ and prove that the y-coordinate at this point is less than y_{lb} . Solving for $\frac{\partial \chi(\gamma_{BR})}{\partial \gamma} = \frac{\chi-1}{2} = 1$, we get,

$$x = \frac{3Mp_1}{16C_1} \text{ and } y = \frac{Mp_1}{16C_1}.$$

For $x \geq \frac{Mp_2}{4C_2}$, we have $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} \leq 0$ and for $y \leq \frac{Mp_1}{4C_1}$, we have $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} \geq 0$. $y_{ub} = \frac{Mp_2}{4C_2} \leq \frac{Mp_1}{4C_1}$. So $x_{ub} = x_{BR}(y_{ub}) \leq x_{BR}\left(\frac{Mp_1}{4C_1}\right) = \frac{Mp_1}{4C_1}$. So we get $y_{lb} = y_{BR}(x_{ub}) \geq y_{BR}\left(\frac{Mp_1}{4C_1}\right)$. As per the first order conditions,

$$\frac{\frac{Mp_1}{4C_1}}{\left(y_{BR}\left(\frac{Mp_1}{4C_1}\right) + \frac{Mp_1}{4C_1}\right)^2} = \frac{C_2}{Mp_2} \iff y_{BR}\left(\frac{Mp_1}{4C_1}\right) = \frac{Mp_1}{4C_1} (2\sqrt{k} - 1).$$

$y_{lb} \geq y_{BR}\left(\frac{Mp_1}{4C_1}\right) = \frac{Mp_1}{4C_1} (2\sqrt{k} - 1) > \frac{Mp_1}{16C_1}$ because $k \geq 0.4$. Therefore, the y coordinate of the point where $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} = 1$ is less than y_{lb} .

Now, let us obtain the coordinates of the point (which turns out to be unique) such that $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = 1$ and prove that the x-coordinate of this point is less than x_{lb} . Solving for $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = 1$, we get,

$$x = \frac{Mp_2}{16C_2} \text{ and } y = \frac{3Mp_2}{16C_2}.$$

Because $\frac{Mp_1}{4C_1} \geq y_{lb} > \frac{Mp_1}{16C_1}$, and $\frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} > 0$ for $y < \frac{Mp_1}{4C_1}$, we get $x_{lb} = x_{BR}(y_{lb}) > x_{BR}\left(\frac{Mp_1}{16C_1}\right) = \frac{3Mp_1}{16C_1} > \frac{Mp_2}{16C_2}$. The last inequality holds because $k \leq 1$. Therefore, the x coordinate at the point where $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} = 1$ is less than x_{lb} .

Thus we have proved that $-1 < \frac{\partial \gamma_{BR}(\chi)}{\partial \chi} \leq 0$ at $x = x_{ub}$ and $\frac{\partial \gamma_{BR}(\chi)}{\partial \chi} < 1$ at $x = x_{lb}$. Therefore, $-1 < \frac{\partial \gamma_{BR}(\chi)}{\partial \chi} < 1$ for all x such that $x_{lb} \leq x \leq x_{ub}$. Also we have proved that $0 \leq \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} < 1$ at $y = y_{lb}$ and $0 \leq \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma}$ at $y = y_{ub}$. Therefore, $-1 < \frac{\partial \chi_{BR}(\gamma)}{\partial \gamma} < 1$ for all y such that $y_{lb} \leq y \leq y_{ub}$.

Therefore, for $\alpha = 1$, the absolute value of the slopes of each of the best response curves inside interval I is less than 1 in the $\chi - \gamma$ coordinates. \square

In order to prove the next proposition, we assume that the absolute value of slope of each of the best response curves is less than 1 in interval I .

Proposition 10. *If the absolute value of slope of each of the best response curves is less than 1 in interval I , then as long as the competitor frequency for each carrier remains in region B , regardless of the starting point, the myopic best response algorithm converges to the unique type BB equilibrium.*

Proof. We have assumed that the absolute value of slope of each of the best response curves is less than 1 in interval I . Also we have proved that as long as the competitor frequency for each carrier remains in region B , regardless of the starting point the myopic best response algorithm will reach a point in interval I in a finite number of iterations and once inside interval I , it will never leave the interval.

Let (χ_{eq}, γ_{eq}) be the type BB equilibrium point in the $\chi - \gamma$ coordinate system. We define a sequence $L(i)$ as follows:

$$L(i) = \begin{cases} |\chi_i - \chi_{eq}| & \text{if } i \text{ is odd} \\ |\gamma_i - \gamma_{eq}| & \text{if } i \text{ is even.} \end{cases}$$

Let us consider any iteration i after the algorithm has reached inside the interval I . We will prove that once inside interval I , $L(i)$ is strictly decreasing.

Let us first consider the case where i is odd. $L(i) = |\chi_i - \chi_{eq}|$. In the $(i+1)^{th}$ iteration, χ value remains unchanged. Only the γ value changes from γ_i to γ_{i+1} .

$$\begin{aligned} L(i+1) &= |\gamma_{i+1} - \gamma_{eq}| = |\gamma_{BR}(\chi_i) - \gamma_{BR}(\chi_{eq})| = \left| \int_{\chi_{eq}}^{\chi_i} \left(\frac{\partial \gamma(\chi)}{\partial \chi} \right) d\chi \right| \\ &\leq \left| \int_{\chi_{eq}}^{\chi_i} \left| \frac{\partial \gamma(\chi)}{\partial \chi} \right| d\chi \right| < \left| \int_{\chi_{eq}}^{\chi_i} 1 d\chi \right| = |\chi_i - \chi_{eq}| = L(i) \end{aligned}$$

We have proved that once inside interval I , $L(i)$ is strictly decreasing for odd values of i . By symmetry, the same is true for even values of i . Moreover, $L(i) = 0$ if and only if $x = x_{eq}$ and $y = y_{eq}$. Therefore, $L(i)$ is a decreasing sequence which is bounded below. So it converges to the unique type BB equilibrium point. \square

Proposition 11. *Regardless of the starting point, the myopic best response algorithm converges to the unique type BB equilibrium as long as the following conditions are satisfied.*

$$\begin{aligned} \frac{\alpha Mp_1}{4C_1} &< x_{th} \\ \frac{\alpha Mp_2}{4C_2} &< y_{th} \\ x_{cr} &< x_{BR}(y_{th}) \\ y_{cr} &< y_{BR}(x_{th}) \\ x_{cr} &< x_{BR}(y_{cr}) \\ y_{cr} &< y_{BR}(x_{cr}) \end{aligned}$$

Proof. First we develop sufficient conditions under which the competitor frequency for each carriers remains in region B for all iterations $i \geq 2$, regardless of the starting point.

As proved earlier, the shape of the best response curve $y_{BR}(x)$ is such that at $x = 0$, $y = \frac{M}{S_2}$. Initially it is strictly decreasing followed by a point of non-differentiability (at x_{cr}) beyond which it is strictly increasing until a local maximum is reached at $x = \frac{\alpha Mp_2}{4C_2}$. Beyond the local maximum, it is strictly decreasing again to a point of discontinuity (at x_{th}), beyond which it takes a constant value 0. For $x \leq x_{th}$, the only candidates for global minima of the best response curve $y_{BR}(x)$ are x_{cr} and x_{th} . The only candidates for global maxima are $x = 0$ and $x = \frac{\alpha Mp_2}{4C_2}$. If the y-coordinate at each of these four important points lies in the range $y_{cr} < y < y_{th}$, then $y_{cr} < y_{BR}(x) < y_{th}$ for all $x < x_{th}$. Similarly, if $x_{BR}(y)$ at $y = 0, y = y_{cr}, y = \frac{\alpha Mp_1}{4C_1}$, and $y = y_{th}$ are all in the range $x_{cr} < x < x_{th}$, then $x_{cr} < x_{BR}(y) < x_{th}$ for all $y < y_{th}$. So for any starting point x_0 such that $x_0 < x_{th}$, the algorithm will remain in the region B of both carriers for all subsequent iterations. The only remaining case is when $x \geq x_{th}$ or $y \geq y_{th}$. This does not pose any problem because for all $x \geq x_{th}$, $x_{BR}(y_{BR}(x)) = x_{BR}(0)$ and $x_{cr} < x_{BR}(0) = \frac{M}{S_2} < x_{th}$. So if the aforementioned conditions are satisfied, then regardless of the starting point, the algorithm will remain in the region B of both carriers for all iterations i such that $i \geq 2$.

For all the aforementioned conditions to be satisfied, it is sufficient to ensure that

the upper bound conditions on the points of local maxima are satisfied and the lower bound conditions on the points of local minima are satisfied. Let us first look at the upper bounds on the points of local maxima. There are 4 such conditions per carrier, namely $\frac{M}{S_1} < x_{th}$, $\frac{M}{S_2} < y_{th}$, $\frac{\alpha M p_1}{4C_1} < x_{th}$ and $\frac{\alpha M p_2}{4C_2} < y_{th}$. $\frac{M}{S_1} < x_{th}$ simplifies to,

$$\frac{C_2}{S_2 p_2} < \frac{S_1}{S_2} \frac{1}{\alpha} (\alpha - 1)^{\frac{\alpha-1}{\alpha}}$$

which is the exact negation of the condition for existence of type CA equilibrium. Because we have assumed that the only unique PSNE in this game is a type BB equilibrium, a type CA equilibrium cannot exist. Hence this condition is automatically satisfied. By symmetry, due to the non-existence of a type AC equilibrium, the condition $\frac{M}{S_2} < y_{th}$ is automatically satisfied.

The remaining six conditions are as follows:

$$\begin{aligned} \frac{\alpha M p_1}{4C_1} &< x_{th} \\ \frac{\alpha M p_2}{4C_2} &< y_{th} \\ x_{cr} &< x_{BR}(y_{th}) \\ y_{cr} &< y_{BR}(x_{th}) \\ x_{cr} &< x_{BR}(y_{cr}) \\ y_{cr} &< y_{BR}(x_{cr}) \end{aligned}$$

If each of these conditions is satisfied then the myopic best response algorithm converges to the unique type BB equilibrium, regardless of the starting point. \square

Out of the 157 records of 2-player cases analyzed for the domestic US segments, 86 segments were such that the only PSNE was a type BB equilibrium. In each and every one of these 86 cases, all the 6 conditions mentioned above were satisfied. Therefore, for each of these 86 cases, the myopic best response dynamic converges to

the unique type BB equilibrium point regardless of the starting point. Thus the data analysis suggests that these conditions are very mild.

4.5.4 Alternative Dynamic

This dynamic is applicable only in the part of region B where the utility function is strictly concave i.e. we will consider the region where $\frac{\alpha-1}{\alpha+1} + \epsilon \leq \left(\frac{x}{y}\right)^\alpha \leq \frac{\alpha+1}{\alpha-1} - \epsilon$, where ϵ is any sufficiently small positive number. This requirement is not very restrictive. This condition is always satisfied at the type BB equilibrium, due to the second order conditions. Moreover, the $\frac{x}{y}$ values satisfying this condition cover a large region surrounding the type BB equilibrium. For example, for $\alpha = 1$, this condition is always satisfied for all values of $\frac{x}{y}$, while for $\alpha = 1.5$, the condition translates approximately to $0.342 \leq \frac{x}{y} \leq 2.924$, which is a large range. In order to provide a complete specification of the player utilities, we will define the player i utility outside this region by means of a quadratic function of a single variable x_i . The coefficients are such that $u_i(x_i)$ and its first and second order derivatives with respect to x_i are continuous.

Multiplying the utility function by a positive real number is an order preserving transformation, which does not affect the properties of the game. We will multiply the utility of player i by $\frac{1}{p_i}$. So $u_i = \frac{\Pi_i}{p_i}$. This dynamic was proposed by Rosen [84]. Under this dynamic, each player changes his strategy such that his own utility would increase if all the other players held to their current strategies. The rate of change of each player's strategy with time is equal to the gradient of his utility with respect to his own strategy, subject to constraints. For the frequency competition game, where each player's strategy space is 1-dimensional, the rate of change of each player's strategy simply equals the derivative of the player's utility with respect to the frequency decision, subject to the upper and lower bound on allowable frequency values. Therefore, the rate of adjustment of each player's strategy is given by,

$$\frac{dx_i}{dt} = \frac{du_i(x)}{dx_i} + b_{min} - b_{max}.$$

The only purpose of the summation term is to ensure that the frequency values stay within the allowable range, $x_{min} \leq x \leq x_{max}$. b_{min} will be equal to 0 for all $x > x_{min}$ and will take an appropriate positive value at $x = x_{min}$ to ensure that the lower bound is respected. Similarly, b_{max} will be equal to 0 for all $x < x_{max}$ and will take an appropriate positive value at $x = x_{max}$ to ensure that the upper bound is respected. As long as the competitor frequencies remain in region B for each carrier, the utilities are given by:

$$u_1(x, y) = M \frac{x^\alpha}{x^\alpha + y^\alpha} - \frac{C_1}{p_1} x \text{ and } u_2(x, y) = M \frac{y^\alpha}{x^\alpha + y^\alpha} - \frac{C_2}{p_2} y.$$

The vector of utility functions $u(x, y)$ is given by: $u(x, y) = [u_1(x, y), u_2(x, y)]$. The vector of first order derivatives of each player's utility with respect to his own frequency is given by: $\nabla u(x, y) = [\frac{\partial u_1(x, y)}{\partial x}, \frac{\partial u_2(x, y)}{\partial y}]$. The Jacobian of ∇u is given by:

$$U(x, y) = \begin{pmatrix} \frac{\partial^2 u_1(x, y)}{\partial x^2} & \frac{\partial^2 u_1(x, y)}{\partial x \partial y} \\ \frac{\partial^2 u_2(x, y)}{\partial y \partial x} & \frac{\partial^2 u_2(x, y)}{\partial y^2} \end{pmatrix}$$

The first order derivatives are given by,

$$\frac{\partial u_1(x, y)}{\partial x} = \frac{M \alpha x^{\alpha-1} y^\alpha}{(x^\alpha + y^\alpha)^2} - \frac{C_1}{p_1}$$

and $\frac{\partial u_2(x, y)}{\partial y} = \frac{M \alpha y^{\alpha-1} x^\alpha}{(x^\alpha + y^\alpha)^2} - \frac{C_2}{p_2},$

and the second order derivatives are given by,

$$\begin{aligned}
[U(x, y)]_{11} &= \frac{\partial^2 u_1(x, y)}{\partial x^2} = \frac{M\alpha x^{\alpha-2} y^\alpha}{(x^\alpha + y^\alpha)^3} ((\alpha - 1) y^\alpha - (\alpha + 1) x^\alpha) < 0 \\
[U(x, y)]_{22} &= \frac{\partial^2 u_2(x, y)}{\partial y^2} = \frac{M\alpha y^{\alpha-2} x^\alpha}{(x^\alpha + y^\alpha)^3} ((\alpha - 1) x^\alpha - (\alpha + 1) y^\alpha) < 0 \\
[U(x, y)]_{12} &= \frac{\partial^2 u_1(x, y)}{\partial x \partial y} = \frac{M\alpha^2 x^{\alpha-1} y^{\alpha-1}}{(x^\alpha + y^\alpha)^3} (x^\alpha - y^\alpha) \\
[U(x, y)]_{21} &= \frac{\partial^2 u_2(x, y)}{\partial y \partial x} = \frac{M\alpha^2 x^{\alpha-1} y^{\alpha-1}}{(x^\alpha + y^\alpha)^3} (y^\alpha - x^\alpha) \\
\Rightarrow [(U(x, y) + U^T(x, y))]_{11} &= 2[U(x, y)]_{11} \\
\text{and } [(U(x, y) + U^T(x, y))]_{22} &= 2[U(x, y)]_{22} \\
\text{and } [(U(x, y) + U^T(x, y))]_{12} &= [(U(x, y) + U^T(x, y))]_{12} = 0.
\end{aligned}$$

Therefore, $(U(x, y) + U^T(x, y))$ is a diagonal matrix with both diagonal elements strictly negative. Therefore, $(U(x, y) + U^T(x, y))$ is negative definite. This is sufficient to prove that the payoff functions are diagonally strictly concave [84]. Therefore, under the alternative dynamic mentioned above, the frequencies of the competing carriers will converge to the unique type BB equilibrium frequencies.

4.6 N-Player Symmetric Game

Now we will extend the analysis to the N-player symmetric case, where $N \geq 2$. By symmetry, we mean that the operating cost C_i , the seating capacity S_i and the fare p_i is the same for all carriers. For the analysis presented in this section, it is sufficient to have $\frac{C_i}{p_i}$ constant for all carriers. However, for computing the price of anarchy in the next section we need the remaining assumptions. We will simplify the notation and denote the operating cost for each carrier as C , seating capacity as S and fare as p . Under symmetry, the necessary and sufficient conditions for the existence of a type BB equilibrium for a 2-player game reduce to a single condition, $\frac{\alpha PS}{C} > 2$. We will assume that this condition holds throughout the following analysis.

Assumption 4. $\frac{\alpha PS}{C} > 2$

Proposition 12. *In an N -player symmetric game, a symmetric equilibrium with excess seating capacity exists at $x_i = \frac{\alpha Mp}{C} \frac{N-1}{N^2}$ for all i if and only if $N \leq \frac{\alpha}{\alpha-1}$ and if it exists, then it is the unique symmetric equilibrium.*

Proof. The utility of each carrier i is given by $u_i(x_i, y_i) = M \frac{x_i^\alpha}{x_i^\alpha + y_i^\alpha} - \frac{C}{p} x_i$, where $y_i = \left(\sum_{j=1, j \neq i}^N x_j^\alpha \right)^{\frac{1}{\alpha}}$ is the effective competitor frequency for player i . From the FOCs, we get $x_i = \frac{\alpha Mp}{C} \frac{x_i^\alpha y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^2}$. In the symmetric game, $\frac{C_i}{p_i}$ is the same for every player i . In general, this symmetric game may have both symmetric and asymmetric equilibria. In a symmetric equilibrium, $x_1 = x_2 = \dots = x_N$. Assume excess seating capacity for each carrier. Substituting in the FOCs we get $y_i = (N-1)^{\frac{1}{\alpha}} x_i$. Therefore, $x_i = \frac{\alpha Mp}{C} \frac{N-1}{N^2}$ for all i is the unique solution. Therefore, we have proved that if an equilibrium exists at this point, then it must be the unique symmetric equilibrium of this game.

In order to prove that this point is an equilibrium point, we need to prove that the SOC is satisfied, the profit at this point is non-negative and seating capacity is at least as much as the demand for each carrier.

The SOC is satisfied if and only if,

$$\frac{\partial^2 U_i}{\partial x_i^2} = \frac{M \alpha x_i^{\alpha-2} y_i^\alpha}{(x_i^\alpha + y_i^\alpha)^3} ((\alpha-1) y_i^\alpha - (\alpha+1) x_i^\alpha) \leq 0 \iff N \leq \frac{2\alpha}{\alpha-1}.$$

The condition on non-negativity of profit is satisfied if and only if,

$$\frac{\alpha Mp}{C} \frac{N-1}{N^2} * C \leq \frac{Mp}{N} \iff N \leq \frac{\alpha}{\alpha-1}.$$

The condition of excess seating capacity is satisfied if and only if,

$$\frac{\alpha Mp}{C} \frac{N-1}{N^2} * S > \frac{M}{N} \iff N > \frac{\alpha \frac{pS}{C}}{\alpha \frac{pS}{C} - 1},$$

which is always true for $\alpha \frac{pS}{C} > 2$.

Thus the symmetric equilibrium exists if and only if $N \leq \frac{\alpha}{\alpha-1}$. \square

Proposition 13. *In a symmetric N -player game, there exists no asymmetric equilib-*

rium where all players have a non-zero frequency and excess seating capacity.

Proof. Let us assume the contrary. For a symmetric N-player game, let there exist an asymmetric equilibrium such that all players have a non-zero frequency and excess seating capacity. Let us define $\beta = \sum_{j=1}^N x_j^\alpha$ and $\omega_i = \frac{x_i^\alpha}{\sum_{j=1}^N x_j^\alpha}$. So $x_i = (\omega_i \beta)^{\frac{1}{\alpha}}$. Substituting in the FOC, we get,

$$\begin{aligned} (\omega_i \beta)^{\frac{1}{\alpha}} &= \frac{\alpha M p}{C} \omega_i (1 - \omega_i) \\ \Rightarrow \frac{C}{\alpha M p} \beta^{\frac{1}{\alpha}} &= \omega_i^{\frac{\alpha-1}{\alpha}} - \omega_i^{\frac{2\alpha-1}{\alpha}} \end{aligned}$$

Let us define a function $h(\omega_i) = \omega_i^{\frac{\alpha-1}{\alpha}} - \omega_i^{\frac{2\alpha-1}{\alpha}}$. The value of $h(\omega_i)$ is the same across all the players at equilibrium. For all $\omega_i > 0$, $h(\omega_i)$ is a strictly concave function. So it can take the same value at at most two different values of ω_i . So all ω_i can take at most two different values. Let $\omega_i = v_1$ for $m (\leq N)$ players, and $\omega_i = v_2$ for the remaining $N - m$ players. Let $v_1 > v_2$, without loss of generality. $h(\omega_i)$ is maximized at $\omega_i = \frac{\alpha-1}{2\alpha-1}$. So $v_2 < \frac{\alpha-1}{2\alpha-1} < v_1$.

At equilibrium, each player's profit must be non-negative. Therefore, the profit for each player i such that $\omega_i = v_2$ is given by $Mp\omega_i - Cx_i$. But $x_i = \frac{\alpha M p}{C} \omega_i (1 - \omega_i)$. So the condition on non-negativity of profit simplifies to, $v_2 \geq \frac{\alpha-1}{\alpha}$. Therefore, $\frac{\alpha-1}{2\alpha-1} > v_2 \geq \frac{\alpha-1}{\alpha}$, which can be true only if $\alpha < 1$. This leads to a contradiction. So we have proved that for a symmetric N-player game, there exists no asymmetric equilibrium such that all players have a non-zero frequency and excess seating capacity. \square

Proposition 14. *In a symmetric N-player game, there exists some n_{min} such that for any integer n with $\max(2, n_{min}) \leq n \leq \min(N - 1, \frac{\alpha}{\alpha-1})$ there exist exactly $\binom{N}{n}$ asymmetric equilibria such that exactly n players have non-zero frequency and all players with nonzero frequency have excess seating capacity. There exists at least one such integer for $N \geq \frac{\alpha}{\alpha-1}$. The frequency of each player with non-zero frequency equals $\frac{\alpha M p}{C} \frac{n-1}{n^2}$.*

Proof. Let us denote this game as G . Consider any equilibrium having exactly n

players with non-zero frequency. Let us rearrange the player indices such that players $i = 1$ to $i = n$ have non-zero frequencies. Let us consider a new game which involves only the first n players. We will denote this new game as G' . An equilibrium of G where only the first n players have a non-zero frequency is also an equilibrium for the game G' where all players have non-zero frequency. As we have already proved, the equilibrium frequencies of each of the first n players must be equal to $\frac{\alpha Mp}{C} \frac{n-1}{n^2}$. This ensures that any of the first n players will not benefit from unilateral deviations from this equilibrium profile. In order to ensure that none of the remaining $N - n$ players has an incentive to deviate, we must ensure that the effective competitor frequency for any player j such that $j > n$ must be at least equal to y_{th} . This condition is satisfied if and only if,

$$\begin{aligned} n^{\frac{1}{\alpha}} \frac{n-1}{n^2} \frac{\alpha Mp}{C} &\geq (\alpha-1)^{\left(\frac{\alpha-1}{\alpha}\right)} \frac{Mp}{\alpha C} \\ \iff n^{\frac{1-\alpha}{\alpha}} - n^{\frac{1-2\alpha}{\alpha}} &\geq \frac{(\alpha-1)^{\frac{\alpha-1}{\alpha}}}{\alpha^2}. \end{aligned} \quad (4.22)$$

LHS is an increasing function of n for $n \leq \frac{\alpha}{\alpha-1}$. Also the RHS is a decreasing function of α (this can be verified by differentiating the log of RHS with respect to α). Also it can be easily verified that at $n = \frac{\alpha}{\alpha-1}$, the inequality holds for every α . Therefore, for any given α value, there exists some $n_{min} \geq 0$ such that for all n such that $\frac{\alpha}{\alpha-1} \geq n \geq n_{min}$, this inequality is satisfied. As proved earlier, the condition for existence of an equilibrium with all players having non-zero frequency in game G' is $n \leq \frac{\alpha}{\alpha-1}$.

So all the conditions for an equilibrium of game G are satisfied if $\max(2, n_{min}) \leq n \leq \min(N-1, \frac{\alpha}{\alpha-1})$. Therefore, any equilibrium of game G' where all players have non-zero frequency is also an equilibrium of game G where all the remaining players have zero frequency and vice versa. The players in game G' can be chosen in $\binom{N}{n}$ ways. Therefore, we have proved that in a symmetric N -player game, for any integer n such that $\max(2, n_{min}) \leq n \leq \min(N-1, \frac{\alpha}{\alpha-1})$, there exist exactly $\binom{N}{n}$ asymmetric equilibria such that exactly n players have non-zero frequency. To show that there

exists at least one such integer n , consider 2 cases. If $\alpha > 1.5$, then it is easy to verify that the inequality (4.22) is always satisfied for $n = 2$. If $\alpha \leq 1.5$, then we see that (4.22) is satisfied by $n = \frac{1}{\alpha-1} = \frac{\alpha}{\alpha-1} - 1$. In either case, $\frac{\alpha}{\alpha-1} > 2$ is always satisfied. So there always exist some such n . The frequency of each player with non-zero frequency equals $\frac{\alpha Mp}{C} \frac{n-1}{n^2}$. \square

From here onwards, we will denote each such equilibrium as an n -symmetric equilibrium of an N -player game.

Proposition 15. *Among all equilibria with exactly n players ($n \leq N$) having nonzero frequency, the total frequency is maximum for the symmetric equilibrium.*

Proof. As proved earlier, any possible asymmetric equilibria with exactly n players having nonzero frequency must involve at least one player with no excess seating capacity. Let player i be such a player with nonzero frequency and no excess seating capacity at equilibrium. So the effective competitor frequency y must be at most equal to y_{cr} and $x_i \geq x_{cr} = \frac{M}{S} \left(1 - \frac{C}{\alpha p S}\right) > \frac{M}{2S}$. Therefore, each such player must carry at least $\frac{M}{2}$ passengers. Therefore, at equilibrium there can be at most one such player. So each of the remaining $n - 1$ players has excess capacity. Using the same argument as the one used in proving proposition 13, we can prove that each player with non-zero frequency and excess capacity will have equal frequency at equilibrium.

Let us denote the equilibrium frequency of the sole player with no excess capacity by x_1 and that of each of the remaining players as x_2 . We will denote the equilibrium market share of the player with no excess capacity as l . Therefore, the total frequency under the asymmetric equilibrium equals,

$$\begin{aligned} (n-1)x_2 + x_1 &= \frac{\alpha Mp}{C} \frac{(n-1)x_2^\alpha}{(n-1)x_2^\alpha + x_1^\alpha} \left(1 - \frac{x_2^\alpha}{(n-1)x_2^\alpha + x_1^\alpha}\right) + \frac{M}{S} \frac{x_1^\alpha}{(n-1)x_2^\alpha + x_1^\alpha} \\ &= \frac{\alpha Mp}{C} (1-l) \left(1 - \frac{1-l}{n-1}\right) + \frac{M}{S} l \end{aligned}$$

Let us assume that there exists an asymmetric equilibrium where the total frequency is greater than that under the corresponding n -symmetric equilibrium, which

equals $\frac{\alpha Mp}{C} \frac{n-1}{n}$. This condition translates into,

$$\frac{\alpha Mp}{C} (1-l) \left(1 - \frac{1-l}{n-1}\right) + \frac{M}{S} l > \frac{\alpha Mp}{C} \frac{n-1}{n},$$

which further simplifies to, $nl(5-n-2l) > 2$. But we know that $n \in \mathbb{I}^+$, $n \geq 2$ and $l \geq \frac{1}{2}$. So $5-n-2l > 0$ only if $n < 5-2l \leq 4$. So $n = 2$ or $n = 3$. For $n = 2$, the conditions for existence of type BC equilibrium in the 2-player case require $\frac{\alpha PS}{C} \leq 2$, which contradicts our assumption. For $n = 3$, we need some l such that $3l^2 - 3l + 1 < 0$, which is true if and only if $3(l-0.5)^2 + 0.25 < 0$, which is also impossible. Thus our assumption leads to a contradiction. So we have proved that among all equilibria with exactly n players ($n \leq N$) having nonzero frequency, the total frequency is maximum for the symmetric equilibrium. \square

Proposition 16. *There exists no equilibrium with exactly n players with non-zero frequency such that $n > \frac{\alpha}{\alpha-1}$.*

Proof. We already proved that if $n > \frac{\alpha}{\alpha-1}$, there exists no equilibrium with all n players having excess capacity. We have also proved that the number of players without excess capacity can be at most one. So consider some equilibrium with one player with no excess capacity. Let the market share of that player be l and let the equilibrium frequency of each of the remaining players be x_2 . Because $n > \frac{\alpha}{\alpha-1}$, therefore $\alpha > \frac{n}{n-1}$.

For non-negative profit at equilibrium we require, $\frac{Mp}{C} \frac{1-l}{n-1} \geq x_2$. From the FOC, we get $x_2 = \frac{\alpha Mp}{C} \frac{1-l}{n-1} \left(1 - \frac{1-l}{n-1}\right)$. Combining the two we get,

$$\begin{aligned} 1 &\geq \alpha \left(1 - \frac{1-l}{n-1}\right) \\ \Rightarrow \frac{n}{n-1} &< \alpha \leq \frac{n-1}{n+l-2} \leq \frac{n-1}{n-1.5} \\ \Rightarrow n &< 2, \end{aligned}$$

which is impossible.

Therefore, we have proved that there exists no equilibrium with exactly n players with non-zero frequency such that $n > \frac{\alpha}{\alpha-1}$. \square

In this section, we proved that for an N -player symmetric game, if $N \leq \frac{\alpha}{\alpha-1}$, then there exists a fully symmetric equilibrium where the equilibrium frequency of each carrier at equilibrium is $\frac{\alpha Mp}{C} \frac{N-1}{N^2}$ and there exists no asymmetric equilibrium with all N players having a non-zero frequency. On the other hand, if $N > \frac{\alpha}{\alpha-1}$, then there exists no equilibrium with all players having non-zero frequency. In either case, there exist exactly $\binom{N}{n}$ n -symmetric equilibria for each integer $n < N$ such that $\max(2, n_{min}) \leq n \leq \min(N-1, \frac{\alpha}{\alpha-1})$ for some $n_{min} \geq 0$. Additionally, there may be asymmetric equilibria such that each asymmetric equilibrium has exactly one player with 100% load factor, $n-1$ more players with non-zero frequency and excess seating capacity and $N-n$ players with zero frequency. We also proved that there always exists at least one equilibrium for an N -player symmetric game. The aforementioned types of equilibria are exhaustive, that is there exist no other types of equilibria. As before, we realize that all the equilibria except those where all players have a nonzero frequency and excess capacity are suspect in terms of their portrayal of reality. So the fully symmetric equilibrium appears to be the most realistic one. In addition, the fully symmetric equilibrium is also the worst case equilibrium in the sense that it is the equilibrium which has the maximum total frequency, as will be apparent in the next section.

We proved that for some $n' < N$, if there exists no symmetric equilibrium for all $n \geq n'$, then there exists no asymmetric equilibrium for all $n \geq n'$ either. We also proved that for any given n , the total frequency at each asymmetric equilibrium having n non-zero frequency players is at most equal to the total frequency at the corresponding n -symmetric equilibrium. These results will help us obtain the price of anarchy in the next section.

4.7 Price of Anarchy

In any equilibrium, the total revenue earned by all carriers remains equal to Mp . The total flight operating cost to all carriers is given by $\sum_{i=0}^n Cx_i = C \sum_{i=0}^n x_i$. On the other hand, if there were a central controller trying to minimize the total operating cost, the minimum number of flights for carrying all the passengers would be equal to $\frac{M}{S}$ and the total operating cost would be $\frac{MC}{S}$. Similar to the notion introduced by Koutsoupias and Papadimitriou [59], let us define the price of anarchy as the ratio of total operating cost at Nash equilibrium to the total operating cost under the optimal frequency. The denominator is a constant and the numerator is proportional to the total number of flights.

A large proportion of airport delays are caused by congestion. Congestion related delay at an airport is an increasing (often nonlinearly) function of the total number of flights. Therefore, the greater the total number of flights, more is the delay. Total profit earned by all the airlines in a market is also a decreasing function of the total frequency. Also, because the total number of passengers remains constant, the average load factor in a market is inversely proportional to the total frequency. Lower load factors mean more wastage of seating capacity. Thus total frequency is a good measure of airline profitability, total operating cost, airport congestion and load factors. Higher total frequency across all carriers in a market means lower profitability, more cost, more congestion and lower average load factor, assuming constant aircraft size. Greater the price of anarchy, more is the inefficiency introduced by the competitive behavior of players at equilibrium.

Proposition 17. *In a symmetric N -player game, the price of anarchy is given by $\frac{\alpha p S}{C} \frac{n-1}{n}$, where n is the largest integer not exceeding $\min(N, \frac{\alpha}{\alpha-1})$.*

Proof. As proved earlier, a symmetric N -player game has $\sum_{n=\max(2, n_{\min})}^{\min(N, \frac{\alpha}{\alpha-1})} \binom{N}{n}$ equilibria (for some $n_{\min} \geq 0$). such that each equilibrium has a set of exactly n players each with frequency $\frac{\alpha M p}{C} \frac{n-1}{n^2}$ and excess capacity, whereas remaining $N - n$ players have zero frequency. Also, for any $n < \min(N, \frac{\alpha}{\alpha-1})$, there may exist equilibria with exactly n players having non-zero frequency and one of them having no excess

capacity at equilibrium. However, the frequency under any equilibrium with exactly n players having non-zero frequency is at most equal to the corresponding n -symmetric equilibrium. In any equilibrium having n players with non-zero frequency, the total flight operating cost is given by $\frac{\alpha M p}{C} \frac{n-1}{n^2}$, which is an increasing function of n . The total cost under minimum cost scheduling would be $\frac{MS}{C}$. Therefore, the ratio of total cost under equilibrium to total cost under minimum cost scheduling is $\frac{\alpha p S}{C} \frac{n-1}{n}$, which is an increasing function of n . Also, no equilibrium exists for $n > \frac{\alpha}{\alpha-1}$. Therefore, the price of anarchy is given by $\frac{\alpha p S}{C} \frac{n-1}{n}$, where n is the greatest integer less than equal to $\min(N, \frac{\alpha}{\alpha-1})$. \square

This expression has several important implications. Greater the α value, more is the price of anarchy. This means that as the market share-frequency share relationship becomes more and more curved, and goes away from the straight line, greater is the price of anarchy. So the S-curve phenomenon has a direct impact on airline profitability and airport congestion. Also, more the airfare compared to the operating cost per seat (i.e. more is the value of $\frac{pS}{C}$), greater is the price of anarchy. In other words, for short-haul, high-fare markets the price of anarchy is greater. Finally, more the number of competitors, greater is the price of anarchy (up to a threshold value beyond which it remains constant).

The equilibrium results from this simple model help substantiate some of the claims mentioned earlier. The price of anarchy increases because of the S-shaped (rather than linear) market share-frequency share relationship. Therefore, similar to the suggestions by Button and Drexler [29] and O'connor [69], the S-curve relationship tends to encourage airlines to provide excess capacity and schedule greater numbers of flights. Total profitability of all the carriers in a market under the worst case equilibrium provides a lower bound on airline profitability under competition. This lower bound is an increasing function of the price of anarchy, which in turn increases with number of competitors. Therefore, similar to Kahn's [56] argument, this raises the question of whether the objectives of a financially strong and highly competitive airline industry are inherently conflicting. In addition, these results also establish the link between airport congestion and airline competition. Airport congestion under

the worst-case equilibrium is directly proportional to the price of anarchy. So greater the number of competitors and more the curvature of the market share-frequency share relationship, greater is the airport congestion and delays.

4.8 Summary

In this chapter, we modeled airline frequency competition based on the S-curve relationship which has been well documented in airline literature. Regardless of the exact value of α parameter, it is usually agreed that market share is an increasing (linear or S-shaped) function of frequency share. Our model is general enough to accommodate somewhat differing beliefs about the market share-frequency share relationship. We characterize the best response curves for each player in a multi-player game. Due to complicated shape of best response curves, we proved that there exist anywhere between 0 to 6 different equilibria depending on the exact parameter values. All the existence and uniqueness conditions can be completely described by 3 unitless parameters (in addition to α) of the game. Only one out of the 6 possible equilibria seemed reasonable in terms of portrayal of reality. This equilibrium corresponds to both players having nonzero frequency and less than 100% load factors. In order to narrow down the modeling effort, realistic parameter ranges were identified based on real world data that come closest to the simplified models analyzed in this chapter. We proposed 2 different myopic learning algorithms for the 2-player game and proved that under mild conditions, either of them converges to Nash equilibrium. For the N-player (for any integer $N \geq 2$) game with identical players, we characterized the entire set of possible equilibria and proved that at least one equilibrium always exists for any such game. The worst case equilibrium was identified. The price of anarchy was found to be an increasing function of number of competing airlines, ratio of fare to operating cost per seat and the curvature of S-curve relationship.

In this chapter we presented two central results. First, there are simple myopic learning rules under which less than perfectly rational players would converge to an equilibrium. This substantiates the predictive power of the Nash equilibrium concept.

Second, the S-curve relationship between market share and frequency share has direct and negative implications to airline profitability and airport congestion, as speculated in multiple previous studies.

Chapter 5

Administrative Mechanisms for Airport Congestion Mitigation

This chapter corresponds to the most significant practical contribution of this thesis. In the previous three chapters of this thesis, we evaluated the impacts of competition and congestion on the three major stakeholder groups. In this chapter, we propose some simple modifications to the existing administrative slot allocation mechanisms at congested airports and assess their impacts on these different stakeholders.

5.1 Introduction

As mentioned in chapter 1, demand-capacity mismatch is responsible for a large proportion of the National Aviation System (NAS) delays. These delays are disproportionately distributed across airports and metropolitan areas in the country. Congestion at a few major airports is responsible for a large proportion of overall delays. An analysis of air traffic patterns and delays by the Brookings Institution [96] suggests that almost 65% of the delayed flight arrivals are concentrated in the 25 largest metropolitan areas. Moreover, operations across an airline's network are interrelated due to linkages in aircraft, crew and passenger movements. Therefore, delays originating at these major airports propagate across the airline networks causing system-wide disruptive impacts. In the summer of 2007, according to the New

York Aviation Rulemaking Committee [68] report, three-quarters of the nationwide flight delays were generated from the air congestion surrounding New York. This suggests that mitigation of demand-capacity imbalance at a handful of congested airports should yield system-wide benefits in terms of delay alleviation.

5.1.1 Demand Management

As mentioned in Chapter 1, in short-to medium-time horizon, demand management strategies are the most promising approaches for alleviating airport congestion. Demand management strategies refer to any administrative or economic policies and regulations that restrict airport access to users. All the demand management strategies proposed in the literature and practiced in reality can be broadly categorized as administrative controls and market-based mechanisms, although various hybrid schemes have also been proposed. The demand management problem involves two types of decisions, namely, (1) slot determination, which involves deciding the total number of slots to be allocated, and (2) slot allocation, which involves the decision on distribution of these slots among the different users. These decisions can be taken either sequentially, such as in an auction or administrative mechanism, or simultaneously, such as in a congestion pricing mechanism.

Administrative Controls

As mentioned in Chapter 1, administrative controls are the most popular form of demand management strategies currently employed in practice. Five major US airports and several airports in Europe and Asia have been slot controlled for several years over the last few decades. However, one fundamental problem with the current administrative slot allocation procedures is that they are economically inefficient because they create barriers to entry by new carriers ([63]) and encourage airlines to over-schedule in order to avoid losing the slots [51]. Another problem, as pointed out by Ball et al. [8], is the implicit need to make a tradeoff between delays and resource utilization. Specifically, current approaches require ascertaining the *declared* capacity

of an airport beforehand even though the actual capacity on the day of operations is a function of prevalent weather conditions. Declaring too high a value for capacity poses the danger of large delays under bad weather conditions (Instrument Meteorological Conditions (IMC)) and declaring too low a value leads to wastage of resources under good weather conditions (Visual Meteorological Conditions (VMC)). Declared capacity, that is, the total number of allocated slots per time period, ultimately determines the congestion and delays at an airport.

Congestion Pricing

Congestion pricing and slot auctions are two of the most popular market mechanisms proposed in the literature. Classical studies such as Vickrey [97], Levine [62] and Carlin and Park [30] proposed congestion pricing based on the marginal cost of delays. Such pricing schemes, in theory, maximize the social welfare through optimal allocation of public resources. Under congestion pricing, the total cost to the user includes the delay cost as well as the congestion price. The notion of equilibrium congestion prices relies on the existence of a demand function, that is, an expression that gives the aggregate demand for airport resources as a function of total cost to the user. Some researchers, such as Morrison [65] and Daniel [40], performed numerical experiments under some specific assumptions about the underlying demand function, while others, like Carlin and Park [30], have acknowledged the problems in estimating demand as a function of congestion prices with any level of reliability because of lack of sufficient data.

Beyond the unavailability of data, however, there is an even more basic issue associated with accurate demand estimation. Under congestion pricing, the aggregate demand for slots at an airport is the sum of the number of slots demanded by each airline. Assuming profit maximizing airlines, the number of slots demanded by an airline can be obtained by equating the incremental profitability of the last slot to the congestion price per slot. In reality, among other factors, the profitability of an airline depends on its own schedule as well as on competitor schedules. It is easy to see that the incremental profitability of having an extra flight in a particular market largely

depends on the number of additional passengers that the airline will be able to carry because of the additional flight, which in turn depends on the schedule of flights offered by the competitor airlines in the same market. So given a set of congestion prices, the total demand for slots should reflect these competitive interactions. Some recent congestion pricing studies by transportation economists such as Pels and Verhoef [80] and Brueckner [25], have modeled competitive effects through Cournot-type [38] models of firm competition. However, these models do not incorporate the inverse dependence of one airline's market share on competitor airlines' frequencies, which is a critical component of such competitive interactions.

Slot Auctions

Previous research has shown the efficiency benefits of conducting an auctioning of airport slots [39, 9, 63, 51]. An auction by itself does not, however, alleviate airport congestion, but rather allocates a fixed set of resources in a more efficient way. So, to that extent, auctions are similar to administrative controls, as they too pose an implicit need to make a tradeoff between delays and resource utilization.

Once the number of slots to be allocated is determined through some procedure, slot auctions, in theory, should maximize the social welfare by allocating the slots to those who value them the most. But the determination of the actual value of a package of slots to an airline is a complicated problem. Harsha (2008) proposed a valuation model for estimating the value of a package of slots. However, the formulation does not capture any effects of airline competition.

In summary, in an auction or administrative mechanism, slot allocation must be explicitly preceded by some process for slot determination. It is this previous step that primarily determines the congestion level. Existing literature has typically focused on the second step and the first step has not received much attention. Furthermore, much of the discussion of the second step excludes any effects of frequency competition. Although congestion pricing tackles both these decisions simultaneously and hence implicitly handles the slot determination step, existing literature on congestion pricing does not capture important elements of frequency competition.

In this chapter, we propose a framework for assessing different administrative slot allocation schemes while explicitly modeling the effects of frequency competition. In our first experiment, we evaluate the impacts of slot determination step in terms of airline profits and passengers carried by varying the total number of allocated slots. In our second experiment, for a fixed number of total slots, we focus on the problem of slot allocation and evaluate the impacts of two different simple strategies for slot allocation on the various stakeholders. In Chapter 6, we evaluate the effectiveness of congestion pricing schemes under airline frequency competition.

5.1.2 Literature Review

The existing body of literature on airline frequency competition can be categorized into three broad groups- econometric, theoretical and computational studies. Studies by transportation economists such as Brander and Zhang [22], Aguirregabiria and Ho [4], Norman and Strandenes [67] etc. employ econometric methods to estimate the parameters in the airline competition models using large datasets and use the calibrated models for gaining critical insights into the competitive behavior of the airlines and for answering policy-related broad questions. These studies do not deal with the issues of existence, computation and empirical validation of the equilibrium predictions. Theoretical studies including Brueckner [25], Brueckner and Flores-Fillol [28], Hendricks, Piccione and Tan [52], Pels, Nijkamp, and Rietveld [79], Hong and Harker [53] etc. investigate analytically solvable game theoretic models of airline competition and derive theoretical results that provide insights into important characteristics of equilibria and the comparative statics. These studies do not deal with real datasets. Computational studies such as Hansen [50], Wei and Hansen [99], [43], Adler ([1], [2]) etc. employ mathematical models and solution algorithms for obtaining Nash equilibria of airline competition games. Our research in this chapter falls within this third category.

Dobson and Lederer [43] model schedule and fare competition as a strategic form game for a sample problem comprising six airports and two airlines. Adler [1] models airline competition on fare, frequency and aircraft sizes as an extensive form game and

presents equilibrium results for a network comprising four airports and two airlines. Subsequently, Adler [2] considers the decisions on hub locations and decisions about fares, frequencies and aircraft sizes in a two-stage extensive form game framework for a reasonably sized-problem consisting of three airlines with two hubs for each airline. None of these studies provides any empirical justification of suitability of Nash equilibrium outcome. Hansen [50] analyzes frequency competition in a hub-dominated environment using a strategic form game model and presents results for a large network of realistic size involving multiple airlines. This study reports significant disparities between model predictions and the state of the actual system. Each of these four studies adopts a successive optimization approach to solve for a Nash equilibrium. In this chapter, we also use a successive optimizations approach for the computation of a Nash equilibrium. We assess the impact of starting point on the equilibrium being reached. We also provide empirical validation of our equilibrium predictions.

Furthermore, in most of the previous research, scheduling decisions on one segment are not constrained by the schedule on other segments. (We define a segment as an origin and destination pair for non-stop flights.) This is a good approximation for a situation where an airport is not congested, and takeoff and landing slots are freely available. But some congested US airports and several major airports in Europe and Asia are slot constrained. With projected passenger demand in the US expected to outpace the development of new airport capacity, there is a possibility of many more airports in the US employing some form of demand management in the future. At a slot constrained airport, increasing the frequency of flights on one segment usually requires the airline to decrease the frequency on some other segment from that airport. To the best of the author's knowledge, no previous study has incorporated slot constraints into airline competition models.

5.1.3 Contributions

The main contributions of this chapter fall into four categories. First, we propose a game-theoretic model of frequency competition as an evaluation methodology for slot

allocation schemes. Second, we provide a solution algorithm with good computational performance for solving the problem to equilibrium. Third, we provide justification of the credibility of the Nash equilibrium solution concept in two different ways, through empirical validation of the model outcome and through convergence properties of the learning dynamics for non-equilibrium situations. Finally, under simple slot allocation schemes, we evaluate system performance from the perspectives of the passengers and the competing airlines, and provide insights to guide the demand management policy decisions.

In the Chapter 2, we solved a large-scale mixed integer optimization problem to obtain delay-minimizing schedules for the air transportation network of the entire United States. We concluded that effective administrative and/or market-based mechanisms for slot control have the potential to reduce delays while satisfying all passenger demand given the available airport capacity. Le [61] showed that the delays at congested airports such as LaGuardia airport at New York are caused in large part due to the inefficient slot controls. Instead of modeling airline competition, this study assumed a hypothetical "single benevolent airline" and proved the existence of profitable flight schedules at LGA that can accommodate the passenger demand while reducing flight delays substantially. Our conclusions in this chapter confirm the findings of Le [61] previous studies. In this chapter, we explicitly model airline frequency competition and propose tangible mechanisms for achieving profitable schedules that accommodate passenger demand and significantly reduce delays.

Market-based mechanisms lead to socially efficient resource allocation. But the problems such as calculating the equilibrium congestion prices or designing an efficient auction are computationally challenging, even without considering any competitive interactions among the carriers. Therefore, we approach the problem in a different way. We do not try to integrate schedule competition into the slot allocations problem. Instead, given a slot allocation, we provide a framework for predicting the airline schedules and estimating the impact on passengers and competing airlines.

The airline planning process involves a large number of decision variables. Considerations such as network effects and demand uncertainty introduce further complica-

tions in the process. More tactical decisions such as pricing and revenue management often interact with these planning decisions and hence should be considered in evaluating an airline's response to any slot allocation scheme. Therefore, any tractable mathematical model of airline decisions involves substantial simplifications and approximations of reality. In this chapter, we present the models of airline competition along with a brief discussion of the underlying assumptions and the extent of their validity. After presenting the numerical results, we analyze and estimate the direction and magnitude of the impacts of the main assumptions on the results. In section 5.2, we provide details of the game-theoretic model of frequency competition under slot constraints. In section 5.3, we describe an efficient algorithm for equilibrium computation. In section 5.4, we provide empirical and learning-based justifications of the Nash equilibrium outcome. Finally, in section 5.5, we consider two different slot allocation schemes and evaluate their performance based on multiple criteria. In section 5.6, we conclude with a summary and discussion of the main results.

5.2 Model

In this section, we describe the relevant notation and formulate the model. In Subsections 5.2.1 and 5.2.1, we present two important extensions to this model.

We will first formulate the frequency planning problem as an optimization problem from a single airline's point of view. Let us consider an airline a . Consider an airport which is slot constrained, that is, the number of flights arriving at and departing from that airport is restricted by slot availability. A slot available to an airline can be used for a flight to or from any other airport, but the total number of slots available to each airline is limited. In this model, we will consider only the flight departures from a slot constrained airport and assume that the airports at the other end are not slot constrained. This assumption is quite reasonable in the US context, where only a handful of airports are slot constrained. The timing of a slot is also an important aspect of its attractiveness from an airline's point of view. In our model, we focus only on the daily allocation of slots while ignoring the time-of-the-day aspects.

We will calculate airlines' operating profits under the full fare assumption, in which it is assumed that the entire fare of a connecting passenger contributes to the operating profits of each of the segments in the passenger's itinerary. In Sub-section 5.5.5, we will analyze the impact of alternate profit calculation methods on our results. To begin with, we will consider frequency planning decisions while assuming that the aircraft sizes remain constant for each segment. We will analyze the impact of this assumption in Sub-section 5.5.3. We propose a multi-player model of frequency competition where each airline's decision problem is represented as an optimization problem. From here onwards, this model will be referred to as the basic model. In this basic model, the only decision variables are the numbers of non-stop flights of airline a on each segment with destination at the slot constrained airport. This *basic model* is applicable for situations where the fares and other factors are similar among the competing airlines and the main differentiating factor between different airlines is the service frequency. We will relax this assumption in model extension 1 proposed in Sub-section 5.2.1.

Let \mathbb{S}_a be the set of potential segments with destination at the slot constrained airport. Let p_{as} be the average fare charged by airline a on segment s . Let Q_{as} be the number of passengers carried by airline a on segment s . In general, a passenger might travel on more than one segment to go from his origin to destination, which in some cases involves connecting between flights at an intermediate airport. However, we will assume segment-based demand, that is, a passenger traveling on two different segments will be considered as a part of the demand on each segment. This assumption is quite reasonable for the airports in New York City area where nearly 75% of the passengers are non-stop [73], but not very accurate for major transfer hubs such as the Chicago O'Hare airport. We will analyze the extent of impact of this assumption in Sub-section 5.5.5. Let the total passenger demand on segment s be M_s . C_{as} is the operating cost per flight for airline a on segment s . S_{as} is the seating capacity of each flight of airline a on segment s . Let α_s be the exponent in the S-curve relationship between the market share and the frequency share on the non-stop segment s . The value of α_s depends on the market's characteristics such as long-haul/short-haul,

proportion of business/leisure passengers, etc. In short-haul markets and in markets dominated by business passengers, the value of α_s is expected to be higher and in long-haul markets and in markets dominated by leisure passengers, the value of α_s is expected to be lower.

The vector of decision variables for airline a is $[f_{as}]_{s \in \mathbb{S}_a}$. Because the destination airport is slot constrained, the maximum number of flights that can be scheduled by airline a is restricted to U_a . Often, under the current set of administrative policies based on use-it-or-lose-it type rules, there are restrictions on the minimum number of slots that must be utilized by an airline in order to avoid losing slots for the next year. So there may be a lower limit on the number of slots that must be used. Let L_a be the minimum number of slots that must be utilized by airline a . Let \mathbb{A} be the set of all airlines and let \mathbb{A}_s be the set of airlines whose set of potential segments include segment s .

As defined by the S-curve relationship, the market share of airline a on non-stop segment s equals $\frac{f_{as}^{\alpha_s}}{\sum_{a' \in \mathbb{A}_s} f_{a's}^{\alpha_s}}$, which provides an upper bound on the number of passengers for a specific carrier on a specific segment. This restriction is imposed by constraint 5.2 in the model that follows. Obviously, the number of passengers on a segment cannot exceed the number of seats. Moreover, due to demand uncertainty and due to the effects of revenue management, the airlines are rarely able to sell all the seats on an aircraft. Assuming a maximum average segment load factor of LF_{max} , the seating capacity restriction is modeled by constraint 5.3. We present results assuming 85% as the maximum average segment load factor value. We test the sensitivity of the results to variations in this value in Sub-section 5.5.1. The objective function 5.1 to be maximized is the total operating profit, which is total fare revenue minus total flight operating cost. We have assumed average fares and deterministic demand. We will analyze the impact of these two assumptions in Sub-section 5.5.4. The overall optimization model is as follows,

$$\text{maximize } \sum_{s \in \mathbb{S}_a} p_{as} Q_{as} - C_{as} f_{as} \quad (5.1)$$

$$\text{subject to: } Q_{as} \leq \frac{f_{as}^{\alpha_s}}{\sum_{a' \in \mathbb{A}_s} f_{a's}^{\alpha_s}} M_s \forall s \in \mathbb{S}_a \quad (5.2)$$

$$Q_{as} \leq LF_{max} S_{as} f_{as} \forall s \in \mathbb{S}_a \quad (5.3)$$

$$\sum_{s \in \mathbb{S}_a} f_{as} \leq U_a \quad (5.4)$$

$$\sum_{s \in \mathbb{S}_a} f_{as} \geq L_a \quad (5.5)$$

$$f_{as} \in \mathbb{Z}^+ \forall s \in \mathbb{S}_a \quad (5.6)$$

The market share available to each airline depends on the frequency of other competing airlines in the same market, which in turn are decision variables of those other airlines. Therefore, this is multi-agent model. The optimization problem given by 5.1 through 5.6 can only be solved for a given set of values of competitors' frequencies.

We now propose two extensions to the *basic model*. The first extension is applicable to segments where the competing carriers differ in terms of fare charged or in some other important way. The second extension is applicable to segments on which only one carrier operates non-stop flights.

5.2.1 Model Extension 1: Fare Differentiation

The *basic model* assumes that the market share on each segment depends solely on the frequency share on that segment. This assumption is reasonable in many markets where the competitor fares are very close to each other and the competing airlines are similar from the perspectives of the passengers in most other ways. However, for markets where the fares are different, the basic S-curve relationship can be a poor approximation of actual market shares. Consider a market where the competing airlines are differentiated in both fare and frequency. Different types of the passengers would react differently to these attributes. While some passengers value lower fares

more, others give more importance to higher frequency and the associated greater flexibility in scheduling their travel. In addition, there could be other airline-specific factors that impact the passenger share. For example, some passengers might have a preference for the big legacy carriers operating wide-body or narrow-body fleets over the regional carriers operating turbo-prop aircraft or small regional jets. To incorporate these effects, we propose an extension of inequality 5.2. Let there be T types of passengers. Let γ_s^t be the fraction of segment s passengers belonging to type t such that $\sum_{t=1}^T \gamma_s^t = 1$. Let α_s^t be the frequency exponent corresponding to the type t passengers for segment s , which serves the same purpose as the exponent α_s of the S-curve in the *basic model*. Let β_s^t be the fare exponent corresponding to the type t passengers for segment s . Obviously, we expect α_s^t to be non-negative and β_s^t to be non-positive. Let θ_a be the airline-specific factor for airline a . Inequality 5.2 can then be extended as,

$$Q_{as} \leq \sum_{t=1}^T \frac{\theta_a f_{as}^{\alpha_s^t} p_{as}^{\beta_s^t}}{\sum_{a' \in \mathbb{A}_s} \theta_{a'} f_{a's}^{\alpha_s^t} p_{a's}^{\beta_s^t}} \gamma_s^t M_s \quad (5.7)$$

The market share of each airline is now a function of the fares, frequencies, and airline specific factors of all competing airlines. This model incorporates the effects of different fares and frequencies on the passenger shares. Also, it can model multiple passenger types such as leisure vs. business, by specifying different exponents for fare and frequency for different types of passengers. Finally, the remaining airline specific factors are captured through the θ_a parameter. The magnitudes of α_s^t values are expected to be high in short-haul markets and in markets dominated by business passengers because the business passengers tend to give particularly high importance to more frequent flights. The magnitudes of β_s^t values are expected to be high in markets dominated by leisure passengers because leisure passengers are more sensitive to fares. The γ_s^t values depend on the business/leisure composition of specific markets, e.g. destinations such as Orlando, FL and Miami, FL are expected to have a higher fraction of leisure passengers than destinations such as Washington, DC and Boston,

MA. The θ_a values can be expected to be higher for airlines which have a bigger brand name and a better track record.

5.2.2 Model Extension 2: Market Entry Deterrence

This second model is similar to the *basic model* except that the player decisions are now sequential rather than simultaneous. The idea of modeling the frequency competition as an extensive form game was proposed by Wei and Hansen [99] where, for contractual or historical reasons, one airline has the privilege of moving first, i.e., deciding the frequency on a segment. The other airline responds upon observing the action by the first player. The *basic model* and the first extension implicitly assumed the existence of at least two competing airlines on a segment. However, frequency decisions in markets with only one existing airline are not completely immune to competition and the incumbent airline must consider of possibility of entry by another competitor while deciding the optimal frequency. Such situations can be modeled using the idea of Stackelberg equilibrium ([90]) or a subgame perfect Nash equilibrium of an extensive form game. In this situation, the incumbent carrier is the Stackelberg leader and the potential entrant is the follower. A potential entrant (a') is assumed to be a rational player. Inequality 5.2 can be extended as,

$$Q_{as} \leq \frac{f_{as}^{\alpha_s}}{f_{as}^{\alpha_s} + f_{a's}^{\alpha_s}} M_s \quad (5.8)$$

$$f_{a's} = \arg \max_{f \in \mathbb{Z}^+} \left(\min \left(\frac{f^{\alpha_s}}{f_{as}^{\alpha_s} + f^{\alpha_s}} M_s, LF_{max} S_{a's} f \right) p_{a's} - C_{a's} f \right) \quad (5.9)$$

5.3 Solution Algorithm

We use the Nash equilibrium solution concept to predict the outcome of this airline frequency competition game. In this section, we describe the solution algorithm used for solving this problem. In section 5.4, we will provide justification for using the Nash equilibrium outcome.

The objective function for each airline is the sum of profits on each segment and the frequencies of an airline on different segments are interrelated through the constraints on the minimum and maximum number of slots. The effect of competitors' frequencies on the profitability of an airline, as described by the *basic model*, can be fully captured through the notion of effective competitor frequency. Let us define the effective competitor frequency for airline a on segment s as $f_{as}^{eff} = \left(\sum_{a' \in \mathbb{A}_s, a' \neq a} f_{a's}^{\alpha_s} \right)^{\frac{1}{\alpha_s}}$. So constraint 5.2 in the *basic model* can be more succinctly expressed as $Q_{as} \leq \frac{f_{as}^{\alpha_s}}{f_{as}^{\alpha_s} + f_{as}^{eff \alpha_s}} M_s \forall s \in \mathbb{S}_a$. In a two-airline market, f_{as}^{eff} for either airline is nothing but the frequency of the other airline in that market. In case of markets with three or more airlines, if there is a dominant competitor, then f_{as}^{eff} tends to be slightly higher than the frequency of the dominant competitor. In such cases, f_{as}^{eff} is highly dependent on the frequency of the dominant competitor. For example, in a market with three competitors with frequencies of 10, 2 and 2 respectively, f_{as}^{eff} equals 11.2 (assuming $\alpha_s = 1.5$). If the frequency of one of the marginal competitors increases from 2 to 3, f_{as}^{eff} changes from 11.2 to 11.6. But if the frequency of the dominant competitor changes from 10 to 11, then the f_{as}^{eff} value changes from 11.2 to 12.1. Furthermore, the dependence of f_{as}^{eff} on the dominant competitor's frequency increases with increasing α_s value. On the other hand, in balanced competitive markets, the dependence of f_{as}^{eff} on the frequencies of all the competitors is comparable.

Figure 5-1 shows the typical form of the segment profit function under the *basic model* for a fixed value of effective competitor frequency, ignoring slot constraints and integrality constraints. Under the same assumptions, Figure 5-2 shows the typical shape of the optimal segment frequency (best response) as a function of effective competitor frequency. The profit function and the best response function get further complicated by slot constraints, integrality constraints, and extensions 1 and 2 to the *basic model*. The optimization problem has discrete variables, and as visible from Figure 5-1, its continuous relaxation is non-convex. In addition, optimal decisions for each airline depend on the frequency decisions by other airlines. Therefore, the problem of computing an outcome of this multi-agent model can be very challenging. The strategy space for a typical problem size for a major airport is very large with the

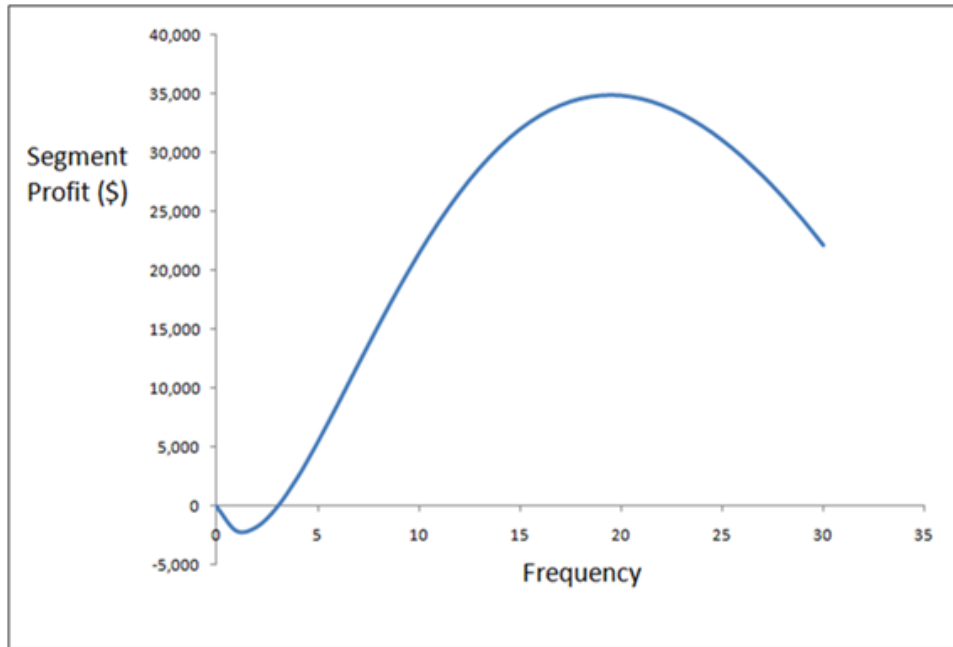


Figure 5-1: Typical shape of the segment profit function

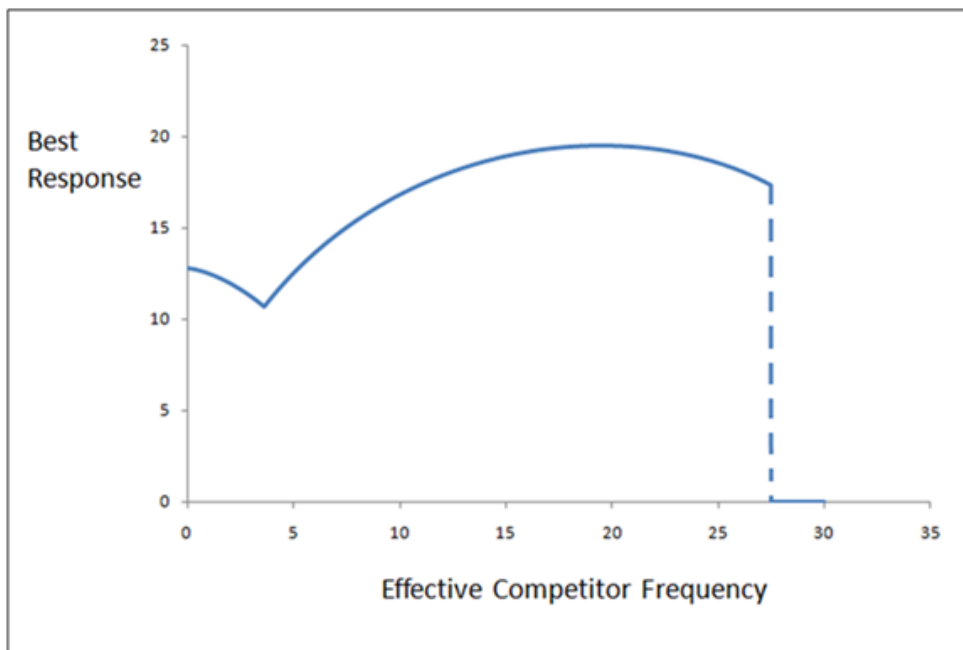


Figure 5-2: Typical shape of the best response function

number of potential candidates for equilibrium solutions being of the order of 10^{50} . To solve this problem, we propose a heuristic based on the idea of myopic best response, which employs successive optimizations, and individual optimization problems are solved to full optimality using a dynamic programming-based technique. In Sub-section 5.3.1, we describe the myopic best response algorithm and in Sub-section 5.3.2 we describe the dynamic programming formulation for individual optimizations.

5.3.1 Myopic Best Response Algorithm

Let $f_a = [f_{as}]_{s \in \mathbb{S}_a}$ be the vector of frequencies for carrier a . Let $f_{-a} = [f_{a'}]_{a' \in \mathbb{A}, a' \neq a}$ be the vector formed by concatenating the frequency vectors of all competitors of airline a . So any outcome of this problem can be compactly denoted as $f = (f_a, f_{-a})$. Then the myopic best response algorithm (a heuristic) is described as follows,

```

while there exists a carrier  $a$  for whom  $f_a$  is not a best response to  $f_{-a}$  do
     $f'_a \leftarrow$  some best response by  $a$  to  $f_{-a}$ 
     $f \leftarrow (f'_a, f_a)$ 
return

```

This heuristic is based on the idea of myopic best response. Some classes of games have certain desirable properties which make them solvable to equilibrium using an algorithm where each player successively optimizes his own decisions while assuming that the decisions of other players remain constant. Obviously, if such a heuristic converges to some outcome, then it must be a Nash equilibrium. In general, there is no guarantee that it will converge. Further, even if such an algorithm converges to some Nash equilibrium, there is no guarantee that the equilibrium will be unique. We discuss issues regarding its convergence, and the existence and uniqueness of equilibrium for the game model under consideration, in Sub-section 5.4.3.

5.3.2 Dynamic Programming Formulation

The main building block of the myopic best response algorithm is the calculation of an optimal response of airline a to the competitors' frequencies. Given the frequencies of all the competing carriers on all the segments, the problem of calculating a best response is an optimization problem. This problem can have a large solution space. For typical problem sizes, the number of discrete solutions in the solution space can be of the order of 10^{10} . As mentioned earlier, this problem is non-convex and discrete. However, this problem has a nice structure. Slot restrictions are the only coupling constraints across different segments and the objective function is additive across segments. Therefore, the problem structure is amenable to solution using dynamic programming.

Let $\Pi_s(n)$ denote the profit from operating n flights on segment s . We order the segments arbitrarily and number them from 1 to $|\mathbb{S}_a|$. Segments are considered in order and each segment corresponds to a stage. Each state, (S, n) corresponds to the combination of the last segment being considered, s , and the cumulative number of flights, n , operated on all the segments considered before and including the last segment being considered. Let $R(S, n)$ be the maximum profit that can be obtained from operating a total of n flights on the first s segments. We initialize $R(0, 0) = 0$ and $R(0, n) = -\infty$ for $n \geq 1$. For $s \geq 1$,

$$R(s, n) = \max_{0 \leq n' \leq n} (R(s-1, n') + \Pi_s(n - n')).$$

The optimal value of total profit for airline a is given by,

$$\max_{L_a \leq n \leq U_a} R(|\mathbb{S}_a|, n).$$

5.4 Validity of Nash Equilibrium Outcome

Similar to our work, almost all the previous studies on airline competition have used the concept of Nash equilibrium (or one of its refinements) for predicting the outcome of a competitive situation. The traditional explanation for Nash equilibrium is that it results from introspection and detailed analysis by the players assuming that the rules of the game, the rationality of the players, and the profit functions of players are all common knowledge. A Nash equilibrium outcome is attractive mainly because of the fact that unilateral deviation by any of the players does not yield any additional benefit to that player. So given an equilibrium outcome, the players do not have any incentive to deviate from the equilibrium decisions. However, in the absence of any apriori knowledge of an equilibrium outcome, given complicated profit functions such as the ones in this case, it isn't immediately clear that airlines would take the equilibrium decisions. In this section, we substantiate the predictive power of the equilibrium outcome using two different approaches in Sub-sections 5.4.2 and 5.4.3 respectively.

5.4.1 Data Sources and Implementation Details

All the numerical results presented in sections 5.4 and 5.5 correspond to LGA airport as the slot controlled airport. For reasons of computational tractability we decided to restrict our analysis to all the segments of all airlines with destination at the LGA airport. LGA is one of the most congested airports in the US. Furthermore, a very high proportion of non-stop passengers on segments to LGA airport makes it comparatively easier to separate the airlines' decisions at LGA from the rest of the network. We discuss the impacts of passenger connections and network effects on our results in more detail in Sub-section 5.5.5.

Flight schedules for the US domestic segments are available on the Bureau of Transportation Statistics (BTS) website [72] for each certified US carrier with at least 1% of total domestic passenger revenue. The data on flight frequencies, aircraft sizes and segment passengers are obtained from the T-100 Segment Database [75].

Average fare values for each market are obtained from the Airline Origin Destination Survey database [73]. Aircraft operating costs for each aircraft type for each carrier are obtained from the Schedule P-5.2 information [74]. Public data on segment passengers and operating expenses is available on a monthly aggregate level, while data on average fares is only available on a quarterly aggregate level. Unfortunately, more disaggregate values of these entities, such as on a daily level, are not available publicly. Also the daily values often tend to fluctuate due to various types of cyclical variations. In order to avoid biases in our model estimates because of choice of certain days over others, and also to circumvent the data unavailability issue, we ran our experiments on quarterly average values. All results in sections 5.4 (except Sub-section 5.4.4) and 5.5 are for the first quarter of 2008. In Sub-section 5.4.4, we verify the robustness of our model’s fit to reality by running our model for the 2nd, 3rd and 4th quarters of 2008.

In order to estimate the flight delay reduction for experiment 2 presented in section 5.5, we use realized values of airport capacity for an entire year, which were made available by Metron Aviation®[®], and actual flight delay data obtained from the Airline On-time Performance Database available on the BTS website [72]. Details of the delay reduction estimation procedure are described in Sub-section 5.5.1.

Our dataset consists of all segment-carrier combinations with destination at LGA operating at least one flight per day on average. Thus our dataset encompasses 96.11% of all the flights destined for LGA. For all segments where only one carrier provides non-stop service, we use the market share function given by model extension 2. We use the market share function given by model extension 1 for segments on which: 1) the competitors’ average fares differ by more than 5%; and/or 2) one or more major carriers operating a narrow- or a wide-body fleet compete against one or more regional carriers operating small jets. For all the other segments, we use the market share function given by inequality 5.2 in the *basic model*.

We conducted several test runs to choose the parameters such that the frequency estimates given by the model match the actual frequency values closely. For the *basic model*, we used different α_s values for different markets. In general, flights

into LGA tend to be short haul flights. In our dataset, there was not a single flight from the west-coast airports. The range of α_s values that we used varied from 1.5 for very short-haul markets to 1.2 for the comparatively long-haul markets. For model extension 1, we considered 2 types of passengers: 1) business passengers, and 2) non-business passengers, i.e. $T = 2$. The value of airline specific factor (θ_a) was taken to be 0.3 for all regional carriers operating turbo-props or small regional jets, and 1.0 for all other carriers. The fraction of passengers belonging to type 1 (business passengers) was taken to be $\gamma_s^1 = 0.3$, and hence, $\gamma_s^2 = 0.7$, for all markets. Also, we used the following values for the exponents in model extension 1: $\alpha_s^1 = 1.3$, $\beta_s^1 = -0.5$, $\alpha_s^2 = 0.3$, and $\beta_s^2 = -1.2$. Given that in most cases, the average fares of competing airlines on each segment into LGA were very close to each other, we assumed the average fare ($f_{a's}$) of the potential entrant in model extension 2 to be the same as the fare charged by the existing operator on that segment. The seating capacity ($S_{a's}$) and the operating cost ($C_{a's}$) of the potential entrant were taken to be those corresponding to the most profitable combination (decided by the minimum ratio of operating cost to seating capacity) available across all the fleet types operated by all airlines into LGA.

Because we ascertained the values of the model parameters using a heuristic process, it is very important to investigate the sensitivity of our results to changes in these parameter values. The results of the sensitivity analyses to various model parameter values are presented in Sub-section 5.5.2. Additionally, in order to avoid over-fitting the model parameters to a certain dataset, we used the same model parameters to compare the error between the model's frequency predictions and the actual frequency values for the 2nd, 3rd, and 4th quarters of 2008. These results are presented in Sub-section 5.4.4.

5.4.2 Empirical Validation

To validate our model against actual frequency data, we compared the equilibrium frequencies predicted by the model against the actual values. At LGA, the maximum number of slots for each airline is restricted and each airline usually wants to make

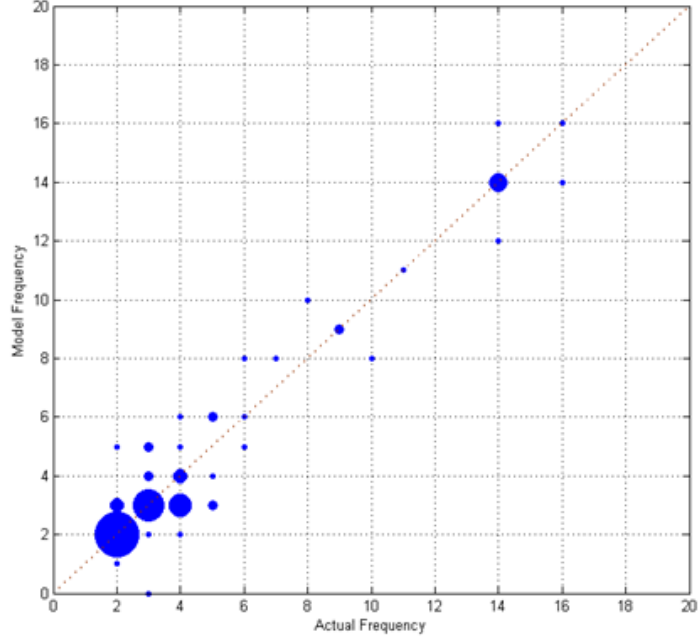


Figure 5-3: Empirical validation of frequency predictions

use of all the slots available to it in order to avoid losing any slots in subsequent seasons. The minimum and maximum numbers of slots available to an airline, that is, L_a and U_a , are assumed to be equal. So the total number of slots allocated to each airline is fixed. The airline needs only to decide the number of slots to allocate to flights to each of its destinations. Let f_{as} be the actual frequency of airline a on segment s and \hat{f}_{as} be the equilibrium frequency as predicted by the model. The model ensures that the total frequency for each airline remains constant. Therefore, when the model overestimates the frequency on one segment it necessarily underestimates the frequency on some other segment corresponding to the same carrier. In order to measure the model fit to reality, we will use Mean Absolute Percentage Error (MAPE) defined as,

$$MAPE = \frac{\sum_{a \in \mathbb{A}} \sum_{s \in \mathbb{S}_a} |\hat{f}_{as} - f_{as}|}{\sum_{a \in \mathbb{A}} \sum_{s \in \mathbb{S}_a} f_{as}}.$$

Figure 5-3 compares the actual frequency and the frequency predicted by the model for each carrier from each origin. The x-axis denotes the actual frequency and

the y-axis denotes the frequency as predicted by the model. The weight of each point in this figure indicates the number of observations corresponding to that point. As shown in the Figure 5-3, most of the observations are on or very close to the 45° line. The overall MAPE was found to be 14.72%. The model predictions thus match actual frequencies reasonably well.

5.4.3 Game Dynamics

Airlines typically operate flights on similar sets of segments year after year. The group of competitors on each segment and the general properties of markets stay constant over long periods of time. Therefore, the airlines have opportunities to adapt their decisions primarily by fine-tuning the frequency values to optimize their profits. Such adjustments can be captured by modeling the dynamics of the game. A simplified version of the frequency competition model used in this chapter was used in Chapter 4 and the convergence of best response dynamics was proved in the two-player case without slot constraints. The key factor responsible for convergence of the myopic best response algorithm was the flat shape of the best response function near equilibrium. In other words, the magnitude of the derivative of the optimal frequency with respect to the effective competitor frequency is very small. Therefore, the best response for a large range of competitor frequency values is very close to the equilibrium frequency, resulting in strong convergence properties of the best response dynamics. The *basic model* of frequency competition used in this chapter is the same as the model presented in chapter 4, except for the addition of slot constraints and integrality constraints. Though the convergence results proved in chapter 4 are not directly applicable to this complicated model, they provide some intuition.

In this chapter, we have used the best response algorithm for computation of an equilibrium. For the results of empirical validation presented in the previous subsection, we used the vector of actual frequency values as the starting point of the best response algorithm and the algorithm converged to an equilibrium in just 2 iterations per player (per airline). Let us term this equilibrium solution as the *base equilibrium*. In this section, we present the impact of variation in the starting point

on the computed equilibrium prediction.

Table 5.1: Stability of algorithm results to starting point perturbations

Maximum Perturbation	MAPE with respect to Actual Frequencies	MAPE with respect to <i>Base Equilibrium</i>
0%	14.72%	0.00%
10%	14.72%	0.00%
20%	14.78%	0.06%
30%	15.04%	0.34%
40%	15.25%	0.56%
50%	15.65%	0.96%
60%	15.81%	1.72%
70%	16.14%	2.50%
80%	16.22%	2.95%
90%	16.43%	3.59%
100%	16.76%	4.27%

For each starting point, the algorithm was run for at most 10 iterations per player. In most of the following cases, the algorithm converged to an equilibrium and terminated in fewer than 10 iterations. However, in the few cases that the algorithm did not converge within 10 iterations, it was terminated after 10 iterations. Starting from the actual frequency values, we perturbed each dimension of the frequency vector uniformly between $-x\%$ to $+x\%$ of the original value. For each x value, we drew 1000 samples of starting points randomly from this uniform distribution. Values presented in Table 5.1 are the average MAPE values across the 1000 runs obtained by comparing the solution computed by the best response algorithm to the actual frequencies as well as to the *base equilibrium*. These results indicate that the model predictions are quite insensitive even to large perturbations in the starting point. The algorithm converges to, or comes very close to, the equilibrium solution within very few iterations, irrespective of the starting point. This also suggests that the best response dynamics displays good convergence properties. Therefore, even assuming less than perfectly rational players, an equilibrium outcome can be reached through a simple myopic learning procedure. Empirical validation results, coupled with these desirable convergence properties, make a strong case for using the Nash equilibrium solution

concept for predicting the outcome of this airline frequency competition game.

5.4.4 Robustness Verification

In Sub-sections 5.4.2 and 5.4.3, we presented empirical results using data from the 1^{textst} quarter of 2008. In this Sub-section, we will verify the robustness of our models by validating them against empirical data from the 2^{textnd}, 3^{textrd}, and 4^{teatth} quarters of 2008.

Table 5.2 presents the discrepancy between the frequency values predicted by the model and the actual frequency decisions taken by the airlines using the MAPE measure of error for each quarter of 2008. Table 5.2 shows that the error in frequency predictions is very stable across different time periods. Table 5.3 presents the stability properties of the algorithm across the four quarters of 2008. We find that the performance is stable across different quarters, which means that the model predictions are robust to large perturbations to the starting point for different time periods.

Table 5.2: Model prediction error across different quarters

Quarter	Mean Absolute Percentage Error
1	14.72%
2	16.12%
3	15.67%
4	14.38%

Table 5.4 presents the MAPE with respect to actual frequencies as the values of key model parameters vary between -25% to +25% of the values listed in Sub-section 5.4.1. The first row lists the changes in MAPE with different percentage changes in α_s values for *all* segments. The next six rows list the variation in MAPE with changes in parameters of model extension 1. The remaining three rows list the variation in MAPE with operating cost, seating capacity and average fare of the potential entrant in model extension 2. As shown in Table 5.4, the MAPE values vary between 13.5% and 19.6% and are reasonably stable to significant variations in model parameters. Additionally, in Sub-section 5.5.2 we present results on sensitivity of the impacts of

Table 5.3: Stability of algorithm results to starting point perturbations across different quarters

Maximum Perturbation	MAPE with respect to Actual Frequencies				MAPE with respect to <i>Base Equilibrium</i>			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
0%	14.72%	16.12%	15.67%	14.38%	0.00%	0.00%	0.00%	0.00%
10%	14.72%	16.12%	15.67%	14.38%	0.00%	0.00%	0.00%	0.00%
20%	14.78%	16.12%	16.05%	14.45%	0.06%	0.20%	0.73%	0.13%
30%	15.04%	16.12%	16.63%	14.57%	0.34%	0.22%	1.90%	0.40%
40%	15.25%	16.12%	17.07%	14.99%	0.56%	0.23%	2.71%	1.28%
50%	15.65%	16.12%	17.64%	15.35%	0.96%	0.34%	3.91%	2.06%
60%	15.81%	16.19%	18.05%	16.32%	1.72%	0.80%	4.91%	4.99%
70%	16.14%	16.28%	18.53%	16.85%	2.50%	1.53%	6.04%	6.55%
80%	16.22%	16.36%	18.84%	17.23%	2.95%	2.29%	6.71%	7.51%
90%	16.43%	16.72%	19.04%	17.88%	3.59%	3.40%	7.44%	8.77%
100%	16.76%	16.90%	19.47%	18.32%	4.27%	4.05%	8.33%	9.58%

slot reduction to variations in these parameters.

Table 5.4: Sensitivity of prediction accuracy (in MAPE) to model parameters

Parameter	-25%	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
α_s	17.2%	16.0%	16.0%	15.3%	15.3%	14.7%	17.2%	17.8%	19.0%	19.0%	19.6%
θ_a	17.8%	17.8%	16.6%	15.3%	15.3%	14.7%	17.8%	17.8%	17.8%	17.8%	18.4%
α_s^1	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	15.3%	15.3%	17.8%	17.8%	16.6%
β_s^1	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	15.3%	15.3%	15.3%	15.3%	15.3%
γ_s^1	15.3%	15.3%	15.3%	15.3%	14.7%	14.7%	15.3%	16.0%	17.8%	17.8%	17.8%
α_s^2	15.3%	15.3%	15.3%	15.3%	15.3%	14.7%	15.3%	15.3%	18.4%	18.4%	16.0%
β_s^2	17.8%	17.8%	17.8%	17.8%	14.7%	14.7%	15.3%	15.3%	16.0%	16.0%	16.0%
$C_{a's}$	19.6%	16.6%	17.8%	14.7%	13.5%	14.7%	14.1%	13.5%	14.1%	14.1%	14.1%
$S_{a's}$	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%	14.7%
$p_{a's}$	14.1%	14.1%	14.1%	14.1%	14.1%	14.7%	13.5%	14.7%	17.8%	17.2%	16.6%

5.5 Evaluation of Simple Slot Reduction Strategies

In this section, we propose two different strategies for allocating the available slots among different airlines and evaluate the performance of each strategy under the Nash equilibrium modeling framework.

Proportionate Allocation Scheme: Under the existing administrative controls, airlines often receive a similar number of slots from year to year. Historical precedent is usually used as the main criterion for slot allocation. There is opposition from the established carriers to any significant redistribution of slots. In the spirit of maintaining much of the status quo, our first slot distribution strategy involves proportionate allocation of slots. We vary the total number of slots at an airport while always distributing them among different carriers in the same ratio as that of actual flight schedules. For example, if the total number of slots at an airport is reduced from 100 to 80 and if the 100 slots were distributed as 40 and 60 between two carriers, then under our proportionate allocation scheme, the 80 slots will be distributed as 32 and 48 between the same two carriers.

Reward-based Allocation Scheme: While the proportionate allocation scheme is likely to be considered more acceptable by major carriers, it ignores the level of efficiency with which an airline utilizes its slots. Airlines differ, often substantially, in the number of passengers carried per flight or per slot. The idea behind the reward-based allocation is to reward those airlines which carry more passengers per slot, due to larger planes and/or higher load factors, and penalize those who carry fewer passengers per slot. Under this scheme, the number of slots allocated to each airline is proportional to the total number of passengers carried by that airline. In the previous example, if the first airline currently carries 140 passengers per slot and the second airline currently carries 120 passengers per slot, then under our reward-based allocation scheme, when the total number of slots is reduced to 80, the first airline will receive 35 slots and the second airline will receive 45 slots.

Next we present results of the impacts of slot reduction assuming that these strate-

gies are implemented as administrative slot controls. Alternatively, these slot allocations can also be as results of some more complicated demand management strategy, such as a market mechanism. If the market mechanism involves monetary payments, the resulting airline profits will have to be adjusted to account for the payments. In Sub-section 5.5.1, we present the impacts of slot reduction on various important metrics using the models presented in section 5.2. Subsequently, in Sub-sections 5.5.2 through 5.5.5, we test the sensitivity of our results to the various parameters and assumptions underlying our models.

5.5.1 Numerical Results

We conducted the following two experiments. In the first experiment, we varied the total number of allocated slots at LGA and studied the impact on two important metrics, namely, the total operating profits of all the airlines and the total number of passengers carried. Figures 5-4 and 5-5 show the changes in total operating profits of all the airlines with slot reductions under the proportionate and reward-based allocation schemes, respectively. Figures 5-6 and 5-7 show the change in the total number of passengers carried, assuming that the aircraft type (and seating capacity) for each airline on each segment remains unchanged upon slot reduction. The total number of passengers carried decreases as the number of slots decreases, but at a much lower rate. For the proportionate allocation scheme, up to a 30% slot reduction, each 1% reduction in slots leads to, on average, just a 0.27% reduction in the total passengers. A 30% reduction in slots leads to approximately 8% reduction in total passengers. Beyond 30%, the rate of decrease in passengers nearly quadruples, with each 1% reduction in slots leading to about 1.12% reduction in total passengers. Also, the total operating profit for the proportionate allocation scheme increases with increasing slot reduction percentage, up to 30% slot reduction. Beyond that point, the operating profit starts to decrease. Very similar patterns are observed for the reward-based allocation scheme. Up to a 40% reduction in slots, each 1% reduction in slots leads to, on average, just a 0.25% reduction in the total passengers. A 40% slot reduction results in about 10.2% reduction in total passengers. However,

beyond that point, each 1% reduction in slots results in about 0.86% reduction in total passengers. Similarly, total operating profit increases up to a 40% reduction and decreases thereafter.

These effects are easy to understand intuitively. Given that aircraft sizes remain constant, the initial reduction in the number of slots results primarily in increases in load factors and hence, under our constant fares assumption, operating costs decrease at a faster rate than the rate of decrease in total revenue. So the operating profit increases. This effect continues until a point where the aircraft size constraint becomes binding and reduces the number of passengers almost proportionally to the number of slots. Therefore the operating revenue decreases at almost the same rate as the operating cost decrease, causing the operating profit to decrease.

Also, the rate of change in total profits and in passengers carried is comparable for both the proportionate and the reward-based strategies at small levels of slot reduction. However, at higher slot reduction percentages, the decrease in total passengers carried is smaller for the reward-based strategy, which makes sense given that the reward-based strategy allocates a greater proportion of slots to carriers who carry more passengers per slot. As a result, the increase in total profits is also greater for the reward-based strategy at higher slot reduction percentages.

In our second experiment, we fixed a particular level of slot reduction and evaluated its system-wide impacts on the airlines (both individually and as a group), and on the passengers, based on multiple metrics. We considered the impact on the following metrics: airline operating profits, average flight delays, average passenger delays, total number of passengers carried, and average schedule displacement for passengers. The airport capacity benchmark report published by the Federal Aviation Administration [44], sets the IMC capacity of LaGuardia airport at approximately 87.7% of its VMC capacity. Currently, the number of operations scheduled at LaGuardia is close to the VMC capacity. We chose to evaluate the case of a 12.3% reduction in slots, which approximately corresponds to scheduling at IMC capacity instead of at VMC capacity. This policy is very similar to that currently followed at many major European airports.

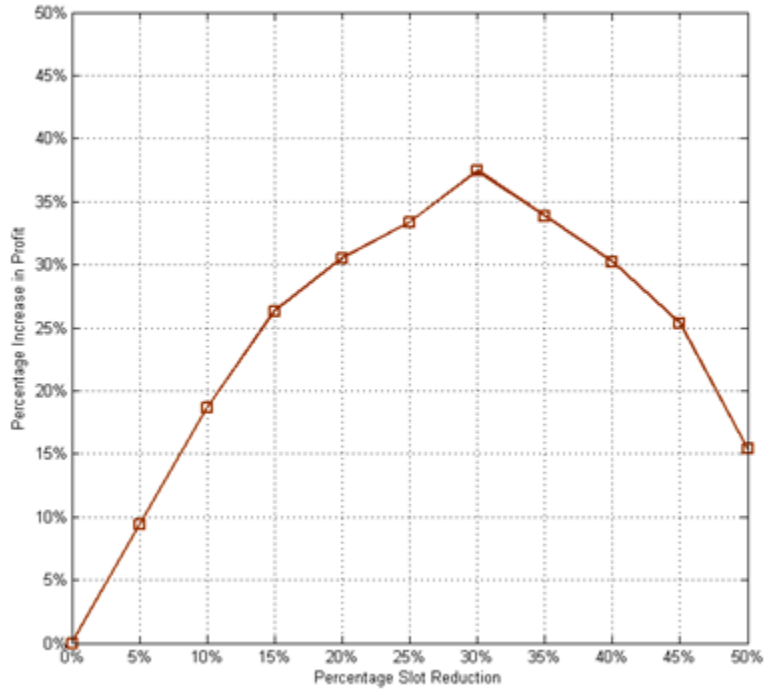


Figure 5-4: Total operating profit as a function of slot reductions under a proportionate allocation scheme assuming constant aircraft sizes

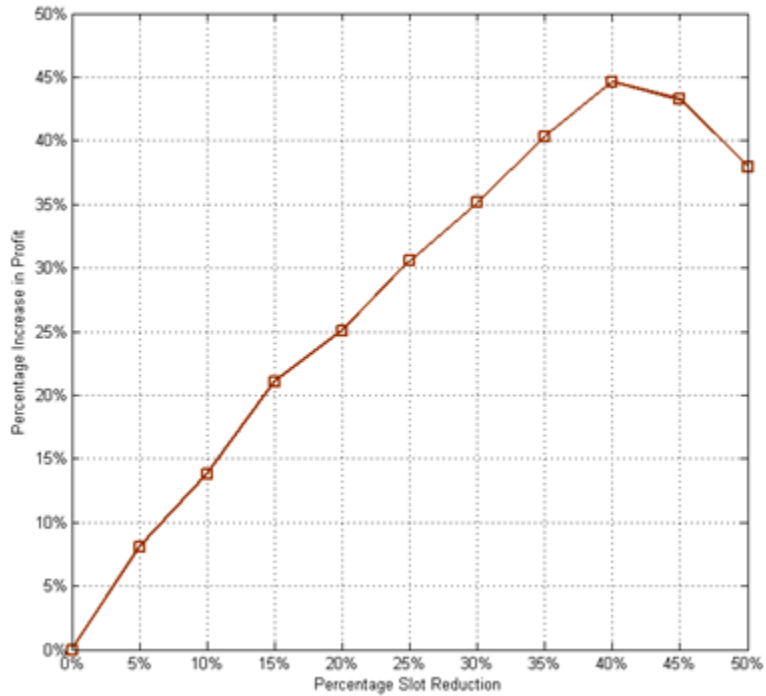


Figure 5-5: Total operating profit as a function of slot reductions under a reward-based allocation scheme assuming constant aircraft sizes

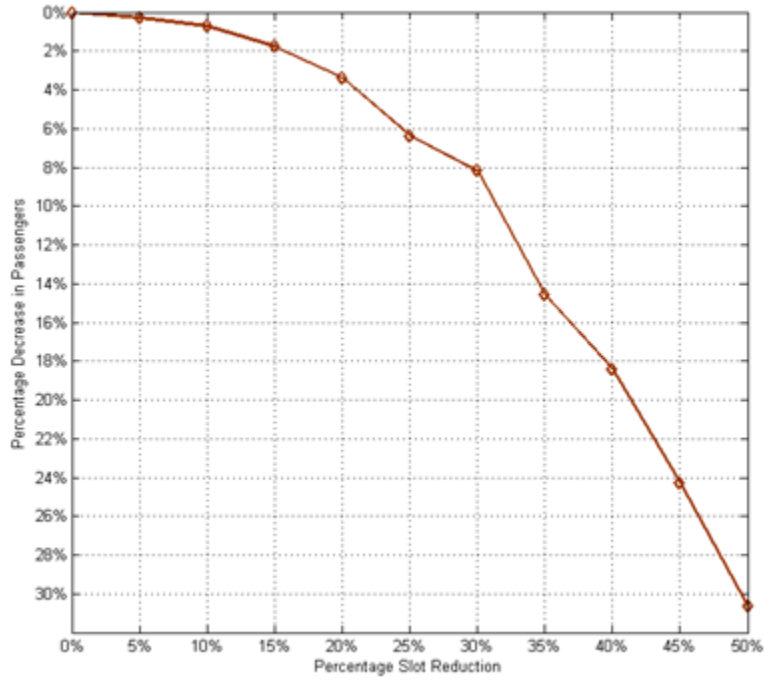


Figure 5-6: Total number of passengers carried as a function of slot reductions under a proportionate allocation scheme assuming constant aircraft sizes

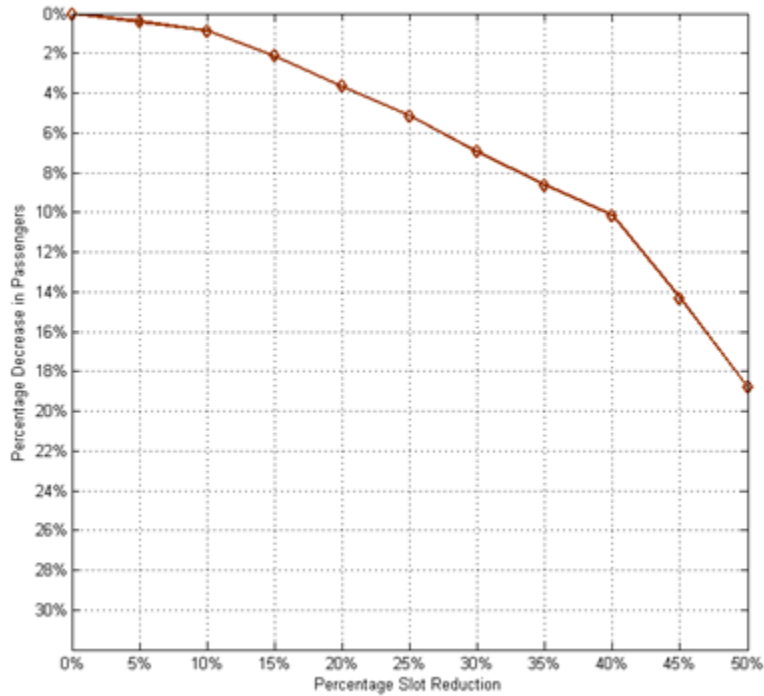


Figure 5-7: Total number of passengers carried as a function of slot reductions under a reward-based allocation scheme assuming constant aircraft sizes

Next, we describe the procedures used to estimate the average flight delays, the average passenger delays and the average schedule displacement. In order to estimate the impact on the average flight delays, we used the information on ground delay programs (GDPs) for an entire year (made available from Metron Aviation (®)) and actual flight delay data (obtained from the airline on-time performance database available on the BTS website [72]). All the delay computations were performed only for the NAS delay component of flight delays. While the number of operations currently scheduled at LaGuardia is close to its VMC capacity, realized capacity drops to the IMC value during bad weather. We assumed that the realized capacity equaled the IMC capacity during the period when a GDP was implemented at LGA, and it equaled the VMC capacity otherwise. We calculated the average NAS delays to flights landing at LaGuardia for both GDP and non-GDP periods for the entire year. We assumed that the average flight delays under IMC capacity after 12.3% slot reduction equal the average flight delays under the VMC capacity before slot reduction. After the 12.3% slot reduction, the average flight delays under VMC capacity will be lower than those under VMC capacity before slot reduction. However, in order to be conservative in our delay reduction estimates, we assumed that the average delays under VMC capacity remain unchanged upon 12.3% slot reduction. Finally, we calculated the overall average flight delay as the expected value of delays under VMC and IMC capacities.

In addition to flight delays, passenger itinerary disruptions due to flight cancellations and missed connections are responsible for a significant component of passenger delays. Barnhart, Fearing and Vaze [12] estimated the ratio of average passenger delay to average flight delay in the domestic US to be 1.97. Their procedure is also described briefly in chapter 3. In this chapter, we used this representative value for computing the average passenger delays from the average flight delays.

The total trip time for the passengers is also affected by what is known as schedule displacement ([17]) or schedule delay. Schedule displacement is a measure of the difference between the time when a passenger wishes to travel and the actual time when he/she can travel given a flight schedule. The higher the daily frequency of

flights, the lower is the schedule displacement. Due to slot reduction, the flight frequency on some segments is expected to reduce, which affects schedule displacement adversely. Schedule displacement is expressed as $\frac{K}{F}$, where F is the flight frequency and K is a constant which depends on the distribution of flight departure times and the distribution of desired times when passengers wish to travel. In this research, we assume both these distributions to be uniform. Let T be the duration of time over which the frequency F is distributed. Under the uniform distribution assumption for flights, if we divide the time T into F intervals of equal size, then there will be one flight scheduled at the midpoint of each interval. So the schedule displacement for all the passengers with desired departure times in that interval will vary uniformly between 0 and $\frac{1}{2}\frac{T}{F}$, with an average value of $\frac{T}{4F}$. Therefore, K equals $\frac{T}{4}$. We assume T to be 16 hours because 97% of all the arrivals at LGA are concentrated in the 16 hour time duration from 8:00 am to midnight [72].

Table 5.5 summarizes the impacts of slot reduction to airlines and passengers based on various metrics. Values in the 3rd and 4th columns correspond to those under the actual frequencies and the base equilibrium before slot reduction, respectively. The results in the 5th and 6th columns correspond to a 12.3% reduction in slots for proportionate and reward-based allocation schemes respectively. The values in parentheses indicate the percentage change in each metric with respect to the base equilibrium before slot reduction. The level of congestion depends on the total number of slots and not on the distribution of these slots among different airlines. Therefore, the delay reduction is the same under both proportionate and reward-based slot allocation schemes. Under either strategy, slot reductions lead to substantial reductions in flight delays as well as passenger delays. The total operating profits across of all carriers increase substantially. There is a small reduction in the total passengers carried. However, this is partly because we have assumed that aircraft sizes on each segment for each airline remain unchanged upon slot reduction. We will investigate the impact of relaxing this restriction in Sub-section 5.5.3. The average schedule displacement increases by just over 2 minutes. The total travel time for passengers arriving at LGA includes not only the schedule displacement and the duration of the flight into LGA,

but also the airport access and egress times, and in cases of connecting passengers, the layover times and duration of the first flight in their itineraries. For flights into LGA airport, the average flight duration itself is 185.38 minutes. Therefore, in comparison, the increase in schedule displacement is negligibly small.

Table 5.5: Effect of a 12.3% slot reduction on system-wide performance metrics

Stakeholder	Metrics	Actual Frequencies	No Reduction Equilibrium	Proportionate Reduction	Reward-based Reduction
Airline	Operating Profits (Excl. Delay Costs)	\$1,228,749	\$1,281,663	\$1,550,565 (20.98%)	\$1,501,100 (17.12%)
	NAS Delay/Flight	12.74 min	12.74 min	7.52 min (-40.97%)	7.52 min (-40.97%)
Passengers	Total Passengers Carried	22,896	22,965	22,678 (-1.25%)	22,661 (-1.32%)
	Avg. Passenger Delay (due to NAS Delays)	25.10 min	25.10 min	14.81 min (-40.97%)	14.81 min (-40.97%)
	Avg. Schedule Displacement	24.23 min	25.05 min	27.68 min (10.50%)	27.32 min (9.06%)

Table 5.6 presents the distribution of operating profits across different carriers. Profits of all carriers that account for at least 1% of the operations at LGA are included. The impacts of a 12.3% slot reduction on the remaining carriers are negligible because their slots do not get reduced under a 12.3% reduction due to integral rounding. The 2nd, 3rd, 4th and 5th columns correspond to actual frequencies, base equilibrium before slot reduction, 12.3% reduction under proportionate allocation scheme, and 12.3% reduction under reward-based allocation scheme respectively. Again, the values in parentheses represent the percentage increases in profits compared with the base equilibrium before slot reduction. When the total number of slots is reduced under either allocation scheme, the operating profit of each carrier is strictly greater compared to that under the no slot reduction scenario. The relative increase in operating profits is largest for the regional carriers operating small regional jets into LGA. This is primarily because they had very low operating profit margins at LGA under the no slot reduction scenario. In fact, for one of regional carriers, the slot reduction under reward-based allocation helps achieve an operating profit instead of an operating loss, which is the case under the no slot reduction scenario. On the other

hand, the network legacy carriers achieve the maximum absolute increase in operating profit per carrier. This is primarily because the average number of slots per day for network legacy carriers (34.50) itself is nearly 27% higher than that for the remaining carriers (27.25), and the average operating profits for the network legacy carriers per day (\$192,276) are much higher than that for the remaining carriers (\$23,037) under the no slot reduction scenario. If we compare against the actual frequencies case, then the profit for all but one carrier increases after slot reduction under both allocation schemes.

Table 5.6: Increase in operating profits due to a 12.3% slot reduction

Carrier	Actual Frequencies	No Reduction Equilibrium	Proportionate Reduction	Reward-based Reduction
Network Legacy Carrier 1	\$349,363	\$390,735	\$452,701 (15.86%)	\$423,466 (8.38%)
Low Cost Carrier 1	\$47,090	\$26,684	\$34,487 (29.24%)	\$31,887 (19.50%)
Network Legacy Carrier 2	\$49,693	\$71,922	\$79,550 (10.61%)	\$75,823 (5.42%)
Network Legacy Carrier 3	\$202,489	\$206,315	\$299,385 (45.11%)	\$282,772 (37.06%)
Low Cost Carrier 2	\$54,000	\$59,927	\$79,766 (33.11%)	\$73,541 (22.72%)
Regional Carrier 1	\$29,836	\$34,461	\$39,480 (14.56%)	\$39,473 (14.54%)
Network Legacy Carrier 4	\$91,772	\$85,708	\$113,736 (32.70%)	\$106,451 (24.20%)
Regional Carrier 2	- \$28,493	- \$28,923	- \$2,227 (n.a.)	\$8,513 (n.a.)
Network Legacy Carrier 5	\$200,796	\$200,796	\$201,786 (0.49%)	\$208,280 (3.73%)
Network Legacy Carrier 6	\$196,346	\$198,180	\$216,043 (9.01%)	\$214,913 (8.44%)
Total: Network Legacy Carriers	\$1,090,459	\$1,153,656	\$1,363,201 (18.16%)	\$1,311,705 (13.70%)
Total: Low Cost Carriers	\$101,090	\$86,611	\$114,253 (31.92%)	\$105,428 (21.73%)
Total: Regional Carriers	\$1,343	\$5,539	\$37,254 (572.61%)	\$47,986 (766.39%)

These results in Tables 5.5 and 5.6 are obtained assuming a maximum average segment load factor (LF_{max}) of 85%. Now, we will present results on the sensitivity of slot reduction impacts to this assumption. We will focus on the sensitivity of the results of the second experiment. Table 5.7 describes the sensitivity of total profits and total number of passengers carried to variations in the maximum average segment load factor value. Obviously, the average flight delays, average passenger delays and average schedule displacements do not change, because they depend only on the scheduled number of flight operations. The increase in total operating profit varies between 16.56% and 21.99% and the decrease in number of passengers varies between 0.49% and 3.71%. With increases in the maximum average segment load factor, we observe a general trend towards smaller reductions in total passengers due to slot reduction, which is intuitively reasonable. Consequently, we also observe a general trend towards greater increases in total profits with increases in the maximum average segment load factor. Due to the integrality constraints on the number of slots, the results don't vary smoothly in some cases.

Table 5.7: Sensitivity of slot reduction impacts to the maximum average segment load factor value under a 12.3% slot reduction

Maximum Load Factor	Increase in Total Profits		Decrease in Passengers Carried	
	Proportionate	Reward-based	Proportionate	Reward-based
75%	20.25%	16.65%	3.71%	3.36%
80%	19.18%	16.56%	2.59%	2.14%
85%	20.98%	17.12%	1.25%	1.32%
90%	21.14%	17.93%	0.94%	0.94%
95%	21.99%	17.86%	0.49%	0.95%

5.5.2 Sensitivity to Model Parameters

Table 5.4 in Sub-section 5.4.4 showed that the model's prediction accuracy is not highly sensitive to variations in model parameters. In this section, we present the sensitivity of slot reduction impacts to variations in the values of various parameters in the *basic model* and in the model extensions 1 and 2. The parameters are

varied within -25% to +25% of their values listed in Sub-section 5.4.1. Tables 5.8 and 5.9 present the percentage increase in total profits and the percentage decrease in passengers carried, respectively, under a 12.3% reduction with the proportionate allocation scheme. Even with significant variations in the parameter values, slot reduction results in at least a 17.3% increase in total operating profits with at most a 1.9% reduction in passengers carried (assuming constant aircraft sizes). We also performed similar sensitivity analyses of our slot reduction results under the reward-based allocation scheme and found the results to be similarly stable, as shown in Tables 5.10 and 5.11. Even with significant variations in the parameter values, under our reward-based based slot allocation scheme, a 12.3% reduction in total number of slots results in at least a 10.9% increase in total operating profits with at most a 2.8% reduction in passengers carried (assuming constant aircraft sizes).

Table 5.8: Sensitivity of increase in total profit due to slot reduction to model parameters (under 12.3% reduction with proportionate allocation)

Parameter	-25%	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
α_s	20.3%	20.9%	20.4%	20.7%	21.5%	21.0%	21.3%	18.9%	18.3%	19.1%	18.7%
θ_a	21.5%	21.2%	21.0%	20.8%	20.5%	21.0%	21.6%	21.0%	22.3%	22.4%	22.4%
α_s^1	20.8%	20.7%	20.7%	20.5%	20.8%	21.0%	20.9%	20.5%	20.3%	21.6%	20.3%
β_s^1	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	20.8%	20.9%	20.9%	20.8%	20.8%
γ_s^1	20.7%	20.7%	20.7%	20.7%	20.8%	21.0%	20.8%	20.6%	20.4%	20.4%	20.4%
α_s^2	20.7%	20.7%	20.7%	20.8%	20.9%	21.0%	20.9%	20.5%	20.3%	20.3%	20.2%
β_s^2	20.3%	21.3%	21.0%	21.0%	21.0%	21.0%	20.8%	20.5%	20.1%	21.1%	20.5%
$C_{a's}$	22.8%	27.7%	19.2%	17.3%	19.3%	21.0%	23.7%	24.2%	24.1%	25.5%	25.8%
$S_{a's}$	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%	21.0%
$p_{a's}$	25.8%	25.8%	25.5%	24.1%	23.7%	21.0%	19.3%	17.3%	20.4%	24.0%	27.7%

5.5.3 Effect of Aircraft Upgauges

Results in section 5.5.1 were obtained under the assumption that, even when the total number of slots available to an airline is reduced, the airline will continue to operate the same-size aircraft as it did in the absence of slot reduction. This assumption might be realistic for very small reductions in the number of slots, but for significant reductions, it is reasonable to expect that the airlines will operate larger aircraft

Table 5.9: Sensitivity of decrease in passengers carried due to slot reduction to model parameters (under 12.3% reduction with proportionate allocation)

Parameter	-25%	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
α_s	1.17%	1.23%	1.41%	1.63%	1.30%	1.25%	1.29%	1.77%	1.71%	1.41%	1.52%
θ_a	1.03%	1.30%	1.13%	1.45%	1.40%	1.25%	0.91%	1.17%	1.30%	1.25%	1.30%
α_s^1	1.29%	1.30%	1.28%	1.33%	1.26%	1.25%	1.22%	1.46%	1.43%	1.43%	1.45%
β_s^1	1.25%	1.25%	1.26%	1.25%	1.25%	1.25%	1.27%	1.23%	1.22%	1.23%	1.23%
γ_s^1	1.30%	1.31%	1.30%	1.30%	1.26%	1.25%	1.23%	1.39%	1.39%	1.39%	1.39%
α_s^2	1.30%	1.30%	1.30%	1.27%	1.26%	1.25%	1.23%	1.46%	1.42%	1.33%	1.46%
β_s^2	1.30%	1.01%	1.14%	1.14%	1.24%	1.25%	1.22%	1.40%	1.40%	1.36%	1.50%
$C_{a's}$	1.37%	0.85%	1.59%	1.90%	1.35%	1.25%	1.03%	0.99%	0.82%	0.78%	0.60%
$S_{a's}$	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%	1.25%
$p_{a's}$	0.60%	0.60%	0.78%	0.82%	1.03%	1.25%	1.35%	1.90%	1.37%	0.88%	0.85%

Table 5.10: Sensitivity of increase in total profit due to slot reduction to model parameters (under 12.3% reduction with reward-based allocation)

Parameter	-25%	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
α_s	16.8%	17.5%	17.1%	17.4%	17.6%	17.8%	18.0%	13.4%	13.7%	13.6%	13.5%
θ_a	16.8%	16.3%	16.7%	16.5%	16.3%	17.1%	18.1%	18.2%	17.6%	17.5%	17.8%
α_s^1	17.4%	17.0%	17.0%	17.0%	17.0%	17.1%	16.9%	16.9%	16.8%	17.4%	16.1%
β_s^1	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	16.9%	16.9%	16.9%	16.9%	16.9%
γ_s^1	17.0%	17.0%	17.0%	17.0%	17.0%	17.1%	16.9%	16.9%	16.2%	16.2%	16.1%
α_s^2	17.0%	17.0%	16.9%	16.9%	17.1%	17.1%	16.9%	16.9%	16.6%	16.2%	16.1%
β_s^2	17.7%	17.7%	17.6%	17.3%	17.2%	17.1%	16.9%	16.0%	15.5%	15.9%	16.6%
$C_{a's}$	19.5%	20.0%	10.9%	11.7%	13.4%	17.1%	17.6%	18.0%	17.1%	17.0%	17.4%
$S_{a's}$	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%	17.1%
$p_{a's}$	17.4%	17.4%	17.0%	17.1%	17.6%	17.1%	13.4%	11.7%	13.2%	19.4%	20.0%

Table 5.11: Sensitivity of decrease in passengers carried due to slot reduction to model parameters (under 12.3% reduction with reward-based allocation)

Parameter	-25%	-20%	-15%	-10%	-5%	0%	5%	10%	15%	20%	25%
α_s	1.07%	1.13%	1.39%	1.39%	1.35%	1.32%	1.58%	2.42%	2.33%	2.35%	2.37%
θ_a	1.19%	1.49%	1.49%	1.45%	1.54%	1.32%	1.21%	1.40%	1.69%	1.76%	1.86%
α_s^1	1.33%	1.34%	1.32%	1.34%	1.33%	1.32%	1.34%	1.34%	1.32%	1.53%	1.55%
β_s^1	1.33%	1.33%	1.33%	1.33%	1.33%	1.32%	1.34%	1.34%	1.34%	1.34%	1.34%
γ_s^1	1.34%	1.35%	1.34%	1.34%	1.33%	1.32%	1.33%	1.33%	1.53%	1.53%	1.53%
α_s^2	1.34%	1.33%	1.33%	1.34%	1.33%	1.32%	1.34%	1.33%	1.36%	1.42%	1.56%
β_s^2	1.20%	1.20%	1.20%	1.20%	1.32%	1.32%	1.33%	1.65%	1.66%	1.55%	1.43%
$C_{a's}$	1.29%	1.63%	2.74%	2.80%	1.85%	1.32%	1.27%	1.26%	1.26%	1.23%	1.06%
$S_{a's}$	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%	1.32%
$p_{a's}$	1.06%	1.06%	1.23%	1.26%	1.27%	1.32%	1.85%	2.80%	2.34%	1.29%	1.63%

on some of the segments in order to accommodate more passengers and therefore increase profit. The main problem with modeling aircraft size decisions is that such decisions depend on the fleet availability. We estimate the impact of aircraft size upgauges by allowing for a limited number of upgauges for each airline. We sort all the available types of aircraft operated into LGA by any of the airlines in increasing order of seating capacity. We allow a certain maximum percentage of an airline's fleet (flying into LGA) to be upgauged to the next bigger-sized aircraft. This constraint indirectly models the fact that an airline cannot arbitrarily increase aircraft sizes due to fleet availability constraints. We calculate the equilibrium frequency solution under the slot reduction scenario as described in section 5.5 and subsequently perform, for each airline, the most profitable flight upgauges subject to the limits on maximum allowable upgauge percentage. As before, we assume a maximum average segment load factor of 85%.

Figure 5-8 describes the impact of a limited number of aircraft upgauges on the reductions in total passengers when the total number of slots is reduced by 12.3%, and the proportionate allocation scheme is used. The maximum allowable upgauge percentage is on the x-axis, which represents the maximum percentage of an airline's flights that can be upgauged to the next bigger aircraft size. The percentage reduction in the total number of passengers varies from 1.25%, assuming 0% upgauges, to

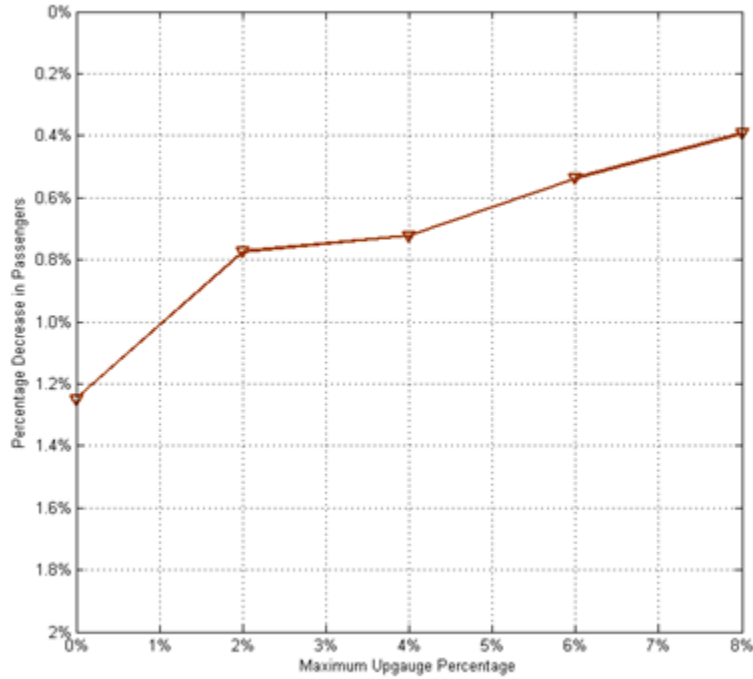


Figure 5-8: Effect of limited upgauging on total number of passengers under a 12.3% slot reduction

0.39% assuming at most 8% upgauges for each airline. This shows that even with a small fraction of flights upgauging to a larger-sized aircraft, most of the reduction in the number of passengers disappears. The remaining reduction in the number of passengers is primarily attributable to the fact that there is only a limited number of aircraft sizes available; and on some segments, the number of passengers who are denied a seat due to a smaller aircraft size is not large enough to justify a profitable upgauge to the next bigger aircraft size.

5.5.4 Effects of Demand Uncertainty and Revenue Management Practices

The results presented in Sub-section 5.5.1 assume deterministic passenger demand and the existence of a fixed fare value for each flight. In reality, passenger demand varies from day-to-day and these variations affect the number of passengers transported. Spill (or spilled passengers) is defined as the portion of passenger demand that cannot

be accommodated because of the limited capacity of the aircraft. Models of spill estimation typically assume some distribution of demand and calculate the expected value of this distribution truncated by the seating capacity to obtain the expected number of passengers carried ([18]). In our models presented thus far, to account for demand stochasticity, we put a hard constraint on the maximum number of seats sold on each segment at 85% of the seating capacity of all flights on the segment. This is a fairly conservative value, considering the fact that in the year 2007, across all 7,452 combinations of segments and carriers in the domestic U.S. with at least 1 flight per day, 923 combinations had an average load factor of greater than 85%. Moreover, the higher the demand factor, the higher is the average load factor, where demand factor is defined as the ratio of average passenger demand to seating capacity. If the total number of seats offered on a segment is reduced, which is likely to be the case under a slot reduction scenario, the demand factor increases further. Therefore, we expect our method to introduce a downward bias in the expected number of passengers carried.

The revenue management methods practiced by the airlines affect the average fares of the spilled passengers by ensuring that the spilled passengers are predominantly the low-fare passengers. Until now, we have ignored this effect. Therefore, our method can also be expected to introduce a downward bias in the average fare values, and therefore, in the operating profit estimates. Moreover, as pointed out by Belobaba and Farkas ([18]), in addition to impacting the average fares of the spilled passengers, revenue management practices also affect the number of spilled passengers. In this section, we estimate the extent and direction of this bias, in the number of passengers carried and in the airline profits, introduced by our simplified assumptions about demand uncertainty and fares.

Under a slot reduction scenario, the total segment seating capacity is reduced, unless there is a substantial increase in aircraft sizes. Revenue management systems used by the airlines are expected to readjust their seat allocation decisions across different fare classes by spilling the low-fare passengers, resulting in some increase in average fares. In order to estimate the effect of demand uncertainty and revenue management on expected spills and average fares, we use passenger spills and average

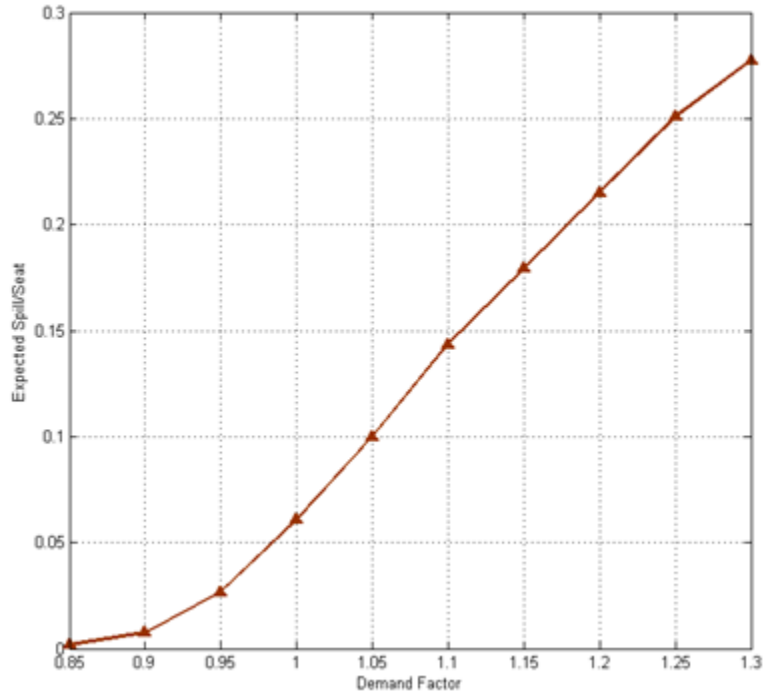


Figure 5-9: Expected spill per seat obtained using the spill modeling approach for a fare discount ratio of 0.7 (Source: Belobaba and Farkas, 1999)

fares of spilled passengers (called spill fares) estimated by Belobaba and Farkas ([18]) using a multiple nested fare class, single booking period model. Belobaba and Farkas ([18]) have estimated the expected spills and spill fares for a 5-fare class example using different values of demand factors and fare discount ratios, where a fare discount ratio is defined as the ratio of the fare of a class to the fare of the immediately higher fare class.

Figures 5-9 and 5-10 show the expected spill and spill fare for different values of demand factors and fare discount ratios estimated by Belobaba and Farkas ([18]). In both figures, different curves correspond to different fare discount factors. The demand factor is on the x-axis in both figures. Figure 5-9 has the ratio of the expected number of spilled passengers to the aircraft seating capacity on the y-axis while Figure 5-10 has the average fare of spilled passengers on the y-axis. Note that Figure 5-10 corresponds to a maximum fare of \$600. For other fare values, the average spill fare can be calculated by simply rescaling the y-axis accordingly.

We will call our default approach as the *deterministic approach* and this new

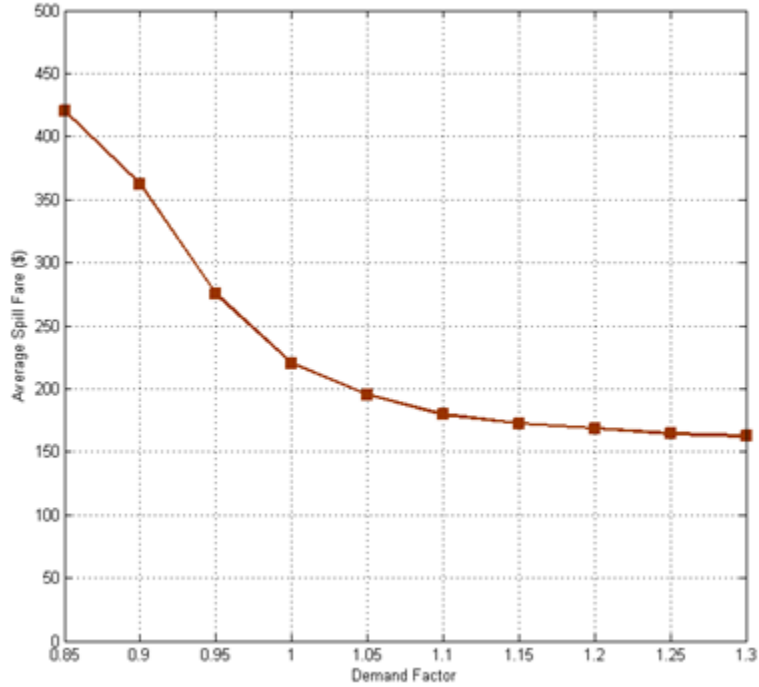


Figure 5-10: Average spill fare obtained using the spill modeling approach for a fare discount ratio of 0.7 (Source: Belobaba and Farkas, 1999)

approach as the *spill modeling approach*. Under the *spill modeling approach*, for a given set of flight frequencies, the demand for each airline on each segment is computed using the appropriate market share model (given by constraint 5.2, constraint 5.7, or by constraints 5.8 and 5.9). The expected number of spilled passengers on each segment is calculated using Figure 5-9 and the average fare of the spilled passengers is calculated using Figure 5-10. The airline’s revenue is obtained by subtracting the product of the expected spill and spill fare from the product of the demand and the overall average fare across the unconstrained demand for each airline on each segment.

Table 5.12: Effect of a 12.3% slot reduction under *spill modeling approach*

Metrics	Actual Frequencies	No Reduction Equilibrium	Proportionate Reduction	Reward-based Reduction
Operating Profit (Excl. Delay Costs)	\$1,263,650	\$1,326,149	\$1,617,158 (21.9%)	\$1,566,471 (18.1%)
Passengers Carried	23,085	23,129	22,927 (-0.87%)	22,897 (-1.00%)

For the equilibrium frequency solution with and without slot reductions, we recalculated the airline profits using the *spill modeling approach*. Table 5.12 confirm that our deterministic approach introduces a slight downward bias in number of passengers carried as well as in the total operating profits. This can be observed by comparing these results with those in Table 5.5. Furthermore, as the demand factor increases, the magnitude of this bias increases, which means that the bias is greater for the slot reduction scenario than for the no reduction scenario. So, our *deterministic approach*, in fact, slightly underestimates the benefits of slot reduction. This is apparent from the greater increase in total operating profits and smaller reduction in passengers carried due to slot reduction as reported in Table 5.12 compared to that in Table 5.5.

The preceding discussion about increases in average fares assumes that the total seating capacity on a segment is reduced due to slot reduction. However, in light of the possibility of aircraft upgauges as discussed in Sub-section 5.5.3, the decrease in seats due to frequency reduction might be partially or completely compensated by increases in seats per flight, depending on the extent of upgauging. Therefore, even the results presented in Table 5.12 are likely to be overestimates of the reduction in passengers carried and underestimates of the increase in total operating profits.

In addition to the effects of demand uncertainty and revenue management, one can consider the possibility of changes in the actual fare values due to a reduction in the total number of seats offered on a non-stop segment. Slot reduction can have an impact on fares in different ways. Borenstein ([21]) noted that an airline's share of passengers on a route and at an airport affects fares. Slot reduction might affect the passenger shares of different airlines differently, thus leading to an increase in fares for some combinations of routes and carriers, and a decrease for some other combinations. However, there is no reason to expect any significant change in the overall fare levels because of this effect. Additionally, the delay reduction resulting from slot reduction strategies can, in turn, be expected to have an impact on fares in multiple ways. Britto, Dresner and Voltes ([24]) found that an increase in flight delays increases airline costs and hence has an increasing effect on fares. According to their results, for every minute of reduction in average flight delays, the fares decrease

by about \$0.04. On the other hand, Forbes ([46]) analyzed a rather extreme delay situation for LGA airport and concluded that an increase in flight delays deteriorates the quality of air service and therefore decreases the fares. According to the findings of this study, the fares increase by about \$1.42 for every minute of reduction in average flight delays. Detailed analysis of these various, often opposite, effects is beyond the scope of this research. However, the possibility of net increases in fares upon slot reduction suggests that our results are likely to be conservative, that is, the actual increase in total profits of the airlines due to slot reduction strategies could be even higher than what we report in Table 5.5.

In addition to the inherent demand stochasticity, the total demand for travel on each segment into LGA (M_s), is likely to be affected by slot controls in a complicated way. A significant proportion of passengers traveling to New York City have other travel alternatives to flying into LGA, e.g. travel by car, travel by rail, and travel by air into other New York area airports such as Newark (EWR) and Kennedy (JFK). Upon slot reduction at LGA, the impact on total passenger demand can be several-fold. Substantial delay reduction is expected to make LGA more attractive while a corresponding reduction in frequency on some routes and a possible increase in fares are expected to make LGA less attractive. While the net effect is difficult to predict, it is reasonable to expect some increase in net attractiveness of LGA unless EWR and/or JFK also implement similar slot reduction strategies. This is yet another reason the passenger demand and actual number of passengers carried into LGA after slot reduction can be expected to be higher than those presented in Table 5.5.

5.5.5 Effects of Passenger Connections

Our models of frequency competition assume segment-based demand. In reality, passengers demand seats on itineraries, which might be combinations of two or more flights. Accordingly, the airlines also charge fares on an itinerary basis rather than on a flight basis. So calculation of the fare revenue generated by a segment is not a straightforward process. Belobaba ([16]) acknowledged that different assumptions made by airline planners can lead to very different estimates of profitability. The

results presented in Sub-section 5.5.1 were calculated under the full fare assumption. One potential issue with full fare assumption is that the fare of a one-stop passenger gets counted twice, once for each flight in the passenger's itinerary. In this research, we have considered only those segments which have one end-point at LGA airport. Because a very small fraction of passengers (5%) connect at LaGuardia [73], there will be very little double-counting of the revenues. Moreover, whatever double counting occurs has a comparable effect on the results under the slot reduction scenario and the scenario with no slot reduction. So the issue of double-counting does not affect the percentage increase in total profits significantly.

Another issue with the full fare assumption is that when a passenger is spilled, the entire fare revenue corresponding to that passenger is assumed to be lost. This fails to capture the possibility of having an additional non-stop passenger on the other segment, which has seats available. For example, when a reduction in the frequency of American Airlines' flights from LGA to Dallas-Fort Worth (DFW) airport results in spilling a passenger traveling from LGA to Los Angeles (LAX) via DFW due to lack of seats on the LGA-DFW segment, there is still a possibility of recovering a part of that revenue by carrying an additional non-stop passenger on the DFW-LAX segment. An alternative method for fare revenue calculation is the complete proration approach, in which the fare is completely prorated based on distance. Under this assumption, if the airline is not able to carry a connecting passenger, it will only lose the revenue equal to a fraction of the passenger's full fare. This is not necessarily the most valid representation of segment profits either, because of the possibility that the seat vacated by that passenger on the other segment may not be filled by another non-stop or connecting passenger. In the aforementioned example, the seat on the DFW-LAX segment vacated by the spilled passenger going from LGA to LAX via DFW may not get filled by another passenger. Moreover, another issue with the distance-based proration is that the fares are not necessarily well correlated with distances.

While neither the full fare assumption nor the complete proration assumption is very accurate, these represent the two extremes of possible methods for calculating segment profitability. Any reasonable method of fare revenue estimation is expected

to lie somewhere in between these two extremes. To measure the difference between these extreme scenarios, we evaluated the impact of slot reduction strategies under the distance-based complete proration assumption. The results are presented in Table 5.13. As expected, the profitability values are considerably lower than those corresponding to the full fare assumption, i.e. those in Table 5.5. However, the absolute increase in total profits due to slot reduction changes only slightly.

Table 5.13: Effect of a 12.3% slot reduction under completely prorated fares approach

Metrics	Actual Frequencies	No Reduction Equilibrium	Proportionate Reduction	Reward-based Reduction
Operating Profit (Excl. Delay Costs)	\$901,346	\$959,944	\$1,247,750 (29.98%)	\$1,174,736 (22.38%)
Passengers Carried	22,895	22,971	22,676 (-1.29%)	22,532 (-1.91%)

In addition to the fare proration issue, another limitation of the S-curve model of market share is that it does not estimate the connecting passenger shares accurately. The model predicts that when the frequency of one carrier decreases compared to that of another competing carrier, a fraction of its passengers will shift to the other carrier. But this does not make sense for connecting passengers if the other carrier does not offer a connecting itinerary to that passenger's final destination. For example, if American Airlines reduces its flight frequency from DFW to LGA, the segment-based S-curve model predicts that some passengers on the DFW-LGA segment will shift to other airlines on DFW-LGA segment. But in reality, a passenger traveling from LAX to LGA via DFW may not shift to the DFW-LGA flights of another airline if that other airline does not offer service from LAX to DFW. In fact, with changes in frequency values, the passengers might shift to traveling through a completely different hub of the same carrier or some other carrier to their final destinations. For example, with a reduction in the frequency of American Airlines' flights from DFW to LGA, a passenger traveling from LAX to LGA on American Airlines through its DFW hub may decide to instead travel on Continental Airlines by connecting at Houston (IAH) or to continue traveling on American Airlines but through the Chicago

O'Hare (ORD) hub rather than DFW. Therefore, such interactions can get extremely complicated.

Of all passengers on flights arriving at LaGuardia, approximately 75% are non-stop passengers, only 20% are connecting passengers with a final destination at LGA and the remaining 5% connect at LaGuardia itself [73]. Of all the connecting passengers with a final destination at LGA, nearly 97% are carried by seven major airlines and nearly 89% of them connect through the two biggest hubs of each of these seven airlines. We expect the effects of connecting passengers to be fairly well distributed across these major airlines, without any obvious advantages or disadvantages to any particular carriers. Furthermore, these connecting passenger effects should have a similar impact on results with and without slot reduction. So we expect that the passenger connections would not affect the main results of this study significantly.

5.6 Summary

Any demand management strategy implicitly or explicitly involves deciding the total capacity to be allocated and the distribution of this capacity among different airlines. In this research, we explicitly consider these two stages separately. Although there is extensive literature on airport demand management strategies, none of the previous studies have captured critical elements of frequency competition among carriers. To the best of the author's knowledge, this is the first study that tries to model airline competition under demand management strategies.

We developed a game theoretic model of airline frequency competition based on the S-curve relationship, which is a popular model of market share in the airline literature. Due to the discreteness of the problem and the non-convexity of its continuous relaxation, the optimization problem for each airline is complicated. Furthermore, due to competitive interactions among different players, the problem becomes one of computing a Nash equilibrium. The large size of the solution space makes it very challenging to solve. We propose an efficient solution algorithm for obtaining a Nash equilibrium. We justify the predictive power of the Nash equilibrium solution con-

cept using empirical validation of the model estimates under existing slot allocation. Irrespective of the starting point, the best response algorithm was found to approach the equilibrium outcome within a very few iterations. This shows that even less than perfectly rational carriers can reach the equilibrium outcome through simple myopic learning dynamics, and thus provides further justification of the predictive power of Nash equilibrium outcome.

We evaluated two simple slot reduction strategies. The results showed that in addition to a substantial reduction in flight and passenger delays, small reductions in total allocated capacity can improve the operating profits of carriers considerably. While the two strategies led to some differences in the actual profitability increases across individual carriers, the aggregate impacts were similar. Under each strategy, slot reduction led to a substantial increase in the operating profits of all carriers across the board, and substantial reductions in flight delays and passenger delays. It also led to a small reduction in the number of passengers carried. However, most of the reduction in total passengers carried was eliminated when the possibility of a limited number of aircraft upgauges was introduced. The increase in schedule displacement due to the slot reduction was negligibly small compared to the overall travel times of the passengers.

These results are obtained based on the various conceptual and numerical assumptions made in developing our Nash equilibrium-based modeling framework. Therefore, we tested the sensitivity of the results to many of our assumptions, parameter values and changes in time periods and datasets. In most of the cases, the model results varied only slightly. Also in most of the cases, our assumptions were found to be conservative, that is, relaxing these assumptions is expected to make the slot reduction schemes even more attractive.

Therefore, a small reduction in the total number of slots at congested airports is beneficial to the carriers, all of whom experience reductions in delay costs as well as increases in planned operating profits. It also benefits the passengers, almost all of whom get transported to their respective destinations, with negligible increases in schedule displacement and significantly lower average passenger delays. It is also

beneficial to the airport operators because congestion and airport delays are reduced substantially. From the perspective of the entire system, slot reduction strategies lead to almost all passengers being transported with many fewer flights and lower total cost. Hence, slot reduction strategies are also attractive from the perspective of overall social welfare.

Chapter 6

Pricing Mechanisms for Airport Congestion Mitigation

As mentioned earlier, Chapter 5 describes a detailed computational experiment which demonstrates the potential stakeholder benefits to be had from an administrative slot reduction mechanism at a congested airport such as LGA. This corresponds to the most significant practical contribution of this thesis. In this chapter, we extend our framework for modeling airline frequency competition to a congestion pricing setting. Using a small, hypothetical network of airlines, we provide a proof-of-concept demonstration of some critical characteristics of airline competition, which when captured appropriately, can critically modify the stakeholder benefits of a congestion pricing mechanism.

6.1 Introduction

With airport capacity being a scarce resource, market-based mechanisms such as congestion pricing and slot auctions are expected to bring demand and supply in balance by placing monetary prices on the airport capacity. These market-based mechanisms rely on the ability of the airlines to assess the economical value of airport slots, while bidding for slots in the case of auctions and for determining the demand for slots at a given level of prices in the case of congestion pricing. Airlines are typically

assumed to be rational decision makers, each driven by its own profit-maximization objective. However, an airline needs to account for competition from other airlines operating in the same markets as it does while ascertaining its own valuation of an airport slot. In this chapter, we model the airline frequency decisions under congestion pricing mechanisms through explicit modeling of competition and assess the dependence of the effectiveness, or lack thereof, of airport congestion pricing mechanisms on the characteristics of the airline markets. Many prior studies have accounted for airline competition under pricing using conventional micro-economic models of firm competition. However, these generalized models fail to capture some essential characteristics of competition which are peculiar to the airline industry and therefore, as we show later in this chapter, tend to underestimate the congestion pricing benefits to the airlines. We capture these characteristics through an industry-specific competition model and generate insights that were not possible with the previous models.

The rest of this chapter is organized as follows. Section 6.2 describes the relevant existing literature on airport congestion pricing. Section 6.3 provides details of our model of frequency competition that captures airline frequency decisions in the presence of delays and congestion prices. Section 6.4 describes our delay model that captures the dependence of flight delays on airline frequency decisions. Section 6.5 presents the algorithm that we used to solve this problem iteratively. Section 6.6 outlines the data sources and experimental setup for the subsequent computational experiments. Section 6.7 provides results of delay function fitting. Section 6.8 provides computational results from the pricing experiments for a small hypothetical network of airlines. Section 6.9 concludes with a summary of the practical implications of this research and a description of directions for future research.

6.2 Literature Review

A user (such as an airline) of a public resource (such as an airport) generates value for itself through the utilization of the resource. Such utilization might sometimes

result in detrimental effects to the other users of the public resource. In particular, an airline operating at a congested airport imposes additional delay costs on the other airlines operating at the same airport. Economists have long been advocating the use of pricing of public resources in the presence of negative externalities such as congestion, wherein each user of the public resource is required to pay a toll equal to the marginal cost imposed by that user on all the other users of the resource [97]. Such prices based on marginal costs have been claimed to achieve the social welfare maximization objective. Not surprisingly, early studies on airport congestion pricing, including Levine [62], and Carlin and Park [30], have advocated marginal cost pricing of airport resources. Levine [62] proposed to implement a system in which each airport user is charged fully for the marginal cost of an additional operation, while Carlin and Park [30] recommended a hybrid system involving pricing and administrative controls due to various practicality issues associated with a full marginal cost pricing scheme.

Airport congestion pricing, however, is fundamentally different from pricing of a resource such as roadway infrastructure which involves a large number of users, each using a very small portion of the capacity of the public resource, otherwise known as atomistic users. Airlines, on the other hand, are considered to be non-atomistic users of airport resources because each airline typically operates more than one flight at an airport, and the number of users of an airport resource is comparatively smaller. So each additional operation by an airline delays the flights of other airlines as well as the other flights of the same airline at that airport. More recent studies such as Daniel [40, 41], Brueckner [25, 26], Pels and Verhoef [80] recognize the fact that airlines automatically internalize a part of the congestion costs they impose. A recent study by Morrison and Winston, however, compared the atomistic (or flat) and non-atomistic pricing policies across 74 commercial US airports in 2005. They found the difference between the net benefits generated by the two congestion pricing policies to be small because the bulk of airport delays are not internalized [66]. In this chapter, we analyze the impacts of various levels of flat pricing (also known as atomistic pricing) as well as the marginal cost pricing (also known as non-atomistic pricing) equilibrium for non-atomistic users.

Daniel [40, 41] modeled the interaction between airport demand, slot prices and delays using detailed queuing theoretic models. These two studies, however, do not capture frequency-based competition for passenger share of a market, even though such competition between airlines is intricately related to the congestion problem at major airports.

Several other existing studies have tackled this problem from a microeconomic perspective. Studies such as Brueckner [25, 26], Pels and Verhoef [80], and a recent one by Perakis and Sun [81] provide a rigorous mathematical treatment of the problem of airport slot pricing under airline frequency competition and derive the Nash equilibrium outcomes of such models mathematically. However, these studies model airline capacity allocation decisions using general micro-economic models of firm competition, which typically assume Cournot-type [38] quantity competition. By assuming constant load factors and constant aircraft seating capacities, they fail to recognize the important distinguishing features of the airline industry for which the quantity produced can be captured by three different entities: number of flights, number of seats and number of passengers carried.

As discussed in Sub-section 5.1.1, the incremental profitability of having an extra flight in a particular market largely depends on the number of additional passengers that the airline will be able to carry because of the additional flight, which in turn depends on the schedule of flights offered by the competitor airlines in the same market. So, given a set of congestion prices, the total demand for slots should reflect these competitive interactions. However, Cournot-type [38] models of firm competition do not incorporate the inverse dependence of one airline's market share on competitor airlines' frequencies, which is a critical component of such competitive interactions.

Furthermore, the assumption of constant load factors and constant aircraft seating capacities means that studies such as Brueckner [25, 26], Pels and Verhoef [80], and Perakis and Sun [81], cannot account for the possibility of increases in average number of passengers per flights (through increased load factors, or increased number of seats per aircraft, or both) as the slots become more expensive under congestion pricing. Consequently, delay cost reductions have often been considered as the only type of

benefit from congestion pricing. Most of the prior studies evaluate congestion pricing benefits in terms of overall societal welfare gain, rather than in terms of the benefits to airlines and passengers. Perakis and Sun conclude that, while congestion pricing leads to the welfare maximization solution, both airlines and passengers are worse off than without congestion pricing because the welfare gain from congestion pricing is in the form of the revenue generated from pricing [81]. Many of the prior congestion pricing studies propose some form of direct or indirect mechanisms for re-distribution of this revenue gain among the airlines if the congestion pricing scheme is to be attractive to the airlines.

Our models are able to explicitly capture the phenomenon of varying number of passenger per flight. In fact, as we show in Section 6.8, a reduction in operating costs is an important driver of the benefits of congestion pricing to the airlines, which has not been considered in any of the prior studies.

Schorr provided a model of airline frequency competition under flat pricing of airport slots [86]. He produced interesting results on the benefits of flat pricing, albeit focusing on symmetric equilibria for the somewhat restrictive case of identical airlines. We model airline frequency competition under congestion pricing using a popular market share model of frequency competition, which is similar to Schorr's model. The main objective of this research is to investigate the role of airline frequency competition under congestion pricing. We consider the general case of non-identical airlines and do not restrict our analysis to symmetric equilibria.

The major contributions of the research in this chapter are threefold. First, we develop a model for airline frequency competition that explicitly accounts for the relationship between the number of flights operated, number of seats flown and the number of passengers carried by an airline under slot pricing. To the best of author's knowledge, this is the first computational study that accounts for this relationship. Second, using a small hypothetical network, we evaluate the impacts of congestion prices on the various stakeholders and investigate the dependence of effectiveness of congestion pricing mechanisms on the different characteristics of airline competition in individual markets. Third, we provide computational results under flat prices, as well

as marginal cost pricing equilibrium. Our results show that variation in the number of passengers per flight plays a vital role in determining the degree of attractiveness of congestion pricing to the airlines. A significant part of impact of congestion pricing could not be accounted for using the earlier models based on assumptions of constant load factors and constant aircraft sizes.

6.3 Model

Our model of airline decision making is an extension of the *basic model* presented in Section 5.2. In all of our computational experiments, we assume that congestion pricing is being considered at a single airport. From here onwards, the airport which is under consideration for implementation of congestion pricing will be simply denoted as *the airport*. Before presenting the model, let us first define the relevant notation, the majority of which is similar to that described in Section 5.2.

- \mathbb{S}_a = Set of potential segments for airline a with destination at the airport
- p_{as} = Average fare charged by airline a on segment s
- Q_{as} = Number of passengers carried by airline a on segment s
- M_s = Total passenger demand on segment s
- C_{as} = Operating cost per flight for airline a on segment s
- S_{as} = Seating capacity of each flight of airline a on segment s
- α_s = Exponent of the S-curve relationship between market share and frequency share on the non-stop segment s
- U_a = Maximum number of slots that can be utilized by airline a at the airport
- L_a = Minimum number of slots that must be utilized by airline a at the airport
- \mathbb{A} = Set of all airlines

- \mathbb{A}_s = Set of all airlines whose set of potential segments includes s
- LF_{max} = Maximum average segment load factor
- c_a = Unit cost of flight delay to airline a (in \$/aircraft-minute)
- \mathbb{O} = Total number of operations at the airport
- $D(\cdot)$ = Average flight delay as a function of total number of operations at the airport
- $T(\cdot, \cdot)$ = toll (in \$) charged to an airline as a function of that airline's demand for operations and the total number of operations at the airport
- f_{as} = Daily frequency of flights for airline a on segment s

Our model of frequency competition, just like the one presented in Section 5.2, we will assume segment-based demand, that is, we will assume that all the passenger demand on a segment is that for non-stop travel on that segment and is independent of demand on other segments. The definition of a segment here is the same as that given in section 2.1, which was, *an origin and destination pair for non-stop flights*. So we will use the terms market and segment interchangeably in the rest of this chapter.

Expressions (6.1) through (6.6) describe the problem of deciding the flight frequencies as an optimization formulation from the perspective of a single airline. The objective function, 6.1, consists of three parts: 1) the difference between the total revenue and operating costs summed across all markets, 2) flight delay cost incurred by the airline, and 3) the congestion prices paid by the airline. Note that the operating cost inside the first summation excludes the cost due to flight delays. Flight delay cost is the product of the unit cost of flight delay (c_a) to that airline, the total number of operations of that airline at the airport ($\sum_{s \in \mathbb{S}_a} f_{as}$), and the average flight delay ($D(\cdot)$), which is a function of the total number of operations from all airlines at the airport ($\sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's}$). The congestion price (T) paid by the airline is decided by the airport administrator. It is reasonable to expect that T will be a function of total number of operations ($\sum_{s \in \mathbb{S}_a} f_{as}$) of airline a at the airport. The greater the

number of operations of airline a at the airport, the higher is the total congestion price paid by airline a . Furthermore, we can expect T to also be a function possibly of the total number of operations by all airlines at that airport. For the same number of operations of airline a at the airport, the greater the number of total operations of all airlines at the airport with congestion prices, the greater is the value of each slot, the greater is the additional delay cost imposed by airline a on other users, and consequently, the higher is the total congestion price paid by airline a . So we consider T to be a function of $\sum_{s \in \mathbb{S}_a} f_{as}$ and $\sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's}$. Note that this framework is general enough and it still accounts for the possibility that T is a constant (a constant function). The constraints (6.2) through (6.6) are same as those defined in Section 5.2 of Chapter 5.

$$\begin{aligned} \text{maximize} \quad & \sum_{s \in \mathbb{S}_a} (p_{as} Q_{as} - C_{as} f_{as}) - c_a * \left(\sum_{s \in \mathbb{S}_a} f_{as} \right) * D \left(\sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's} \right) \\ & - T \left(\sum_{s \in \mathbb{S}_a} f_{as}, \sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's} \right) \end{aligned} \quad (6.1)$$

$$\text{subject to:} \quad Q_{as} \leq \frac{f_{as}^{\alpha_s}}{\sum_{a' \in \mathbb{A}_s} f_{a's}^{\alpha_s}} M_s \forall s \in \mathbb{S}_a \quad (6.2)$$

$$Q_{as} \leq LF_{max} S_{as} f_{as} \forall s \in \mathbb{S}_a \quad (6.3)$$

$$\sum_{s \in \mathbb{S}_a} f_{as} \leq U_a \quad (6.4)$$

$$\sum_{s \in \mathbb{S}_a} f_{as} \geq L_a \quad (6.5)$$

$$f_{as} \in \mathbb{Z}^+ \forall s \in \mathbb{S}_a \quad (6.6)$$

In this chapter, we present two types of experiments. In the first type of experiments, we compute the impacts of continuously varying slot prices. In such experiments we will assume flat prices, that is, an equal price per slot is charged to all airlines. Under flat congestion prices, there is a constant price per slot paid by each airline regardless of the total number of operations of that airline. We use c_F to de-

note the flat congestion price per slot. Then the total congestion price paid by airline a is given by,

$$T\left(\sum_{s \in \mathbb{S}_a} f_{as}, \sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's}\right) = c_F \sum_{s \in \mathbb{S}_a} f_{as}. \quad (6.7)$$

In such experiments, we will need to compute the frequency decisions by all airlines such that the frequencies of each airline are optimal with respect to the frequencies of all the other airlines at that airport. We will compute one such competitive equilibrium at each c_F value. Let us denote these experiments as *type I experiments*.

On the other hand, in the second type of experiments, we will compute an equilibrium between prices and demand. In such experiments, the marginal delay cost imposed by each airline on other airlines equals the total congestion price paid by that airline. Mathematically, at an equilibrium,

$$T\left(\sum_{s \in \mathbb{S}_a} f_{as}, \sum_{a' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a'}} f_{a's}\right) = \sum_{a' \in \mathbb{A}, a' \neq a} \left(c_{a'} * \left(\sum_{s \in \mathbb{S}_{a'}} f_{a's} \right) \left(D \left(\sum_{a'' \in \mathbb{A}} \sum_{s \in \mathbb{S}_{a''}} f_{a''s} \right) - D \left(\sum_{a'' \in \mathbb{A}, a'' \neq a} \sum_{s \in \mathbb{S}_{a''}} f_{a''s} \right) \right) \right). \quad (6.8)$$

In such experiments, we will need to compute the frequency decisions by all airlines such that the frequencies of each airline are optimal with respect to the frequencies of all other airlines at that airport, as well as the corresponding prices T , that satisfy equation (6.8). Thus the demand-price equilibrium involves equilibrium decisions by all the airlines and by the airport administrator. Let us denote such experiments as *type II experiments*.

In our discussion so far, we have assumed the delay model $D(\cdot)$ as a given. In the next section, we describe our choice of delay model in detail.

6.4 Flight Delay Model

Flight delays at a congested airport are dependent largely on the utilization ratio, which is the ratio of flight demand to airport capacity. However, some part of flight delay is independent of congestion at that airport. Such delays are due to other effects such as propagated delays, delays due to mechanical failures, absence of crews etc. Some prior studies have developed detailed queuing theoretic models of delays as a function of the number of operations and solved them through simulation or numerical methods [40, 41]. Such detailed simulation models are beyond the scope of this research. We are interested in a simple delay function that captures the critical queuing theoretic elements. Many existing studies have used simplified assumptions for modeling delay as a function of utilization [30, 25, 80, 26]. Carlin and Park [30] as well as Pels and Verhoef [80] assumed delays to be an increasing linear function of the number of operations. A linear delay function is not very realistic given that it is well known that delays increase with utilization and the rate of increase itself increases very fast as the utilization ratio approaches 1.0. Brueckner [25, 26] assumed the delay cost to be a general non-decreasing and convex function of the number of airport operations. Zhang and Zhang suggested four standard conditions that a delay function must satisfy. Morrison [65], and Zhang and Zhang [100] used delay functions derived from steady-state queuing theory. The expression that we chose for the delay function is given in equation (6.9), with ρ being the utilization ration, that is, the ratio of the total number of scheduled operations to the airport capacity.

$$D = a \frac{1}{1 - \rho} + b \quad (6.9)$$

Here, a and b are parameters of the model that have to be estimated using actual delay data. This expression has a number of favorable properties. It is non-decreasing and convex in the number of operations as assumed by Brueckner [25, 26]. Also it satisfies all the four conditions specified by Zhang and Zhang [100]. The functional form is somewhat different from the one used by Morrison [65], and Zhang and Zhang

[100]. We considered using the exact functional form used by these two studies, but decided in favor of the chosen form because it gave a much better fit to the empirical delay data, as shown in Section 6.7.

6.5 Solution Algorithm

The solution algorithm used to solve the congestion pricing equilibrium problem builds on the myopic best response algorithm described in Sub-section 5.3.1. The solution algorithm used to obtain a solution to the *Type I experiments* is described in Sub-section 6.5.1 and that for the *Type II experiments* is described in Sub-section 6.5.2. In fact, the algorithm to obtain the solution for the *type I experiments* serves as a component, which is run multiple times, to obtain the solution for the *type II experiments*.

6.5.1 Solution Algorithm for *Type I Experiments*

For a given value of flat prices c_F , the total congestion price paid by an airline depends only on the total number of operations of that airline. Therefore, for a given level of average flight delay D , the objective function of each airline is the same as the objective function given by 5.1 in Chapter 5 except that the operating cost per flight, c_{as} is now replaced by $c_{as} + c_F + c_a * D$. Therefore the problem of computing the Nash equilibrium solution is the same as the problem presented in Section 5.2 of Chapter 5, for a given level of average flight delay D and for given flat prices d_F per slot. However, the added complication comes from the fact that the average flight delay itself is an increasing function of the total number of flight operations. Thus the equilibrium frequencies affect the average flight delay and average flight delay affects the equilibrium frequencies. In order to solve this problem iteratively, we employ an algorithm similar to that presented in Section 5.3, but we update the average flight delay after each iteration of the best response heuristic. The algorithm ends when 1) the flight frequency of each airline on each segment is optimal with respect to the frequency of competing airlines for the given level of average flight delays, and 2) the

change in the number of operations from iteration to iteration is within a pre-defined tolerance level.

Module I presented below is implemented to obtain solutions to all *type I experiments*. $f_a = [f_{as}]_{s \in \mathbb{S}_a}$ is the vector of frequencies for carrier a and $f_{-a} = [f_{a'}]_{a' \in \mathbb{A}, a' \neq a}$ is the vector formed by concatenating the frequency vectors of all competitors of airline a . So the set of frequencies of all airlines can be compactly denoted as $f = (f_a, f_{-a})$. The average flight delay is updated using moving averages so as to smooth any sharp fluctuations in the value of the sum of flight frequencies from iteration to iteration. We use $\tilde{\mathbb{O}}$ to denote the moving average estimate of the total number of operations at the airport. $\tilde{\mathbb{O}}$ is used to compute the average flight delay after each iteration. We use, for each airline a , the vector of actual frequency values $f_a^0 = [f_{as}^0]_{s \in \mathbb{S}_a}$ as the starting values of the flight frequencies for the best response algorithm. Also, we assume $\tilde{\mathbb{O}} = \sum_{s \in \mathbb{S}} \sum_{a \in \mathbb{A}_s} f_{as}^0$ at the starting point. We set the tolerance level $\epsilon_1 = 0.5$.

Module I

```

while (there exists a carrier  $a$  for whom  $f_a$  is not a best response to  $f_{-a}$ )
  or  $\left( \left| \tilde{\mathbb{O}} - \left( \sum_{s \in \mathbb{S}} \sum_{a \in \mathbb{A}_s} f_{as} \right) \right| \geq \epsilon_1 \right)$  do
     $f'_a \leftarrow$  some best response by  $a$  to  $f_{-a}$ 
     $f \leftarrow (f'_a, f_a)$ 
     $\tilde{\mathbb{O}} \leftarrow \left( \left( \sum_{s \in \mathbb{S}} \sum_{a \in \mathbb{A}_s} f_{as} \right) + \tilde{\mathbb{O}} \right) / 2$ 
return

```

6.5.2 Solution Algorithm for *Type II Experiments*

Module I, as described above, is used for obtaining a solution to all the *type I experiments*; and as a component for obtaining solutions to all the *type II experiments*.

At a solution to a marginal cost pricing (*type II*) experiment, all the conditions

mentioned in the *type I experiment* must be satisfied. Additionally, the total congestion price paid by each airline must equal the delay cost imposed by that airline on all the other airlines operating at the airport. For a given level of congestion prices, *module I* produces a set of flight frequencies such that the each airline's frequencies are optimal with respect to the frequencies of all other airlines and the average flight delay. However, the flight frequencies produced by a *module I* solution might lead to a different set of values of marginal delay costs imposed by each airline on the other airlines. Thus, the *module I* solution affects the congestion prices and the congestion prices affect the *module I* solution. Therefore, we solve this problem by an outer iterative algorithm, where we update the estimates of the total number of operations by each airline after every run of *module I*. The algorithm ends when the change in the number of operations from iteration to iteration is within a pre-defined tolerance level.

The full algorithm for obtaining solutions to all the *type II experiments*, is presented below. Updates are performed using moving averages so as to smooth large jumps in the congestion prices from iteration to iteration. We use $\hat{\mathcal{O}}_a$ to denote the moving average estimate of the total number of operations of airline a at the airport. We assume $\hat{\mathcal{O}}_a = \sum_{s \in \mathbb{S}_a} f_{as}^0 \forall a \in \mathbb{A}$ at the starting point. We set the tolerance level $\epsilon_2 = 0.1$.

Module II

while $\exists a \in \mathbb{A}$ such that $|\hat{\mathcal{O}}_a - \left(\sum_{s \in \mathbb{S}_a} f_{as}\right)| \geq \epsilon_2$ **do**

Run **Module I**

$$\hat{\mathcal{O}}_a \leftarrow \left(\left(\sum_{s \in \mathbb{S}_a} f_{as} \right) + \hat{\mathcal{O}}_a \right) / 2 \quad \forall a \in \mathbb{A}$$

return

6.6 Experimental Setup and Data Sources

We build a mathematical framework for evaluating the impacts of congestion pricing mechanisms on the various stakeholders, while explicitly modeling the competition between airlines. Using this framework, we generate insights into the extent to which the effectiveness of congestion pricing mechanisms depends on the various characteristics of the airline markets. In order to allow extensive analysis of the relationships between congestion pricing and the market characteristics, we opt for a simple experimental setup consisting of 3 airports and up to 5 airlines. We will denote the airport which is under consideration for the implementation of a congestion pricing mechanism as $AP0$. In order to have balanced operations, on average, an airline operates approximately the same number of flights per day in both directions on a segment. Therefore, in our experiments we focus on only the flights arriving at $AP0$ and not on those departing from $AP0$. Our setup includes two more airports, namely, $AP1$ and $AP2$. As mentioned before, our model assumes segment-based demand. Therefore, we assume that passengers demand non-stop service from $AP1$ to $AP0$ and from $AP2$ to $AP0$. Thus we have $AP1$ to $AP0$ and $AP2$ to $AP0$ as the two markets under consideration. Our experiments consist of two airlines, denoted as $AL1$ and $AL2$, operating in each of these two markets.

Our experiments are loosely based on data from two big markets into LaGuardia (LGA) airport at New York, namely, Logan (BOS) airport to LaGuardia (LGA) and Reagan (DCA) airport to LaGuardia (LGA). Furthermore, the data is loosely based on two major airlines, namely, Delta Airlines (DL) and US Airways (US) operating in each of these two markets. We obtained data on average fares, seating capacities, operating costs and passenger flows through the Bureau of Transportation Statistics (BTS) website. We obtained the average fares from the DB1B Market database [73]. We retrieved the operating cost values from the Form 41 financial data reported by the airlines in Schedules P-5.2, and Schedule P-7 [74, 76]. Aircraft seating capacities and passenger flows were obtained from the T100 Segment database [75]. Actual flight frequencies were obtained from the ASQP database [72]. The airport capacities for

estimation of the delay model were obtained from the FAA’s airport capacity benchmark report [44]. All the data used in our computational experiments corresponds to the 1st quarter of 2008.

In order to generate broader insights into the effectiveness of congestion pricing mechanisms, we varied important characteristics of our markets (e.g. sensitivity of the passengers to frequency, average fares, number of competitors in the markets, etc.) and tested their impacts on the effectiveness of the congestion pricing mechanism.

6.7 Delay Function Fitting

As mentioned in Section 6.4, the delay function represents the relationship between airport utilization and average delays to flights. To model this relationship, we used data including average flight delays, number of operations and the expected values of airport capacities from 34 major airports in the continental US. We tried a variety of functional forms and selected the one described by equation (6.9), because of its fit and intuitive appeal based on its similarity with queuing theoretic results.

As mentioned in Section 6.6, the expected values of airport capacities were obtained from the FAA’s airport capacity benchmark report [44]. The number of airport operations is obtained from the Aviation System Performance Metrics (ASPM) database maintained by the FAA [3]. The average utilization rate for each airport is calculated as the ratio of the average total number of operations (takeoffs and landings) that took place at that airport in the 18-hour time period from 6:00 am to midnight across all days of the 1st quarter of 2008, to the product of the expected value of hourly capacity of that airport (as given by the benchmark report) and 18. The average flight delay is computed as the average of delays to all the flights of the ASQP-reporting airlines landing and taking off from that airport during the 18-hour time period, from 6:00 am to midnight across all days of the 1st quarter of 2008.

Parameters a and b are estimated by using simple linear regression with average flight delays as the dependent variable and $\frac{1}{1-\rho}$ as the independent variable. Figure 6-1 shows the regression data and model fit. The regression analysis gave a strong

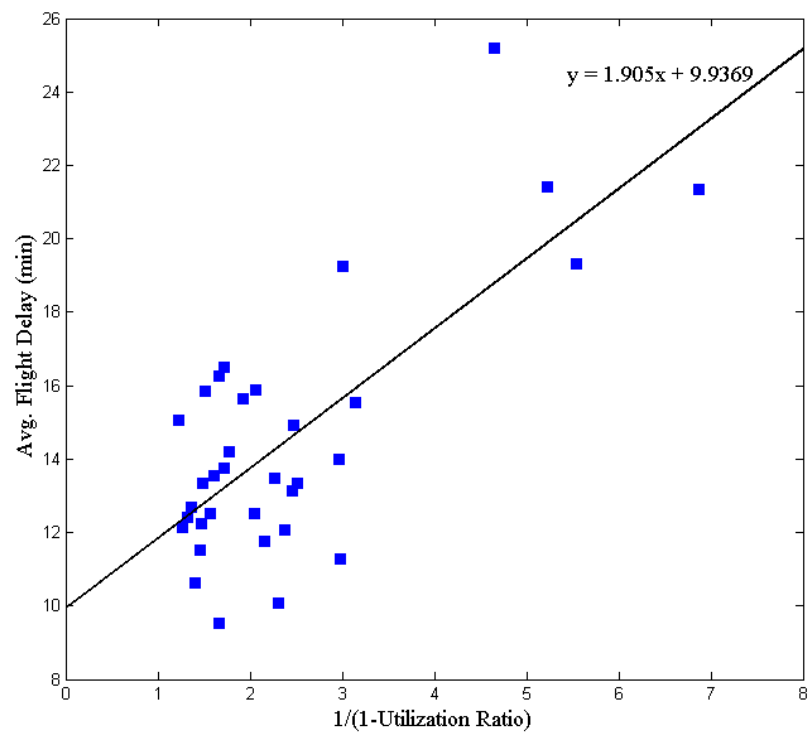


Figure 6-1: Estimation of delay function parameters using linear regression

goodness of fit, with an R^2 value of 53.37%. We will use this delay function extensively in our experiments described in Section 6.8 for calculating: 1) average flight delays; and 2) the marginal delay cost imposed by any one airline on others at the airport.

6.8 Numerical Results

Using the experimental setup described in Section 6.6, we conducted a series of computational experiments, some involving flat pricing of slots and others involving computation of a marginal cost pricing equilibrium. In Sub-section 6.8.1, we present the flat pricing results and in Sub-section 6.8.2, we present the marginal cost pricing equilibrium results.

6.8.1 Flat Pricing Results

Experiment 1: Administrative controls and zero slot prices

Let us first analyze the *base case* scenario where we assume that the slots at AP0 airport are restricted under administrative slot controls and no congestion pricing mechanism has been implemented. In particular, we assume that the upper (U_a) and the lower (L_a) bound on the total number of slots available to an airline are equal, and there is no congestion price being charged to the airlines. Mathematically, $U_a = L_a$ and $T = 0, \forall a \in \mathbb{A}$. In our experiment, the total number of flights operated by AL1 to AP0 from either AP1 or AP2 is 30 and that for AL2 it is 32. So the only decision to be taken by each airline is how to distribute the available slots across the two markets. For this base case, we assume AP1-AP0 and AP2-AP0 to be short-haul, business-intensive markets, similar to BOS-LGA and DCA-LGA. Therefore, we assume that the S-curve parameter α takes on a value of 1.5.

In Table 6.1, the frequency predictions based on the output of our solution algorithm are given in column titled *Model Frequency*.

Table 6.1: Model results for the *base case*

Market	Carrier	Avg. Fare	Model Freq.	Seats/Flight	Passengers	Operating Cost (\$)	Revenue (\$)	Profit (\$)
AP1	AL1	161	15	134	726	110,813	116,886	6,073
AP1	AL2	163	16	124	797	110,984	129,911	18,927
AP2	AL1	152	15	134	741	110,269	112,632	2,363
AP2	AL2	160	16	124	813	108,747	130,080	21,333

Experiment 2: No administrative controls and zero slot prices

In this experiment, we assume that there are neither any administrative slot controls nor any slot prices. Mathematically, $U_a = \infty, L_a = 0$, and $T = 0, \forall a \in \mathbb{A}$. We find that the model predictions in this experiment match the frequencies predicted in the previous experiment exactly. Both of these are represented in Table 6.1. The results presented in Table 6.1 will be referred to as the *base case* and will be used as a reference point for our remaining experiments, all of which involve congestion prices.

In Section 4.7 of Chapter 4, we proved that the level of congestion introduced by airline competition is an increasing function of three factors, namely, 1) the S-curve parameter α (which is nothing but the sensitivity of passenger demand to frequency), 2) the market’s gross profit margin (the ratio of average fare to operating cost per seat), 3) the number of competitors. The higher the value of any of these three factors, the greater is the incentive for the airlines to schedule more frequent flights, and hence the greater the adverse impact of competition on congestion. In the absence of congestion prices, airline competition leads to congestion. Thus, it is reasonable to expect that the success of a congestion pricing mechanism depends directly on the extent to which the congestion prices can discourage the airlines from scheduling frequent flights. Therefore, each of these three factors is expected to play a critical role in determining the success of a congestion pricing mechanism. In each of the next three experiments, we analyze the impact of one of these three factors. In particular, in order to analyze the impact of gross profit margin (which is the ratio of average fare to operating cost per seat), we will vary the average fares.

In the next three experiments, we evaluate the impacts of varying the slot prices

assuming a flat pricing scheme. Obviously, an exceedingly high value of congestion price per slot would result in airlines no longer being able to operate flights profitably. So in each case we make sure that we do not reach such high levels of congestion prices per slot.

Experiment 3: Effect of sensitivity of passenger demand to frequency

As mentioned earlier, an airline provides more frequency in a market to attract a higher share of the market. The extent to which the distribution of market share is affected by frequencies is what we call the sensitivity of passenger demand to frequency. In our S-curve model (as shown in equation 6.1), sensitivity of passengers to frequency is represented by parameter α . In the *base case*, we assumed $\alpha = 1.5$. In this experiment, we vary the congestion price per slot and evaluate the impacts on three important system performance metrics, namely, demand for airport operations, total delay cost to passengers and total operating profits of the airlines for six different values (1.5, 1.4, 1.3, 1.2, 1.1, and 1.0) of parameter α .

Figures 6-2, 6-3 and 6-4 show the variation in these three performance metrics with increasing slot prices for different α values. The flat slot price in \$/slot is on the x-axis. In Figures 6-2 and 6-3, y-axis shows the normalized value of demand for airport operations and total delay costs to the passengers respectively. The normalization is performed such that the value for zero slot prices equals 100. Figure 6-4 has the change in operating profit margin percentage for the airlines on the y-axis. Operating profit margin percentage is defined as the ratio of total operating profit earned by both airlines in both markets to total fare revenue generated by both airlines in both markets.

Each line in each of these three figures corresponds to a different value of α . The plots are not smooth because of the integrality constraints on the number of flights in each market. Each time the slot price exceeds a certain threshold value, it abruptly becomes unprofitable to operate the last flight being operated by an airline. So the demand drops in a lumpy fashion, resulting in non-smooth trends in the performance metrics. Therefore, rather than looking at any single slot price for comparison across

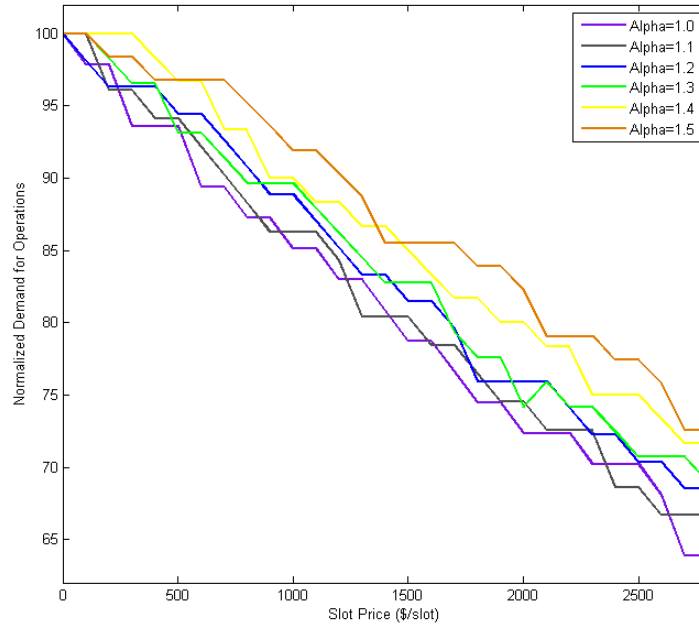


Figure 6-2: Demand for airport operations as a function of slot prices for different values of α

different α values, we base our conclusions on the overall trends that can be observed from Figures 6-2, 6-3 and 6-4.

As shown in Figure 6-2, with an increase in slot price, the total demand for airport operations falls and consequently, the total passenger delay cost decreases, both of which are intuitively reasonable. The impact of increasing slot prices on change in operating profit margin percentage is more complicated. Due to increasing slot prices, the airlines are incentivized to reduce their flight frequency, thus increasing load factors. Therefore, the airlines benefit from lower operating costs, as well as from a reduction in flight delay costs. Depending on whether these benefits partially or fully offset the total congestion price paid by the airlines, the airline profits increase or decrease as a result of congestion prices.

For a given slot price, the total demand for airport slots increases with an increase in α value. At higher values of α , passengers are more sensitive to frequency, which means that for a given level of slot price, airlines' demand larger numbers of slots. As a result, for a given slot price, the higher the α value, the greater is the slot demand

and the larger is the total delay costs to the passengers. Also, because airlines are comparatively more reluctant to reduce flight frequency at a higher α value, airline profits are lower at higher α . Thus, the curvature of the S-curve (α) plays a crucial role in determining the effectiveness of slot prices. At high values of α , slot prices result in very little reduction in demand and delays and a significant reduction in airline profits. However, at lower α values, airline profits increase under congestion pricing, due to a significant reduction in operating costs and in delay costs. Furthermore, a greater reduction in passenger delays opens up the possibility of some increase in average fares. The airlines could monetize a part of the passengers' gains through increased fares, resulting in further increases in airline profits.

As shown in Figure 6-4, for flat pricing in the absence of average fare increases, operating profits are decreasing with increasing slot prices for α values of 1.5, 1.4, 1.3, and 1.2. For α values of 1.1 and 1.0, there is a slight increase in operating profits in some cases. But the operating profit increase is never more than 1% or 2% across different values of α and across different values of slot prices. In Sub-section 6.8.2, however, we show that under marginal cost pricing, airline profits could actually increase with congestion pricing even under the constant average fares assumption. An important reason for this difference is the fact that marginal cost pricing explicitly accounts for the non-atomistic nature of the airport congestion pricing problem. We look at this phenomenon in more details in Sub-section 6.8.2.

Experiment 4: Effect of average fare

Apart from the sensitivity of passenger demand to frequency, the ratio of average fare to operating cost per seat (which we term as *gross profit margin* or GPM) determines the effectiveness of congestion pricing. Markets with high GPM are the markets where fares are relatively high compared to the operating costs, which indicates that the passengers are willing to pay more for a given travel distance.

For markets with higher GPM, the passengers are also more valuable to the airlines, as they provide more revenue compared to operating cost per seat. So the airlines have an even greater incentive to acquire more market share and hence, are

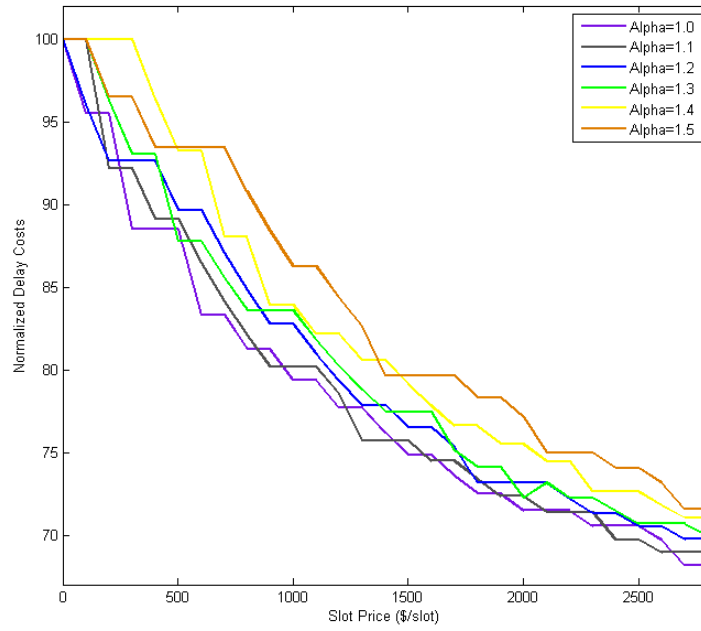


Figure 6-3: Passenger delay costs as a function of slot prices for different values of α

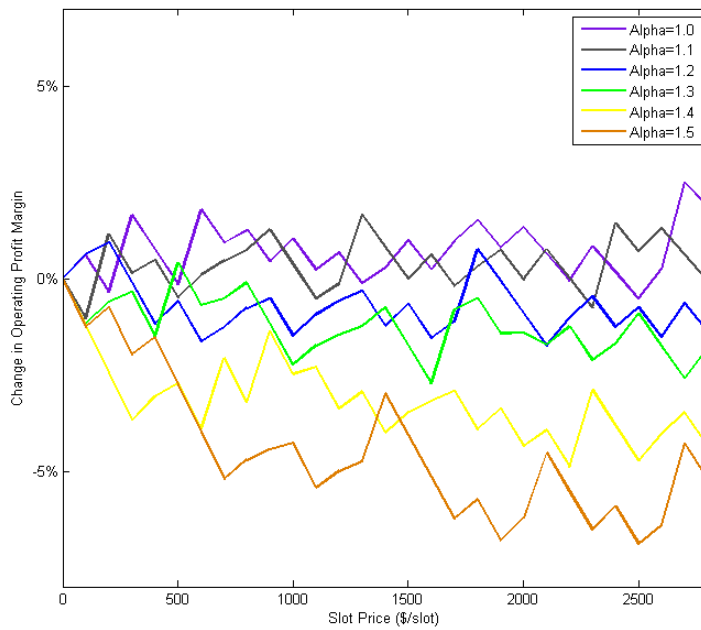


Figure 6-4: Total operating profits of the airlines as a function of slot prices for different values of α

more reluctant to give up market share by decreasing the number of flights even under congestion pricing. This hypothesis is confirmed by the trends shown in Figures 6-5 and 6-6. In these two figures (as well as in Figure 6-7), we vary the price per slot (on the x-axis) and plot the corresponding variation in the respective performance metric on the y-axis. The entities on the y-axis in Figures 6-5, 6-6 and 6-7 are the same as those for Figures 6-2, 6-3 and 6-4 respectively. We vary the average fare (thus varying the GPM) on each segment from -20% to +30% of the *base case* value in increments of 10% each. There are 6 lines in each of these three figures, one line corresponding to each value of GPM.

For Figures 6-5, 6-6, and 6-7, we assume $\alpha = 1.3$, and hold the operating cost and flight seating capacities equal to those in the *base case* for each segment for each airline. For this experiment, we considered the possibility of using the *base case* value of $\alpha = 1.5$, but decided against it because of the following issues. In this experiment, we vary the average fares above and below the *base case* values. For a high *alpha* value (such as 1.4 or 1.5) and for a low average fare value, (such as 0.8 times the *base case* fare), the combination of low fares and extreme sensitivity of passenger demand to frequency, makes it impossible for the airlines to continue operating profitably even at moderately high congestion price per slot, resulting in discontinuation of service. So the range of slot prices under consideration gets reduced. Therefore, in order to improve the expository power of our analysis and the following discussion, we decided in favor of using an α value of 1.3 instead of 1.5 for this particular experiment.

As expected, at a given slot price, there is a smaller decrease in airport slot demand and in passenger delay costs for a higher value of GPM than for a lower value of GPM. Also, at a higher value of GPM, a given slot price yields a smaller reduction in flight frequencies, thus leading to a smaller increase (or greater decrease) in operating profits. By the same reasoning as given in the explanation of results from the previous experiment, airline operating profits could be higher if airlines are able to monetize a part of the passengers' gain (in terms of lower passenger delay costs) resulting from congestion prices through increased fares. In the absence of such fare increases, Figure 6-7 shows that the airline profits decrease (or increase by

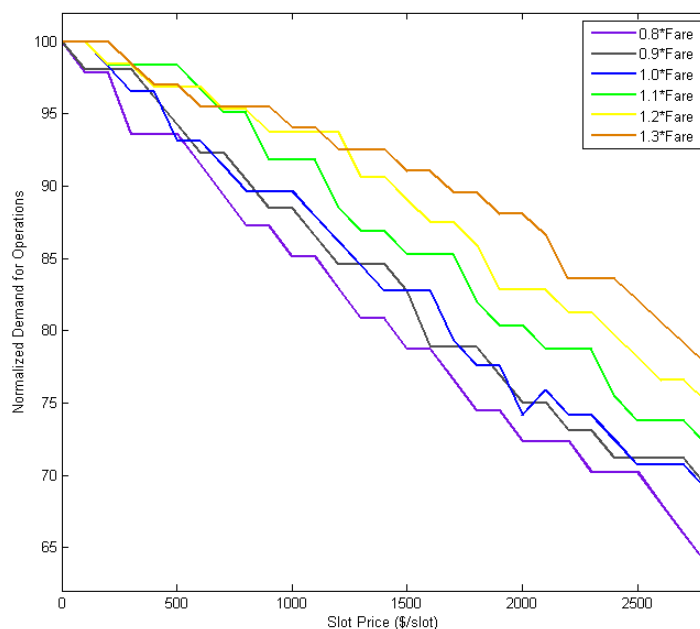


Figure 6-5: Demand for airport operations as a function of slot prices for different fare levels

less than 2%) across all slot price levels and across all the 6 levels of GPM considered here. However, as shown in Sub-section 6.8.2, we find that the profit increases under a marginal cost pricing equilibrium compared to that without congestion pricing. We explain this phenomenon in detail in Sub-section 6.8.2.

Experiment 5: Effect of the number of competitors

As shown in Section 4.7 of Chapter 4, apart from the sensitivity of passenger demand to frequency (α) and the gross profit margin (GPM), the number of competing airlines in a market also affects the extent of congestion introduced by competition. However, the effect of the number of competitors on the extent of congestion is not as strong as that of α or GPM. Figures 6-8, 6-9, and 6-10 show the impact of variation of the congestion price per slot on the demand for airport operations, total passenger delay costs, and the total operating profits to the airlines for different numbers of competitors. The congestion price per slot is on the x-axis. The entities on the y-axes of Figures 6-8, 6-9, and 6-10 are the same as those for Figures 6-2, 6-3 and 6-4

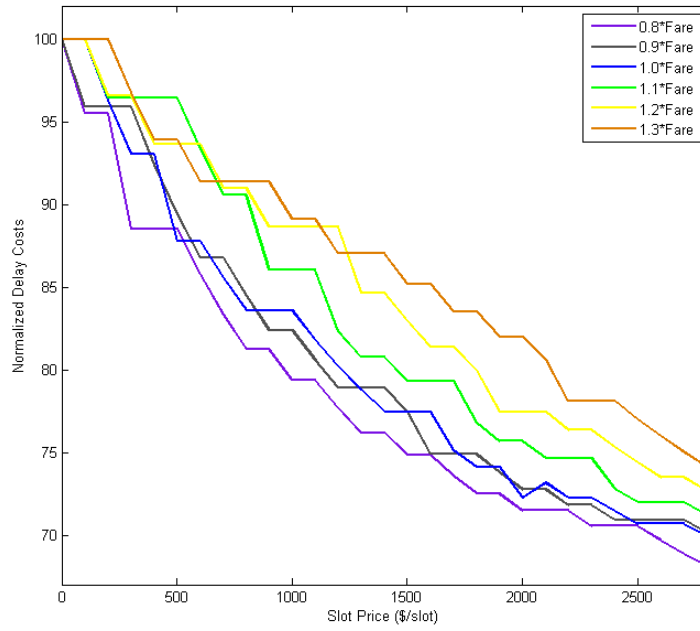


Figure 6-6: Passenger delay costs as a function of slot prices for different fare levels

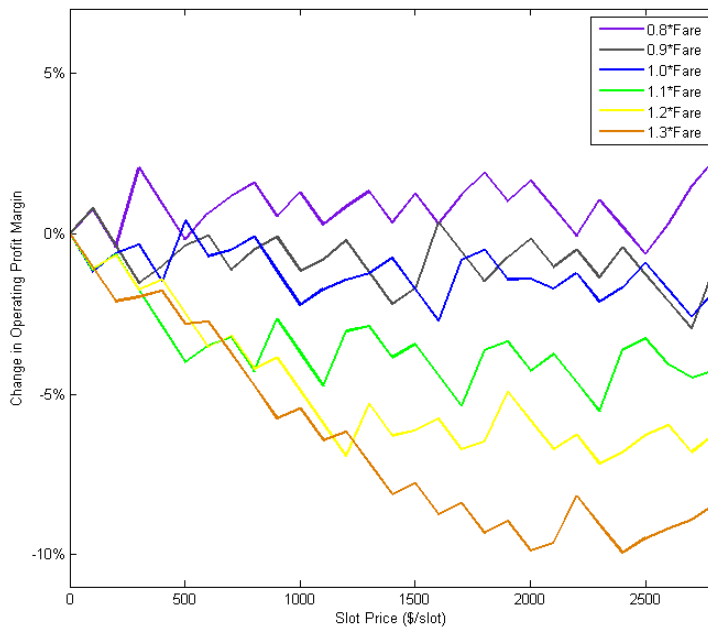


Figure 6-7: Total operating profits of the airlines as a function of slot prices for different fare levels

respectively.

As shown in Section 4.7, for a symmetric game, the maximum number of competitors which can have a non-zero frequency at a Nash equilibrium, cannot exceed $\frac{\alpha}{\alpha-1}$. Extending the same intuition to asymmetric games, we conclude that at higher values of α , the maximum number of competitors with non-zero frequencies at a Nash equilibrium will be low. We considered different values of α for this computational experiment and the computational results confirmed our intuition. In this experiment, we vary the number of competitors up to 5 and hence we need to use a lower value of α for enhancing the expository power of our analyses. So we decided in favor of using $\alpha = 1.0$ for this experiment. The operating cost, average fare and the seating capacities used for this experiment were the same as those for the *base case* on each segment operated by AL1 and AL2.

We vary the number of competitors from 2 to 5. For the 3-, 4-, and 5-competitor cases, we assume that respectively 1, 2, and 3 additional competitors compete with AL1 and AL2 in each market. All additional competing airlines are assumed to have average fares, seating capacities and operating costs on each segment equal to the average values for DL and US on that respective segment.

We proved, in Section 4.7, that the price of anarchy for the airline frequency competition game, is given by $\alpha * \frac{ps}{C} * \frac{n-1}{n}$ where n is the largest integer not exceeding $\min(N, \frac{\alpha}{\alpha-1})$. Here, N is the number of competing airlines. Thus, the level of congestion introduced by competition (given by $\alpha * \text{GPM} * \frac{N-1}{N}$ assuming $N < \frac{\alpha}{\alpha-1}$) increases 1) linearly with α , 2) linearly with GPM, but 3) slower than linearly with the number of competitors (N). The trends in Figures 6-8, 6-9, and 6-10 are consistent with these earlier results from Section 4.7.

For any given slot price, the effect of the number of competitors on the demand for airport operations and the passenger delay costs is not as high as that of α or GPM. But the reduction in demand for airport operations and the reduction in passenger delay costs does decrease with an increase in the number of competitors for the same slot prices. As shown in Figure 6-10, the effect of the number of competitors on airline profit is more obvious. As the number of competing airlines increases, the operating

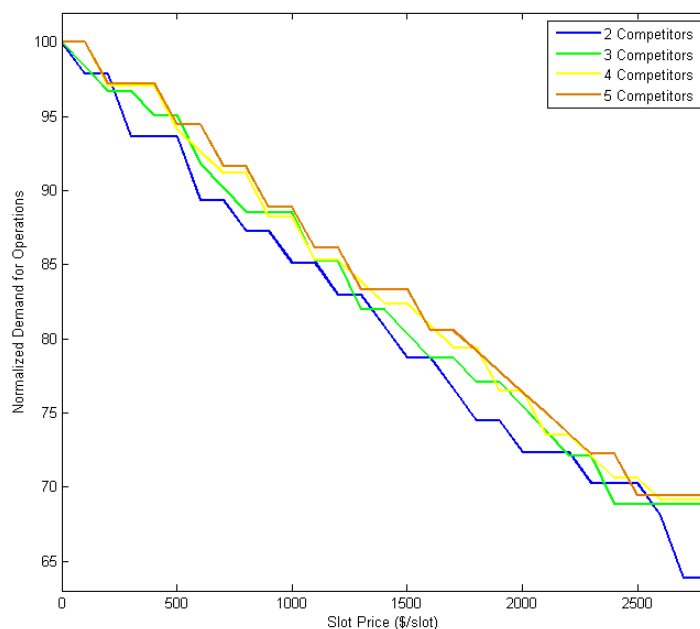


Figure 6-8: Demand for airport operations as a function of slot prices for different number of competitors

profit margin decreases, for the same slot prices.

Partial monetization of passenger delay reduction gains through increases in average fares can increase airline profits beyond the values shown in Figure 6-10. However, assuming constant average fares, we observe that the airlines' operating profits decrease with increasing flat slot prices, just as we observed in Experiment 3 and 4. We will contrast these results with those in Sub-section 6.8.2.

6.8.2 Marginal Cost Pricing Results

Experiment 6: Marginal cost pricing equilibrium

In Experiments 3, 4, and 5, we assumed flat prices per slot across different airlines, and analyzed the impacts on three different performance metrics. According to the microeconomic theory, social welfare is maximized if each user of a public resource, such as airport capacity, pays exactly the marginal cost imposed by that user on the remaining users [97, 62]. However, in the airport congestion pricing literature, such

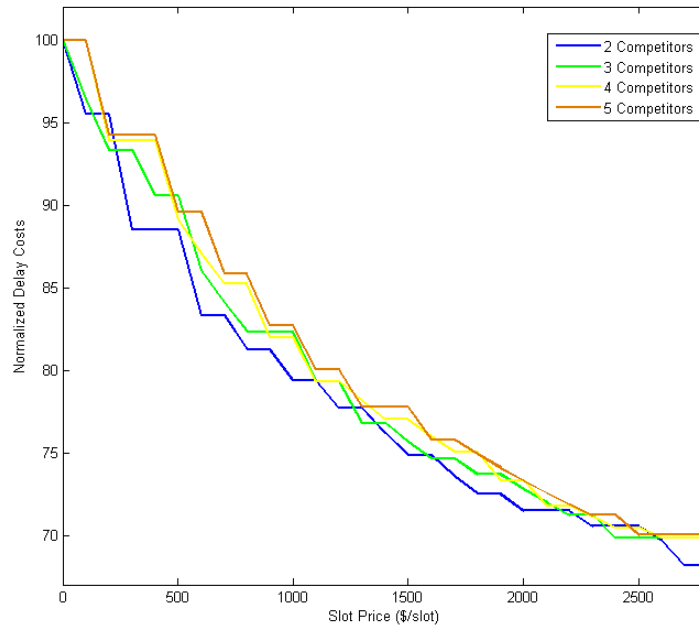


Figure 6-9: Passenger delay costs as a function of slot prices for different number of competitors

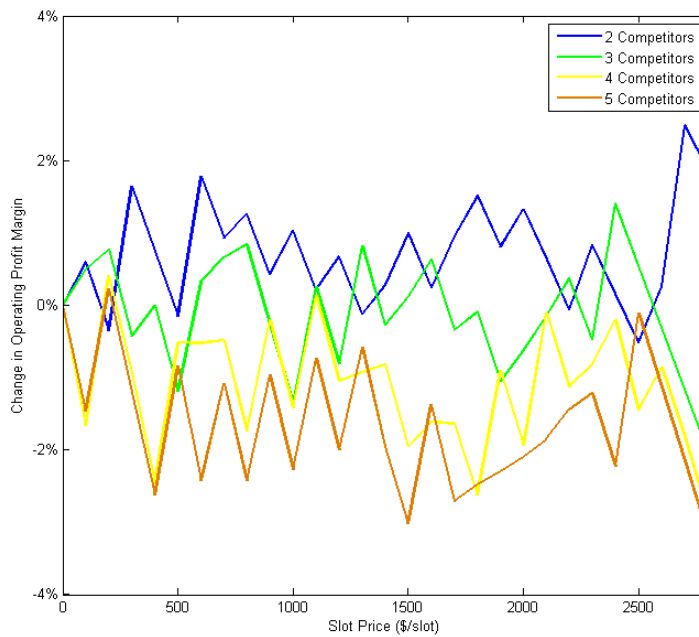


Figure 6-10: Total operating profits of the airlines as a function of slot prices for different number of competitors

as Daniel [40, 41], Brueckner [25, 26], Pels and Verhoef [80] etc., it is recognized that airlines being non-atomistic users of the airport resources, internalize a part of the airport congestion cost and in order to achieve the social optimal, each user must be charged a congestion toll which may be different across users. In this experiment, we compute the marginal cost pricing equilibrium and evaluate the various performance metrics at the point of equilibrium, where the marginal delay cost imposed by each user on all other users equals the congestion price paid by that user.

Table 6.2: Marginal cost pricing equilibrium results (Experiment 6)

	Case	% Change in Number of Operations	% Change in Passenger Delay Costs	Change in Operating Profit Margin			
				Due to Tolls	Due to Delays	Due to Op. Cost	Overall
6a	$\alpha = 1.0$	-10.64%	-16.67%	-2.58%	2.25%	4.69%	4.35%
	$\alpha = 1.1$	-9.80%	-15.90%	-2.97%	2.36%	4.69%	4.08%
	$\alpha = 1.2$	-9.26%	-15.22%	-3.23%	2.40%	4.86%	4.02%
	$\alpha = 1.3$	-8.62%	-14.44%	-3.61%	2.45%	4.86%	3.70%
	$\alpha = 1.4$	-8.33%	-14.14%	-3.81%	2.50%	4.89%	3.58%
	$\alpha = 1.5$	-6.45%	-11.61%	-4.42%	2.15%	4.09%	1.82%
6b	0.8*Base Fare	-10.64%	-16.67%	-3.23%	2.81%	5.86%	5.44%
	0.9*Base Fare	-9.62%	-15.56%	-3.37%	2.61%	5.43%	4.67%
	1.0*Base Fare	-8.62%	-14.44%	-3.61%	2.45%	4.86%	3.70%
	1.1*Base Fare	-6.56%	-11.84%	-3.94%	1.97%	3.68%	1.71%
	1.2*Base Fare	-6.25%	-11.37%	-3.87%	1.82%	3.40%	1.35%
	1.3*Base Fare	-5.97%	-10.93%	-3.46%	1.69%	2.78%	1.01%
6c	2 Competitors	-10.64%	-16.67%	-2.58%	2.25%	4.69%	4.35%
	3 Competitors	-11.48%	-17.67%	-5.26%	3.08%	6.59%	4.41%
	4 Competitors	-11.76%	-18.01%	-7.30%	3.50%	7.34%	3.53%
	5 Competitors	-12.50%	-18.70%	-8.42%	3.82%	8.36%	3.76%

Table 6.2 summarizes the results from this experiment. We conduct three Sub-experiments (6a, 6b, and 6c), one each for understanding the impacts of variation in α , GPM, and the number of competitors, on the effectiveness of marginal cost congestion pricing. The first column provides the Sub-experiment designation, while the second column describes the exact case including the entity being varied in each Sub-experiment. The next two columns list the percentage change in the number of

operations and in the passenger delay costs respectively. All changes are reported with respect to the zero price case. The next three columns describe the change in operating profit margin due to each cause. The main three causes of change in profit margin are the congestion tolls paid by the airlines (which reduces the profit margin), a reduction in delay costs (which increases the profit margin) and a reduction in operating costs due to a reduction in the number of flights operated (which also increases the profit margin). The last column reports the overall change in operating profit margin. All the changes in operating profit margin are computed as a percentage of total fare revenue. For example, for the first row of Sub-experiment 6a, the marginal cost equilibrium reduces the total demand for airport operations by 10.64%, reduces the total passenger delay cost by 16.67% and results in a 4.35% increase in the operating profit margin for the airlines. Out of this 4.35% increase, 2.25% is due to a reduction in delay costs to the airlines, 4.69% is due to an operating cost reduction due to fewer flights, while the congestion tolls paid by the airlines reduce the profit margin by 2.58%.

In sub-experiment 6a, we vary the α value between 1.0 and 1.5. As the α value decreases, a greater percentage reduction is achieved in the number of operations as well as in the passenger delay costs at equilibrium. Furthermore, percentage improvement in airline profits at equilibrium is also greater at lower values of α . Thus, congestion pricing can be greatly beneficial in markets with lower sensitivity of demand to frequency. These results are consistent with the results from Experiment 3. Reduction in the number of operations, reduction in the passenger delays and increase in airline profits, are all greater at lower values α .

In Sub-experiment 6b, we vary the average fares between 0.8 times to 1.3 times the base fares. Just as in Experiment 4, we assumed $\alpha = 1.3$ for this Sub-experiment and used the operating costs and seating capacities equal to those in the *base case* experiment. Intuitively, in markets with a higher ratio of average fare to operating cost per seat, airlines continue to find it profitable to operate high frequency service even if it means paying congestion prices. Loosely speaking, in such markets, an airline's monetary gains from additional flights are more than the marginal delay cost

they impose on other users. Thus, congestion pricing has a more positive effect in markets where the ratio of average fare to operating cost per seat is lower rather than higher. These results are consistent with our flat-price results in Experiment 4.

In Sub-experiment 6c, we vary the number of competing airlines from 2 to 5. We assumed $\alpha = 1.0$, just as we did in Experiment 5. The average fares, operating costs and seating capacities are assumed to be those in the *base case* experiment. As shown in Chapter 4, the extent of congestion introduced by competition increases with an increase in the number of competitors, but the increase is slower than linear. Consequently, as shown in Experiment 5, for a given slot price, reduction in airport operations, reduction in passenger delays, and increase in airline profits are all greater for a smaller number of competitors. But the effect is not as strong as the effect of sensitivity of demand to frequency or that of gross profit margin (as shown in Experiments 3 and 4, respectively). But as shown in Table 6.2, with increases in the number of competitors, the percentage reduction in operations and the percentage reduction in delays increase while the increase in operating profit margin shows no clear trend.

On the face of it, the results in this Sub-experiment appear to be inconsistent with those in Experiment 5, but the disparity can be easily explained by noting that these results are for a marginal cost pricing equilibrium, while those in Experiment 5 are for a given level of slot prices. As the number of competitors increases, the marginal delay cost imposed by each airline on all other airlines also increases, which in turn increases the slot prices under marginal cost pricing, leading to a greater reduction in operations and delays. Thus, for greater number of competitors, the airlines benefit more from reduction in operating costs and delays, but at the same time have to pay higher slot prices. Thus, the net effect of an increase in the number of competitors on the operating profit margin is complicated and no clear trend is observed. This phenomenon cannot be observed under the flat pricing regime (in Experiment 5), leading to the apparent inconsistency between results in Experiment 5 and in Sub-experiment 6c.

Beyond these factors affecting the effectiveness of congestion prices, results pre-

sented in Table 6.2 show airline profit increases due to marginal cost pricing, while the results from the flat pricing experiments (Experiments 3, 4, and 5) show the operating profits to be either decreasing, or very modestly increasing with congestion prices in most cases. This is another interesting manifestation of the difference between flat prices and marginal cost prices. Under marginal cost pricing, an airline has to pay a congestion price equal to the cost of the increase in delays to other airlines because of the operations of that airline. Therefore, the incremental price of a marginal slot to an airline is often substantially greater than the average price being paid by each airline. So under marginal cost pricing, the additional price of an extra operation becomes prohibitively high at a level of demand where the actual average congestion price per slot being paid by the airlines is still relatively low. The result is that airlines are discouraged from adding extra flight frequencies even though they continue to pay a relatively small congestion price per slot for the flights they operate. In this way, marginal cost pricing can discourage airlines from increasing airport operations without penalizing them with an exceedingly high congestion price per slot, leading to a lower level of congestion and higher profits for the airlines. This is a ramification of the fact that some of the delay is internalized by the airlines.

Consider a concrete example of the phenomenon described above. Specifically, consider the case in the sixth row below the header row of Table 6.2 where $\alpha = 1.5$, AL1 and AL2 are the only two competitors, and the average fares, operating costs and seating capacities are the same as those in the *base case*. Under the marginal cost pricing equilibrium, the total number of operations is reduced by 6.45% (from 62 to 58) and passenger delay costs are reduced by 11.61%. At equilibrium, the price of each additional slot is approximately \$1046, averaged across the two airlines (\$1118/slot for AL1 and \$974/slot for AL2). However, the average congestion price being paid by the airlines is approximately \$367. As a result, the total operating profit margin increases by 1.82% compared to the base case. Under flat pricing case, in order to achieve the same 6.45% reduction in operations, the marginal as well as the average congestion price paid by the airlines equals \$900 per slot. As a result, the total operating profit margin decreases by 4.43%. Alternatively, at a flat congestion price

of \$367 per slot, airline operations are reduced only by 3.23%, resulting in passenger delay cost reductions of just 6.54%, and a 1.12% decrease in total operating profit margin of the airlines.

It is important to note that these results are for a relatively small number of airlines at the airport; 2 in most experiments and 3, 4, or 5 in the remaining experiments. Thus the large differences between flat pricing and marginal cost pricing results obtained in our experiments, are partly owing to the small number of airlines, which internalize a large part of the delays. It should be noted that for airports with many airlines each contributing a smaller part of the operations at that airport, the difference between the flat and marginal cost pricing results is expected to be lower.

Finally, it must be noted that a large proportion of the congestion pricing benefits to the airlines come from a reduction in operating costs because of operating a smaller number of flights. In fact, in many cases in the Table 6.2, the benefits due to delay reduction are more than compensated by the congestion tolls. Hence operating cost reduction due to a smaller number of flights is a prime reason behind the profit increases.

6.9 Summary

Congestion pricing has already been implemented in practice fairly widely for congestion reduction in other contexts, such as roadway infrastructure, which involve a large number of individual users such that 1) each user uses only a small portion of the capacity of the public resource (atomistic users), and 2) the users do not directly compete with each other (non-competing users). The former point implies that the users of such resources internalize a negligibly small part of the delay costs and the latter point implies that beyond the congestion costs that the users impose on each other, the value derived by the users through their usage of the public resource does not depend on the usage by other users. Either of these points is not true for the utilization of airport resources by airlines. In order to understand the impacts of airport congestion pricing, it is critical to model the airline frequency competition

and the partial internalization of delay costs imposed by an airline. In this chapter, we model airline frequency competition under congestion prices and investigate the differences between the atomistic and non-atomistic pricing. We identify a variety of characteristics of airline markets that critically determine the effectiveness of airport congestion pricing mechanisms.

Our model of frequency competition under slot pricing is an extension of the models developed in Chapters 4 and 5, which in turn are consistent with a popular characterization of the relationship between market share and frequency share. In this chapter, we developed and used an efficient algorithm that iteratively solves the problem of equilibrium computation under congestion pricing.

Our results showed that the frequency sensitivity of passenger demand (or the exponent in the S-curve relationship), a measure of the gross profit margin (or the ratio of average fare to operating cost per seat), and the number of competitors in a market, critically affect the effectiveness of a congestion pricing mechanism. As expected, slot prices reduce congestion by reducing the number of operations at the airport. But the impact of slot prices on airlines' operating profit margin is not that straightforward. Airlines benefit from reduction in operating costs because of fewer flights and higher load factors, and also benefit from the delay cost reduction. The net impact of airline profit margin depends on whether these benefits are sufficient to offset the slot prices paid by the airlines.

While flat pricing has the advantage of being comparatively easier to understand and implement, we found that the marginal cost pricing (non-atomistic pricing) is more effective in reducing congestion without penalizing the airlines with exceedingly high congestion toll payments. Marginal cost pricing discourages the airlines from scheduling additional operations through high incremental price for an additional slot, while keeping the average price for the purchased slots relatively low. On the other hand, for flat pricing, the incremental and average congestion prices are equal by definition. As a result, compared to flat pricing, marginal cost pricing results in higher operating profit margins for the airlines for the same level of congestion reduction. However, it must be noted that these differences between flat and marginal

cost pricing paradigms were amplified because our example network involved a small number of airlines.

The main aim of this research was to develop a model of congestion pricing under airline frequency competition and to generate insights into the critical factors that affect the effectiveness of a congestion pricing mechanism. But it must be noted that our research in this chapter was conducted for a small hypothetical network, consisting of 2 markets and 2, 3, 4 or 5 airlines. In order to fully quantify the effects of congestion pricing, it is necessary to develop a full-scale case study of a congested airport. Furthermore, we made a number of assumptions including a segment-based demand, constant average fares and constant aircraft sizes. In order to conduct a full investigation of the impacts of congestion pricing, the extent of validity of these assumptions needs to be assessed and the effects of relaxing these assumptions need to be quantified.

Although, this evidence based on a small hypothetical network is insufficient to conclude whether the net effect of congestion pricing on airline profits will be positive or negative, the results clearly show that appropriately capturing the variation in number of passengers per flight could have a decisive impact on the answer to this vital question. Therefore, an interesting followup study would be a more detailed experiment with a much larger real dataset. Our results and insights based on a small network provide sufficient motivation for a full-fledged analysis of airline frequency competition under congestion pricing, and the models and algorithms developed by us in this research will serve as useful tools for this followup analysis.

Chapter 7

Conclusions and Recent Events

In this thesis, we modeled the interactions between the various decision-makers in the National Aviation System (NAS) of the United States as a system of multiple autonomous agents. We evaluated the dynamics between competition and congestion in the NAS from the perspectives of these different agents and proposed measures for congestion mitigation that are beneficial to these various stakeholders. In this chapter, we discuss the main conclusions of our study especially in the context of recent events in the US airline industry. Section 7.1 focuses on our main conclusions and contributions in each chapter of this thesis. These conclusions are based on our theoretical as well as computational analyses. All computational results are based on data either from the year 2007 or from the early part of year 2008. Section 7.2 discusses some of the more recent changes in the US airline industry, and the applicability and validation of our conclusions in the context of these changes.

7.1 Conclusions and Contributions

Scheduling and operational decisions by airlines require them to balance the often-conflicting objectives of minimizing the cost of operating their schedules and maximizing their attractiveness to the passengers relative to the attractiveness of their competitors' schedules. We found that airline competition is an important determinant of the level of congestion and this relationship between competition and congestion

critically affects the level of efficiency in the NAS, the level-of-service experienced by the passengers and the profits earned by the airlines. Furthermore, accurate understanding and modeling of this relationship is vital for evaluating the impacts of any congestion mitigation strategy on the different stakeholders of the system.

The major contributions of this thesis can be divided into two broad categories. In Chapters 2, 3, and 4 we evaluate the impacts of competition and congestion from the perspectives of three major groups of stakeholders, and in Chapters 5 and 6, we propose and assess mechanisms for congestion mitigation that are beneficial for the different stakeholders.

In the next 5 sub-sections, we summarize the main conclusions and contributions of each chapter, from Chapter 2 through 6, sequentially.

7.1.1 Chapter 2: Minimization of System-wide Delays in the Absence of Competition

Efficiency is one of the most important objectives from the perspectives of the system operators. In Chapter 2, we measure the extent to which airport capacity in the US domestic air transportation network is being inefficiently utilized because of airlines' competitive scheduling practices. By comparing the delays in a hypothetical network of a single, delay-minimizing airline with the delays in the existing network of multiple, competing, profit-maximizing airlines, we demonstrated that airline competition introduces a large degree of inefficiency in NAS resource utilization. Delays could be substantially lower in the absence of airline competition. Obviously, getting rid of (or even reducing the extent of) competition between airlines is a highly unrealistic strategy in the real world. It is also not the point of the research presented in this chapter. However, these results show that there is a significant room for improvement in the level of congestion even with the existing airport infrastructure. Given the available capacity, efficient administrative controls and/or market-based mechanisms can potentially lead to substantial reductions in airport congestion and delays. These results strongly motivate our work on demand management strategies for congestion

mitigation in Chapters 5 and 6.

The main contributions of this chapter are threefold. First, we propose a novel, optimization-based approach for attributing the congestion-related delays in the NAS to two different causes, namely, delays due to insufficient airport capacity and delays due to inefficient utilization of available capacity due to airline competition. Second, we develop an aggregated, integrated airline scheduling model with a proxy objective function for delay minimization and an elaborate heuristic-based approach for an approximate solution of this large-scale (non-binary) mixed-integer programming problem. Finally, and most importantly, this is the first study which proves that there is a significant room for reducing the level of congestion even with the existing airport infrastructure without compromising the passengers' level-of-service, if we can control the negative impacts of airline competition through efficient demand management strategies. In this chapter, we make a case for the strong potential for demand management-based mechanisms for congestion mitigation.

7.1.2 Chapter 3: Quantification and Analysis of Passenger Delays and Disruptions

Passenger delays and disruptions are an important manifestation of airline competition and congestion, and in turn, affect the overall level-of-service experienced by the passengers. Airlines tune their networks, schedules and operations in order to attract a larger share of the market while minimizing costs, but in the process, critically modify the passenger delays and disruptions. In Chapter 3, we present a framework that combines data mining and statistical modeling techniques for generating disaggregate data on passenger travel and delays. We subsequently use this rich dataset to carry out a sequence of analyses that provide insights into the impacts of airline scheduling and operational decisions related to network structures, hub locations, connecting bank structures, flight frequencies, departure schedules, flight cancellations, passenger re-bookings etc on passenger delays and disruptions. Beyond the analyses and findings in this chapter, we foresee several other applications of this passenger delays

framework for passenger-centric decision-making in airline scheduling, air traffic flow management, and aviation policy.

The three main contributions of the research in this chapter are as follows. First, we develop a detailed approach for disaggregating publicly available aggregate passenger flow data which, among other applications, facilitates the usage of a pre-existing passenger delay calculation heuristic to a much wider dataset. Second, we analyze the spatio-temporal patterns in passenger delays using these estimated disaggregate passenger flows and present numerous insights into the factors affecting passenger delays. Such insights could not be generated in any of the prior studies due to a lack of comprehensive passenger itinerary flow data. Third, we investigate the causes of passenger travel disruptions by applying data analysis and statistical modeling to historical flight and passenger data. We believe that our research in this chapter has opened a whole new avenue for the air transportation research concerning passenger-related issues.

7.1.3 Chapter 4: Implications of Airline Frequency Competition for Airline Profitability and Airport Congestion

Competition affects airlines' seating capacity allocation decisions, which in turn have a strong impact on airline profitability and on airport congestion. In Chapter 4, we study this relationship using game-theoretic models of airline frequency competition and prove several important properties of this relationship. We find the Nash equilibrium solutions to be particularly suitable for modeling such competitive situations because of their attractive convergence and stability properties. We conclude that the worst-case degree of inefficiency and congestion introduced by competition is a direct and increasing function of intensity of competition, profit margins, and number of competing airlines. This is the first study, to the best of the author's knowledge, which actually proves that the S-curve has direct and adverse implications on airline profitability and airport congestion, as has been speculated in multiple previous studies. Furthermore, these results provide the intuition behind our framework for

modeling the frequency competition under administrative slot controls in Chapter 5 and under congestion pricing mechanisms in Chapter 6.

The main contributions of the research in this chapter are threefold. Ours is the first study that models the S-curve-based airline frequency competition using game-theoretic tools. The S-curve has been mentioned in many empirical studies and has also been an important part of the airline-industry lore. Second, we provide credibility to the idea of using Nash equilibrium as a means of modeling airline frequency competition by proving the convergence of 2 alternative simple frequency adjustment rules (otherwise known as *myopic learning dynamics*) to a Nash equilibrium. Finally, using the idea of Nash equilibrium, we prove that the S-shaped relationship between market share and frequency share has direct and adverse implications on airline profitability and airport congestion. These results make a strong case for careful incorporation of airline frequency competition in subsequent research on modeling the impacts of demand management mechanisms (as presented in Chapters 5 and 6).

7.1.4 Chapter 5: Administrative Mechanisms for Airport Congestion Mitigation

Some varieties of administrative controls have been in place at some of the busiest US airports and yet these airports often suffer from large delays and congestion. In Chapter 5, we propose simple slot reduction mechanisms which fall under the umbrella of administrative slot controls and evaluate their benefits in the presence of airline frequency competition. We model the competitive airline frequency planning decisions under slot constraints using game-theoretic models and solve the model to a Nash equilibrium. We propose a successive optimizations heuristic, wherein individual optimization problems are solved to full optimality using dynamic programming. Empirical validation of model results shows that a Nash equilibrium outcome to our model is able to describe the actual airline decisions with a reasonable level of accuracy. We evaluate two simple slot reduction strategies. Under the assumptions of our model, the results show that, in addition to a substantial reduction in flight and

passenger delays, small reductions in total allocated capacity (i.e. slot reductions) can improve the operating profits of carriers considerably. Our sensitivity analyses of major results to many of our assumptions, parameter values, changes in time periods and datasets show that our results are robust and our original estimates of benefits from slot reduction are conservative in most cases. Thus, we show that a small reduction in the total number of slots at a congested airport is beneficial to all the major stakeholders including the airlines, passengers and system operators.

The most significant practical contributions of this thesis are presented in this chapter. The main contributions of this chapter fall into four categories. First, we propose a game-theoretic model of frequency competition under slot constraints as an evaluation methodology for slot allocation schemes. Second, we provide a solution algorithm with good computational performance for solving the problem to a Nash equilibrium. Third, we provide justification of the credibility of the Nash equilibrium solution concept in two different ways, through empirical validation of the model outcome and through a computational demonstration of the convergence properties of the learning dynamics for non-equilibrium situations. Finally, under simple slot allocation schemes, we evaluate system performance from the perspectives of the passengers and the competing airlines, and provide insights to guide the demand management policy decisions. Our administrative slot allocation-based strategies are shown to be beneficial to all the major NAS stakeholders at the same time.

7.1.5 Chapter 6: Pricing Mechanisms for Airport Congestion Mitigation

Airport slot pricing mechanisms have often been proposed in literature for efficient utilization of airport capacity. In Chapter 6, we propose a modeling framework for analyzing the effectiveness of slot pricing in the presence of airline competition. We develop an equilibrium model and solve it using iterative algorithms. Our preliminary results using a small hypothetical network highlight some important differences between marginal cost pricing and flat pricing. We find that the effectiveness of a

congestion pricing mechanism critically depends on three essential characteristics of frequency competition in individual markets. These are the same three parameters that affect the level of congestion introduced by competition as described in Chapter 4. We show that a marginal cost pricing mechanism is able to deter the airlines from scheduling very frequency flights without penalizing them with very high congestion toll payments. Most importantly, we prove that in addition to delay reduction benefits, a significant part of congestion pricing benefits to the airlines are in the form of reduction in operating costs due to a greater number of passengers per flight. Our models of competition are able to capture this important effect which could not be captured by previous studies.

The major contributions of the research in this chapter are threefold. First, we develop a model for airline frequency competition that explicitly accounts for the relationship between the number of flights operated, number of seats flown and the number of passengers carried by an airline under slot pricing. To the best of author's knowledge, this is the first computational study that accounts for this relationship. Second, using a small hypothetical network, we evaluate the impacts of congestion prices on the various stakeholders and investigate the dependence of effectiveness of congestion pricing mechanisms on different characteristics of airline competition in individual markets. Third, we provide computational results under flat prices, as well as under a marginal cost pricing equilibrium. Our results show that variation in the number of passengers per flight plays a vital role in determining the degree of attractiveness of congestion pricing to the airlines. A significant part of the impact of congestion pricing could not be accounted for using the earlier models based on the assumptions of constant load factors and constant aircraft sizes. These results, based on a small hypothetical network, provide an important proof-of-concept demonstration of the potential benefits of congestion pricing.

7.2 Recent Events

All the results and conclusions in this thesis are based on data either from the year 2007 or from the 1st quarter of 2008. This is primarily because the year 2007 and early part of 2008 were amongst the worst times, in terms of congestion and delays, in the history of US aviation industry. Since then, a variety of recent events have modified the congestion situation in the NAS. In this section, we look at some recent noteworthy trends and discuss the validity of our conclusions in this thesis in the light of these recent events.

As shown in Table 1.1 in Chapter 1, the period from 2002 to 2007 saw a sustained growth in passenger demand and the number of flights, which paralleled the growth in various other sectors of the US economy. However, the year 2008 was marked by two patterns of significant economic consequences. First, a sharp rise in global crude oil prices in the first half of 2008 resulted in a steep hike in jet fuel prices. As shown in Figure 7-1, in the 5-month period from February to July 2008, average jet fuel prices increased by 45%, which corresponds to approximately an 8% monthly rate of increase [78]. This resulted in a sharp increase in airlines' unit operating costs.

As shown in Figure 7-1, the jet fuel prices reached their peak value of \$3.69 per gallon in July 2008. From August onwards, the prices started decreasing. However, before the industry could recover from the shock of fuel price hike, the next important trend began to manifest. Following the financial meltdown and the credit crunch on Wall Street, the US economy drifted towards an economic recession. These economic troubles were accompanied by a significant decrease in passenger demand in the US domestic air transportation markets. Figure 7-2 shows the trend in the total number of domestic passengers carried by all the US carriers from January 2000 to January 2011 [75]. The number of passengers in each month of a year is normalized such that the number equals 100 for each month of 2007. In other words, the number of passengers in each month of a year is divided by the number of passengers in the same month in year 2007 and then multiplied by 100. This normalization is performed to segregate the seasonal fluctuations in demand from year-to-year decreases in the

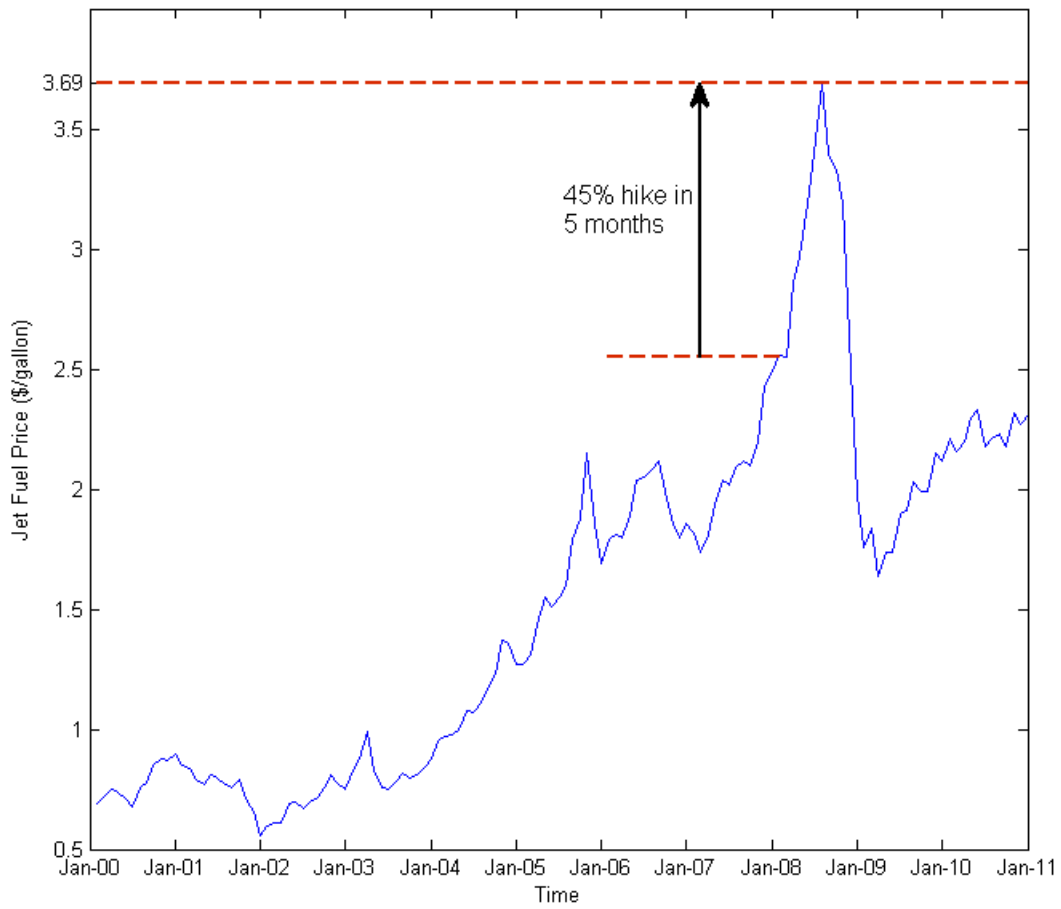


Figure 7-1: Trend in jet fuel prices

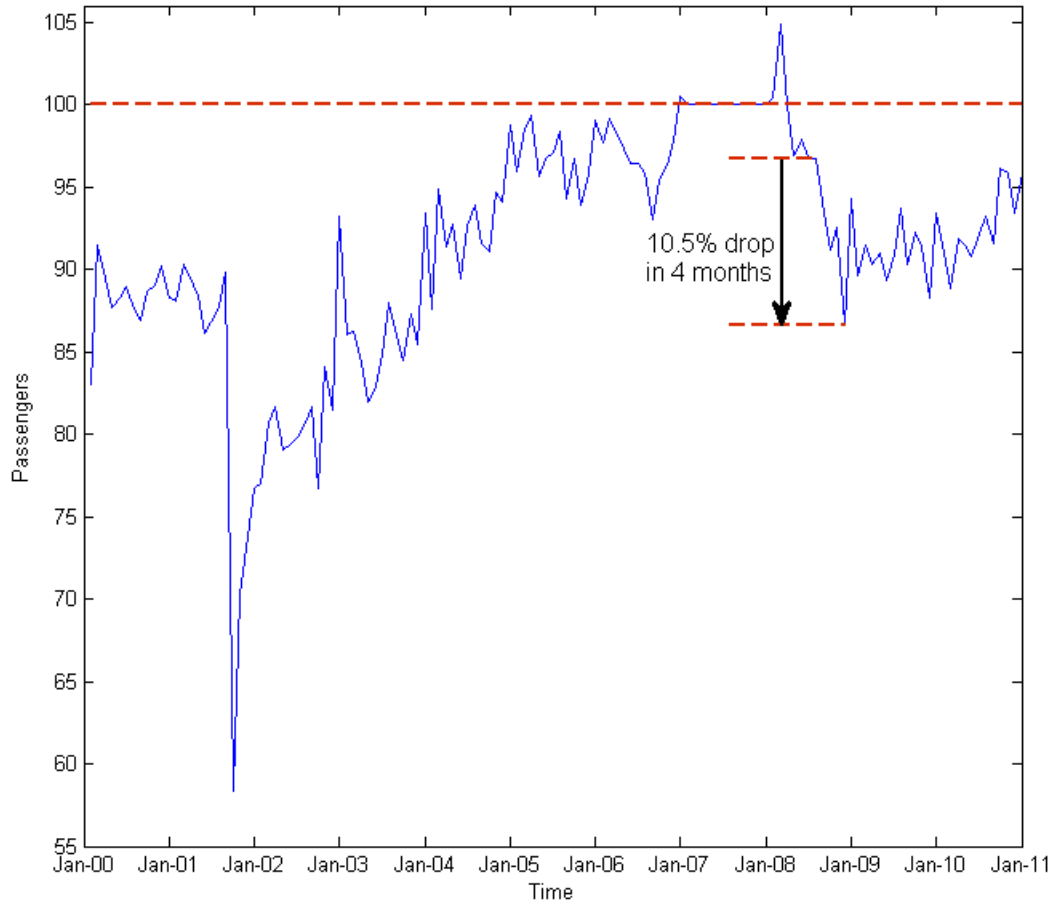


Figure 7-2: Trend in the number of passengers

numbers of passengers. Note that 2007 corresponded to the highest annual number of domestic passengers flown in the decade. As shown in Figure 7-2, immediately after the peak jet fuel prices, the normalized passenger demand began to fall precipitously. The number of passengers fell by approximately 10.5% by November 2008 compared to the number of passengers in July 2008, which corresponds to an average monthly rate of decrease of approximately 2.7% over that 4-month period.

The number of flights scheduled by an airline is obviously an increasing function of the passenger demand and, as shown by the game-theoretic analysis presented in Section 4.7, it is also a decreasing function of the operating costs. The combination of a sharp increase in fuel prices and a sharp decline in demand resulted in intense

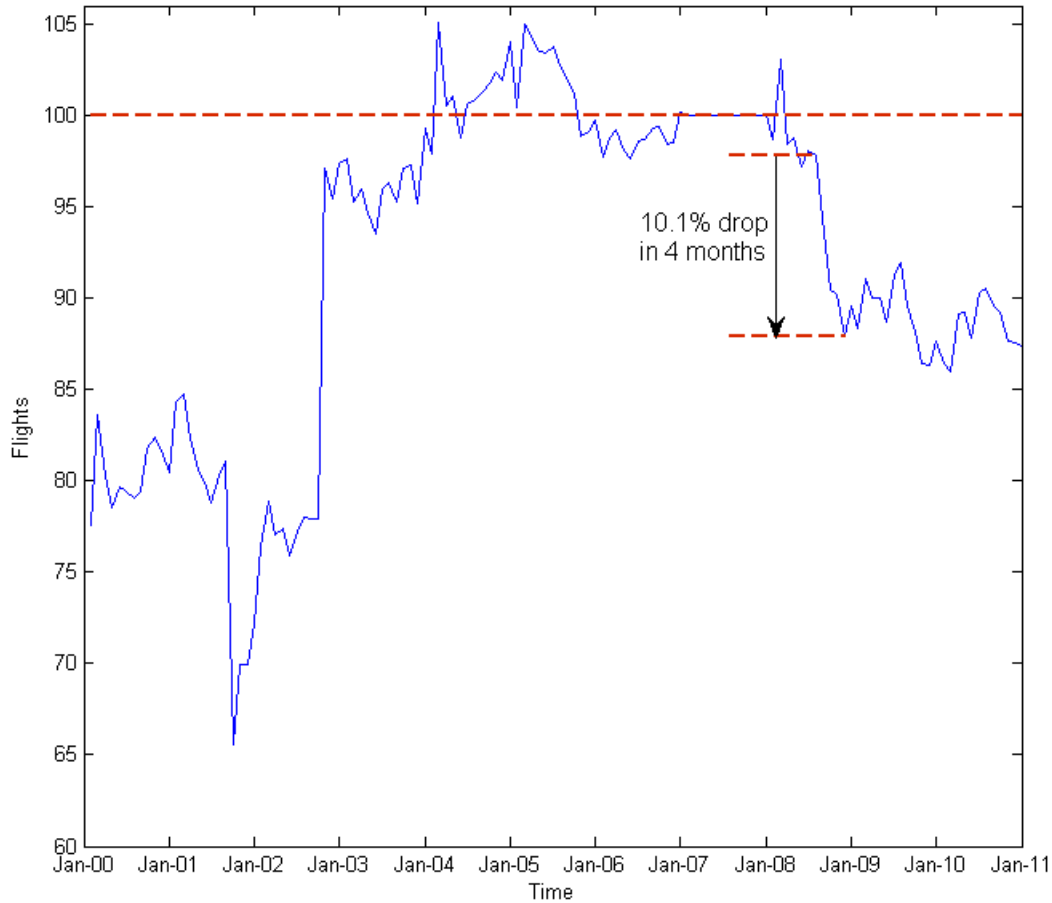


Figure 7-3: Trend in number of flights

pressure on airlines to cut down on the number of flights. Figure 7-3 shows the trend in the monthly number of flights operated by US carriers in the domestic markets from January 2000 to January 2011 [75]. Again the numbers are normalized in the same way as the number of passengers shown in Figure 7-2. As shown in Figure 7-3, the normalized number of flights operated by the US carriers in the domestic markets were down by approximately 10.1% by November 2008 as compared to July 2008, a 2.6% average monthly rate of decrease in that 4-months period.

This reduction in the number of flights resulted in a substantial reduction in average flight delays. The dotted blue line in Figure 7-4 shows the trend in the normalized average arrival delays in the US domestic markets for the flights of all

the ASQP-reporting carriers (that is, certified U.S. air carriers that account for at least one percent of domestic scheduled passenger revenues) [74]. As before, the numbers are normalized with respect to the corresponding month in the year 2007. Average arrival delays depend not only on the number of flights scheduled but also on the realized values of airport capacities on the day of operations, which are directly affected by weather. As a result of the variability in weather patterns, the monthly average values of arrival delays fluctuate considerably. Therefore, in order to allow for an easy visual inspection of the trends in average arrival delays, we also plotted their 6-monthly moving averages. The solid black line shows the 6-monthly moving average values of the normalized average arrival delays. As expected, the solid black line is much smoother than the dotted blue line and has some visible lag compared to the dotted blue line. The average arrival delays were approximately 32.2% lower in the first quarter of 2009 as compared to the first quarter of 2008.

After the fuel price shock in mid-2008, the jet fuel prices came down very fast as well. Over the subsequent 2 years, while jet fuel prices never reached the record levels of mid-2008, it is interesting to note that the number of passengers, the number of flights, and the average arrival delays all remained considerably lower than the record levels of 2007 and early months of 2008. In the year 2009, the number of passengers was 9.0% lower, the number of flights was 10.9% lower and the average arrival delays were 24.4% lower than 2007. While in 2010, the number of passengers was 7.3% lower, the number of flights was 11.6% lower and the average arrival delays were 26.7% lower than 2007. By the second half of 2010, the US domestic passenger demand had started displaying some signs of growth; the number of passengers in the fourth quarter of 2010 was 4.4% higher than that in the fourth quarter of 2009. The fourth quarter of 2010 was also the first quarter after 2007 that saw an increase (albeit a very small increase of 0.8%) in the number of flights compared to the same quarter a year before.

Given that our analysis in Chapters 5 and 6 specifically focussed on the LaGuardia Airport (LGA) in New York, we are especially interested in investigating the recent trends in the number of operations and delays at LGA. Figure 7-5 shows the recent trend in the seasonally normalized values of the total number of operations by all

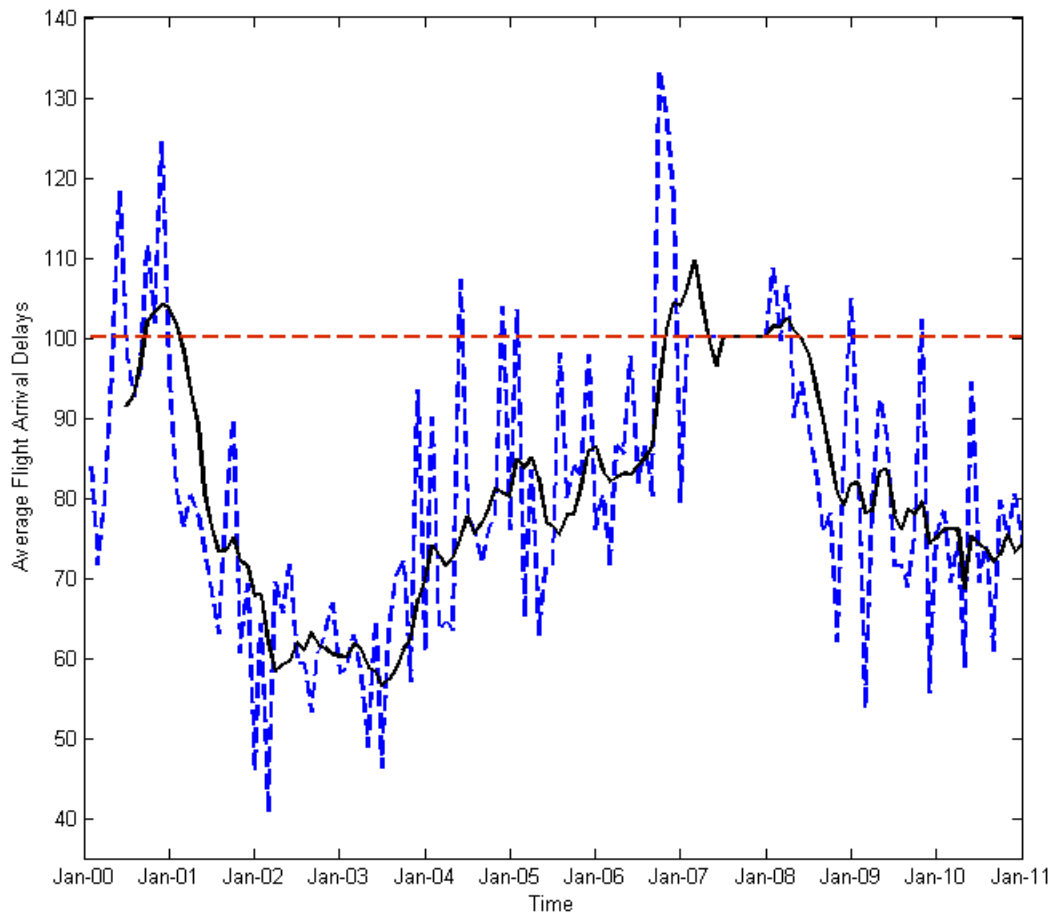


Figure 7-4: Trend in average arrival delays

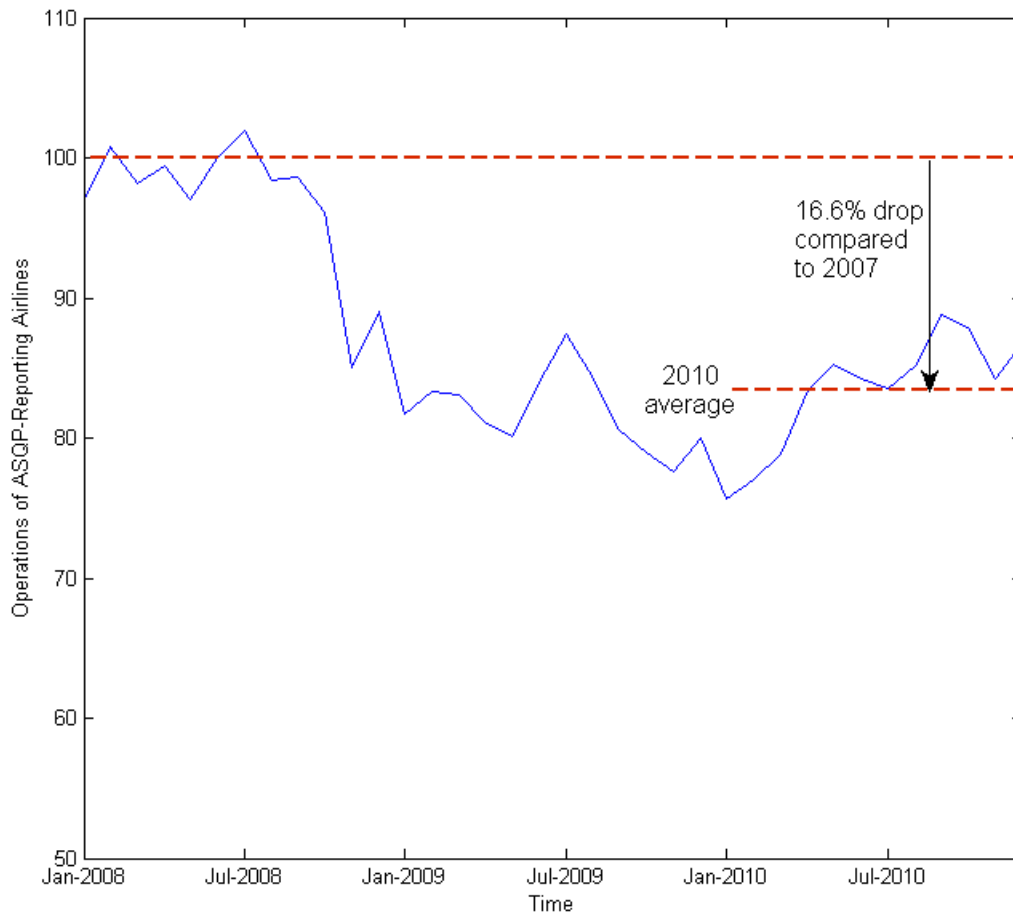


Figure 7-5: Trend in number of operations for ASQP reporting airlines at LGA

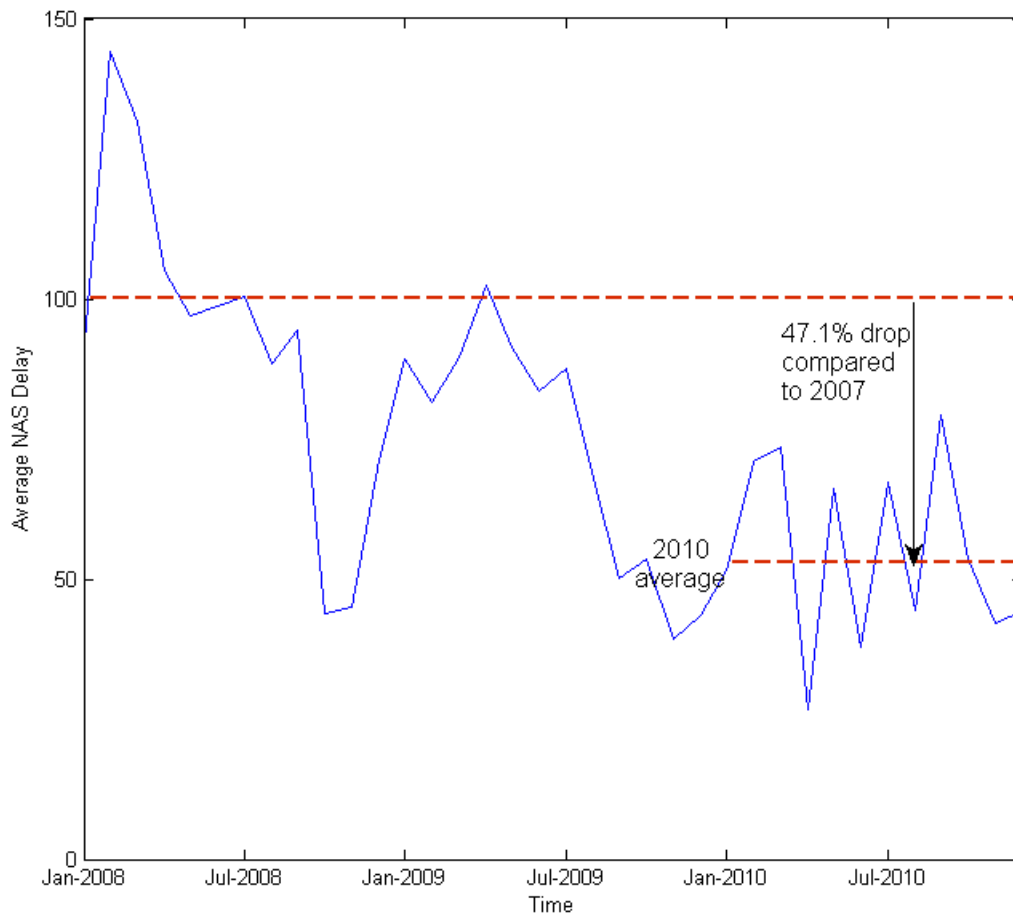


Figure 7-6: Trend in NAS delays per flight at LGA

the ASQP-reporting airlines at LGA [74]. The values are normalized such that the corresponding values for each month in the year 2007 equal 100. Figure 7-6 shows the recent trend in the average NAS delays per flight at LGA [74]. The average value of NAS delay is calculated across all flights either departing from or arriving at LGA in a particular month. Finally, Figure 7-7 shows the corresponding trend in the normalized values of the average arrival delays per flight for all flights either departing from or arriving at LGA [74].

For the year 2010, the total operations of ASQP-reporting airlines at LGA had reduced by 16.6%, the average NAS delays at LGA were lower by 47.1% and the average arrival delays at LGA were lower by 39.9% compared to the respective values

for the year 2007.

In Chapters 5 and 6, we evaluated the impacts of administrative controls and congestion pricing mechanisms for managing the demand for flight operations at LGA. In Section 5.5, we concluded that a 12.3% reduction in the number of operations resulted in approximately a 41.0% reduction in average NAS delay per flight. After 2007, October 2010 was the month when the percentage reduction in the number of operations at LGA was the closest to the 12.3% reduction proposed by our administrative slot reduction mechanisms. From Figures 7-5, 7-6 and 7-7, it is especially interesting to note that in the month of October 2010, the number of operations reduced by 12.1% resulting in a 46.4% reduction in average NAS delays per flight and a 40.1% reduction in average flight delays at LGA. These numbers are consistent with our estimate of a 41.0% reduction due to a 12.3% reduction in the number of slots, which was based on a somewhat conservative methodology.

Furthermore, for the domestic US flights, the average load factor increased by 2.9%, and the average number of passengers per flight increased by 4.6% from 2007 to 2010 [77]. If we focus only on the flights at LGA airport, then the average load factor increased by 3.1%, and the average number of passengers per flight increased by 4.1% from 2007 to 2010 [77]. The percentage reduction in the number of flights has been considerably greater than the percentage reduction in passengers, both at LGA and in domestic US flights overall. Thus, the primary drivers for delays reduction in 2010 compared to 2007 are the increased load factors and increases in average aircraft sizes, which are identical to the ones we demonstrated to be the major drivers of delay benefits in Chapter 5.

In Chapter 5, we analyzed the problem of airport congestion and delays from a system perspective. By modeling the problem as a multi-agent game, we replicated the complexity of airline decision-making under competition with reasonable accuracy. Subsequently, based on this game-theoretic model, our computational experiments predicted that a delay reduction of over 40% can be achieved by reducing the operations at an administratively controlled airport through slot reduction of just over 12%. The benefits of slot reductions were achieved through increased load factors

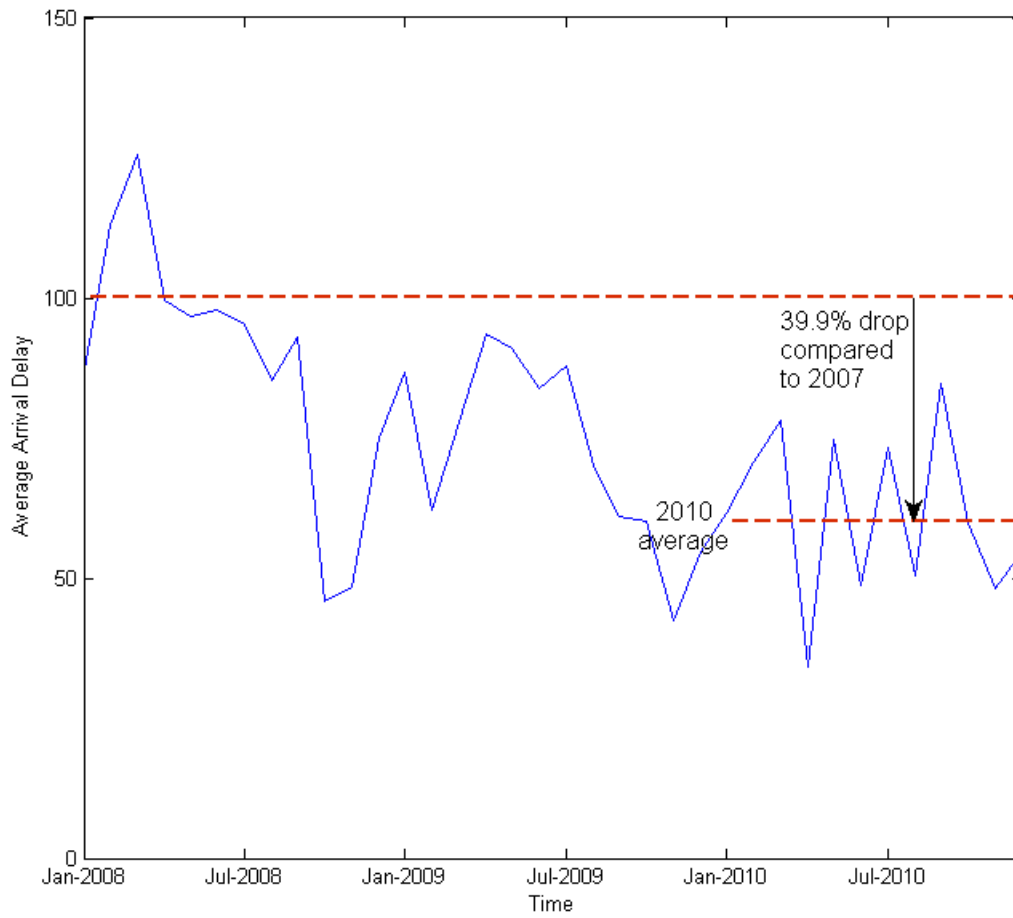


Figure 7-7: Trend in average arrival delay at LGA

and increases in average aircraft sizes.

These results have been validated by the series of events that happened in the US airline industry over the last 3 years. Due to the various changes in the US airline industry over the past 3 years, we have seen double digit reductions in the number of operations by the major carriers at LGA airport (as well as at many other airports across the NAS). That the percentage reduction in operations at LGA turned out to be similar to those proposed by our administrative mechanisms is an interesting coincidence. But more importantly, it is not a coincidence that the resultant reductions in delays are also very much comparable to those estimated by our data analysis based on the 2007 delays data. Furthermore, the main drivers of these delay savings are also identical to those in our computational results in Chapter 5. Thus, these recent trends are a further testament to the validity our conclusions, especially those derived in Chapter 5.

Because of these recent events in the US airline industry, we have been able to provide a real-life validation of the computational experiment we performed. Rather than just being a queuing-theoretic claim that reducing operations will reduce delays, our computational experiments and models provided a behavioral description of the complex interactions between the airline decisions and passenger decisions in a competitive industry landscape. It is remarkable to note that our conclusions based on this system-modeling approach are nicely validated because of the recent events serving as a real-life case study.

However, it should be noted that these recent reductions in delays are brought about primarily because of the reduction in passenger demand for air transportation, which in turn is due to the economic downturn that has affected the US economy for a major part of the last 3 years. Thus the temporary respite from flight delays was a result of the economic recession. However, large delays are expected to return once the economic crisis subsides [96]. Therefore, in the future, the administrative and/or pricing mechanisms proposed and evaluated in this thesis will become increasingly important and relevant for efficient management of demand for airport resources in the National Aviation System.

Appendix A

Abbreviations

A.1 Carrier Abbreviations

Table A.1: Carrier names and IATA codes

IATA Code	Carrier Name
9E	Pinnacle Airlines
AA	American Airlines
AQ	Aloha Airlines
AS	Alaska Airlines
B6	JetBlue Airways
CO	Continental Airlines
DL	Delta Air Lines
EV	Atlantic Southeast Airlines
F9	Frontier Airlines
FL	AirTran Airways
HA	Hawaiian Airlines
MQ	American Eagle Airlines
NW	Northwest Airlines
OH	Midwest Airlines
OO	SkyWest Airlines
UA	United Airlines
US	US Airways
WN	Southwest Airlines
XE	ExpressJet Airlines
YV	Mesa Airlines

A.2 Airport Abbreviations

Table A.2: Airport names and IATA codes

IATA Code	Airport Name	IATA Code	Airport Name
ABQ	Albuquerque International Sunport	LGA	New York LaGuardia
ATL	Hartsfield-Jackson Atlanta International	MCI	Kansas City International
AUS	Austin-Bergstrom International	MCO	Orlando International
BNA	Nashville International	MDW	Chicago Midway International
BOS	Boston Logan International	MEM	Memphis International
BWI	Baltimore Washington International	MIA	Miami International
CLE	Cleveland Hopkins International	MSP	Minneapolis - St. Paul International
CLT	Charlotte Douglas International	OAK	Oakland International
CMH	Columbus Regional	ONT	Ontario International
CVG	Cincinnati / Northern Kentucky International	ORD	Chicago O'Hare International
DAL	Dallas Love Field	PDX	Portland International
DCA	Reagan National	PHL	Philadelphia International
DEN	Denver International	PHX	Phoenix Sky Harbor International
DFW	Dallas / Fort Worth International	PIT	Pittsburgh International
DTW	Detroit Metro	RDU	Raleigh-Durham International
EWR	Newark Liberty International	SAN	San Diego International
FLL	Fort Lauderdale - Hollywood International	SAT	San Antonio International
HNL	Honolulu International	SEA	Seattle - Tacoma International
HOU	Houston Hobby	SFO	San Francisco International
IAD	Washington Dulles International	SJC	Mineta San Jose International
IAH	Houston George Bush	SLC	Salt Lake City International
IND	Indianapolis International	SMF	Sacramento International
JFK	John F Kennedy International	SNA	John Wayne
LAS	Las Vegas - McCarran International	STL	Lambert-St. Louis International
LAX	Los Angeles International	TPA	Tampa International

Bibliography

- [1] N. Adler. Competition in a deregulated air transportation market. *European Journal of Operational Research*, 129(2):337–345, 2001.
- [2] N. Adler. Hub-spoke network choice under competition with an application to western europe. *Transportation science*, 39(1):58–72, 2005.
- [3] Federal Aviation Administration. Aviation system performance metrics(aspm)-complete. Retrieved November 28, 2010, from Operations and Performance Data: <http://aspm.faa.gov/>, 2010.
- [4] V. Aguirregabiria and C. Y Ho. A dynamic oligopoly game of the US airline industry: Estimation and policy experiments. *Manuscript. University of Toronto*, 2009.
- [5] ATA. 2008 economic report. Technical report, Air Transport Association, 2008.
- [6] BAC. Decision window path preference methodology description. Technical report, Boeing Airplane Company, Seattle WA, 1997.
- [7] M. Ball, C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, B. Zou, et al. Total delay impact study. *NEXTOR Draft Final Report prepared for the Federal Aviation Administration*, 2010.
- [8] M. Ball, C. Barnhart, G. Nemhauser, and A. Odoni. Air transportation: Irregular operations and control. *Transportation*, 14, 2006.
- [9] M. Ball, G. Donohue, and K. Hoffman. Auctions for the safe, efficient, and equitable allocation of airspace system resources. *Combinatorial auctions*, 1, 2006.
- [10] C. Barnhart. Airline schedule optimization. In P. Belobaba, A. Odoni, and C. Barnhart, editors, *The Global Airline Industry*, chapter 7, pages 183–211. Wiley, 2009.
- [11] C. Barnhart, N. L Boland, L. W Clarke, E. L Johnson, G. L Nemhauser, and R. G Sheno. Flight string models for aircraft fleet and routing. *Transportation Science*, 32(3):208–220, 1998.

- [12] C. Barnhart, D. Fearing, and V. Vaze. Modeling passenger travel and delays in the national air transportation system. Under review, 2010.
- [13] C. Barnhart, E. L Johnson, G. L Nemhauser, M. W.P Savelsbergh, and P. H Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations research*, pages 316–329, 1998.
- [14] C. Barnhart, T. S Kniker, and M. Lohatepanont. Itinerary-based airline fleet assignment. *Transportation Science*, 36(2):199–217, 2002.
- [15] R. Baseler. Airline fleet revenue Management-Design and implementation. *Handbook of Airline Economics*, 2002.
- [16] P. Belobaba. The airline planning process. In P. Belobaba, A. Odoni, and C. Barnhart, editors, *The Global Airline Industry*, chapter 6, pages 153–181. Wiley, 2009.
- [17] P. Belobaba. Overview of airline economics, markets and demand. In P. Belobaba, A. Odoni, and C. Barnhart, editors, *The Global Airline Industry*, chapter 3, pages 47–72. Wiley, 2009.
- [18] P. P Belobaba and A. Farkas. Yield management impacts on airline spill estimation. *Transportation Science*, 33(2):217–232, 1999.
- [19] U. Binggeli and L. Pompeo. Does the s-curve still existt? Technical report, McKinsey & Company, 2006.
- [20] P. A Bonnefoy. *Scalability of the air transportation system and development of multi-airport systems: a worldwide perspective*. PhD thesis, Massachusetts Institute of Technology, 2008.
- [21] S. Borenstein. Hubs and high fares: dominance and market power in the US airline industry. *The RAND Journal of Economics*, 20(3):344–365, 1989.
- [22] J. A Brander and A. Zhang. Dynamic oligopoly behaviour in the airline industry. *International Journal of Industrial Organization*, 11(3):407–435, 1993.
- [23] S. Bratu and C. Barnhart. An analysis of passenger delays using flight operations and passenger booking data. *Air Traffic Control Quarterly*, 13(1), 2005.
- [24] R. Britto, M. Dresner, and A. Voltes. The impact of flight delays on passenger demand and consumer welfare. In *12th World Conference on Transport Research, Lisbon, Portugal*, July 2010.
- [25] J. K Brueckner. Airport congestion when carriers have market power. *The American Economic Review*, 92(5):1357–1375, 2002.
- [26] J. K Brueckner. Internalization of airport congestion: a network analysis. *International Journal of Industrial Organization*, 23:599–614, 2005.

- [27] J. K Brueckner. Schedule competition revisited. *Journal of Transport Economics and Policy*, 44(3):261–285, 2010.
- [28] J. K Brueckner and R. Flores-Fillol. Airline schedule competition. *Review of Industrial Organization*, 30(3):161–177, 2007.
- [29] K. Button and J. Drexler. Recovering costs by increasing market share: An empirical critique of the S-Curve. *Journal of Transport Economics and Policy*, 39(3):391–410, 2005.
- [30] A. Carlin and R. E Park. Marginal cost pricing of airport runway capacity. *The American Economic Review*, pages 310–319, 1970.
- [31] L. W. Clarke, C. A. Hane, E. L Johnson, and G. L Nemhauser. Maintenance and crew considerations in fleet assignment. *Transportation Science*, 30(3):249–260, 1996.
- [32] F. J Cohas, P. P Belobaba, and R. W Simpson. Competitive fare and frequency effects in airport market share modeling. *Journal of Air Transport Management*, 2(1):33–45, 1995.
- [33] A. M Cohn and C. Barnhart. Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, pages 387–396, 2003.
- [34] G. M Coldren and F. S Koppelman. Modeling the competition among air-travel itinerary shares: GEV model development. *Transportation Research Part A: Policy and Practice*, 39(4):345–365, 2005.
- [35] G. M Coldren, F. S Koppelman, K. Kasturirangan, and A. Mukherjee. Modeling aggregate air-travel itinerary shares: logit model development at a major US airline. *Journal of Air Transport Management*, 9(6):361–369, 2003.
- [36] US Code Collection. Title 49, sections 41714-41719,, 2005.
- [37] J. F Cordeau, G. Stojkovic, F. Soumis, and J. Desrosiers. Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation science*, 35(4):375–388, 2001.
- [38] A. A Cournot and I. Fisher. *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan, 1897.
- [39] P. Cramton, M. O Ball, L. M Ausubel, F. Berardino, G. Donohue, M. Hansen, and K. Hoffman. Market-based alternatives for managing congestion at new york’s laguardia airport. In *Proc. of AirNeth Annual Conference*, 2007.
- [40] J. I Daniel. Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues. *Econometrica*, 63(2):327–370, 1995.
- [41] J. I Daniel. Distributional consequences of airport congestion pricing. *Journal of Urban Economics*, 50:230–258, 2001.

- [42] J. Desrosiers, Y. Dumas, M. M Solomon, and F. Soumis. Time constrained routing and scheduling. *Handbooks in operations research and management science*, 8:35–139, 1995.
- [43] G. Dobson and P. J Lederer. Airline scheduling and routing in a hub-and-spoke system. *Transportation Science*, 27(3):281–297, 1993.
- [44] FAA. Airport capacity benchmark report. Technical report, Federal Aviation Administration, Washington DC, 2004.
- [45] T. P Fan and A. R Odoni. The potential of demand management as a short-term means of relieving airport congestion. In *Proceedings of EUROCONTROL-FAA Air Traffic Management R&D Review Seminar, Santa Fe, NM*, 2001.
- [46] S. J Forbes. The effect of air traffic delays on airline prices. *International journal of industrial organization*, 26(5):1218–1232, 2008.
- [47] D. M Grether, R. M Isaac, and C. R Plott. *The allocation of scarce resources: Experimental Economics and the problem of allocating airport slots*. Westview Press, 1989.
- [48] D. M Grether, R. M Isaac, C. R Plott, Polinomics Research Laboratories, and United States. Civil Aeronautics Board. *Alternative methods of allocating airport slots: Performance and evaluation*. Polinomics Research Laboratories, Inc., 1979.
- [49] C. A Hane, C. Barnhart, E. L Johnson, R. E Marsten, G. L Nemhauser, and G. Sigismondi. The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming*, 70(1):211–232, 1995.
- [50] M. Hansen. Airline competition in a hub-dominated environment: An application of noncooperative game theory. *Transportation Research Part B: Methodological*, 24(1):27–43, 1990.
- [51] P. Harsha. *Mitigating airport congestion: market mechanisms and airline response models*. PhD thesis, Massachusetts Institute of Technology, 2009.
- [52] K. Hendricks, M. Piccione, and G. Tan. Equilibria in networks. *Econometrica*, 67(6):1407–1434, 1999.
- [53] S. Hong and P. T Harker. Air traffic network equilibrium: Toward frequency, price and slot priority analysis* 1. *Transportation Research Part B: Methodological*, 26(4):307–323, 1992.
- [54] T. L Jacobs, E. L. Johnson, B. C. Smith, and R. Darrow. O&D FAM: incorporating passenger flows into the fleeting process. In *Presentation to AGIFORS Symposium, New Orleans, LA*, 1999.

- [55] H. Jiang. *Dynamic airline scheduling and robust airline schedule de-peaking*. PhD thesis, Massachusetts Institute of Technology, 2006.
- [56] A. E Kahn. Change, challenge, and competition: A review of the airline commission report. *Regulation*, 3:1–10, 1993.
- [57] D. Klabjan, E. L Johnson, G. L Nemhauser, E. Gelman, and S. Ramaswamy. Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications*, 20(1):73–91, 2001.
- [58] D. Klabjan, E. L Johnson, G. L Nemhauser, E. Gelman, and S. Ramaswamy. Airline crew scheduling with time windows and plane-count constraints. *Transportation Science*, 36(3):337–348, 2002.
- [59] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. *Computer Science Review*, 3(2):65–69, 2009.
- [60] R. C Larson and A. R Odoni. *Urban operations research*. Dynamic Ideas, 2007.
- [61] L. T Le. *Demand management at congested airports: How far are we from utopia?* PhD thesis, George Mason University, 2006.
- [62] M. E Levine. Landing fees and the airport congestion problem. *Journal of Law and Economics*, 12(1):79–108, 1969.
- [63] Dot Econ Limited. Auctioning airport slots: A report for hm treasury and the department of the environment, transport and the regions. Technical report, DotEcon Ltd, United Kingdom, 2001.
- [64] M. Lohatepanont and C. Barnhart. Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, 38(1):19–32, 2004.
- [65] S. A Morrison. The equity and efficiency of runway pricing. *Journal of Public Economics*, 34(1):45–60, 1987.
- [66] S. A Morrison and C. Winston. Another look at airport congestion pricing. *The American Economic Review*, 97(5):1970–1977, 2007.
- [67] V. D Norman and S. Strandenes. *Deregulation of Scandinavian Airlines: a case study of the Oslo-Stockholm route*. University Of Chicago Press, 1994.
- [68] NYARC. New york aviation rulemaking committee report. Technical report, US Department of Transportation, Washington DC, 2007.
- [69] W. E O’Connor. *An introduction to airline economics*. Praeger Publishers, 2001.

- [70] A. Odoni, N. Pyrgiotis, and K. Malone. The propagation of delays in a network of airports: The approximate network delays model. In *12th World Conference on Transport Research*, 2010.
- [71] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved January 3, 2010, from Airline On-Time Performance and Delay Causes: http://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp, 2010.
- [72] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved May 10, 2011, from On-Time: On-Time Performance: http://www.transtats.bts.gov/Fields.asp?Table_ID=236, 2010.
- [73] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved January 3, 2010, from Airline Origin and Destination Survey: http://www.transtats.bts.gov/Fields.asp?Table_ID=247, 2010.
- [74] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved January 3, 2010b, from Air Carrier Financial: Schedule: P-5.2: http://www.transtats.bts.gov/Fields.asp?Table_ID=297, 2010.
- [75] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved January 3, 2010, from Air Carrier: T-100 Domestic Segment (U.S. Carriers): http://www.transtats.bts.gov/Fields.asp?Table_ID=259, 2010.
- [76] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved January 3, 2010b, from Air Carrier Financial: Schedule: P-7: http://www.transtats.bts.gov/Fields.asp?Table_ID=278, 2010.
- [77] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved May 10, 2011, from Airlines and Airports: http://www.transtats.bts.gov/Data_Elements.aspx?Data=1, 2011.
- [78] Bureau of Transportation Statistics. (n.d.). Transtats. Retrieved May 10, 2011, from Airlines Fuel Cost and Consumption: http://www.transtats.bts.gov/Data_Elements.aspx?Data=1, 2011.
- [79] E. Pels, P. Nijkamp, and P. Rietveld. Airport and airline competition for passengers departing from a large metropolitan area. *Journal of Urban Economics*, 48(1):29–45, 2000.
- [80] E. Pels and E. T Verhoef. The economics of airport congestion pricing. *Journal of Urban Economics*, 55(2):257–277, 2004.
- [81] G. Perakis and W. Sun. Congestion pricing for airport efficiency. In *Seventh Triennial Symposium on Transportation Research, Tromso, Norway*, June 2010.
- [82] S. J Rassenti, V. L Smith, and R. L Bulfin. A combinatorial auction mechanism for airport time slot allocation. *The Bell Journal of Economics*, 13(2):402–417, 1982.

- [83] B. Rexing. *Airline fleet assignment with time windows*. PhD thesis, Massachusetts Institute of Technology, 1997.
- [84] J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.
- [85] N. Rupp and M. Holmes. Why are so many flights canceled? *Working Papers*, 2003.
- [86] R. A Schorr. *Essays on airport and airway congestion*. PhD thesis, Massachusetts Institute of Technology, 2006.
- [87] C. E. Schumer and C. B. Maloney. Your flight has been delayed again: flight delays cost passengers, airlines, and the US economy billions. *The US Senate Joint Economic Committee*, 2008.
- [88] L. Sherry and G. Donohue. US airline passenger trip delay report (2007). Technical report, Center for Air Transportation Systems Research, George Mason University, Fairfax, VA., 2008.
- [89] R. W Simpson. *A market share model for US domestic airline competitive markets*. Massachusetts Institute of Technology, Flight Transportation Laboratory, 1970.
- [90] H. Von Stackelberg. *The theory of the market economy*. William Hodge, 1952.
- [91] N. K Taneja. *The commercial airline industry*. DC Heath, 1976.
- [92] N. K Taneja, United States. Office of High-Speed Ground Transportation, and Massachusetts Institute of Technology. Flight Transportation Laboratory. *Airline competition analysis*. Massachusetts Institute of Technology, Flight Transportation Laboratory, 1968.
- [93] G. Theis, T. Adler, J. P Clarke, and M. Ben-Akiva. Risk aversion to short connections in airline itinerary choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1951(-1):28–36, 2006.
- [94] S. L Tien, A. M Churchill, and M. O Ball. Quantifying the relationship between airline load factors and flight cancellation trends. *Transportation Research Record: Journal of the Transportation Research Board*, 2106(-1):38–46, 2009.
- [95] S. L. Tien, B. Subramanian, and M. Ball. Constructing a passenger trip delay metric: an aggregate level approach. In *Proceedings of 2008 ICRAT Conference*, George Mason University, 2008.
- [96] A. Tomer, R. Puentes, and Brookings Institution. Metropolitan Policy Program. *Expect delays: an analysis of air travel trends in the United States*. Metropolitan Policy Program at Brookings, 2009.

- [97] W. S. Vickrey. Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260, 1969.
- [98] W. Wei and M. Hansen. Impact of aircraft size and seat availability on airlines’ demand and market share in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review*, 41(4):315–327, 2005.
- [99] W. Wei and M. Hansen. Airlines’ competition in aircraft size and service frequency in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review*, 43(4):409–424, 2007.
- [100] A. Zhang and Y. Zhang. Airport capacity and congestion when carriers have market power. *Journal of Urban Economics*, 60:229–247, 2006.
- [101] Y. Zhu. Evaluating airline delays: the role of airline networks, schedules, and passenger demands. Master’s thesis, Massachusetts Institute of Technology, 2009.