# UC Berkeley
## UC Berkeley Electronic Theses and Dissertations

**Title**

Flight Delays, Capacity Investment and Welfare under Air Transport System Equilibrium

**Permalink**

https://escholarship.org/uc/item/19069773

**Author**

Zou, Bo

**Publication Date**

2012

Peer reviewed|Thesis/dissertation

Flight Delays, Capacity Investment and Welfare under Air Transport
System Equilibrium

by

Bo Zou

A dissertation submitted in partial satisfaction of the

requirement for the degree of

Doctor of Philosophy

in

Engineering — Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Mark Hansen, Chair
Professor Samer Madanat
Professor James Powell

Spring 2012

Abstract

Flight Delays, Capacity Investment and Welfare under
Air Transport System Equilibrium

by

Bo Zou

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Mark Hansen, Chair


Infrastructure capacity investment has been traditionally viewed as an important means to mitigate congestion and delay in the air transportation system. Given the huge amount of cost involved, justifying the benefit returns is of critical importance when making investment decisions. This dissertation proposes an equilibrium-based benefit assessment framework for aviation infrastructure capacity investment. This framework takes into consideration the interplays among key system components, including flight delay, passenger demand, flight traffic, airline cost, and airfare, and their responses to infrastructure capacity investment. We explicitly account for the impact of service quantity changes on benefit assessment. Greater service quantity is associated with two positive feedback effects: the so-called Mohring effect and economies of link/segment density. On the other hand, greater service quantity results in diseconomies of density at nodes/airports, because higher traffic density at the airport leads to greater airport delays. The capacity-constrained system equilibrium is derived from those competing forces.

Two approaches are developed to investigate air transport system equilibrium and its shift in response to infrastructure capacity expansion. In Chapter 2, we first view the system equilibrium from the airline competition perspective. We model airlines' gaming behavior for airfare and frequency in duopoly markets, assuming that airlines have the knowledge of individuals' utility structure while making decisions, and that delay negatively affects individuals' utility and increases airline operating cost. The theoretical airline competition model developed in Chapter 2 provides analytical insights into the interactions among various system components. Under a symmetric Nash equilibrium, we find that the presence of flight delay increases passenger generalized cost and discourages air travel. Airlines would not pass delay cost entirely onto passengers through higher fare, but also account for the impact of service degradation on passenger willingness-to-pay and consequently passenger demand. To avoid exorbitant flight delays, airlines would use larger aircraft, meanwhile taking advantage of economies of aircraft size. The resulting unit cost reduction partially offsets operating delay cost increase. The equilibrium shift triggered by capacity expansion reduces both schedule delay and flight delay, leading to lower passenger generalized cost and higher demand, despite slightly increased airfare.

1

Airlines will receive larger profit, and consumer welfare will increase, as a result of the expansion. Although delay reduction is less than expected because of induced demand, the overall benefit, which encompasses reduction in both schedule delay and flight delay, would be much greater than estimated from a purely delay-based standpoint.

The equilibrium analysis can be alternatively approached from a traveler-centric perspective. The premise of an air transport user (i.e. traveler) equilibrium is that each traveler in the air transportation system maximizes his/her utility when making travel decisions. The utility depends upon market supply and performance characteristics, consisting of airfare, flight frequency, and flight delay. The extent of airline competition is implicitly reflected in the determination of airfare and flight frequency. Given the limited empirical evidence of the delay effect on air transportation system supply, two econometric models for airfare and flight frequency are estimated in Chapter 3. We find positive delay effect on fare, which should be interpreted as the net effect of airlines' tendency to pass delay cost to passengers while also compensating for service quality degradation. Higher delay discourages carriers from scheduling more flights on a segment. Both delay effects, however, are relatively small. The estimated fare and frequency models, together with passenger demand and airport delay models presented in Chapter 4, are integrated to formulate the air transport user equilibrium as fixed point and variational inequality problems. We prove that the equilibrium existence is guaranteed; whereas equilibrium uniqueness cannot be guaranteed. We apply the user equilibrium to a fully connected, hypothetical network with the co-existence of direct and connecting air services. Using a simple, heuristic algorithm, we find that the equilibrium is insensitive to initial demand values, suggesting that there may be a single equilibrium for this particular model instance. Hub capacity investment attracts spoke-spoke passengers from non-stop routes, and generates new demand on hub-related routes. At the market level, hub capacity expansion would result in greater total demand and consequently passenger benefits in almost all markets—except for ones where a predominant portion of passengers choose non-stop routes due to extremely high circuity for one-stop travel. In the latter set of markets, after capacity expansion passenger demand and benefits would be both reduced. This counter-intuitive result carries important implications that capacity increase does not necessarily benefit everyone in the system. Similar to the findings from the airline competition model, with changes in flight delay, schedule delay, airfare, and total demand, the user equilibrium model yields much higher passenger benefits from capacity investment than the conventional method; whereas hub delay saving is offset by traffic diversion and induced demand. With continuous capacity investment, the air transportation network will witness substantial changes in service supply and traffic patterns.

Dedicated to my parents,

Xiong Ping and Zou Huaqiao

# Table of Contents

# List of Tables

# List of Figures

# Acknowledgements

Coming to Berkeley is one of the most important landmarks in my life. I feel very blessed with the help and company of many people, because of whom my academic pursuit and personal adventure have been a truly enriching and rewarding experience. The complete list of people to whom my thanks should be delivered may require a few extra pages. Here I would like to convey my special gratitude to some of them.

First and foremost, I am deeply indebted to Professor Mark Hansen, who has been my advisor ever since I began my graduate studies at Cal. I was attracted to Berkeley largely by an earlier series of his papers, but have only felt more fortunate to have the chance to take my PhD journey under his supervision. I feel immensely grateful to him for leading me to the world of transportation economics, exposing me to a variety of research topics, and always standing with me to cope with difficult situations and share joyful moments in his unique, jocular manner. I am totally impressed by the width and depth of his knowledge and insights about transportation, his critical thinking, creativity, and inspiration in identifying and shaping research problems. This is not to mention a comprehensive, interactive, five-year long elective course on the philosophies of research, humor, and life that I have taken with him inside and beyond McLaughlin 107. Mark, no matter whether I get an A or B+ upon graduation, your special course has already been carved into my unforgettable Cal memory. The influence of your special class is deemed profound and will permeate through the rest of my academic life. Much alike one of the important findings in my dissertation, any conventional view of such influence would only lead to substantial underestimation of its value.

I would like to thank the other two members in my dissertation committee, Professors Samer Madanat and James Powell. Samer is not only an excellent teacher but a thoughtful research mentor. His Infrastructure Systems Management course is one of my all-time favorite classes. Thank you Samer for encouraging me to deepen my research, and offering me valuable suggestions on career development. Special thanks to Professor James Powell for his Econometric class, from which I have benefited greatly in building my methodological foundation for conducting empirical research. I am also grateful to him for being always considerate.

My gratitude further extends to other faculty members within and outside of Berkeley. My research ideas were shaped by several conservations with Professor Adib Kanafani in the early stage of my PhD study. The advice from Professor Carlos Daganzo is very helpful in clarifying some methodological issues. I feel very fortunate to have worked with Joan Walker as a graduate student instructor for her CE 93 class. I always enjoyed chatting with her and appreciate her enthusiastic support in the latter part of my PhD study. I would also like to thank Professor Kenneth Train for his excellent ECON 244 class, and the enlightening discussion on the discrete choice model issues I encountered in my dissertation. I feel fortunate to have known Professor Michael Ball. Collaboration

Thank you Mom and Dad for your profound love and eternal encouragement. Without the many years that you unconditionally spent raising, educating, and supporting me, nothing would have been possible. I hope this dissertation can be dedicated to you as both a gift and compensation for being away from you over the years.

# 1. Introduction

## 1.1 Background and Problem Statement

Flight delay is a serious and widespread problem in the United States. Between 2002 and 2007, flight traffic increased by about 22%, but the number of late-arriving flights more than doubled. One in every four flights was delayed for more than 15 minutes in 2007 in the National Airspace System (NAS) (BTS, 2012). The total economic impact of flight delays accruing to passengers, airlines, and the remaining of the economic system amounted to $32 billion in 2007 (Ball et al., 2010). Although traffic and delay have declined somewhat over the past few years due to the economic recession, the Federal Aviation Administration (FAA) expects demand growth to resume, with total Revenue Passenger Miles (RPMs) forecast to increase at an average annual rate of 3.2%, from 814.6 billion in 2011 to 1.57 trillion in 2032 (FAA, 2012a). This will impose unprecedented challenges on operational punctuality in the future NAS.

A major cause of flight delay is demand-capacity imbalance in the air transportation system, particularly in the airport terminal area. At many major US airports, air traffic is often scheduled close (sometimes even above) the maximum airfield capacity. Capacity constraints can result from persistent shortfalls in physical aviation infrastructure, such as runways, or as the result of a capacity drop under inclement weather conditions. The imbalance between demand and capacity implies that the flight delay problem can be approached from the perspective of either demand or capacity. Classic demand-side solutions include introducing congestion fees and limiting the number of slots at congested airports. The other solution is to add more capacity to the existing aviation infrastructure, either through deploying new technologies , or building new physical infrastructure. In the US, the major program for technology deployment is FAA's $160 billion-dollar Next Generation Air Transportation System (NextGen), (GAO, 2010), which aims to transform the NAS from a ground-based into a satellite-based system, through the deployment of seven major technology elements (FAA, 2012b). Physical infrastructure improvement refers to the reconstruction and extension of runways, which may involve concurrent taxiway and airspace reconfiguration. According to FAA (2007a), four major airports in the US currently need more capacity. The number will grow to more than two dozen by 2025. Between 2010 and 2015, the reported investment needs for airport capacity expansion amounts to $19 billion (Hansen, 2010). In order to estimate the return on this investment, appropriate benefit assessment methodologies are of critical importance.

This research proposes a novel, equilibrium-based framework to assess benefits from aviation infrastructure capacity investment. Different from the conventional approach which often focuses exclusively on flight delay reduction, our work takes a more holistic view of aviation system response to capacity investment. We recognize delay as one of

1

the several integral components in the air transport supply-demand equilibrium, and that capacity investment triggers further interactions among system components, leading to an equilibrium shift. System equilibrium, equilibrium shift and the associated benefits are modeled from both airline competition and user behavior perspectives. Our research fills an important gap in the literature and can lead to improvements of the current benefit assessment methodologies for aviation capacity investment.

## 1.2   Current Practices, Existing Studies, and Research Framework

While conventional aviation capacity investment analysis are conducted at the airport level (FAA, 1999), over the years the ability to assess the economic value of aviation infrastructure investment has become increasingly sophisticated. Considerable strides have been made in NAS-wide simulation tools, such as NASPAC, ACES, and LMINET, to incorporate flight trajectories, weather, en-route and airport capacity constraints to perform capacity investment analysis at the system level (Post, 2006; Post et al., 2008). On the other hand, the measurement of benefits is still largely delay based, with delay reduction benefits taking the form of passenger travel time and airline cost savings (e.g. Steinbach and Giles, 2005). The underlying assumption is that the only changes arising from a capacity investment that need to be considered are the increase in capacity and consequent reduction in delay. Very limited attention has been paid to the mechanisms through which airlines and air travelers respond to flight delay reduction. This delay-centric approach sometimes causes problems. In practice, delays predicted by the simulation models can appear too high to be realistic. To cope with this, analysts resort to ad hoc remedies such as smoothing the flight schedule, or simply truncating delay estimates. Thus, despite the added capability to simulate operations, the lack of comprehensive understanding of the system response may lead to a biased and incomplete representation of the investment benefits.

Air transport system response to capacity investment has been considered more in the academic arena, but with the majority of the studies conducted at the airport level. There is a long tradition in theoretical studies to model passenger demand as a function of either delay alone (Morrison and Winston, 1983), or one part of the passenger generalized cost (Morrison and Winston, 1989, 2007; Oum et al., 2004; Pels and Verhoef, 2004; Zhang and Zhang, 2006). But the systemic impacts of flight delay impact certainly extend beyond that. Jorge and de Rus (2004) highlight the inclusion of delay savings for both existing and diverted passenger traffic in airport investment benefit analysis. With delays aircraft have to spend more time either on the ground or in the air, increasing airline operating expenses. The impact of delay on airline cost has been empirically tested and quantified in Hansen et al. (2001) and Zou and Hansen (2012). Part of these delay-induced expenses will be passed onto passengers through higher fare, which in turn leads to lower demand, as modeled in Miller and Clarke (2008). However, theoretical and empirical knowledge about the extent of delay cost passage is still lacking. As a result, modeling of the percentage of delay cost transfer in Miller and Clarke (2008) has to rely on simulation techniques.

2

In addition to demand and fare responses, an important piece that has not been adequately captured in aviation capacity investment analysis is service quality. The aviation system, like urban transit systems, offers scheduled transportation services. Flight delay is certainly one dimension in the service quality. Another is the quantity of services provided, which has been extensively studied in the urban transit context (e.g. Mayworm et al., 1980; Frankena, 1983; Else, 1985). In air transportation research, metrics for system-wide service quantity include Available Seat Miles (ASM), Available Plane Miles (APM), and Revenue Passenger Miles (RPM); at the more disaggregate, flight segment level, service quantity is measured in flight frequency. Flight frequency determines passenger schedule delay (Panzar, 1979; Abrahams, 1983; Hansen, 1990; Ghobrial and Kanafani, 1995), which measures the difference between a passenger's preferred departure time and the nearest flight departure time, and is therefore part of an individual's total trip time. Most studies assume that flight traffic is invariably proportional to passenger demand before and after investment, with only a few exceptions. Jorge and de Rus (2004) argue that new capacity would enable increase in departure frequency, and encourage the use of smaller aircraft. Hansen and Wei (2006) perform an empirical investigation on the impact of a major capacity expansion at Dallas-Fort Worth airport. They find that the delay reduction benefit may be partially offset by flight demand inducement and airline schedule adaptations. Nonetheless, neither of them gives explicit calculation of passenger schedule delay change.

Associated with flight frequency adjustment is a feedback loop between frequency and passenger demand, first found and commonly known in public transit as the Mohring effect (Mohring, 1972), which refers to the fact that frequency increase reduces schedule delay, and travelers' generalized cost. As a consequence, more demand will be generated, leading to an even higher frequency. This positive feedback effect has been generalized to other scheduled transportation services, including air transport (Hansen, 1995; Pels and Verhoef, 2002). However, only limited attention to this effect has been paid in aviation infrastructure investment research (Martin and Socorro, 2007). The Mohring effect becomes even more important in hub-and-spoke air transport networks, where frequency increase on a hub-spoke segment will benefit passengers on many routes that share the same segment. The resulting demand increase on these many routes then leads to higher segment passenger traffic and further frequency increase. The enhancement of Mohring effect in hub-and-spoke networks has been rarely recognized in the literature.

Perhaps equally important to the Mohring effect is economies of density, another intrinsic feature that has not been fully considered in the aviation infrastructure investment analysis. Economies of density refers to airlines' declining average cost from denser traffic within a given network. The existence of the economies of density has been empirically identified at both airline (Cave et al., 1984; Gillen et al., 1985, 1990) and route (Brueckner and Spiller, 1994; Brueckner et al., 2011) levels. As pointed out above, aviation infrastructure investment reduces delay and invites more traffic. This results in reduced airline average cost. Lower unit cost allows airlines to offer passenger less expensive fares. As a result, more passengers will be induced. Like the Mohring effect, the economies of density induces positive feedback loop.

3

While reduced flight delay and schedule delay improve the quality of the air service product, plausibly resulting in higher airfare, it is our belief that the overall effect of flight delay reduction, schedule delay decrease, and the economies of density is lower traveler generalized cost. We illustrate this in Figure 1.1, where travelers generalized cost, measured in $/mile, is plotted against total demand in the system, quantified as passenger-miles. With no congestion, the traveler supply curve will be a downward slopping curve $S_0$ because of the Mohring effect and the economies of density, intersecting with the demand curve at point $G$. When congestion occurs, the new constrained traveler supply curve, $S_1$, will track the unconstrained one $S_0$ until delay appears in the system, after which $S_1$ starts to veer upwards, intersecting the demand curve at point $B$. The purpose of investing in infrastructure capacity is to shift the point at which this deviation occurs to the right, and hence downward, such that the new constrained traveler supply curve will intersect with the demand curves at point $C$. Infrastructure investment then leads to a reduction in traveler generalized cost and higher system demand, denoted respectively by the distance $AD$ and $CE$. The area $ABCD$ represents passenger welfare gains from the investment. However, because profit change involves changes in passenger demand, airfare, flight frequency, and unit operating cost, it is difficult to discern graphically airlines' (producers') surplus.
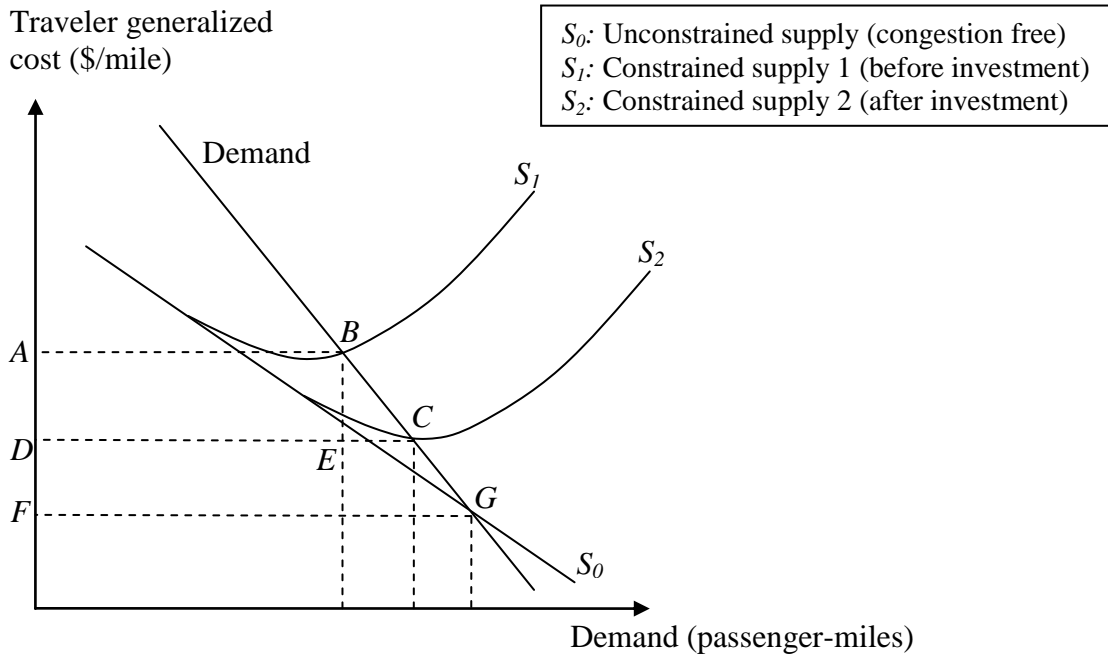


Figure 1.1: Traveler generalized cost as a function of system demand

To explicate the various causal relationships mentioned above, we propose the following equilibrium-based framework to model system response to aviation infrastructure investment, as shown in Figure 1.2. We consider flight delay, passenger demand, airfare, flight traffic, and airline cost as five endogenous system components. An equilibrium is characterized by a set of consistent values of the system components. We hypothesize that, once infrastructure investment is made, increased capacity leads to lower flight

4

delay, which induces more travelers to use the air transportation system, encourages airlines to schedule more flights, and reduces airline unit operating cost. With unit cost reduction, airlines will make necessary fare adjustment in their profit maximization process, which also depends upon the market structure. New airfare and flight traffic affect traveler generalized cost, and therefore passenger demand. Other demand-influencing factors, including socio-economic characteristics, competition from other modes, can be reasonably regarded as exogenous to system responses. Change in demand in turn affects airfare, again through airlines' profit maximization, and the amount of flight traffic. Like airfare, the determination of flight traffic depends further upon the market structure. New flight traffic suggests a new level of flight delay in the system. The changes in flight traffic and delay enter airlines' production process, in which input prices are assumed exogenous, resulting in an updated airline cost. It is clear that, once capacity investment is made, it will trigger a complicated set of interactions among the system components, the final outcome of which is characterized by an equilibrium shift. We use a dashed line to represent the potential feedback from delay to investment decisions, in that the response of capacity investment is on a much longer time scale than those of the boxed system components. For this reason, investment decisions are assumed exogenous in the dissertation.
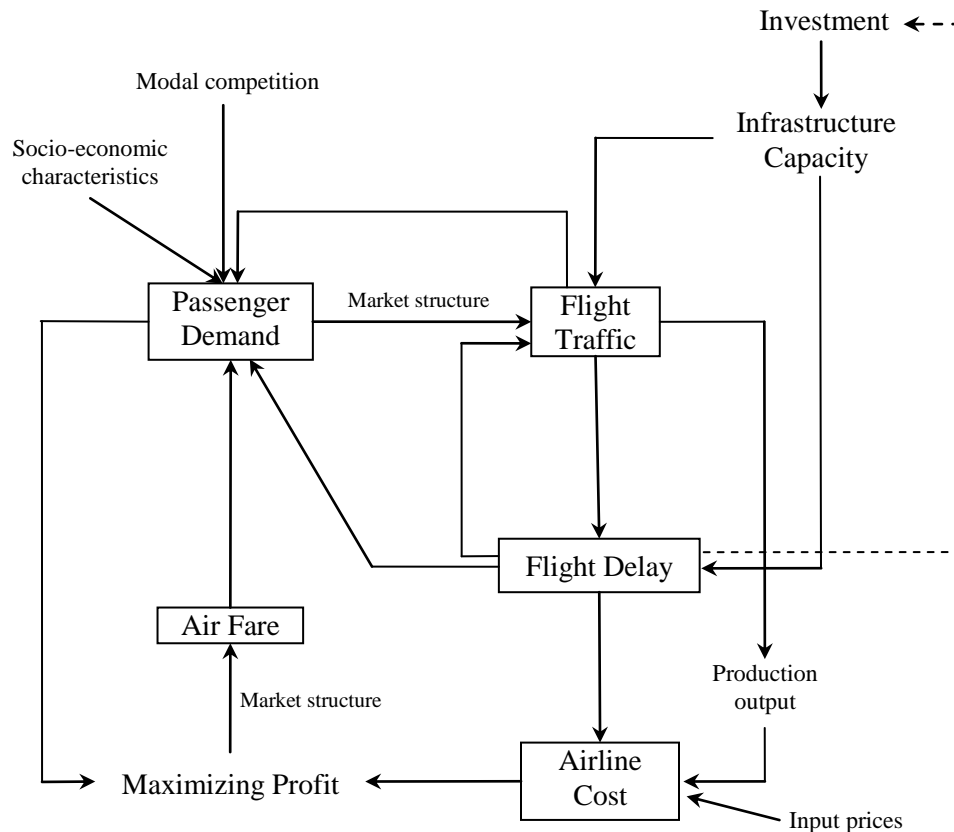


Figure 1.2: Equilibrium framework

## 1.3 Research Methods and Organization of the Dissertation

Based upon the dual relationship between supply and demand, we can model the air transport system equilibrium from two complementary perspectives. On the one hand, the system equilibrium is the result of airlines' profit maximization. Airlines seek their best pricing and scheduling strategies taking into account the reaction of passenger demand, which is determined by the utility structure of travelers. In particular, we assume that airlines, in maximizing profit, explicitly recognize the delay impact on passenger utility structure as well as their own operating cost. Since it is likely that multiple airlines compete in an air transportation system, the achievement of the equilibrium will involve gaming behavior. Following this logic, in the first approach an analytical model will be developed to study the air transport equilibrium from the airline competition perspective. While airport capacity constraints have been considered in some recent airline game-theoretic modeling (e.g. Evans, 2010; Li et al., 2010; Vaze and Barnhart, 2012), our airline competition model deals with simultaneous price and frequency competition specifically in the context of capacity investment. The analytical nature of our model also provides some useful theoretical insights into the interplays among the system components.

Alternatively, the equilibrium problem can be viewed from travelers' vantage point. It is reasonable to assume that every traveler tries to maximize her/his utility when making travel decisions, with full knowledge about the market supply (i.e. fare and frequency) and performance (i.e. flight delay) characteristics, which change with the passenger demand pattern. An equilibrium is achieved when no traveler can improve her/his utility by unilaterally changing the demand choice. This characterizes the user equilibrium condition. In the second approach, we examine the air transport user equilibrium with system components all based upon empirical models. We will demonstrate that the air transport user equilibrium combines features from both classic supply-demand equilibrium and demand-performance equilibrium, the latter of which widely applied to the urban traffic context. To our knowledge, it is the first time to introduce user equilibrium in the air transport context. We will show that our key findings from both approaches are largely consistent.

The remainder of this dissertation is organized as follows. In Chapter 2, we first present the formulation of an analytical, airline competition equilibrium model. Comparative static and numerical analyses of equilibrium shift in response to capacity change are then discussed. We estimate the associated benefit gains and compare them with those from the conventional approach. The discussions then switch to the modeling of air transport user equilibrium. In Chapter 3 we perform an empirical investigation of the delay impact on airfare and flight frequency, the two most important supply-side characteristics. Chapter 4 then proceeds to the full investigation of the air transportation user equilibrium. Equilibrium components, the formulation of the equilibrium, and solution algorithm are discussed in order. We apply the user equilibrium concept to a hypothetical air transport network, in which—similar to Chapter 2—we examine the initial equilibrium and equilibrium shift, and compare benefit gains from the equilibrium approach with

estimates using the conventional method. We offer conclusions and point out directions for future research in Chapter 5.

# 2. An Analytical Airline Competition Equilibrium Model

## 2.1 Introduction

In this chapter, we first apply the research framework proposed in Section 1.2 to an airline competition model to explore the capacity-related air transport supply-demand equilibrium and how the equilibrium shifts in response to capacity expansion. We assume that airlines determine their fare and frequency in a competitive environment, taking into account individuals' utility structure. Flight delay affects both travel utility of individuals and operating cost of airlines. Despite the existence of a large body of theoretical literature analyzing the economics of airline competition behavior, so far relative few efforts have been devoted to airline behavior vis-à-vis infrastructure capacity constraints and investment. The following analytical model intends to provide some helpful insights into the interplays among passenger demand, air fare, airline cost, flight traffic and delay, from a microscopic, airline competitive point of view.

## 2.2 Model setup

### 2.2.1 Demand

We consider a duopoly city-pair airline market, a special case of oligopolistic markets. Two carriers are engaged in price and frequency competition. Following most theoretical and applied literature of this kind (e.g. Schipper et al., 2003; Brueckner and Girvin, 2008; Brueckner and Zhang, 2010), we restrict our attention to the symmetric equilibrium, i.e. the two airlines are identical, to preserve analytical tractability. As previously discussed, travelers consider both fare and service quality when making travel decisions. In the absence of capacity constraints, the primary service quality dimension is schedule delay, defined as the difference between a traveler's desired departure time and the closest scheduled departure time of all flights. Although individual passengers are concerned about their specific departure time, it is reasonable to use frequency to capture the overall schedule delay effect when market demand is concerned. Empirical studies often use the inverse of frequency (Eriksen, 1978; Abrahams, 1983), which is intuitive if we consider a situation where flight departures and passenger demand are uniformly distributed along a time circle of length $T$. Then the expected schedule delay equals $T/4f$, with flight frequency being $f$ (flights). The schedule delay cost is the expected schedule delay multiplied by some cost parameter $\gamma > 0$. This kind of treatment is adopted by many similar studies (e.g. Richard, 2003; Brueckner and Flores-Fillol, 2007; Brueckner and Girvin, 2008).

In the absence of traffic delay, a representative consumer will face two generalized costs (prices) corresponding to the services provided by two airlines: $\overline{P}_i = P_i + \dfrac{\gamma}{f_i}$, for $i=1,2$. We assume the representative consumer has the following utility function:

$$U(q_0,q_1,q_2) = q_0 + \frac{\alpha_{00}}{\alpha_{01}-\alpha_{02}}(q_1+q_2) - \frac{1}{2}\frac{1}{\alpha_{01}^2-\alpha_{02}^2}(\alpha_{01}q_1^2 + 2\alpha_{02}q_1q_2 + \alpha_{01}q_2^2) \qquad (2.1)$$

where $q_0$ represents the numeraire good. $\alpha_{00}, \alpha_{01}, \alpha_{02}$ are positive parameters. The concavity condition requires $\alpha_{01} > \alpha_{02}$. The representative consumer maximizes $U(q_0,q_1,q_2)$, subject to the following income (budget) constraint:

$$q_0 + \overline{P}_1 q_1 + \overline{P}_2 q_2 = I \qquad (2.2)$$

where $I$ denotes income. The first-order conditions of the corresponding Lagrangian $L$, $U(q_0,q_1,q_2) - \lambda(q_0 + \overline{P}_1 q_1 + \overline{P}_2 q_2 - I)$ with $\lambda$ being the Lagrange multiplier, are

$$\frac{\partial L}{\partial q_0} = 1 - \lambda = 0 \qquad (2.3.1)$$

$$\frac{\partial L}{\partial q_1} = \frac{\alpha_{00}}{\alpha_{01}-\alpha_{02}} - \frac{\alpha_{01}}{\alpha_{01}^2-\alpha_{02}^2}q_1 - \frac{\alpha_{02}}{\alpha_{01}^2-\alpha_{02}^2}q_2 - \lambda\overline{P}_1 = 0 \qquad (2.3.2)$$

$$\frac{\partial L}{\partial q_2} = \frac{\alpha_{00}}{\alpha_{01}-\alpha_{02}} - \frac{\alpha_{02}}{\alpha_{01}^2-\alpha_{02}^2}q_1 - \frac{\alpha_{01}}{\alpha_{01}^2-\alpha_{02}^2}q_2 - \lambda\overline{P}_2 = 0 \qquad (2.3.3)$$

$$\frac{\partial L}{\partial \lambda} = -(q_0 + \overline{P}_1 q_1 + \overline{P}_2 q_2 - I) = 0 \qquad (2.3.4)$$

The second-order conditions are guaranteed since the Hessian is negative semi-definite given the concavity of the utility function. Substituting (2.3.1) into (2.3.2) and (2.3.3) yields the following system of linear inverse demand functions:

$$\overline{P}_i = \frac{\alpha_{00}}{\alpha_{01}-\alpha_{02}} - \frac{\alpha_{01}}{\alpha_{01}^2-\alpha_{02}^2}q_i - \frac{\alpha_{02}}{\alpha_{01}^2-\alpha_{02}^2}q_{-i}, \quad i=1,2 \qquad (2.4)$$

where the subscript $–i$ denotes the competing airline. Incorporating the generalized cost expression and solving (2.4) for $i=1,2$ lead to the following "symmetric" demand function

$$q_i = \alpha_{00} - \alpha_{01}P_i + \alpha_{02}P_{-i} - \frac{\alpha_{01}\gamma}{f_i} + \frac{\alpha_{02}\gamma}{f_{-i}}, \quad i=1,2 \qquad (2.5)$$

The market-level airline demand functions, $Q_i$ ($i=1,2$), are obtained by aggregating $q_i$'s over all consumers

9

$$Q_i = \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1 \gamma}{f_i} + \frac{\alpha_2 \gamma}{f_{-i}}, \quad i = 1,2 \tag{2.6}$$

where $\alpha_0 = n\alpha_{00}, \alpha_1 = n\alpha_{01}, \alpha_2 = n\alpha_{02}$, with $n$ being the number of consumers in the market. Obviously $\alpha_1 > \alpha_2$, suggesting that the services provided by the two airlines are imperfect substitutes. The above demand function presents a general carrier-level demand functional form, which differs from a recent paper studying airport congestion by Flores-Fill (2010), where a fixed total demand is assumed. From one perspective, the assumption of fixed total demand is a nice property for analytical tractability since the focus of their study is on congestion. On the other hand, under our demand setup, an increase in ticket price of airline 1 will divert some passengers to airline 2. Our specification further allows some passengers who would have chosen airline 1 if price were not increased to not travel by either airline–they may choose alternative modes, or not traveling at all. Likewise, if airline 1 increases its frequency, then it can not only draw passengers from firm 2 but also generate additional demand. In effect, this market-level demand response presents another important phenomenon caused by congestion.

When congestion emerges due to limited capacity, passengers will suffer directly from flight delay because they value the extra trip time. This adds a new component into the generalized cost. We assume the congestion cost to passengers is identical across passengers regardless of which airline was chosen. We use the average flight delay $L$ and multiply it by a cost factor $k$ to represent the contribution of delay to passenger generalized cost. Following the same derivations as above, the new demand function can be written as

$$Q_i = \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1 \gamma}{f_i} + \frac{\alpha_2 \gamma}{f_{-i}} - \mu L, \quad i = 1,2 \tag{2.7}$$

where $\mu = k \cdot n \cdot (\alpha_{01} - \alpha_{02}) = k \cdot (\alpha_1 - \alpha_2)$ is the coefficient indicating the unit impact of delay on demand. Previous studies model $L$ at the airport level and to be a function of total traffic volume and capacity (e.g. Morrison and Winston, 2008; Zhang, 2010). As one city pair is considered here, we assume $L$ to be a function of the larger of the traffic volume/capacity ratios from the two airports in the city pair. The airport with the larger ratio is defined as the "focal" airport. In the subsequent analysis, we assume the arrival end of the city pair presents the focal airport, which is the terminus of $N$ identical markets, and is the only airport with a significant capacity limitation. We further assume that the decision-making of each market is independent. Then the total traffic volume of arriving flights at the focal airport is $N(f_1 + f_2)$.[1] The traffic volume/capacity ratio is $N(f_1 + f_2)/K$,

---

[1] Since at an airport departure and arrival traffic volumes are almost equivalent, it would suffice to only consider the arrival traffic volume in modeling airport delay. In effect, Morrison and Winston (2008) find that no significant difference would result from considering total flight operations and departures/arrivals separately. For other airport delay studies, the primary concern is often flight arrival delays (e.g. Hansen, 2002; Hansen et al., 2010). Therefore, in this study we focus on the arrival traffic volume at the focal airport, and the term traffic volume in the rest of the chapter refers specifically to traffic volume of arrivals.

with $K$ denoting the arrival capacity at the focal airport. Given a fixed capacity and the number of markets, $L$ is simply a function of $f_1 + f_2$, i.e. $L = L(f_1 + f_2)$.

## 2.2.2 Supply

We follow Brueckner and Flores-Fillol (2007), by assuming that an airline operates aircraft with size $s$ and a load factor of 1 (in fact, for the latter all we require is a constant load factor). A flight's operating cost is given by $c_0 + \tau s$, where $c_0$ is a positive fixed cost independent of aircraft size and $\tau$ the marginal cost per seat. This specification reflects in part the economies of density on the supply side,[2] as cost per passenger is decreasing with aircraft size. For airline $i$ ($i$=1,2), flight frequency ($f_i$), aircraft size ($s_i$), and demand ($Q_i$) are related by the equation $Q_i = f_i \cdot s_i$. Additional expenses will be generated when flight delay occurs, as it is associated with more fuel burn, additional crew cost, etc. These are incorporated in a third term in the flight operating cost:

$$C_i = c_0 + \tau s_i + \eta s_i L \tag{2.8}$$

where $\eta$ is a cost factor associated with a unit time of delay per seat. The delay cost per flight is assumed to be a function of aircraft size ($s_i$) and the length of delay ($L$). Given $L$, a larger plane requires more extra fuel consumption and higher crew cost than a smaller one.

## 2.2.3 Competition and equilibrium

In this duopoly market, airlines compete on fare and frequency to maximize profits. The profit function for each airline is:

$$\pi_i = P_i \cdot Q_i - f_i \cdot C_i = P_i(\alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1 \gamma}{f_i} + \frac{\alpha_2 \gamma}{f_{-i}} - \mu L) - f_i(c_0 + \tau s_i + \eta s_i L)$$

$$= (P_i - \tau - \eta L)(\alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1 \gamma}{f_i} + \frac{\alpha_2 \gamma}{f_{-i}} - \mu L) - f_i c_0, \quad \text{for } i = 1,2$$

$$\tag{2.9}$$

Depending on the assumptions made, the competition between the two airlines can follow different game models. We consider the case that flight frequency and fare can be adjusted simultaneously in a Nash fashion. The reasoning rests on the fact that typically airlines adjust schedules every 3 month (Ramdas and Williams, 2008) and travelers may also purchase tickets months in advance. The first order conditions (FOC) for airline 1 are:

---

[2] From carriers' perspective, the economies of density includes four aspects: the use of larger and more efficient aircraft, higher load factors, more intensive use of fixed ground facilities, and more efficient aircraft utilization (Brueckner and Spiller, 1994). In this chapter as load factor is assumed to be 1, economies of density on the supply side are primarily embodied in the first aspect.

$$\frac{\partial \pi_1}{\partial P_1} = (\alpha_0 - \alpha_1 P_1 + \alpha_2 P_2 - \frac{\alpha_1 \gamma}{f_1} + \frac{\alpha_2 \gamma}{f_2} - \mu L) - \alpha_1 (P_1 - \tau - \eta L) = 0 \qquad (2.10.1)$$

$$\frac{\partial \pi_1}{\partial f_1} = (P_1 - \tau - \eta L)(\frac{\alpha_1 \gamma}{f_1^2} - \mu \frac{\partial L}{\partial f_1}) - c_0 - \eta \frac{\partial L}{\partial f_1}(\alpha_0 - \alpha_1 P_1 + \alpha_2 P_2 - \frac{\alpha_1 \gamma}{f_1} + \frac{\alpha_2 \gamma}{f_2} - \mu L) = 0 \quad (2.10.2)$$

Note that $Q_1 = \alpha_0 - \alpha_1 P_1 + \alpha_2 P_2 - \frac{\alpha_1 \gamma}{f_1} + \frac{\alpha_2 \gamma}{f_2} - \mu L > 0$. The fact that airlines should make a positive profit implies $(P_1 - \tau - \eta L) > 0$. Since $L$ increases with frequency, $\frac{\alpha_1 \gamma}{f_1^2} - \mu \frac{\partial L}{\partial f_1} = (c_0 + \eta \frac{\partial L}{\partial f_1} Q_1)/(P_1 - \tau - \eta L) > 0$ according to (2.10.1) and (2.10.2). For the delay function $L$, we further expect marginal delay increase is greater when traffic is at a higher level, i.e. $\partial^2 L / \partial f^2 > 0$. Then the second-order derivatives

$$\frac{\partial^2 \pi_1}{\partial P_1^2} = -2\alpha_1 \qquad (2.11.1)$$

$$\frac{\partial^2 \pi_1}{\partial f_1^2} = (P_1 - \tau - \eta L)(-\frac{2\alpha_1 \gamma}{f_1^3} - \mu \frac{\partial^2 L}{\partial f_1^2}) - 2\eta \frac{\partial L}{\partial f_1}(\frac{\alpha_1 \gamma}{f_1^2} - \mu \frac{\partial L}{\partial f_1}) - \eta \frac{\partial^2 L}{\partial f_1^2}(\alpha_0 - \alpha_1 P_1 + \alpha_2 P_2 - \frac{\alpha_1 \gamma}{f_1} + \frac{\alpha_2 \gamma}{f_2} - \mu L)$$

$$(2.11.2)$$

are easily seen to be negative. The remaining of the second-order condition (i.e. negative definitiveness of the Hessian matrix of $\pi_1$) is assumed to hold.[3]

The first and second order optimality conditions also apply to airline 2. The FOCs are obtained by interchanging subscripts 1 and 2 in (2.10.1) and (2.10.2). Given the symmetry set-up, under equilibrium $P_1 = P_2 = P$, $f_1 = f_2 = f$. Replacing fare and frequency by $P$ and $f$ in the FOC of the fare equation (10.1), we have

$$P = \frac{\alpha_0 - \frac{(\alpha_1 - \alpha_2)\gamma}{f} - \mu L + \alpha_1 (\tau + \eta L)}{2\alpha_1 - \alpha_2} \qquad (2.12)$$

Substituting the above into the FOC frequency equation (2.10.2) yields

---

[3] In our case, this requirement reduces to $-2\alpha_1 \frac{\partial^2 \pi_1}{\partial f_1^2} - (\frac{\gamma_1}{f_1^2} - \mu \frac{\partial L}{\partial f_1} + \alpha_1 \eta \frac{\partial L}{\partial f_1})^2 > 0$. These 2nd order conditions are always satisfied in the following numerical analyses.

$$\frac{\alpha_0 - \dfrac{(\alpha_1-\alpha_2)\gamma}{f} - (\alpha_1-\alpha_2)\tau}{2\alpha_1 - \alpha_2} \cdot \frac{\alpha_1\gamma}{f^2} = c_0 + (\mu + \alpha_1\eta)\frac{\alpha_0 - \dfrac{(\alpha_1-\alpha_2)\gamma}{f} - \mu L - (\alpha_1-\alpha_2)(\tau + \eta L)}{2\alpha_1 - \alpha_2} \cdot \frac{\partial L}{\partial f} + \frac{[\mu + (\alpha_1-\alpha_2)\eta]L}{2\alpha_1 - \alpha_2} \cdot \frac{\alpha_1\gamma}{f^2}$$

$$(2.13)$$

In order to discern potential frequency changes when delay occurs, Equation (2.13) needs to be simplified. The last term on the right hand side (RHS) of (2.13) is positive, as $\alpha_1 > \alpha_2$. So is the second-to-last term on the RHS, since substituting (2.12) into this term yields $(\mu + \alpha_1\eta)(P - \tau - \eta L)(\partial L/\partial f)$, which is greater than zero following the FOC discussion. Then the RHS of (2.13) is positive. Note that all terms except $c_0$ on the RHS are due to the presence of congestion. For simplicity we denote them by $D$. The RHS then becomes $c_0 + D$. The left hand side (LHS) is only a function of $f$.

The increase on the RHS due to congestion leads to an equivalent increase on the LHS, through changing the value of $f$. To study the monotonicity of the LHS, we define a new function $F = [\alpha_0 - \frac{(\alpha_1 - \alpha_2)\gamma}{f} - (\alpha_1 - \alpha_2)\tau]/f^2$. Taking its first order derivative with respect to $f$, we obtain

$$\frac{\partial F}{\partial f} = \{3(\alpha_1 - \alpha_2)\gamma - 2f[\alpha_0 - (\alpha_1 - \alpha_2)\tau]\}/f^4 \qquad (2.14)$$

Our *a priori* expectation is that airlines tend to schedule fewer flights when delay occurs. This suggests that $F$ be monotonically decreasing, or $\partial F/\partial f < 0$, which is equivalent to:

$$\frac{(\alpha_1 - \alpha_2)\gamma}{f} < \frac{2}{3}[\alpha_0 - (\alpha_1 - \alpha_2)\tau] \qquad (2.15)$$

Empirical evidence suggests that it is plausible for (2.15) to hold. More details are provided in appendix A. Therefore, $\partial F/\partial f < 0$ and the LHS of (2.13) is a monotonic decreasing function. When traffic delay occurs, the RHS of (2.13) is increased by $D$. Consequently, the equilibrium frequency should adjust downwards. Let $f_0$ and $\tilde{f}_0$ denote the optimal frequency with and without delay. We have $f_0 < \tilde{f}_0$. This fact will serve as the starting point to derive a set of other results in the ensuing comparative static analysis section.

## 2.3 Comparative static analysis

### 2.3.1 Impact on air fare, passenger generalized cost and demand

The primary objective of this section is to further our qualitative insight into the impact of capacity constraint on air transportation service, by comparing the equilibrium values

with and without congestion. When congestion occurs, according to (2.12) air fare will respond in two different ways: reduced frequency (represented by $-(\alpha_1 - \alpha_2)\gamma/[f(2\alpha_1 - \alpha_2)]$) and flight delay (represented by $-\mu L/(2\alpha_1 - \alpha_2)$) degrade the service quality and therefore reduce the willingness-to-pay (out of their pocket) of travelers. Therefore, the new equilibrium fare tends to be lower. On the other hand, congestion imposes $\eta L$ on airline operating cost for each passenger carried. The term $\alpha_1 \eta L/(2\alpha_1 - \alpha_2)$ in (2.12) shows that airlines would pass $\alpha_1/(2\alpha_1 - \alpha_2)$ portion of their delay-induced operating cost to passengers. This term also implies that, when the substitution effect between the two airlines is stronger (that is, as $\alpha_2 \to \alpha_1$), airlines tend to pass a larger portion of their delay cost to passengers. In normal cases, the portion should be greater than ½ since $0 < \alpha_2 < \alpha_1$. Overall, the two opposing tendencies of price response blur the changes in ticket price. The changes in fare will be explored numerically in the next section.

Recall that the generalized cost to each passenger consists of air fare, frequency, and traffic delay. The demand can be written as a function of a single generalized cost $\bar{P}$.

At equilibrium, demand for each carrier is

$$Q_i = \alpha_0 - (\alpha_1 - \alpha_2)[P + \frac{\gamma}{f} + \frac{\mu L}{\alpha_1 - \alpha_2}] = \alpha_0 - (\alpha_1 - \alpha_2)\bar{P}, \quad i = 1,2 \tag{2.16}$$

Recall in section 3.1 that the contribution of delay to each passenger's generalized cost is $kL$, and $\mu$ is defined as $k \cdot (\alpha_1 - \alpha_2)$. Substituting (2.12) into $P$ above, the generalized cost under equilibrium, $\bar{P}_0$, becomes

$$\bar{P}_0 = \frac{\alpha_0 + \alpha_1(\tau + \eta L)}{2\alpha_1 - \alpha_2} + \frac{\gamma}{f_0}\frac{\alpha_1}{(2\alpha_1 - \alpha_2)} + \frac{\alpha_1}{(\alpha_1 - \alpha_2)(2\alpha_1 - \alpha_2)}\mu L \tag{2.17}$$

When there is no delay, generalized cost equals

$$\bar{\tilde{P}}_0 = \frac{\alpha_0 + \alpha_1 \tau}{2\alpha_1 - \alpha_2} + \frac{\gamma}{\tilde{f}_0}\frac{\alpha_1}{(2\alpha_1 - \alpha_2)} \tag{2.18}$$

Comparing (2.17) with (2.18), two delay-related terms are added in (2.17) when congestion occurs: $\alpha_1 \eta L/(2\alpha_1 - \alpha_2)$ and $\alpha_1 \mu L/(\alpha_1 - \alpha_2)(2\alpha_1 - \alpha_2)$. The first term corresponds to the aforementioned delay cost transfer from carriers; the second term denotes the passenger delay cost, which is the net of direct passenger delay cost ($\mu L/(\alpha_1 - \alpha_2)$) and the price drop due to delay ($-\mu L/(2\alpha_1 - \alpha_2)$) described before. Considering further that $f_0 < \tilde{f}_0$, it is easy to show $\bar{P}_0 > \bar{\tilde{P}}_0$, i.e. generalized cost will increase.

14

A direct consequence of passenger generalized cost increase is suppressed demand for each airline and in the market. Alternatively, airline demand can be expressed as only a function of frequency, by substituting (2.12) for $P$ into the demand function (2.16)

$$Q_{0,i} = \frac{\alpha_1}{2\alpha_1 - \alpha_2}[\alpha_0 - (\alpha_1 - \alpha_2)\tau - \frac{(\alpha_1 - \alpha_2)\gamma}{f_0}] - \frac{\alpha_1}{2\alpha_1 - \alpha_2}[\mu + (\alpha_1 - \alpha_2)\eta]L, \quad i = 1,2 \qquad (2.19)$$

When there is no delay, the corresponding $\tilde{Q}_{0,i}$ equals

$$\tilde{Q}_{0,i} = \frac{\alpha_1}{2\alpha_1 - \alpha_2}[\alpha_0 - (\alpha_1 - \alpha_2)\tau - \frac{(\alpha_1 - \alpha_2)\gamma}{\tilde{f}_0}], \quad i = 1,2 \qquad (2.20)$$

Given $f_0 < \tilde{f}_0$ and the additional delay effect term ( $\alpha_1/(2\alpha_1 - \alpha_2) \cdot [\mu + (\alpha_1 - \alpha_2)\eta]L$ ) in (2.19), demand for each airline becomes less when delay occurs, i.e. $Q_{0,i} < \tilde{Q}_{0,i}, i = 1,2$.

### 2.3.2 Impact on aircraft size and unit operating cost

Although aircraft size is not considered as a decision variable, in our model context it is implicitly determined by passenger demand and the number of flights scheduled. Since flight load factor is assumed to be 1, the aircraft size is obtained by dividing (2.19) by $f_0$:

$$s_0 = \frac{\alpha_1}{2\alpha_1 - \alpha_2} \frac{[\alpha_0 - (\alpha_1 - \alpha_2)\tau - \dfrac{(\alpha_1 - \alpha_2)\gamma}{f_0}]}{f_0} - \frac{\alpha_1}{2\alpha_1 - \alpha_2} \frac{[\mu + (\alpha_1 - \alpha_2)\eta]L}{f_0} \qquad (2.21)$$

For the first term on the RHS, both the denominator and numerator become smaller when traffic delay is considered. Nonetheless, it is plausible for the first order derivative to be negative.[4] This just confirms that demand is inelastic with respect to frequency. However, the second term presents an opposite effect, the effect of delay on suppressing demand. Therefore, the changing direction of aircraft size is inconclusive. The change in the unit operating cost $\tau + \dfrac{c_0}{s_0} + \eta L$ is also left indeterminate as a consequence.

### 2.3.3 Changes in consumer welfare

The increase in passengers' generalized cost and the reduction in demand that result from delay are shown in Figures 2.1 and 2.2, for airlines 1 and 2 respectively, where the abscissa and ordinate denote airline passenger demand and generalized cost. Because demand for one airline also depends upon the generalized cost of the other airline, both

---

[4] $\partial\{[\alpha_0 - (\alpha_1 - \alpha_2)\tau - (\alpha_1 - \alpha_2)\gamma/f]/f\}/\partial f = \{2(\alpha_1 - \alpha_2)\gamma/f - [\alpha_0 - (\alpha_1 - \alpha_2)\tau]\}/f^2$. Focusing on the numerator, as $\tau < P$ we have $2(\alpha_1 - \alpha_2)\gamma/f - [\alpha_0 - (\alpha_1 - \alpha_2)\tau] < 2(\alpha_1 - \alpha_2)\gamma/f - [\alpha_0 - (\alpha_1 - \alpha_2)P] = (\alpha_1 - \alpha_2)\gamma/f - Q$. Since in general $\varepsilon_f^0$ is less than 1, the RHS of the above is negative. Therefore, it is plausible that the first term on the RHS of (18) is a decreasing function of $f$.

demand curves move outward when delay takes place. The overall outcomes are equilibrium shifts from *B* to *A* and from *F* to *E*, for airlines 1 and 2.

To measure changes in consumer welfare, the classical tool is consumer surplus. Since the utility function is specified as quasi-linear, consumer surplus is also an exact measure of consumer welfare (Varian, 1992). When delay occurs, CS loss arises from increase in both airlines' generalized cost. Despite the many potential paths realizing this generalized cost change, the fact that $\frac{\partial q_1}{\partial p_2} = \frac{\partial q_2}{\partial p_1}$ guarantees the calculation of CS to be path independent (Mishan, 1977; Turnovsky, 1980). Here we choose the following two-step path, as indicated in Figures 2.1 and 2.2. In the first step, we increase the generalized cost of airline 1 from $\widetilde{\overline{P}}_{0,1}$ to $\overline{P}_{0,1}$, with the generalized cost of airline 2 being provisionally unchanged. The corresponding CS loss is the area $\widetilde{\overline{P}}_{0,1}\overline{P}_{0,1}DB$ in Figure 2.1. As a direct result of the rise in airline 1's generalized cost, the demand curve for airline 2 now moves outward from $D_2^0$ to $D_2^1$. Following the adjustment, in the second step the generalized cost of airline 2 rises from $\widetilde{\overline{P}}_{0,2}$ to $\overline{P}_{0,2}$, with the further loss of CS given by the area $\widetilde{\overline{P}}_{0,2}\overline{P}_{0,2}EH$ in Figure 2.2. Concurrent with this is the horizontal move of airline 1's demand curve from D to A (Figure 2.1). The total CS loss is calculated by adding together the two areas: $\widetilde{\overline{P}}_{0,1}\overline{P}_{0,1}DB$ and $\widetilde{\overline{P}}_{0,2}\overline{P}_{0,2}EH$ , in which loss for foregone demand consists of two triangular areas: *DBJ* and *EHG*. If this is considered as an infrastructure investment problem with reduced delay after capacity enhancement, then the sum of $\widetilde{\overline{P}}_{0,1}\overline{P}_{0,1}DB$ and $\widetilde{\overline{P}}_{0,2}\overline{P}_{0,2}EH$ is the overall CS gain, and the areas *DBJ* plus *EHG* represent the CS gain for induced demand. Given the symmetry setup, the sum of $\widetilde{\overline{P}}_{0,1}\overline{P}_{0,1}DB$ and $\widetilde{\overline{P}}_{0,2}\overline{P}_{0,2}EH$ is equal to twice the area of the trapezoid $\widetilde{\overline{P}}_{0,1}\overline{P}_{0,1}AB$ (or trapezoid $\widetilde{\overline{P}}_{0,2}\overline{P}_{0,2}EF$ ), and the two triangles *DBJ* and *EHG* are of equal size.
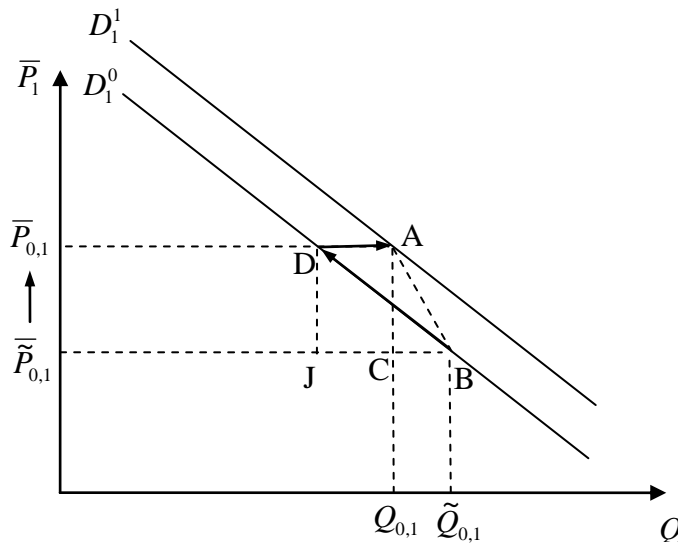


Figure 2.1: Demand as a function of generalized cost for airline 1
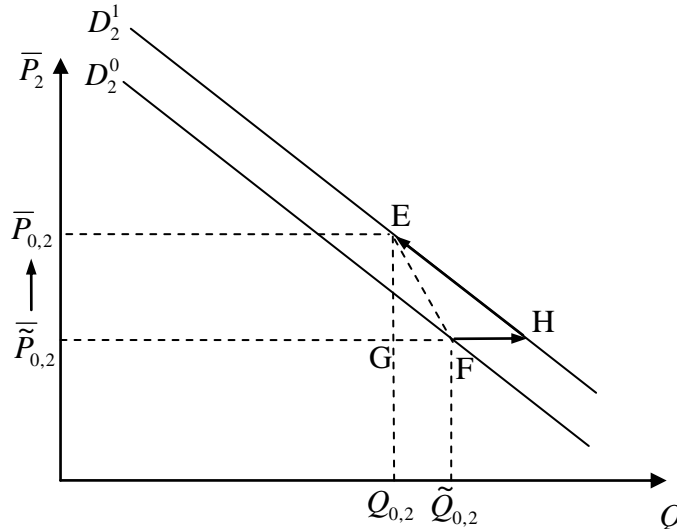
Figure 2.2: Demand as a function of generalized cost for airline 2

The welfare changes on the supply side remains analytically indeterminate due to the opposing effects of delay on ticket price, aircraft size, and flight operating cost. The ensuing section extends the comparative static analysis by numerically exploring the response of both demand and supply sides under a number of capacity scenarios.

## 2.4 Numerical analysis

To gain further insights into the supply-demand equilibrium, especially those elements that are left indeterminate in the preceding comparative static analysis, this section performs a set of numerical analyses. The direction–and to some extent magnitude as well–of the delay effects on the various elements in the equilibrium are examined. We first look at how the congestion-free equilibrium will shift when airport capacity constraint appears. We also investigate the sensitivity of the equilibrium to different capacity levels, including changes in both the supply-demand characteristics and welfare. Furthermore, since the equilibrium approach is not incorporated in the current practice of investment analysis, the differences in benefit assessment from using the conventional and equilibrium methods are compared, which shows that the equilibrium method yields more realistic and plausible estimates.

In conducting numerical analyses, the first step is to determine the parameter values of the model. Many parameter values are based on literature; some assumptions are made when empirical numbers are not available. In this section, we consider a market of roughly 1000 passengers per day in each direction, with 10 daily flights serving the market. Therefore each airline schedules approximately 5 flights per day. One-way fare is set to be $100. In light of the estimated elasticity values in literature (Oum et al., 1993; Jorge-Calderón, 1997; Gillen et al., 2002; Hsiao, 2008), price elasticities are set to be -1.25 and -2.5, at market and airline level respectively. The market frequency elasticities

17

are assumed to be 0.6. Based on the above elasticities and baseline market assumptions, the values for $\alpha_0, \alpha_1, \alpha_2, \gamma$ can be derived.[5] The travel distance is assumed to be 1000 miles, with nominal trip time being 2 hours. According to GRA (2004), aircraft operate at $4000 per hour, in which the fixed part holds $1000. Following this, the fixed operating cost $c_0$ equals $2000 per flight. The unit variable operating cost $\tau = \$3000 \times 2/100 = \$60$ per seat. We adopt an estimate cited in Barnett et al. (2001) for the average aircraft delay cost (measured in $/hr), when inflated to current value, equal to about $3000/hr. As a result $\eta$ =$3000/(60 \times 100)=$0.5/seat-min. The value of delay parameter $\mu$ is inferred from passenger value of travel time. Recall the generalized cost:

$$\bar{P} = P + \frac{\gamma}{f} + \frac{\mu L}{\alpha_1 - \alpha_2} \tag{2.22}$$

*Ceteris paribus*, a one-minute delay increases one passenger's generalized cost by $\mu/(\alpha_1 - \alpha_2)$ in the market. We use the value of travel time to approximate this amount. Using a value of $37.5/hr as in US DOT (2003, updated to 2007 value), $\mu = 37.5(\alpha_1 - \alpha_2)/60 = 37.5 \times (12.5\text{-}6.25)/60 = 3.9$ passenger/min. We choose a power function to depict increasing delay growth as traffic volume increases: $L = d[N(f_1 + f_2)/K]^\theta$, where $d$ and $\theta$ are parameters. This functional form also implies that the persistence of some level of delay even when traffic volume is low. We assume there are $N$=60 city-pair markets connected to the focal airport under study. This number of connections roughly corresponds to a medium-sized hub in the US. $d$ and $\theta$ are set to be 10 and 5 respectively. The parameter values are summarized in Table 2.1. In the subsequent analysis, all variables are treated as continuous.[6]

Table 2.1 Derived parameter values.

| Parameter | Value | Unit | Parameter | Value | Unit |
|---|---|---|---|---|---|
| $\alpha_0$ | 1300 | Passengers | $\mu$ | 3.9 | Passenger/min |
| $\alpha_1$ | 12.5 | Passengers/$ | $\eta$ | 0.5 | $/seat-min |
| $\alpha_2$ | 6.25 | Passengers/$ | $n$ | 60 | Markets |
| $\gamma$ | 240 | $\cdot$ flight | $\theta$ | 5 | (-) |
| $\tau$ | 60 | $/seat | $d$ | 10 | Min |
| $c_0$ | 2000 | $ | | | |

---

[5] Certainly, the market demand and flight frequency under equilibrium will be different from the ones used to determine the parameter values. The presumed numbers above are just to derive plausible parameter values for the numerical analysis. Also note that the elasticities are not constant according to the demand function form. Using these parameter values, in the subsequent analysis we find the majority of elasticities calculated under various equilibria are within the range of existing estimates from literature.

[6] One may argue it may not be very realistic. However, this assumption should have little impact on illustrating the qualitative insights. In fact, this type of treatment has been seen in transportation research literature of this type, for example, Schipper et al. (2003) and Brueckner and Girvin (2008).

## 2.4.1 Equilibrium shift when congestion occurs

We first look at the ideal case of infinite capacity and no congestion. All the terms involving $L$ in Equation (2.13) become zero. We find the equilibrium solution with the second-order conditions satisfied. The first line in Table 2.2 reports flight frequency, air fare, passenger generalized cost, demand for each airline, aircraft size, flight operating cost, and the traffic/capacity ratio under this equilibrium.

If some airport capacity constraint exists, the above results will be changed. We set the airport capacity for arriving flights, $K$, to be 720 aircraft per day (if assuming the airport operates 18 hrs per day, then this is equivalent to 40 arrivals/hr).[7] Solving Equation (2.13) yields a new set of equilibrium values (the second line in Table 2.2). Compared to the ideal case, delay results in smaller frequency, higher passenger generalized cost, and reduced demand, confirming our analytical conclusions.

Table 2.2 Comparison of scenarios with and without capacity constraint.

| Scenario | Frequency | Air fare | Generalized cost | Airline Demand | AC size | Unit operating cost | Traffic/ capacity ratio | Average delay |
|---|---|---|---|---|---|---|---|---|
| K=∞ | 7.6 | 98.9 | 130.3 | 485.7 | 63.6 | 91.4 | 0 | 0 |
| K=720 | 5.6 | 96.0 | 143.2 | 405.0 | 71.9 | 91.5 | 0.94 | 7.3 |

The results also indicate a lower air fare, suggesting the effect of passengers' reduced willingness-to-pay due to degraded service quality dominates over the effect of airlines passing part of delay cost to passengers. Larger aircraft will be chosen, suggesting in (2.21) the effect of frequency reduction outweighs the effect of delay on suppressing demand. The use of larger aircraft takes advantage of the economies of aircraft size. Due to the delay cost added, however, the overall flight operating cost is slightly increased.

## 2.4.2 Sensitivity of equilibrium to different capacities

The previous sub-section shows the response of the equilibrium when traffic delay appears, by comparing the extreme case of infinite capacity and a finite capacity. More intriguing is to see how sensitive the equilibrium is to different capacity levels. In what follows we examine the response of relevant supply-demand elements to an equal amount of capacity increase at a range of baseline levels escalating in a 36-operation increment, from 540 to 1260 arrival operations per day. Corresponding welfare gains are also gauged at these different capacity levels.

---

[7] As a reference, we provide here the actual arrival capacity (measured by airport acceptance rate, or AAR, in terms of the number of arrivals per day) as well as the number of connections at four US hub airports in August, 2007: Newark (EWR, AAR: 718, No. connections: 84), Philadelphia (PHL, AAR: 799, No. connections: 50), Denver (DEN, AAR: 1948, No. connections: 106), St Louis (STL, AAR: 1042, No. connections: 47).

*2.4.2.1 Changes in the supply-demand characteristics*

Holding the market potential constant, capacity increase reduces the traffic volume/capacity ratio and delay. Figure 2.3 shows more significant average delay reduction (as measured by the slope of the average delay curve) at lower baseline capacity levels. Delay reduction induces new demand in the market, at a decreasing rate as shown in Figure 2.4. Despite the additional demand and associated new traffic, incremental delay savings—measured as the product of delay savings per flight and the number of flights at the respective baseline capacity level follows a diminishing trend as well.
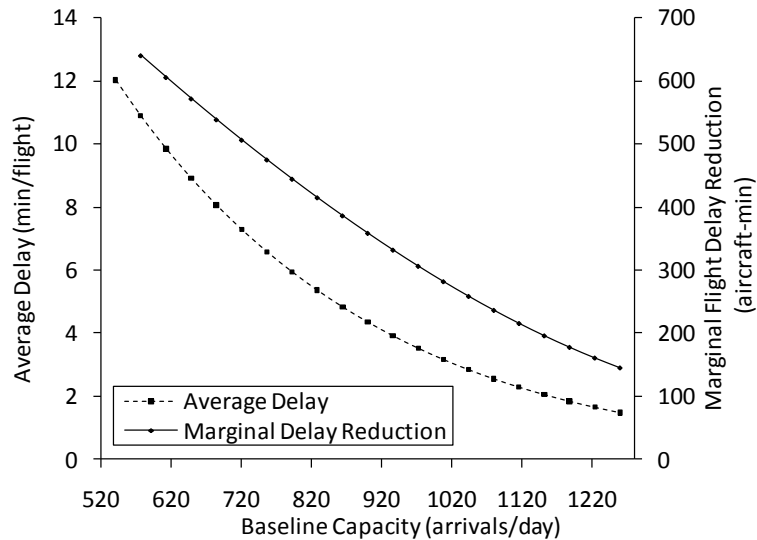


Figure 2.3: Delay and volume/capacity ratio vs. airport capacity

With increasing airport capacity and continuing rise in passenger demand, airlines tend to schedule more flights (Figure 2.4). Frequency seems more sensitive to capacity level than does passenger demand, because airlines also decrease aircraft size as traffic increases. Figure 2.4 shows that, the equilibrium aircraft size continuously decreases. The decrease is moderate in the beginning, due to the concern of incurring higher congestion, as delay remains large in the system. As capacity increases, the impact of flight delay becomes secondary, whereas frequency competition plays a major role. The primary source of aircraft size change now comes from the $1^{st}$ term on the RHS of (2.21). As capacity further increases, the rate of frequency increase slows, presumably because of diminishing returns from schedule delay savings and more limited induced passenger demand. Concomitant with this is a less strong tendency to reduce aircraft size (Figure 2.5).

Capacity augmentation also leads to a lower unit operating cost per seat. When capacity constraint is tight, delay savings contribute more substantially to reducing unit cost than does smaller aircraft size to increasing it. As capacity increases, the cost impact from delay reduction becomes less significant. As this point unit cost increases because the benefits from using smaller aircraft and offering more frequent service so as to attract

more passengers offset the loss of economies of aircraft size. Airlines gain more profits despite some slight increase in unit operating cost (Figure 2.5).
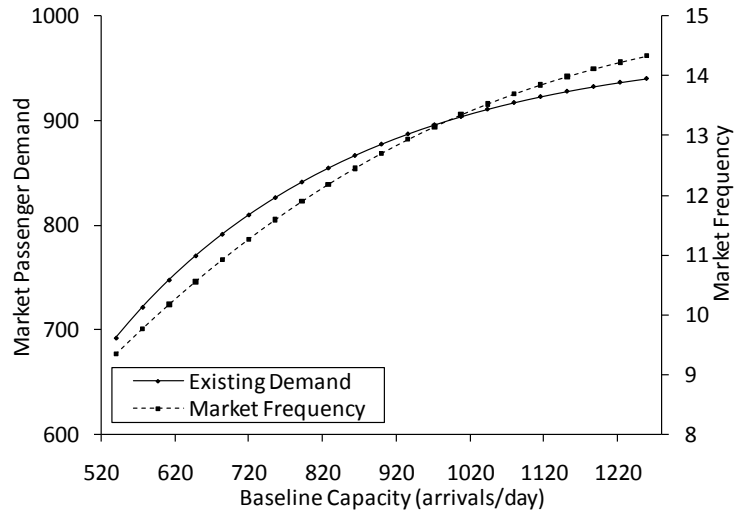


Figure 2.4: Demand and market frequency vs. airport capacity

Figure 2.6 shows that, as capacity increases, airlines raise fares. Since the other two parts (schedule delay and flight delay) continue to decrease, the fare component holds an increasingly important portion in the overall passenger generalized cost. The effect is modest, however, since competition and demand elasticity limit airlines' incentive to increase prices. From the passengers' vantage point, capacity increase enables passengers to enjoy a more substantive reduction in generalized cost. These effects diminish as airport congestion eases.
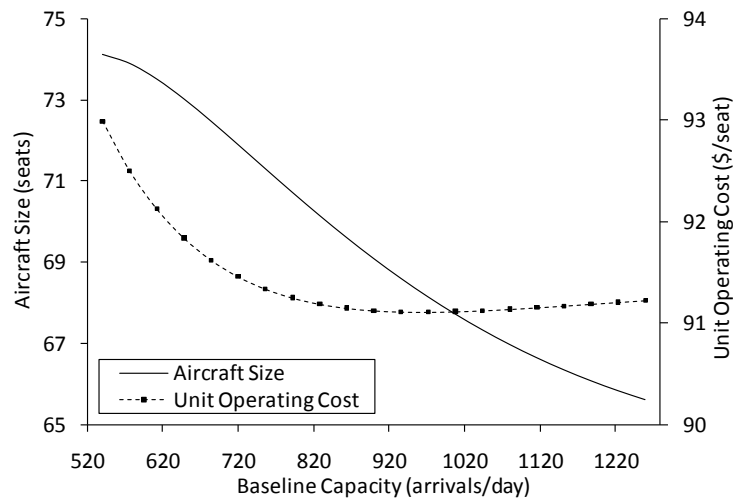


Figure 2.5: Aircraft size and unit operating cost vs. airport capacity
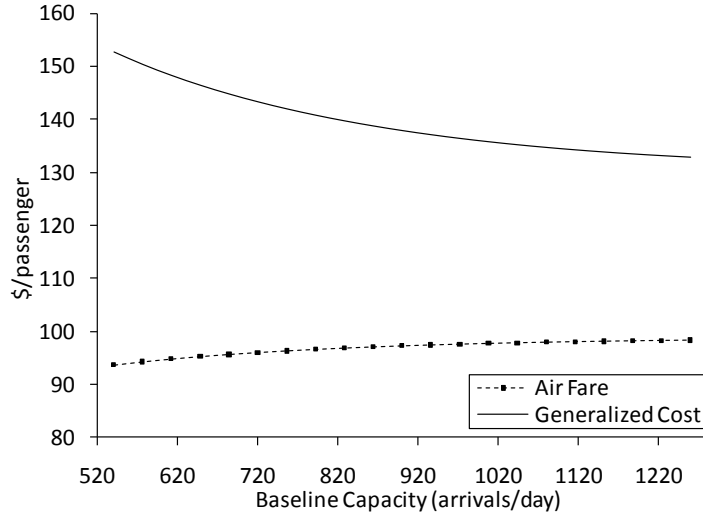
Figure 2.6: Air fare and generalized cost vs. airport capacity

*2.4.2.2   Changes in welfare*
The changes in equilibrium supply-demand characteristics analyzed above imply the importance of baseline capacity to assessing welfare gains. The following experiment makes this explicit. For the range of baseline capacity levels chosen (i.e. from 540 to 1224 daily operations), an investment enhancing capacity by 36 arrival operations per day is made. Following section 2.3.3, at each baseline capacity, we calculate total CS change as $\frac{1}{2}(\overline{P}_1^0 - \overline{P}_1^1)(Q_1^0 + Q_1^1) + \frac{1}{2}(\overline{P}_2^0 - \overline{P}_2^1)(Q_2^0 + Q_2^1)$, where superscripts 0 and 1 denote the states before and after capacity change, and subscripts 1 and 2 indicate airlines. Given the symmetry, the two products are equal; therefore only the calculation of $(\overline{P}_1^0 - \overline{P}_1^1)(Q_1^0 + Q_1^1)$ is needed. By the same token, CS gain for induced demand is obtained as $(\overline{P}_1^0 - \overline{P}_1^1)(Q_1^1 - Q_1^2)$, in which $Q_1^2$ is defined as the hypothetical demand for airline 1 under the old generalized cost of its own and the new generalized cost of airline 2. Note that $(\overline{P}_1^0 - \overline{P}_1^1)(Q_1^1 - Q_1^2)$ equals twice the illustrative area *DBJ* in Figure 2.1. CS gain for existing passengers is then the difference between the total CS change and the CS change for induced demand. On the producer side, PS change is the change in airlines' profit. The estimates for a single market are multiplied by *N* to approximate the aggregate effect across markets. All numbers are on a yearly basis. Figure 2.7 shows the results.

Among the three welfare components, the largest gain comes to CS gains for existing passengers, followed by airlines' profit. For the induced demand, the welfare gain is substantially lower, playing only a secondary role in investment analysis. This is not surprising, since the induced passenger demand only accounts for 0.4 to 4 percent in the total demand for each capacity increment in our analysis. The percentage diminishes as the imbalance between airport capacity and flight demand becomes less severe, which is reflected in the decreasing average delay shown in Figure 2.3. Similar to this and the results obtained in the previous sub-section, we observe decreasing welfare increment in all three components as baseline capacity increases, confirming the conventional wisdom

22

that investment is more beneficial when capacity is more seriously constrained. This gives rise to the question of investment timing. While beyond the scope of the present research, it is important to recognize that investing in capacity will not bring significant benefit–at least immediately–when capacity shortage is not a serious problem. By contrast, although investment at times when there is already severe congestion seems to generate much larger benefit, this must be weighed against the huge delay cost that already occurred due to delayed decisions on expanding capacity.
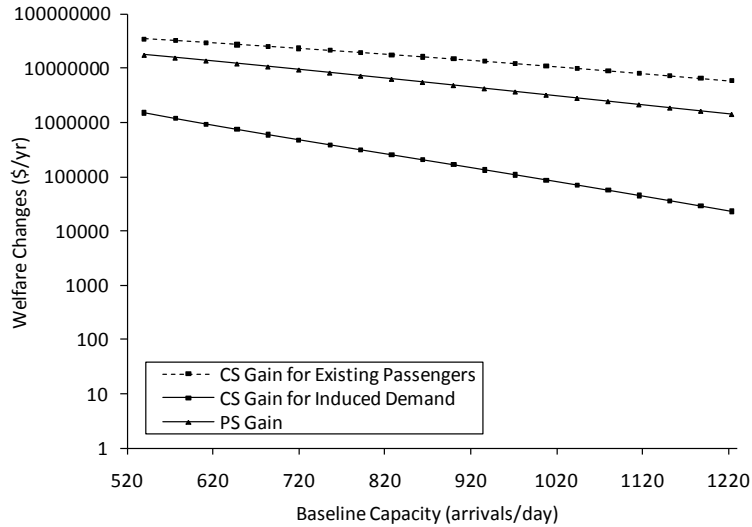


Figure 2.7: Welfare gain under different baseline capacity levels for a fixed capacity increment

### 2.4.3   Benefit assessment using equilibrium and conventional methods

Benefit assessment by incorporating the supply-demand equilibrium would generate different results than the conventional method which is commonly employed in practice. In the conventional method, response from the supply side is usually absent, and in many cases the induced demand part is not considered either. In this sub-section we compare the benefit assessment using the two different methods. Suppose the baseline capacity is 720 operations/day, and some capacity expansion is just completed which increases the capacity by 50%. The evaluation time frame is set to be 10 years, with a 3% discount rate per year. Along the timeline, accounting for the effects of socio-economic development on demand is necessary, and they are primarily embodied in income increase, population growth, and taste variation. Given the quasi-linear utility set-up, income effect is not present in the demand model. Population growth is materialized by simultaneously increasing the values of $\alpha_0, \alpha_1, \alpha_2$ by $100\delta$ percent each year. On top of that, we further allow $\alpha_0$ to increase by another $100\Delta$ percent, to reflect the fact that people increasingly place more importance on air travel. In the following analysis, we set both $\delta$ and $\Delta$ to be 0.01.

We assume in the conventional method, demand is invariant to capacity change. In the starting year, the conventional method calculates delay savings using the same delay

23

function *L* defined above. Under the original capacity, the average delay is 7.29 min/flight; after the capacity increase, the "new" delay becomes only 0.96 min/flight. The difference between the two is multiplied by passengers' value of time[8] and delay-related unit operating cost ($\eta$ = \$0.5/seat-min), and then by passenger demand, to obtain the savings of passenger and carrier delay cost respectively. For the subsequent nine years, the conventional method assumes annual demand increases as result of population growth, which amounts to $\delta$ portion of the previous year's demand, as well as taste variation, whose contribution is $\alpha_0(1+\delta)\Delta$.[9] Since demand increase directly translates into higher frequency, when passenger demand is very large, delay becomes excessively high. In practice, airports experiencing severe delays will not be able to accommodate rising demand for air service. Practical guidance, such as the one issued by FAA (1999), suggests using adjusted traffic levels which reflect a flat or only slightly escalating rate of growth once delay reaches a certain threshold. The FAA guidance states that average delay per operation of 10 minutes or more may be considered severe; at a 20 minutes average delay, growth in operations at the airport will largely cease. In light of this, we cap delay at 20 minutes under the conventional method.

Because of different capacity levels, however, such capping will occur at different times with and without capacity investment. The demand levels will differ starting from the year that demand is capped in the low capacity alternative. FAA (1999) attributes the demand difference to "the availability to airport users of alternative actions to simply waiting in line" (FAA, 1999). Jorge and de Rus (2004) define a similar term of "deviated users", who divert to a substitute in the baseline scenario but switch back when capacity is expanded. Unfortunately, how to cope with this demand inconsistency in benefit analysis is rarely mentioned. In Jorge and de Rus (2004) the delay saving benefits per deviated and existing user are treated as identical. We follow their approach here: to calculate passenger benefits delay saving minutes is multiplied by the number of passengers in the larger capacity case. We use the same approach to calculate airline cost savings. Performing this generates an estimate of annual (present value) benefits for passengers and airlines, which amounts to \$1.52 and \$1.21 billion respectively, or a total of \$2.73 billion over the entire 10-year period.

Using the equilibrium method, benefit assessment requires the calculation of equilibrium values with and without capacity expansion. Following the same procedure as described in section 2.4.2.2, consumer and producer surplus gains are calculated. The present values of gains in PS, CS for existing passengers, CS for induced demand over the 10-year horizon are \$0.68, \$1.52, and \$0.21 billion respectively, with a total at 2.41 billion dollars. Although the overall welfare estimate does not depart substantially from the total

---

[8] We use the same passenger value of time (\$37.5/hr), the one used to determine the parameter $\mu$ before.

[9] Suppose demand for airline 1 in year *k* equals $Q_{1,k} = \alpha_0 - \alpha_1 \overline{P}_{1,k} + \alpha_2 \overline{P}_{2,k}$. According to our treatment of socio-economic impact on parameters, airline 1's demand in the following year becomes $Q_{1,k+1} = \alpha_0(1+\delta)(1+\Delta) - \alpha_1(1+\delta)\overline{P}_{1,k+1} + \alpha_2(1+\delta)\overline{P}_{2,k+1}$. Since response from the supply side is not considered, $\overline{P}_{1,k+1} = \overline{P}_{1,k}, \overline{P}_{2,k+1} = \overline{P}_{2,k}$ and $Q_{1,k+1}$ can be re-expressed as $(1+\delta)Q_{1,k} + \alpha_0(1+\delta)\Delta$, where the second term corresponds to the additional demand resulting from taste variation effect.

benefit using the conventional method, the temporal patterns are very different. As shown in Figure 2.8, the equilibrium approach yields more consistent welfare gains over the timeline. In contrast, when delay capping becomes active, benefits using the conventional method continue to shrink. Therefore, one might expect a total benefit from the conventional method to be even smaller than from the equilibrium approach with a longer planning horizon.

Further interpretation of the results is accompanied by delay savings and changes in demand resulting from the capacity increase, as shown in Figure 2.9. Looking at the first year, delay savings are greater using the conventional method since it disregards passenger and flight frequency adjustment. The equilibrium method predicts more flights because of induced demand. This reduces schedule delay for passengers, and adds to the benefit gain for existing passengers. On the carrier side, although the induced demand allows for additional revenue, the adjustment in fare and flight operating cost produces a total airline profit very similar to the one obtained from the conventional method.

In the successive years, we observe a steady growth of welfare under the equilibrium method, for both airlines and passengers. This results from the growth of market size and the ability of the equilibrium method to internalize passenger and airline adjustment facing delays, which keeps delay at a reasonable level (we observe the average delay at equilibrium is always less than 10 minutes). Failing to incorporate this adjustment, the conventional method provides a distorted delay saving picture. Following a more substantial delay reduction, the welfare gains increases at a much faster rate after the 1st year. The conventional method then avoids excessive delays through a delay cap, which results in reduced delay savings in the latter years. Nevertheless, delays saving estimates remain greater than those from the equilibrium method throughout the 10-year period.
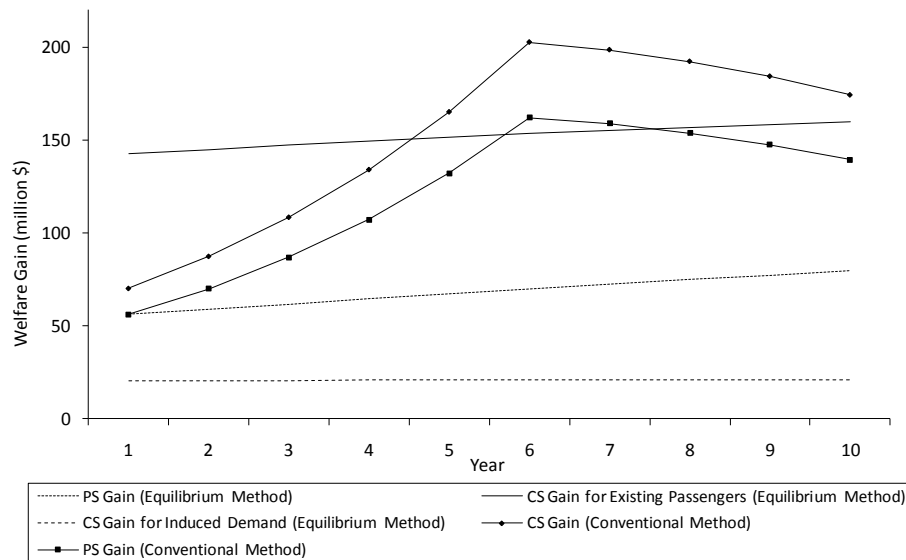


Figure 2.8: Welfare gain using conventional and
equilibrium methods (in present values)
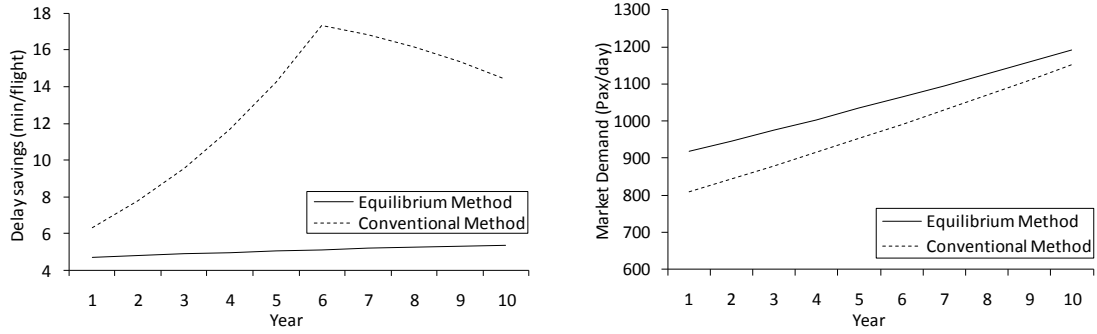
Figure 2.9: Delay savings and demand after capacity increase
using conventional and equilibrium methods

As a final remark, the equilibrium method contributes to a more plausible demand forecast. Compared to the conventional method, the equilibrium predicts a high demand in the beginning due to demand inducement, but a relative slow growth afterwards (Figure 2.8). As illustrated before, the equilibrium permits demand to self-adjust so that exceedingly high delay can be prevented.

## 2.5  Summary

Appropriate assessment methodology for aviation infrastructure investment has become increasingly critical with growing demand and delay in the air transportation system. Recognizing that infrastructure capacity change would trigger a supply-demand equilibrium shift, this chapter proposes a new assessment framework that takes into consideration the interplay among passenger demand, air fare, flight frequency, aircraft size, and flight delay. By developing and analyzing an airline competition model, we find that capacity constraint suppresses demand and increases passenger generalized cost. Facing delays, passengers' willingness-to-pay is reduced; airlines tend to lower frequency and pass part of the delay cost they bear to passengers. In addition to scheduling fewer flights, our numerical analyses further reveal that airlines respond to delay by using larger aircraft and reducing fares. The extent of equilibrium shift depends on how capacity is constrained. The marginal effect of increasing capacity on equilibrium shift and benefit gain diminishes as the imbalance between capacity and demand is mitigated. Through comparing the benefit assessment using the equilibrium and conventional methods, we conclude that the equilibrium method generates more plausible estimates, and prevents the occurrence of unrealistically high delays which often present an issue in the conventional approach.

This chapter presents a first step towards incorporating competitive supply-demand equilibrium into aviation infrastructure investment. There are many opportunities to extend this work. In the model presented here, a simultaneous price-frequency game is assumed. It may be interesting to examine the results under alternative market conditions, such as sequential competition or monopoly. Certainly, empirical investigation of the

findings and benefit assessment simulation using real world data are important next steps, and will be incorporated into our future work.

# 3. Empirical Investigation of Flight Delay Impact on Air Transportation Supply

## 3.1 Introduction

Understanding the supply in the air transportation system has traditionally been an important area in empirical airline economics research. Despite the large body of literature in this area, and growing concerns from airlines and the general public about flight delay and demand for more capacity, the responses of two most critical supply components—airfare and flight frequency—to flight delay remain a less well understood subject. The lack of understanding is in sharp contrast with the record high delay experienced by the air transportation system, especially in the US, as a consequence of ever increasing travel demand. The phenomenon of air traffic congestion and delay will likely become more prominent given the projected demand growth in the coming decades (Boeing, 2011).

Flight delay makes airline operations more expensive, but airlines do not necessarily transfer the delay cost entirely to passengers through higher fare. As shown in the theoretical analysis from the previous chapter, airlines strike a balance between delay cost recovery and maintaining demand despite travelers' decreased willingness-to-pay due to service quality degradation. Decisions on frequency reduction to avoid excessive delays must be weighed against aircraft size economics, higher pricing power, the positive feedback between demand and frequency, and the fear of market share loss. It is of critical importance for policy makers to be cognizant of these interplays, and able to quantitatively gauge airlines' pricing and frequency scheduling responses to delay and delay mitigation strategies.

The objective of this chapter is to enrich the current knowledge base of such responses. We conduct a systematic examination of the delay impact on route airfare and segment flight traffic using data from the US air transportation system. Structural fare and frequency models are estimated, providing empirical evidence that confirms our theoretical conjecture in the previous chapter.

Together with existing insights about the delay impact on passenger demand, which will be presented in Chapter 4, the results from this chapter form a comprehensive basis for future policy analysis and decision making to mitigate delays in the system. We proceed first with a review of the key factors in the determination of airfare and frequency, based on which we the contribution made in this chapter is highlighted. Fare and frequency models are specified in Sections 3.3 and 3.4, respectively, with discussion also covering

data, econometric issues, and estimation results. Summary of key findings and directions for future research conclude the chapter.

## 3.2 Literature review and contribution of the research

### 3.2.1 Fare

The majority of the previous research on airline pricing behavior has been focused on the relationship between average fare and market structures (Gillen and Hazledine, 2011). Specific factors considered include individual airlines' market share, route and endpoint airport concentration (impacts of which are sometimes referred to as "hub premiums"), and low-cost carrier (LCC) competition (e.g. Baily et al., 1985; Borenstein, 1989; Dresner et al., 1996; Morrison, 2001; Hofer et al., 2008; Goolsbee and Syverson, 2008; Chi and Koo, 2009; Brueckner et al., 2011; Zou et al., 2011, to name a few). In addition to market structure, demand and cost characteristics are also considered in structural fare model specifications. Instrumental variable regression and simultaneous equation estimation are the most commonly used techniques to account for the simultaneity between demand and airfare. On the cost side, the straightforward link between fare and distance has been widely acknowledged. Researchers have also paid attention to the existence of the economies of density and its impact on airfare, hitherto at the airline-route level (Brueckner and Spiller, 1994; Berry et al., 1996; Brueckner et al., 2011).

On the other hand, very limited empirical attempts have been made on investigating how delay affects airfare. Theoretically, flight delay causes aircraft to spend more time either on the ground or in the air, increasing fuel consumption and crew time, resulting in additional operating expenses. Anticipating delays, airlines may pad extra time in their published schedules, resulting in less efficient aircraft utilization and therefore higher capital cost. It is natural for airlines to pass these additional costs onto passengers by charging higher fares. Meanwhile, airlines also have to weigh in the demand response to delays, in order to maximize profit. Focusing on short-haul (<500 miles) routes, Britto et al. (2012) find that, controlling for route passenger demand, delay has an upward impact on airfare. Forbes (2008) studies fare response to exogenous delay shock created by the Aviation Investment and Reform Act for the 21st Century (AIR-21), and finds falling airline prices in response to longer flight delays. The airfare reduction found by Forbes should be construed as the net effect of a marginal operating cost increase, combined with reduced demand and passenger willingness-to-pay. This is different from the empirical framework in Britto et al. (2012) which explicitly controls for passenger demand. In addition to the econometric models, there have been recent efforts (e.g. Evans, 2010) that examine price response to delay from the airline gaming perspective.

This chapter takes a comprehensive and novel view to investigate fare determination and in particular the impact of delay on airfare. It is comprehensive because supply, demand, market structure, and delay factors are all considered in the empirical model specification. The novelty lies on the development of separate models for direct and connecting routes. This attempt is, to our knowledge, the first of its kind. Our approach recognizes the

intrinsic differences in unit fare (yield) between the two types of routes (Belobaba, 2009a), and allows for the impact of economies of segment and hub airport density, hub congestion, and route circuity, an important feature of one-stop routes that in effect penalizes airlines for exploiting economies of density, to be explicitly investigated. Our fare models are built upon a more inclusive set of routes which covers a broader range of market types than in previous studies, thus providing a more complete picture of delay impact on airline pricing. The separate consideration of direct and connecting routes provides more detailed insights into the extent to which airlines transfer operating cost increase due to delay to passengers through high fare.

### 3.2.2 Frequency

In airline practice, the decision on frequency and aircraft choice are closely intertwined and entails multiple stages, starting from the more strategic fleet planning process, often performed 2-5 years in advance, to frequency planning beginning a year or more ahead of flight departure, to more tactical fleet assignment process 2-6 months prior to departure date (Belobaba, 2009b). The relationship between frequency and aircraft size has been extensively studied in airline economics literature. Passenger demand growth on a flight segment may lead airlines to adapt by increasing either the aircraft size, or flight frequency, or both.[10] Which option to choose reflects the tradeoffs between schedule delay reduction and economies of aircraft size, as well as shorter run fleet constraints. In principle, airlines may have the incentive to upgauge aircraft because the unit operating cost may become smaller when larger aircraft is utilized (Douglas and Miller, 1974; Morrison and Winston, 1986; Hansen and Kanafani, 1989; Wei and Hansen, 2003). This size economies effect may be offset by pilot wage structures (Meyer and Oster, 1984), which are partially responsible for airlines' choice of smaller aircraft especially on short-haul, high-density markets (Wei and Hansen, 2003).

On the other hand, both historic data and industry outlook suggest that passenger demand growth will be primarily accommodated by more frequent service, with only slight increase in aircraft size (Wei and Hansen, 2003; Givoni and Rietveld, 2009; Boeing, 2011). Higher frequency reduces the average passenger schedule delay, contributing to reduced passenger generalized cost and resulting in higher travel demand, which induces further frequency increase. This positive feedback relationship, known as the Mohring effect, provides one important explanation that airlines prefer to increase flight frequency to accommodate demand growth. In addition, reduced passenger schedule delay leads to an upward shift of travelers' willingness to pay, enabling airlines to charge higher fares, as shown analytically in both monopoly (Brueckner and Zhang, 2001) and duopoly (see the preceding chapter) cases. Further incentive for airlines to increase frequency stems from its association with market power: airlines' market share is superproportional to airlines' frequency share when frequency goes above a certain level, or the S-curve effect (Wei and Hansen, 2005). Schipper et al. (2002) find that the number of airlines on a segment has a significant positive effect on frequency. In effect, the competitive pressure

---

[10] While in principle demand growth can be accommodated by higher load factor, airlines commonly assume some targeted load factors in their planning process. So the load factor is of minor concern in airlines' decision making.

gives airlines little choice but to operate smaller-capacity aircraft with higher frequency and unit operating costs (Belobaba, 2009b).

It is part of our hypothesis, as set out in the research framework, that airlines may choose to cut back the schedule and use larger aircraft to avoid excessively high delays. Under the assumption of constant demand, schedule reduction and aircraft upgauging would be essentially equivalent. The delay effect can be regarded as imposing additional penalty on airlines if using smaller aircraft, since doing so incurs higher delay and airline delay cost. On the other hand, airlines have been observed to progressively pad extra minutes in their published schedules in order to make operations more robust to delays and improve their published on-time performance statistics (Zou and Hansen, 2012). This reduces aircraft utilization, i.e. the number of hours or legs an aircraft can fly in a day. Empirical studies reveal a 0.37 correlation between total airport delay and average aircraft size in Europe (Reynolds-Feighan and Button, 1999) whereas the number of runways at airports has virtually no effect on aircraft size (Givoni and Rietveld; 2009). Using US data at the airline-segment level, Pai (2010) finds that every one minute delay increase at the origin and destination airports result in 2 and 3 fewer flights per month. However, the model does not account for the frequency response to segment passenger demand and airline competition (aside from the presence of LCCs).

In what follows we contribute to the existing literature by specifying and estimating an airlines frequency model that more comprehensively captures the causal relationship between frequency and its influencing factors. The model provides quantitative insights into airlines' choice between frequency and aircraft size facing passenger demand change, examines various competition effects on frequency at the segment, airport, and metropolitan area pair levels, and offers up-to-date evidence about the impact of congestion on flight frequency—through the inclusion of both airport delays and other indirect effects.

## 3.3   Effects of delay on fare

In this section we specify and estimate non- and one-stop route fare models. Air travelers making two stops are only a small fraction of system total in the U.S. (1.1-1.4% between 2004 and 2008), and are therefore not considered in the fare modeling.  Each fare model is a function of cost characteristics, demand, competition, and flight delay on the route. For one-stop routes, the impact of segment passenger density and connecting airport characteristics are further incorporated in the fare model.

### 3.3.1   Empirical model specification

For a given route, we employ the average yield, i.e. the total revenue generated from the route across all airlines divided by the total passengers-miles flown on the route, as the dependent variable in the fare models. Yield is modeled as a function of a set of route-, segment-, market-, and airport-specific variables. The models have log-linear forms, i.e. all continuous variables take logarithmic values. This specification is to approximate the

non-linear relationship between the yield and its explanatory variables, and helps reduce the impact of outlying observations and heteroskedasticity. The resulting coefficients represent yield elasticities with respect to the continuous explanatory variables. As the model is estimated using data aggregated across carriers, the coefficients may be viewed as an average of the underlying, carrier-specific coefficients.

The two econometric models to explain yields are specified as follows:

Non-stop routes:

$$
\begin{aligned}
\ln Y_{it} = {} & \alpha_0 + \alpha_1 \ln(RoutePax)_{it} + \alpha_2 \ln(Dist)_{it} + \alpha_3 \ln(OriginL4Delay)_{it} \\
& + \alpha_4 \ln(DestL4Delay)_{it} + \alpha_5 \ln(RouteHHI)_{it} + \alpha_6 \ln(MarketHHI)_{it} \\
& + \alpha_7 \ln(OriginHHI)_{it} + \alpha_8 \ln(DestHHI)_{it} + \alpha_9 RouteLCC_{it} + \alpha_{10} AdjacentRouteLCC_{it} \quad (3.1) \\
& + \alpha_{11} SlotControl_{it} + \alpha_{12} Vacation_{it} + \varepsilon_{it}
\end{aligned}
$$

One-stop routes:

$$
\begin{aligned}
\ln Y_{jt} = {} & \beta_0 + \beta_1 \ln(RoutePax)_{jt} + \beta_2 \ln(Dist)_{jt} + \beta_3 \ln(SumDensity)_{jt} + \beta_4 \ln(Circuity)_{jt} \\
& + \beta_5 \ln(OriginL4Delay)_{jt} + \beta_6 \ln(DestL4Delay)_{jt} + \beta_7 \ln(ConnectL4Delay)_{jt} \\
& + \beta_8 \ln(LegMeanHHI)_{jt} + \beta_9 \ln(MarketHHI)_{jt} \\
& + \beta_{10} \ln(OriginHHI)_{jt} + \beta_{11} \ln(DestHHI)_{jt} + \beta_{12} \ln(ConnectHHI)_{jt} \quad (3.2) \\
& + \beta_{13} AirportPairLCC_{jt} + \beta_{14} AdjacentRouteLCC_{jt} \\
& + \beta_{15} SlotControl_{jt} + \beta_{16} Vacation_{jt} + \upsilon_{jt}
\end{aligned}
$$

where subscript i, j and t denote a non-stop route, a one-stop route, and a time period, respectively. $\varepsilon_{it}$ and $\upsilon_{it}$ are error terms. Table 3.1 below provides a description of all variables in the two models.

Table 3.1 Definition of fare model variables.

|  | Definition |
|---|---|
| Common variables | |
| Y | Average yield, calculated as total revenue (dollars) on a given route divided by total passenger miles flown a given route and time period. |
| RoutePax | Total number of passengers on the route per quarter. |
| Dist | Airport-to-airport non-stop distance in statute miles. |
| OriginL4Delay | Average flight delay at the origin airport (four quarters lagged). |
| DestL4Delay | Average flight delay at the destination airport (four quarters lagged). |
| MarketHHI | Market-level Herfindahl-Hirschman Index (HHI), measured as the sum of squared market shares of all carriers flying in the market. A market encompasses all routes that connect the corresponding metropolitan area pair. |

| | |
|---|---|
| OriginHHI | HHI at the origin airport. |
| DestHHI | HHI at the destination airport. |
| AirportPairLCC | Dummy variable, equal to 1 if at least one low cost carrier (LCC) is present on the origin-destination airport pair. |
| AdjacentRouteLCC | Dummy variable, equal to 1 if LCC's are present on at least one of the adjacent route. An adjacent route is defined as one of which both origin and destination airports are within the same metropolitan area as the origin and destination airports, respectively. |
| SlotControl | Dummy variable, equal to 1 if at least one of the origin and destination airports is slot controlled. |
| Vacation | Dummy variable, equal to 1 if at least one of the origin and destination airports is in the States of Florida, Nevada, and Hawaii. |

| Variables only in the fare model for non-stop routes | |
|---|---|
| RouteHHI | HHI on the route. |

| Variables only in the fare model for one-stop routes | |
|---|---|
| Circuity | Ratio of total itinerary miles over O-D miles |
| SumDensity | Sum of the number of passengers on the two segments that were used by the route under study |
| ConnectL4Delay | Average flight delay at the connecting airport (four quarters lagged). |
| LegMeanHHI | Geometric mean of the two flight segments' HHI's |
| ConnectHHI | HHI at the connecting airport |

As pointed out in the outset of this sub-section, the explanatory variables encompass cost, demand, competition, and flight delay effects. Airport O-D distance (Dist) represents the major cost-side effect on fare. Even though yield is expressed in terms of revenue per passenger-mile, it remains a function of the airport O-D distance because ending point operations such as takeoffs and landings have significant cost not related to distance (Hurdle et al., 1989). Therefore economies of distance may exist (Wei and Hansen, 2003). For connecting routes, the actual traveling distance is different from O-D distance, and involves an additional takeoff and landing. The Circuity variable, which measures the ratio between the route distance and O-D distance, is introduced. Higher Circuity raises production costs and lower product quality. The former effect would tend to increase price, while the latter would tend to lower it (Borenstein, 1989).

The existence of economies of density suggests that operating cost on a route may depend upon route demand density (RoutePax) (Graham et al., 1983; Baily et al., 1985; Hurdle et al., 1989; Dresner and Trethway, 1992; Dresner et al., 1996). Meanwhile, RoutePax would affect airfare through the demand side effect (Dresner and Trethway, 1992). On non-stop routes, the economies of density effect may be better captured by using segment passenger volume. However, we find very high correlation between RoutePax and segment passenger volume in the dataset (correlation coefficient: 0.88). Only RoutePax is included to avoid multi-collinearity. In contrast, RoutePax on connecting routes is often much smaller than segment passenger volume, because the two flight segments are used by many other connecting routes and two local, non-stop O-D routes. RoutePax on one-stop routes would no longer be a valid proxy for the density effect. Segment passenger

33

density is therefore included in the one-stop fare model. We follow Brueckner and Spiller (1994) by specifying the sum of segment passenger volumes (SumDensity) to capture the density effect. The RoutePax variable, then, would only convey the demand side effect on airfare.

Competition manifests itself in many ways in fare determination. We first use the Herfindahl-Hirschman Index (HHI), the sum of squared shares from all incumbent airlines, to characterize the concentration. In general, lower HHI's means more intense competition and therefore lower airfare. Route-, market-, and airport-specific HHI's (RouteHHI, MarketHHI, OriginHHI, DestHHI) are included in the non-stop fare model. A market consists of all routes connecting the corresponding metropolitan area pair.[11] Multiple non-stop routes may exist when Multiple Airport Systems (MAS) present at either origin or destination metropolitan area or both. A connecting route involves transferring at an intermediate airport which is often an airline's hub, resulting in very high route concentration (RouteHHI close to 1). It is therefore less sensible to incorporate RouteHHI in the one-stop fare model. However, each of the two flight segments could involve substantial competition. As an example, while American Airlines may be the only carrier operating on SFO-DFW-ATL route, on DFW-ATL segment both American and Delta would schedule many flights, since the two ending airports are their respective hubs. We include the geometric mean of segment HHI's (LegMeanHHI) in the one-stop route fare model. While higher concentration at the origin and destination airports would allow the dominant carriers to charge higher fare,[12] the effect at the connecting airport may be different. The market effect may be dominant as an airline intends to attract passengers to use the circuitous routes and connect at the hub. In addition, consolidation of operations at the hub reduces unit operating cost, and therefore airfare.

Besides different HHI's, LCC's present another dimension in the competition effect (Dresner et al., 1996; Morrison, 2001; Brueckner et al., 2011). Here we introduce two dummy variables, AirportPairLCC and AdjacentRouteLCC, to capture the competition effect on the same airport pair and on adjacent routes. The latter case occurs when at least one ending airport of a route is within an MAS. An adjacent route is defined as a route whose ending airport is either the same one as that of the route under study, or in the same metropolitan region.

Since airlines cannot predict flight delay prior to departure, a key hypothesis we make in the study is that, based upon the delay they experienced in the past, airlines tend to pass part of the delay cost onto passengers through higher fare. As the time unit for fare observations is one quarter, the most relevant delay experience would be from the previous quarter as well as one year before due to the same season. While it is tempting

---

[11] The coverage of each metropolitan area in the present study is adopted from the definitions of metropolitan statistical areas (MSA), micropolitan statistical areas, combined statistical areas (CSA), and metropolitan divisions in the Bureau of Economic Analysis (2011).

[12] The fringe competitors might either increase airfare because of the "umbrella" effect or be forced to charge lower fare in order to gain foothold at the airport. As the measured fare is weighted averaged of fares from dominant airlines and these fringe competitors, the average fare would likely be higher at airports with higher concentration.

to include both one- and four-quarter lagged airport departure and arrival delays in the fare model, we observe high correlation among these four delay variables. Again, to avoid multi-collinearity, only four-quarter lagged airport arrival delays are included in the fare models.

Finally, two dummy variables, SlotControl and Vacation, are introduced to capture the effects on fare of a route being slot controlled or connected to a vacation destination. SlotControl equals 1 if at least one ending airport is slot controlled. The coefficient for the SlotControl variable should be interpreted as the effect of this policy after controlling for flight delays. Slot controlled airports increases the resource scarcity, which may effectively ration air travel, creating the potential for price increase on relevant routes (Swaroop et al., 2012). The Vacation variable is expected to account for the lower ratio of business and leisure travelers. One would therefore expect a negative impact on yield (Dresner et al., 1996).

### 3.3.2  Data

The data for fare model estimation are based on airport O-D pairs among top 100 U.S. airports (based on passenger enplanement in the 4th quarter of 2004), for all quarters from 2004 to 2008. These years represent periods where travel demand ratcheted up from the 9/11 terrorism attacks, reached its record high in 2007, and slumped in the following year because of hiking oil prices, therefore providing sufficient variability in passenger demand and flight delay. The routes included in the dataset also cover a wide range of route characteristics such as traffic volume, distance, the extent of delay and competition.

The dataset for model estimation is compiled from several sources. Passenger demand, fare, and HHIs at route, segment, market, and airport levels are either directly obtained, or constructed using the U.S. Bureau of Transportation Statistics (BTS) Airline Origin and Destination Survey (DB1B), a 10% sample of all U.S. domestic air travel tickets. Yield is converted into constant dollars based on the 2nd quarter of 2004. Information about segment passenger density and distance between airports is extracted from BTS T100 Domestic Segment Traffic Database. The average airport delay is calculated using the BTS Airline On-Time Performance Database, which details flight activities for each scheduled domestic flight by major U.S. carriers. AirTran Airways, American West Airlines, ATA Airlines, Frontier Airlines, JetBlue Airways, Southwest Airlines, Spirit Airlines, and Sun Country Airlines are considered as low cost carriers.[13] Between 2004 and 2006, DCA, LGA, JFK, and EWR were slot controlled airports under the High Density Rule (HDR). Information about airport slot control after the expiration of HDR, starting from January 1, 2007, is collected from FAA (2008, 2009a). The coverage of metropolitan areas is adopted from the definitions of metropolitan statistical areas, micropolitan statistical areas, combined statistical areas, and metropolitan divisions in the Bureau of Economic Analysis Regional Economic Accounts (BEA, 2011). We also extract population and income per capita information for each metropolitan area from the same source. The definition of MAS follows those in Hansen and Weidner (1995), and

---

[13] American West Airlines merged with US Airways in January 2006. ATA Airlines ceased its service starting from April 2008.

Hsiao (2008). The correspondence between MAS and metropolitan areas included in our dataset is presented in Appendix B.

To ensure reliable data for model estimation, the dataset is filtered by applying several rules. We preserve routes with average yield greater than or equal to three cents per mile. Segments with flight frequency less than 60 flights per quarter are excluded from the dataset. Also removed are very short-haul routes for which both origin and destination airports are within the same metropolitan area. For one-stop travel, we restrict connecting itineraries to those that through one of 30 major airports.[14] In total, the dataset contains 67,443 non-stop route-quarter and 1,206,282 connecting route-quarter observations from 4,035 non-stop and 90,242 connecting routes. Descriptive statistics of the variables are provided in Tables 3.2 and 3.3.

Table 3.2 Descriptive statistics for variables in the non-stop fare model (N=67,443).

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Y ($/passenger-mile) | 0.255 | 0.236 | 0.040 | 2.795 |
| Dist (miles) | 873.124 | 627.956 | 55 | 4962 |
| RoutePax (passengers) | 20321.530 | 26247.460 | 10 | 272710 |
| RouteHHI | 0.780 | 0.227 | 0.166 | 1 |
| MarketHHI | 0.466 | 0.192 | 0.133 | 1 |
| OriginHHI | 0.259 | 0.153 | 0.067 | 0.942 |
| DestHHI | 0.259 | 0.153 | 0.067 | 0.942 |
| AirportPairLCC | 0.661 | 0.473 | 0 | 1 |
| AdjacentRouteLCC | 0.119 | 0.323 | 0 | 1 |
| OriginL4Delay (min/flight) | 12.640 | 3.870 | 4.584 | 31.020 |
| DestL4Delay (min/flight) | 12.641 | 3.870 | 4.584 | 31.020 |
| SlotControl | 0.094 | 0.291 | 0 | 1 |
| Vacation | 0.221 | 0.415 | 0 | 1 |

Table 3.3 Descriptive statistics for variables in the one-stop fare model (N=1,206,282).

---

[14] The airports are: Atlanta Hartsfield International Airport (ATL), Logan International Airport (BOS), Baltimore–Washington International Airport (BWI), Charlotte Douglas International Airport (CLT), Cincinnati–Northern Kentucky International Airport (CVG), Ronald Reagan National Airport (DCA), Denver International Airport (DEN), Dallas–Ft. Worth International Airport (DFW), Detroit Metro-Airport (DTW), Newark International Airport (EWR), Washington Dulles International Airport (IAD), George Bush Intercontinental Airport (IAH), John F. Kennedy International Airport (JFK), McCarran International Airport (LAS), Los Angeles International Airport (LAX), LaGuardia Airport (LGA), Orlando International Airport (MCO), Memphis International Airport (MEM), Miami International Airport (MIA), Minneapolis–St. Paul International Airport (MSP), Chicago O'Hare International Airport (ORD), Philadelphia International Airport (PHL), Phoenix International Airport (PHX), Pittsburgh International Airport (PIT), San Diego International Airport (SAN), Seattle–Tacoma International Airport (SEA), San Francisco International Airport (SFO), Salt Lake City International Airport (SLC), Lambert–St. Louis International Airport (STL), and Tampa International Airport (TPA). The inclusion of hub airports is consistent with Hsiao (2008) and the set of airports in FAA 2001 airport capacity benchmarking (the Honolulu International Airport is removed from the list as only trips in the continental US is considered).

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Y ($/passenger/mile) | 0.130 | 0.105 | 0.035 | 4.022 |
| Dist (miles) | 1354.045 | 771.292 | 45 | 5107 |
| RoutePax (passengers) | 340.193 | 532.186 | 10 | 10400 |
| Circuity | 1.304 | 0.348 | 0.996 | 24.438 |
| SumDensity (passengers) | 1232079 | 887438 | 30930 | 7041750 |
| MarketHHI | 0.354 | 0.152 | 0.124 | 1 |
| OriginHHI | 0.219 | 0.131 | 0.067 | 0.942 |
| DestHHI | 0.219 | 0.131 | 0.067 | 0.942 |
| ConnectHHI | 0.337 | 0.137 | 0.093 | 0.634 |
| LegMeanHHI | 0.644 | 0.195 | 0.193 | 1 |
| AirportPairLCC | 0.686 | 0.464 | 0 | 1 |
| AdjacentRouteLCC | 0.071 | 0.257 | 0 | 1 |
| OriginL4Delay (min/flight) | 12.649 | 3.648 | 4.584 | 31.020 |
| DestL4Delay (min/flight) | 12.647 | 3.641 | 4.584 | 31.020 |
| ConnectL4Delay (min/flight) | 13.015 | 4.463 | 4.584 | 31.020 |
| SlotControl | 0.097 | 0.296 | 0 | 1 |
| Vacation | 0.253 | 0.435 | 0 | 1 |

### 3.3.3   Estimation, results and discussion

In estimating the fare models specified above, accounting for potential endogeneity is important. An important source of endogeneity arises from the RoutePax variable. The models are estimated using the two-stage least square (2SLS) method to correct for the potential endogeneity bias from using the Ordinary Least Square (OLS) estimation. For non-stop route fare model, we introduce the income and population in the origin and destination metropolitan areas, the number of connected airports and the ratio of connecting to local O-D passengers at the origin and destination airports (based on outbound traffic), all taken logarithmic values, as additional instruments. The socio-economic instruments are clearly exogenous and related to demand generation on the non-stop route. The number of segment connections at the ending airports is often highly correlated with the size and strength of economic activities of the local metropolitan area, therefore affecting air travel passenger demand. The ratio between connecting and local passengers at airports reflects the extent of hubbing at the airports, which affects non-stop demand through competition between connecting and local passengers for aircraft seating capacity provided on the non-stop route (segment). On the other hand, because any one route accounts for only a small share of the total traffic at an airport, those airport-specific instruments are largely exogenous to fare changes on any given route and will be determined by the incumbent carriers' entire routing structure.

The choice of instruments in the one-stop fare model follows a similar rationale. Instruments excluded from the structure equation for the RoutePax variable consist of the income and population at the origin and destination metropolitan areas, the maximum and

minimum flight frequency on the two segments, the number of segments connected at the intermediate airport, and non-stop O-D passenger volumes on the two segments, all in logarithmic forms. The income and population variables are the natural choices for instruments for O-D demand. Flight frequencies on the two segments affect one-stop passenger demand, but the frequencies are segment characteristics and one such segment serves many routes. As a consequence, segment flight frequencies are largely exogenous to fare change on one connecting route. In much the same way, the number of segments connected at the intermediate (hub) airport, which captures the connection attractiveness of the airport, is determined by airlines' routing structure. Finally, the local O-D passengers affect one-stop route demand because of the competition between the two types of travelers for the seating capacity provided on the shared segment. Following the similar argument that one segment serves many connecting routes, local, direct O-D demand can be reasonably assumed to be uncorrelated with the random shocks of airfare on the connecting route.[15]

One might argue that the RouteHHI variable in the non-stop fare model may also be endogenous, since a carrier's share of passengers on a route could be a function of the price it charges. On the other hand, as suggested by Brueckner and Spiller (1994) and Bamberger and Carlton (2003), airlines' decisions on entry and exist are usually a network-wide decision rather than a decision based on characteristics of the individual markets. Baily et al. (1985) argue that, while technology and demand are the key determinants of market structure, if average costs are flat over a wide range of outputs, there may be a wide range of viable scales of operation and hence a wide range of possible industry structures. The observed structure may thus reflect a history of random shocks that determine the relative sizes of the existing firms. Following these considerations we still treat RouteHHI as exogenous in the present study. We also note that no consensus has been achieved as to whether concentration variables should be treated as endogenous. In addition, finding proper instruments for RouteHHI is not straightforward, and could even lead to counter-intuitive coefficient estimates (Baily et al., 1985).

Tables 3.4 and 3.5 present the estimation results for the direct and connecting route fare models respectively. Results from OLS estimation are also reported to serve as a point of reference. In both models, observations are clustered by market, in order to account for the dependence of unobservables among routes within a market. Consequently, the standard errors are robust to heteroskedasticity, serial correlation, and market clustering effect.[16]

---

[15] It is possible that one-stop route demand could inversely affect local, O-D demand; then local O-D demand may itself be endogenous and depend on the one-stop fare. However, as the segment serves many connecting routes, we believe that the effect of the demand on one single connecting route on the local, O-D demand is rather marginal. Any potential endogeneity, therefore, would not be significant.

[16] As the non-stop and connecting route fares are estimated separately, each cluster in the non-stop fare model only contains non-stop routes servicing the corresponding market. The same for clusters in the connecting fare model. While clustering both non- and one-stop routes would further improve the estimation efficiency, this would require simultaneous estimation approach which is beyond the scope of the current models.

Table 3.4 Estimation results for the non-stop route fare model.

| Variable | OLS | | 2SLS | |
|---|---|---|---|---|
| | Est. | Std. Err. | Est. | Std. Err. |
| RoutePax | -0.0827*** | 0.0043 | -0.0218** | 0.0096 |
| Dist | -0.6535*** | 0.0068 | -0.6762*** | 0.0083 |
| OriginL4Delay | 0.0459*** | 0.0113 | 0.0459*** | 0.0116 |
| DestL4Delay | 0.0516*** | 0.0112 | 0.0505*** | 0.0114 |
| RouteHHI | -0.0590*** | 0.0114 | 0.0379* | 0.0195 |
| MarketHHI | 0.0701*** | 0.0124 | 0.0381*** | 0.0144 |
| OriginHHI | 0.0248*** | 0.0067 | 0.0118* | 0.0071 |
| DestHHI | 0.0280*** | 0.0068 | 0.0147** | 0.0072 |
| AirportPairLCC | -0.2883*** | 0.0111 | -0.3676*** | 0.0159 |
| AjacentRouteLCC | -0.1468*** | 0.0129 | -0.2138*** | 0.0161 |
| SlotControl | 0.1042*** | 0.0133 | 0.0685*** | 0.0139 |
| Vacation | -0.1582*** | 0.0076 | -0.1813*** | 0.0089 |
| Constant | 3.5055*** | 0.0643 | 3.1225*** | 0.0760 |
| Number of Observations | 67,443 | | 67,443 | |
| $R^2$ | 0.8579 | | 0.8505 | |
| First stage partial F-stat | | | 78.1848 | |
| Partial $R^2$ | | | 0.2474 | |

*** p<0.01, ** p<0.05, * p<0.10

Table 3.5 Estimation results for the one-stop route fare model.

| Variable | OLS | | 2SLS | |
|---|---|---|---|---|
| | Est. | Std. Err. | Est. | Std. Err. |
| RoutePax | -0.0781*** | 0.0011 | 0.0446*** | 0.0036 |
| SumDensity | 0.0874*** | 0.0023 | -0.0140*** | 0.0037 |
| Circuity | -0.7558*** | 0.0092 | -0.3171*** | 0.0165 |
| Dist | -0.6889*** | 0.0047 | -0.6435*** | 0.0055 |
| OriginL4Delay | -0.0008 | 0.0065 | 0.0174** | 0.0071 |
| DestL4Delay | 0.0042 | 0.0063 | 0.0198*** | 0.0069 |
| ConnectL4Delay | 0.0014 | 0.0037 | 0.0248*** | 0.0040 |
| LegMeanHHI | 0.0662*** | 0.0039 | 0.0908*** | 0.0044 |
| MarketHHI | 0.0499*** | 0.0048 | 0.0638*** | 0.0053 |
| OriginHHI | 0.0020 | 0.0032 | 0.0355*** | 0.0035 |
| DestHHI | 0.0074** | 0.0032 | 0.0410*** | 0.0036 |
| ConnectHHI | -0.0464*** | 0.0029 | -0.1152*** | 0.0042 |
| AirportPairLCC | -0.2234*** | 0.0040 | -0.2565*** | 0.0045 |
| AdjacentRouteLCC | -0.1078*** | 0.0065 | -0.0942*** | 0.0068 |
| SlotControl | 0.0203*** | 0.0043 | 0.0397*** | 0.0044 |
| Vacation | -0.0829*** | 0.0045 | -0.1036*** | 0.0048 |
| Constant | 2.2144*** | 0.0442 | 2.5176*** | 0.0461 |
| Number of Observations | 1,206,282 | | 1,206,282 | |
| $R^2$ | 0.5593 | | 0.5135 | |
| First stage partial F-stat | | | 743.888 | |
| Partial $R^2$ | | | 0.0754 | |

*** p<0.01, ** p<0.05, * p<0.10

Overall, the coefficients obtained from 2SLS have the expected signs and most of them are significant. In contrast, some counter-intuitive results would appear if OLS is employed. The non-stop fare model has a better goodness of fit than the one-stop model. The fairly large values for the first stage F-statistics of the added instruments and partial $R^2$, the latter of which measures the relevance of the instruments to the endogenous variable purged of the effect from exogenous variables included in the fare models, suggest the chosen instruments are sufficiently strong.

The coefficient for RoutePax in the non-stop model has a negative sign, indicating that the cost effect due to economies of density dominates the demand side effect. However, the estimated elasticity of -0.0218 is relatively small: all else being equal, fare on a non-stop route with 25 passengers per day is about 6 percent higher than another non-stop route with 500 passengers per day. On the one-stop route, the cost side effect is stripped off from RoutePax by the SumDensity variable. The positive coefficient for RoutePax, which represent only the demand side effect, is expected. *Ceteris paribus*, a one-stop route with 100 passengers per day would have airfare that is 11% higher than a thinner route that only has 10 daily passengers. The negative coefficient of the SumDensity variable clearly suggests the existence of economies of density. While the estimate is about one fourth that for RoutePax, SumDensity often has a much greater value than RoutePax since each segment consolidates passengers on many routes, including the local, non-stop route which often transports more passengers than any single connecting route. As a consequence, one may expect the positive effect of RoutePax to be neutralized to a great extent by that from SumDensity.

We observe that yield falls with O-D distance, with close coefficients in the two models. The Dist coefficients imply that 10% O-D distance increase would reduce yield by 6.762% and 6.435%, leading to 2.6% and 2.9% ticket price increase. These estimates are comparable to the yield elasticity with respect to distance in the early deregulation era (around -0.5, as in Baily et al., 1985). Controlling for O-D distance, if one connecting route is 20% more circuitous than another, fare on the first route would increase by 12.4%, suggesting the associated cost increase effect is stronger than that of reduced product quality.

The coefficients for the HHI's at route, segment, market, and airport levels support the view that higher concentration enable airlines to charge higher fare on average. This is consistent with the negative relationship between yields and competition as found in the large body of airline literature. For non-stop travel, the results implies that the fare with four equal-sized competitors (i.e. RouteHHI=0.25) on a route would be 2.8% lower than on a monopoly route. Higher concentration at the airports also increases fare, but the effect is smaller than at the route level. Airfare on one-stop routes seems to be more sensitive to variations in concentration. The price elasticity with respect to market HHI is 0.0638 as compared to 0.0381 on direct routes. The most significant effect of HHI on fare comes from LegMeanHHI, which, analogous to the RouteHHI variable in the non-stop model, captures the route level competition effect. Higher HHI's at the origin and destination airports contribute to higher fare, but the effects are smaller than at the market and route levels. A strong, negative coefficient is associated with ConnectHHI,

confirming our conjecture that airlines deliberately lower one-stop fares to attract connecting traffic at highly concentrated hubs in order to reduce unit operating cost. Reduced unit operating cost then allows airlines to charge lower fare. The presence of LCC's imposes drastic downward pressure on fares. The average fare on a non-stop route will be reduced by 37% and 21% if an LCC is present on the same and adjacent routes respectively. This is very close to recent estimates of 34% and 19% in Brueckner et al. (2011). On one-stop routes, the effects of LLC are smaller in magnitude but are still significant. This could be attributed to the fact that the two segments in the one-stop route serve many other markets; legacy carriers as a consequence are less concerned about passengers being attracted by their low cost rivalries. On-route LCC lowers fare by 25.7% whereas adjacent competition from LCC results in a fare reduction of only 9.4%.

Routes involving slot controlled airports have airfares that are 6.9% and 4.0% higher, respectively, on non- and one-stop routes. This reflects the scarcity value of airport slots. Finally, because there are more leisure travelers on vacation routes, airfares on such routes are 18.1% lower on non-stop routes, and 10.4% lower on one-stop routes, than on comparable non-vacation routes.

Turning now to the airport delay coefficients, which are the focus of the present study, we find that all coefficients are positive and significant, confirming our hypothesis that airlines pass part of their delay cost onto travelers through higher fare. The delay elasticities are small, and differ on the two types of routes. If average flight arrival delay at their origin or destination airport involved in a direct route is increased from 5 min to 20 min, then fare would increase by 6.6% and 7.3% respectively. In contrast, one would observe only about 2-3.5% fare increase for the same delay increase at the origin, destination, or connecting airport on a connecting route. The difference may be explained by several reasons. First, it is likely that non-stop routes serve passengers who have higher time values and are less sensitive to price change. Airlines thus feel more comfortable to transfer delay cost to passengers on such routes. In addition, non-stop passengers travel more than connecting passengers during peak, congested times, when airlines are also prone to charging higher price premium. As a consequence, non-stop passengers are likely to see greater fare increase than one-stop passengers for a given amount of delay increase. Airlines' lower willingness to pass delay costs onto connecting passengers may be further associated with the cost-side effect on the connecting routes. In order to be economically competitive against non-stop routes, airlines are more concerned about maintaining high load factors and taking advantage of economies of traffic density on the connecting routes. High load factors due to passenger consolidation would lead to low delay cost borne by each individual passenger.

Figures 3.1 and 3.2 plot the predicted yield as a function of destination airport delay observed in the datasets, at 25th, 50th, and 75th percentile O-D distance values, for direct and connecting routes. All other explanatory variables take their sample means. An increase in delay from the lower to the upper extreme (4.58 min/flight vs. 31.02 min/flight), would cause fare to increase by $15 for a long-haul flight and $6 for a short-haul, contributing 9.2% and 5.8%, respectively, to the total fare at the highest delay level.

On average, one minute delay change would only be associated with $0.40-0.55 fare variation on non-stop routes, and $0.16-0.21 on connecting routes.
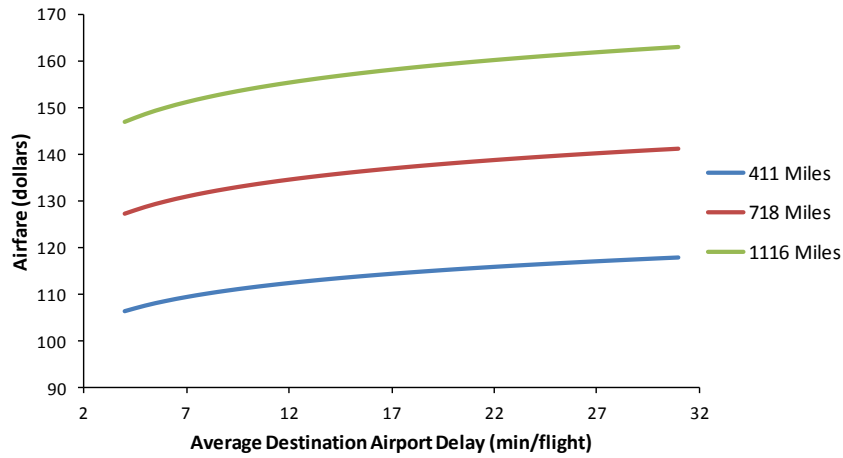
Figure 3.1: Non-stop airfare as a function of delay at the destination airport at 25th, 50th, and 75th percentiles of route O-D distance (411, 718, 1116 miles)
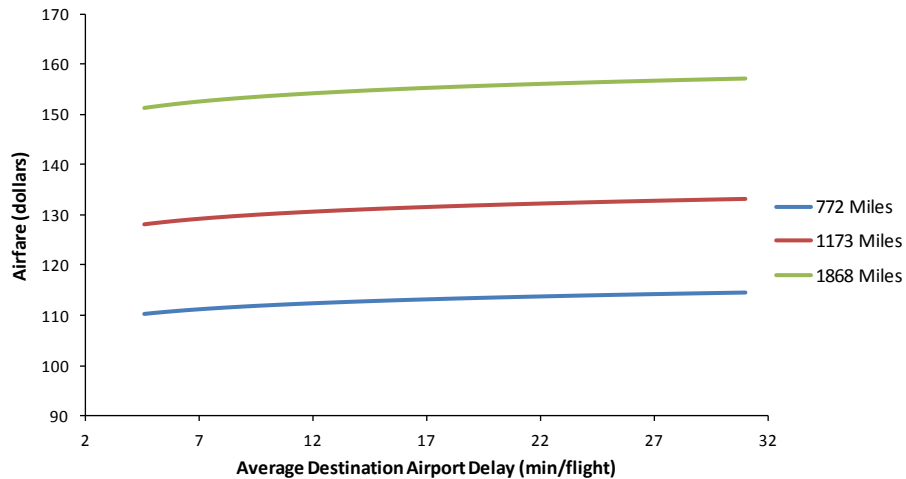
Figure 3.2: One-stop airfare as a function of delay at the destination airport at 25th, 50th, and 75th percentiles of non-stop O-D distance (772, 1173, 1868 miles)

## 3.4 Delay impact on flight frequency

### 3.4.1 Empirical model specification

The dependent variable in the frequency model is the total number of flights on a segment, summed over all carriers. Frequency is modeled as a function of passenger demand on the segment, segment distance, competition, delay, and among other variables. Similar to the fare models, all continuous variables take the logarithmic form. The

42

coefficients denote the frequency elasticities with respect to the explanatory variables. The model is specified as follows:

$$
\begin{aligned}
\ln f_{it} = {} & \gamma_0 + \gamma_1 \ln(\text{SegmentPax})_{it} + \gamma_2 \ln(\text{Dist})_{it} + \gamma_3 \ln(\text{Segment HHI})_{it} \\
& + \gamma_4 \ln(\text{OriginHHI})_{it} + \gamma_5 \ln(\text{DestHHI})_{it} + \gamma_6 \ln(\text{OriginL4Delay})_{it} \\
& + \gamma_7 \ln(\text{DestL4Delay})_{it} + \gamma_8 \ln(\text{Vacation})_{it} + \gamma_9 \ln(\text{Hawaii})_{it} + \gamma_{10}\text{SlotControl}_{it} \qquad (3.3) \\
& + \gamma_{11}\text{OriginMAS}_{it} + \gamma_{12}\text{DestMAS}_{it} + \gamma_{13}\text{LCC}_{it} + \gamma_{14}\ln(\text{PortionLCC}+1)_{it} \\
& + \gamma_{15}\text{MASPair}_{it} + v_{it}
\end{aligned}
$$

where subscripts i and t are indicators of a segment and a time period. $v_{it}$ the error term. The meanings of all variables are described in Table 3.6.

Table 3.6 Definition of frequency model variables.

| Name | Definition |
|---|---|
| f | Total number of flights flown for a given flight segment and time period. |
| SegmentPax | Total number of passengers on the flight segment. |
| Dist | Flight segment distance. |
| SegmentHHI | HHI on the flight segment. |
| OriginHHI | HHI at the origin airport. |
| DestHHI | HHI at the destination airport. |
| OriginL4Delay | Average flight delay at the origin airport (four quarters lagged). |
| DestL4Delay | Average flight delay at the destination airport (four quarters lagged). |
| Vacation | Dummy variable, equal to 1 if at least one of the origin and destination airports is in the States of Florida and Nevada. |
| SlotControl | Dummy variable, equal to 1 if at least one of the origin and destination airports is slot controlled. |
| OriginMAS | Dummy variable, equal to 1 if the origin airport is in a multiple airport system (MAS). |
| DestMAS | Dummy variable, equal to 1 if the destination airport is in an MAS. |
| LCC | Dummy variable, equal to 1 if LCC's are present on the flight segment. |
| PortionLCC | The portion of passengers transported by LCC's on the flight segment. |
| MASPair | The number of segments whose origin and destination share the same metropolitan areas as the origin and destination of the flight segment under study (including the segment itself). |

The most important variable explaining frequency variation is segment passenger volume. Since in principle airlines have the flexibility of changing aircraft size, we expect the frequency to be inelastic to passenger volume change. Distance also plays a significant role in determining frequency, as the least-cost aircraft type varies by stage length: in shorter markets smaller aircraft often have lower costs; as market distance increases, so does the size of the least-cost aircraft (Baily et al., 1985). Controlling for segment passenger demand, longer distance then suggests lower flight frequency. An alternative explanation could be that only larger aircraft can fly on longer-haul segments, resulting in lower frequency.

43

Similar to the fare model specification, competition effects are captured in multiple facets. HHI variables are included to characterize the extent of concentration at the segment (SegmentHHI) and airport (OriginHHI and DestHHI) levels. Lower concentration implies more severe frequency competition, and higher number of flights on the segment, everything else being equal. We further include two variables to capture the impact from LCCs: a dummy (LCC) indicating the presence of LCCs, and the portion of segment passengers transported by LCCs (PortionLCC), which captures the market power of LCC.[17] Competition would further arise when there are parallel segments, i.e. segments whose origin and destination are within the same MAS's. We expect higher frequency on these segments not only because of the competition effect, but also possibly the higher incomes and therefore time values for the metropolitan areas involved. The presence of MAS's may further suggest additional congestion effect because of the complex terminal airspace associated with multiple airport systems; this could partially countervail the aforementioned frequency increase effect due to competition and higher time values of travelers. To capture the parallel segment effect we introduce two dummies, OriginMAS and DestMAS, indicating respectively whether the origin and destination are in MAS's, and MASPair, which denotes the number of total segments servicing a metropolitan area pair.

To test the hypothesis that congestion and delay discourage airlines from scheduling more flights on a segment, we include in the model the origin and destination airport delays with four quarters' lag, which is consistent with the lead time in the frequency planning process mentioned in Section 3.2.2. We do not include the one-quarter lag delay variables because of the similar collinearity concern discussed in the context of the fare model specification (correlation with their respective four-quarter lagged delay variables above 0.7). When an ending airport is slot controlled, we expect the frequency on the segment to be further restricted.

Lastly, everything else being equal, we expect vacation segments to have fewer flights, because of a higher portion of leisure passengers with lower travel time values. This effect will be captured by the Vacation dummy.

### 3.4.2 Data

We use a panel dataset, by segment-quarter, spanning from the first quarter of 2004 to the fourth quarter of 2008 for model estimation. The dataset consists of domestic flight segments connecting the top 100 airports by throughput in the contiguous US (based on the 4th quarter of 2004) and having at least 60 flights per quarter. The average airport delay variables are constructed using the BTS Airline On-Time Performance Database. Again, airport slot control information is extracted from FAA (2008, 2009). Definitions about MAS and LCC follow those in the fare models. Metropolitan area population and income information, used for constructing instruments in model estimation, is obtained from the Bureau of Economic Analysis Regional Economic Accounts (BEA, 2011). All the remaining variables can be either directly obtained from or calculated based on data

---

[17] Because PortionLCC takes the logarithmic form, we use ln(PortionLCC+1) to avoid the occurrence of ln(0).

in the BTS T100 Domestic Segment Traffic Database. The final dataset for model estimation contains 3,858 segments and 65,033 segment-quarter observations. Table 3.7 provides descriptive statistics of all the variables.

Table 3.7 Descriptive statistics for variables in the frequency model (N=65,033).

| Variable | Mean | Std. Dev. | Min | Max |
| --- | --- | --- | --- | --- |
| f (flights/quarter) | 523.987 | 459.087 | 60 | 3899 |
| SegmentPax (passengers) | 43131.07 | 47055.93 | 223 | 376616 |
| Dist (miles) | 836.929 | 571.339 | 36 | 2724 |
| SegmentHHI | 0.748 | 0.256 | 0.170 | 1 |
| OriginHHI | 0.260 | 0.154 | 0.067 | 0.942 |
| DestHHI | 0.260 | 0.153 | 0.067 | 0.942 |
| OriginL4Delay (min/flight) | 12.666 | 3.872 | 4.584 | 31.020 |
| DestL4Delay (min/flight) | 12.672 | 3.872 | 4.584 | 31.020 |
| Vacation | 0.203 | 0.403 | 0 | 1 |
| SlotControl | 0.095 | 0.294 | 0 | 1 |
| OriginMAS | 0.305 | 0.461 | 0 | 1 |
| DestMAS | 0.304 | 0.460 | 0 | 1 |
| LCC | 0.380 | 0.485 | 0 | 1 |
| PortionLCC | 0.240 | 0.379 | 0 | 1 |
| MASPair | 2.010 | 1.805 | 1 | 14 |

### 3.4.3 Estimation, results and discussion

The potential simultaneity between flight frequency and SegmentPax suggests the necessity to use instruments for the SegmentPax variable and that model be estimated using 2SLS. While the intuitive instruments include those socio-economic variables such as population and income (Schipper et al., 2002), it is equally important to include instruments that capture connecting passenger traffic on that segment, which often account for an important portion in total segment passenger traffic. To account for both sources of passenger traffic we use the log of total income in the origin and destination metropolitan areas, and the ratio of connecting to local O-D passengers at origin and destination airports (based on outbound traffic) as the instruments. The portions do not take the logarithmic form as some segments are overwhelmingly dominated by local, O-D passengers. One further concern is the potential endogeneity of SegmentHHI, in that frequencies provided by incumbent carriers would determine the concentration through the S-curve phenomenon. We argue that, in airline practice, determination of flight frequency on a segment involves system-wide considerations, especially given aircraft rotation and hubbing constraints. As a consequence, we still treat SegmentHHI as exogenous. Table 3.8 below reports estimation results using both OLS and 2SLS by instrumenting only the SegementHHI variable. Standard errors are clustered by metropolitan area pair.

Table 3.8 Estimation results for the frequency model.

| Variable | OLS | | 2SLS | |
|---|---|---|---|---|
| | Est. | Std. Err. | Est. | Std. Err. |
| SegmentPax | 0.6400*** | 0.0052 | 0.6561*** | 0.0118 |
| Dist | -0.3652*** | 0.0072 | -0.3662*** | 0.0073 |
| SegmentHHI | -0.3250*** | 0.0102 | -0.3087*** | 0.0137 |
| OriginHHI | -0.0349*** | 0.0082 | -0.0415*** | 0.0092 |
| DestHHI | -0.0385*** | 0.0082 | -0.0451*** | 0.0092 |
| OriginL4Delay | -0.0220* | 0.0120 | -0.0222* | 0.0119 |
| DestL4Delay | -0.0238** | 0.0121 | -0.0238** | 0.0121 |
| Vacation | -0.1395*** | 0.0108 | -0.1435*** | 0.0111 |
| SlotControl | 0.0106 | 0.0130 | 0.0118 | 0.0130 |
| OriginMAS | 0.1087*** | 0.0136 | 0.1069*** | 0.0136 |
| DestMAS | 0.1105*** | 0.0137 | 0.1088*** | 0.0137 |
| LCC | 0.0252** | 0.0113 | 0.0124 | 0.0140 |
| PortionLCC | -0.3722*** | 0.0237 | -0.3643*** | 0.0239 |
| MASPair | -0.0408*** | 0.0116 | -0.0451*** | 0.0118 |
| Constant | 1.7242*** | 0.0775 | 1.5611*** | 0.1294 |
| Number of Observations | 65,033 | | 65,033 | |
| $R^2$ | 0.8949 | | 0.8947 | |
| First stage partial F-stat | | | 160.577 | |
| Partial $R^2$ | | | 0.1606 | |

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

The results from OLS and 2SLS are remarkably close in both coefficient estimates and their significance levels—except for the instrumented SegmentPax which is natural because instruments are to generate predicted, less precise value of SegmentPax. This suggests that the endogeneity effect of SegmentPax might not be very strong.[18] All the coefficients have the expected signs and most of them are significant. SegmentPax has a coefficient of 0.66, comparable to previous estimates of 0.65 in Givoni and Rietveld (2009) and 0.75 in Schipper et al., (2002), implying frequency is inelastic to passenger demand. Nonetheless, this also suggests that the majority of segment passenger increase will be absorbed by frequency increase. Under the assumption of constant load factor, aircraft size would respond to demand only with an elasticity of 0.34. As expected, longer distance entails larger least-cost aircraft type and a smaller set of available aircraft types. Therefore frequency tends to be lower. Alternatively, the negative coefficient may be interpreted as the result of diminishing contribution of schedule delay in total travel time, and less intense competition from surface modes in longer distance.

The negative coefficients for HHI's indicate that higher concentration leads to lower frequency. Frequency on a segment with four equally-sized competitors would be 23% higher than a segment operated by one single carrier everything else being equal. This effect, however, is much weaker at the airport level. When LCC's operate on one segment, frequency will increase by 2.54% (OLS). This effect is much smaller and insignificant

---

[18] We also note similar conclusions drawn in Givoni and Rietveld (2009) when regressing aircraft size on market size.

when 2SLS is employed. Frequency decreases with LCC share, offsetting any positive effect of LCC presence. The coefficient for PortionLCC suggests that a segment with 50% LCCs' share would have 22.3% fewer flights than an otherwise same segment but with 25% LCCs' share. This is because LCCs tend to use larger LCC aircraft size and lower frequency, the latter of which because of a higher portion of point-to-point service in LCC's network. If both origin and destination regions have two airports each (i.e. OriginMAS and DestMAS equal one), and four parallel segments exist, then segment frequency would increase at 15% compared to a non-MAS case. This may also be attributed to the likely higher portion of business passengers and income level when MAS's are involved. Similar to the LCC effect, after controlling for OriginMAS and DestMAS the positive effect of parallel routes attenuates with the increase in the number of parallel segments, as shown by the negative coefficient for MASPair.[19]

The estimated coefficients clearly indicate that high delay at either origin or destination airport tends to reduce the number of scheduled flights. If the average delay per flight at the destination airport increases from 5 min to 20 min, then frequency would be reduced by about 3.3%. This rather insensitive response to delay increase reflects airlines' concern about losing market share and therefore reluctance to cut back their schedules. The insignificant SlotControl coefficient, which has a positive sign, might be interpreted by the higher travel time values for passengers in the relevant metropolitan areas. As a result, airlines tend to schedule more flights despite operation restrictions imposed by slot controls; or alternatively, grandfather and "use it or lose it" rules for allocating runway capacity at these airports encourage airlines to increase frequency to maintain their share of runway capacity (Givoni and Rietveld, 2009). The potential airspace congestion effect due to the presence of MAS may not be a major concern in airline scheduling, given the positive coefficients of OriginMAS and DestMAS.

As suggested by the negative coefficient for the Vacation dummy, when one ending airport is either in Nevada or Florida, frequency on the segment will be about 14% lower than on an otherwise identical segment.

The predicted segment frequencies are plotted as a function of average flight delay at the destination airport, for three segment distances (25th, 50th, and 75th percentiles in the dataset) and with all other variables taking the sample average (Figure 3.3). Comparing the highest and lower delay scenarios (4.58 min/flight and 31.02 min/flight) as appearing in the dataset, daily frequency would only be reduced by up to 0.3 flights (in the 25th percentile case), or 4.4%. The resulting schedule delay increase would be hardly perceivable to passengers.

---

[19] For all possible cases in the number of parallel segments, we find that the overall MAS effect on frequency is always positive.
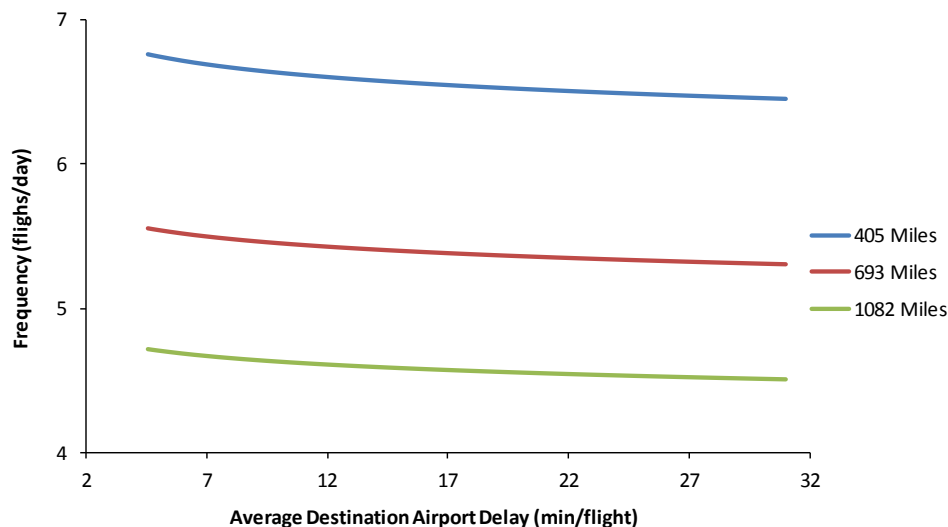
Figure 3.3: Segment frequency as a function of delay at the destination airport at 25th, 50th, and 75th percentiles of segment distance (405, 693, 1082 miles)

## 3.5  Summary

Understanding airlines' behavior in the presence of congestion has become increasingly important in the U.S., given the unprecedented delay in the past decade and the projected future travel demand growth. This chapter has empirically examined the pricing and scheduling responses of U.S. airlines. We find clear evidence that airlines tend to increase fare and decrease frequency facing delay in the air transportation system. The delay effect on airfare is found to be small, and particularly so on connecting routes. This is because, as has been explicitly shown in Chapter 2, airlines, when passing part of the delay cost onto passengers through higher fare, also consider passengers' reduced willingness-to-pay due to degraded air service quality in maximizing the profit. Flight frequency is even less sensitive to delays. Airlines' upgauging possibility facing delay is constrained by the loss of pricing power, the Mohring effect, and frequency competition pressure. Airlines, as a consequence, seem willing to maintain a "robust" schedule to delays despite operating cost increase. It could also imply that the cost penalty imposed by delays is not sizable enough to alter airlines' scheduling decision making.

One should bear in mind that the previous discussion of the delay impact on air transport supply rests on the assumption that fare and frequency respond only to delay change. A more complete evaluation of system response to delay change, for instance, from airport runway expansion, necessitates additional considerations of travelers' perception of delay in their total generalized travel cost, and the interactions between supply and demand. Travelers may choose to switch routes to avoid congested airports, or be discouraged from taking air travel at all. New fare and frequency resulting from delay change would also affect passenger demand. Demand adjustment then feeds back to the supply side. From the model estimates, one would expect major change occurs to flight frequency as

driven by the new passenger demand. Upgauging is more likely to happen on segments with higher passenger traffic rather than higher delay. Variation in fare would be much limited, given its low elasticities with respect to both delay and demand. The feedback then goes from supply to demand, and continues until an exact match between new system demand and supply is achieved. More detailed presentation of the interactions is deferred to the ensuing chapter.

This chapter presents one of the very few attempts to date that provide systematic insights about the delay impact on air transportation supply. The empirical modeling framework can be extended in several ways in future research. First, further efficiency gains can be achieved by jointly estimating airfare on non- and one-stop routes. Given the endogeneity of frequency and aircraft size, simultaneous estimation of frequency and aircraft size would not only increase efficiency, but offer further insights about the split among frequency, aircraft size, and load factor facing passenger demand increase. Another direction is to fine-tune the measurement of delay. With increasing availability of on-time statistics to the traveling public, airlines may adopt different pricing strategies for flights prone to delays from those that are more punctual. Capturing the heterogeneity of pricing strategies would require segmentation of markets/routes. To further investigate the non-linear delay effect on airfare, it may be worthwhile to consider alternative model specification, such as piecewise functional forms. It would also be interesting to extend the investigation to the impact of schedule padding, since buffer reduces the "discernible" delay incurred by flights. Airlines' operational performance can be measured using alternative metrics, for instance, metrics that describe the coherence between the actual flight time and the schedule (Zou and Hansen, 2012).

# 4. User Equilibrium Model

## 4.1 Introduction

In Chapter 2, the supply-demand equilibrium is formulated from the airline competition perspective. There profit-maximizing airlines make decisions on price and frequency in a competitive environment, with full knowledge about travelers' utility structure. The dual relationship between air travelers and airlines can be approached in a reverse manner: utility-maximizing travelers make air travel decisions based upon supply characteristics, such as airfare, flight frequency, and delay, which are directly or indirectly determined by airlines. While an individual's decision can only alter system supply to a very minimum extent, decisions made by the population as a whole would lead to perceivable, systematic changes in airlines' pricing and scheduling behavior, which in turn shapes individuals' travel decision making.

The discussion of system equilibrium in this chapter centers on travelers' decision making while taking into consideration constraints and feedback from system supply. The travelers, each viewed as an economic agent, interact in the air transportation system such that, under equilibrium no one can unilaterally change her/his decision to be better off. This user (traveler)-centered equilibrium presents a special instance of Nash equilibrium with a large number of players (travelers). Once capacity investment is made, the resulting delay reduction will lead travelers to reinteract and adjust their decisions, which change airfare and flight frequency, and subsequently delay in the system. Travelers further interact in response to the updated supply. The process continues until an exact match between demand and supply is achieved, featuring the completion of the system shift to a new equilibrium state.

Building upon this rationale, the remainder of this chapter provides a detailed exposition of the air transport user equilibrium (ATUE). We first discuss the equilibrium components, based on which ATUE is formulated in both fixed point and variational inequalities contexts. Equilibrium solution properties are then investigated, followed by a discussion about the differences between ATUE and the reminiscent Urban Traffic Stochastic User Equilibrium (UTSUE). We propose a solution procedure from the fixed point perspective in Section 4.3, and apply the algorithm to solve for ATUE in a hypothetical network setting in Section 4.4, where we present in detail the initial equilibrium, equilibrium shift in response to airport capacity change, associated passenger welfare gains, as well as some sensitivity analysis of capacity investment. We summarize our analysis and discuss potential extensions in Section 4.5.

## 4.2 Equilibrium formulation

The formulation of ATUE requires, in the first step, specifying equilibrium components. On the supply side, the fare and frequency components have already been empirically determined in the previous chapter. Following the equilibrium framework proposed in Chapter 1, there remain two missing components: passenger demand, which is the center piece in ATUE, and flight delay. With full specification of the equilibrium components, ATUE can be formulated as a fixed point problem, or alternatively, a variational inequality problem with some mild restrictions. While ATUE bears some analogy with the Stochastic User Equilibrium in urban transportation, recognizing the fundamental differences is important. We discuss these issues in sequence in this section.

### 4.2.1 Equilibrium component

*4.2.1.1 Demand*
The centerpiece in ATUE is travelers' trip decision making, which consists of whether to take air travel, and if so, which air travel product to choose. At the aggregate level, these decisions correspond to demand generation and assignment respectively. While in general the demand assignment process entails the choice of both routes and airlines, here we only focus on routing choice, since infrastructure investment analysis does not necessarily require airline-specific information. On the other hand, compared to modeling only route choice, simultaneous consideration of routing and carrier choice would involve a considerable amount of extra modeling work and impose much heavier computational burden in the equilibrium process. Such a price may not be worth paying given the purpose of our equilibrium analysis.

To model the air travel decision process we employ a previously estimated three-level Nested Logit (3NL) model by Hsiao and Hansen (2011). The nesting structure is shown in Figure 4.1. The model assumes a maximum number of potential trips, or saturated demand, for each individual on a given metropolitan-area-pair market. At the top level of the nest, an individual first decides on whether to travel by air or not. Aggregated over all individuals, this gives the number of realized air trips in the saturated demand. The two lower-level choice decisions deal with demand assignment: once air travel is chosen, the individual selects an O-D airport pair. This is relevant when at least one of the end-point metropolitan areas involves an MAS. Given the O-D airport pair choice, at the bottom level the traveler chooses a specific route, which can be non-stop, or connecting through a specific hub airport. This aggregate passenger demand model was estimated using air travel records in the US air transportation system between 1995 and 2004. The estimation results of the 3NL model are summarized in Table 4.1. Interested readers can refer to Hsiao and Hansen (2011) for further details.
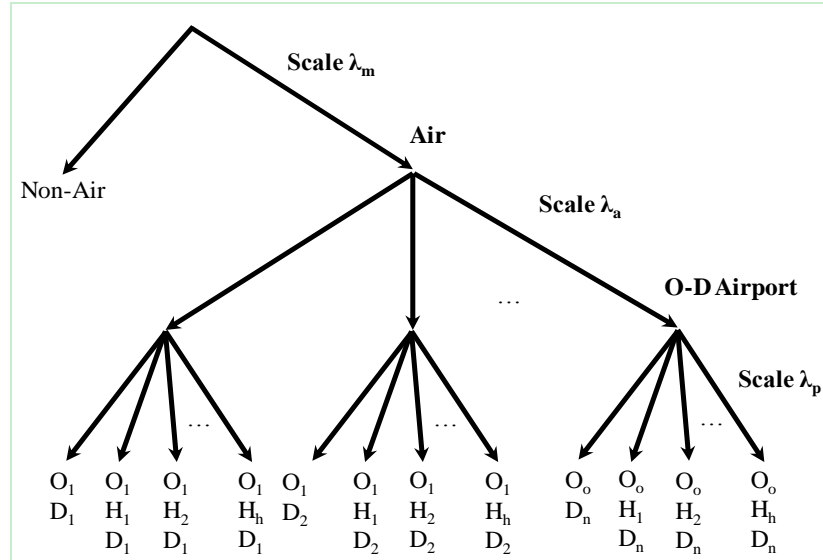
Figure 4.1: Nesting structure of the 3NL model

Table 4.1 Main estimation results of the 3NL model.

| Variable | Est. |
| --- | --- |
| **Level 3** | |
| Fare (hundreds of 2004 dollars) | -1.546*** |
| ln(Frequency)—Direct (flights per quarter) | 1.240*** |
| ln(Max frequency of two segments)— Connecting (flights per quarter) | 0.627*** |
| ln(Min frequency of two segments)— Connecting (flights per quarter) | 0.957*** |
| Scheduled flight time—Direct (minutes) | -0.004 |
| Scheduled flight time— Connecting (minutes) | -0.006** |
| Dummy for direct routes | 6.066*** |
| Positive hub arrival delay (1 quarter lag, minutes per flight) | -0.006*** |
| Positive hub arrival delay (4 quarters lag, minutes per flight) | -0.007*** |
| Constant (level 3) | -0.005 |
| **Level 2** | |
| Inclusive value of level 3 | 0.664*** |
| Constant (level 2) | 0.007 |
| **Level 1** | |
| Inclusive value of level 2 | 0.795*** |
| Inclusive value of level 2*market distance | -0.012*** |
| Market distance (hundreds of miles) | -0.024*** |
| ln(market distance) | 1.575*** |
| Per capita personal income of market (in 000, 2004 dollars) | 0.038*** |
| Constant (level 1) | -16.229*** |

*** p<0.01, ** p<0.05, * p<0.10

*4.2.1.2 Supply*

As mentioned in the beginning of this section, airfare, flight frequency, and flight delay are considered to be endogenous supply-side components in the ATUE. The

determination of fare and frequency is governed by the econometric models in Chapter 3. Here we briefly present the specification and estimation of the delay model.

We consider flight delay at the airport level, and focus on arrival delay, because airport arrival delay has the most direct impact on air travelers, and also to be consistent with the delay specification in the demand, fare and frequency models. Among the primary influencing factors of airport arrival delay are total traffic volume, airport capacity, and weather conditions, whose effect may be non-linear. In light of these, we specify the following semi-log delay function:

$$
\begin{aligned}
\ln(\text{Delay}_{it}) = {} & \delta_0 + \delta_1 \text{IFR}_{it} + \delta_2 \text{IFR}_{it}^2 + \delta_3 \text{Wind}_{it} + \delta_4 \text{VC}_{it} + \delta_5 \text{VC}_{it}^2 + \delta_6 \text{AAR}_{it} \\
& + \delta_7 \text{Peakedness}_{it} + \delta_8 \text{Connection}_{it} + \delta_9 \text{Temp}_{it} + \sum_j \eta_j \text{A}_j + \sum_k \mu_k \text{M}_k + \omega_{it}
\end{aligned} \tag{4.1}
$$

where subscripts i and t are indicators of an airport and a day. $\omega_{it}$ is the error term. Each observation therefore requires flight delay, operational, and weather information for an airport-day pair. Data are collected from the FAA ASPM (Aviation System Performance Metrics) database for US OEP 35 airports in 2007.[20] Since the data has a panel form, we introduce airport dummies as individual airports in the panel could have features that consistently increase or decrease delay. Similarly, monthly dummies are included to capture the time fixed effect. The meanings of the dependent and explanatory variables are provided in Table 4.2 below.

Table 4.2 Description of variables in the airport delay model.

| Name | Definition |
| --- | --- |
| Delay | Average daily delay (min/flight) at the airport. |
| IFR | Portion of time during a day in which the airport operated under Instrument Flight Rules (IFR) conditions. |
| $\text{IFR}^2$ | Square term of IFR. |
| Wind | Average wind speed in a day at the airport. |
| VC | Ratio of daily total flight traffic volume to runway capacity at the airport. |
| $\text{VC}^2$ | Square term of VC. |
| AAR | Daily airport arrival acceptance rate. |
| Peakedness | Standard deviation of scheduled arrival operations throughout the quarter-hours for the airport and in a day (except for hours between 12am and 6am). |
| Connection | Number of airports with non-stop flight segments connecting to the airport under study. |
| Temp | Average temperature at the airport in a day. |

---

[20] The OEP 35 (Operational Evolution Partnership) airports are commercial U.S. airports serving major metropolitan areas and as airline hubs with significant flight activities. More than 70 percent of passengers move through these airports (FAA, 2009b). A full list of OEP 35 airports can be found in http://aspmhelp.faa.gov/index.php/OEP_35.

| | | | | | |
|---|---|---|---|---|---|
| $A_j$ | Airport dummy, equal 1 if the observation corresponds to airport j (TPA as the base airport). | | | | |
| $M_k$ | Monthly dummy, equal to 1 if the observation corresponds to month k (Dec as the base month). | | | | |

Several econometric issues are considered in the model estimation. First, since airport operations are interdependent in the National Airspace System (NAS), it is important to account for this interdependency in estimating the model. Second, errors in econometric delay models are often found to be heteroskedastic (Wei and Hansen, 2006). Third, serial correlation among error terms may exist because, among other reasons, delay at the end of a day could possibly affect the operations of the next day. A Prais-Winsten regression is performed by allowing a first-order autocorrelation between observations for the same airport. We use panel corrected standard errors, in which error terms are assumed heteroskedastic and contemporaneously correlated across panels (i.e. errors are correlated across airports at a given point in time). Estimation results are reported in Table 4.3.

Table 4.3 Estimation results for the airport delay model.

| Variable | Est. | Std. Err. | Variable | Est. | Std. Err. |
|---|---|---|---|---|---|
| IFR | 1.1283*** | 0.0682 | DEN | -0.7216*** | 0.1334 |
| IFR$^2$ | -0.7089*** | 0.0738 | DFW | -0.4513*** | 0.1613 |
| Wind | 0.0168*** | 0.0019 | DTW | -1.3578*** | 0.1262 |
| VC | 1.0192*** | 0.2265 | EWR | -0.6812*** | 0.0772 |
| VC$^2$ | 1.1747*** | 0.1758 | FLL | -0.2309*** | 0.0372 |
| AAR | -0.0017*** | 0.0002 | HNL | -0.3350*** | 0.0708 |
| Peakedness | 0.1774*** | 0.0124 | IAD | -0.6532*** | 0.0639 |
| Connection | 0.0065*** | 0.0020 | IAH | -1.1405*** | 0.1293 |
| Temp | -0.0030*** | 0.0009 | JFK | -0.6409*** | 0.0702 |
| Constant | 2.1330*** | 0.1327 | LAS | -0.5996*** | 0.0771 |
| Jan | -0.3102*** | 0.0625 | LAX | -0.5346*** | 0.1110 |
| Feb | -0.0649 | 0.0640 | LGA | -0.7632*** | 0.0811 |
| Mar | -0.1115* | 0.0640 | MCO | -0.1446*** | 0.0394 |
| Apr | -0.3013*** | 0.0639 | MDW | -0.6656*** | 0.0410 |
| May | -0.2875*** | 0.0651 | MEM | -0.7108*** | 0.0548 |
| Jun | 0.1396** | 0.0672 | MIA | 0.3761*** | 0.0416 |
| Jul | 0.0554 | 0.0679 | MSP | -1.2095*** | 0.1214 |
| Aug | 0.0083 | 0.0681 | ORD | -0.9874*** | 0.2053 |
| Sep | -0.4559*** | 0.0666 | PDX | -0.3270*** | 0.0508 |
| Oct | -0.3530*** | 0.0642 | PHL | -0.8576*** | 0.0821 |
| Nov | -0.4470*** | 0.0644 | PHX | -0.8567*** | 0.0895 |
| ATL | -1.2723*** | 0.2576 | PIT | 0.1381*** | 0.0532 |
| BOS | -0.5232*** | 0.0582 | SAN | -1.0321*** | 0.0557 |
| BWI | -0.6002*** | 0.0376 | SEA | -0.8436*** | 0.0598 |

| | | | | | |
|---|---|---|---|---|---|
| CLE | -0.9261*** | 0.0445 | SFO | -0.5518*** | 0.0557 |
| CLT | -1.1863*** | 0.0920 | SLC | -1.1125*** | 0.0949 |
| CVG | -1.3299*** | 0.1079 | STL | -0.1563*** | 0.0325 |
| DCA | -0.8784*** | 0.0525 | | | |
| Number of Observations | | | 12,605 | | |
| $R^2$ | | | 0.5064 | | |

*** p<0.01, ** p<0.05, * p<0.10

In general, the estimated coefficients are consistent with *a priori* expectations. Both the ratio between airport flight traffic volume and airport capacity (VC) and its square term ($VC^2$) have positive coefficients and are highly significant, suggesting that, for a given airport capacity, average flight delay would rise at a highly nonlinear, increasing rate with flight traffic. Greater prevalence of IFR conditions results in high delay, although the negative coefficient for the quadratic term suggests that at extremely high IFR values predicted delay would be slightly lower. Higher delay values are associated with stronger winds and lower average temperature, as also found in Hansen and Hsiao (2005), and Hansen and Kwan (2010). Larger AAR seems to reduce average delay, because high AAR's tend to be set more conservatively—that is, at a lower level relative to the absolute maximum throughput—than low AAR's (de Neufville and Odoni, 2003). The positive coefficient for Connection suggests that greater connectivity would complicate aircraft turnaround operations and increase the exposure of the airport to delay propagated from other airports, therefore making the airport more susceptible to delays. *Ceteris paribus*, the months of February, July, and August would experience the same level of delays as that in December because of their statistically insignificant dummy coefficients. Airport delay in June would be on average 15% higher than in December; whereas delay would be lower in the remaining months than in December—ranging from 10% in March to 37% in September. Interestingly, the bulk of airport dummy coefficients are negative and significant, implying that, all else equal, delays at most airports will be lower than at TPA.

## 4.2.2  Equilibrium formulation

### 4.2.2.1  ATUE as a fixed point problem

With full specification of the equilibrium component models, attention can now be directed toward formalizing ATUE. In this chapter we consider a generic, strongly connected air transportation network (N, A), where N is the set of airports and A the set of segments. Let K be the set of feasible routes. Recall the equilibrium framework in Chapter 1. It is assumed that, except for passenger demand, airfare, frequency, and airport delay, all other variables in the component models are predetermined and exogenous to the equilibrium process. We further assume that equilibria considered in this chapter are in steady states. As a consequence, lagged delays in demand, fare, and frequency models would be indifferent from contemporary delays. The resulting equilibrium predicts the long-run system behavior. With the above assumptions, the equilibrium component models can be succinctly expressed in vector forms:

Route passenger demand $\quad$ $\mathbf{D} = \mathbf{\textit{G}_1(P, f, d)}$ $\hspace{4cm}$ (4.2)

Route airfare $\hspace{2.9cm}$ $\mathbf{P} = \mathbf{\textit{G}_2(D, Q, d)}$ $\hspace{4cm}$ (4.3)

Segment flight traffic $\hspace{1.6cm}$ $\mathbf{f} = \mathbf{\textit{G}_3(Q, d)}$ $\hspace{4.6cm}$ (4.4)

Airport delay[21] $\hspace{2.6cm}$ $\mathbf{d} = \mathbf{\textit{G}_4(v)}$ $\hspace{5cm}$ (4.5)

where

| | |
|---|---|
| **D** | Route passenger volume vector ($K \times 1$) |
| **P** | Route airfare vector ($K \times 1$) |
| **f** | Segment flight frequency vector ($A \times 1$) |
| **Q** | Segment passenger volume vector ($A \times 1$) |
| **v** | Airport flight traffic vector ($N \times 1$) |
| **d** | Airport delay vector ($N \times 1$) |

Equilibrium exists when the spatial distribution of passenger flows on different routes satisfies (4.2). It is a user equilibrium because underlying (4.2) is utility maximization of individual travelers. Utilities are modeled as stochastic because they cannot be fully perceived by researchers. Under the user equilibrium, no traveler can be better off by unilaterally changing her/his air travel decision.

The level of route passenger demand, as denoted by (4.2), depends upon the supply-side inputs **P**, **f**, and **d**, whose values are determined by (4.3)-(4.5). In addition to **D**, **P**, **f**, and **d**, two additional endogenous vectors, **Q** and **v**, are involved in (4.3)-(4.5). They can be conveniently expressed as functions of **D** and **f**, following the regular network constraints:

$$\mathbf{Q} = \Delta \mathbf{D} \hspace{8cm} (4.6)$$

$$\mathbf{v} = \Phi \mathbf{f} \hspace{8.2cm} (4.7)$$

where $\Delta$ is a $|A| \times |K|$ route-segment incidence matrix in which the element in the $a$th row and $r$th column, $\delta_{ar}$, equals 1 if route $r$ uses segment $a$ and 0 otherwise. $\Phi$ is a $|N| \times |A|$

---

[21] Recall that the airport delay model also includes the Peakedness variable, which may be endogenous in the equilibrium process. A simplification is made that Peakedness varies in proportion to airport flight traffic volume.

segment-airport incidence matrix where the element in the $n$th row and $a$th column, $\varphi_{na}$, equals 1 if segment $a$ goes into airport $n$ and 0 otherwise.

The user equilibrium (4.2), combined with relationships (4.3)-(4.5) determining supply-side inputs and network flow constraints (4.6)-(4.7), equates to solving for the unknowns $\mathbf{D}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{f}$, $\mathbf{v}$, $\mathbf{d}$ in the simultaneous equation system. Stacked together, (4.2)-(4.7) can be viewed as a fixed point problem:

$$\mathbf{Y} = S(\mathbf{Y}) \tag{4.8}$$

where $\mathbf{Y} = (\mathbf{D}^\mathrm{T}, \mathbf{P}^\mathrm{T}, \mathbf{Q}^\mathrm{T}, \mathbf{f}^\mathrm{T}, \mathbf{v}^\mathrm{T}, \mathbf{d}^\mathrm{T})^\mathrm{T}$. This fixed point formulation can be further reduced to only involve $\mathbf{D}$ and $\mathbf{d}$ through the following substitutions:

i) Substitute (4.6) into (4.3):

$$\mathbf{P} = G_2(\mathbf{D}, \Delta\mathbf{D}, \mathbf{d}) = \overline{G_2}\,(\mathbf{D}, \mathbf{d}) \tag{4.9}$$

ii) Substitute (4.6) into (4.4):

$$\mathbf{f} = G_3(\Delta\mathbf{D}, \mathbf{d}) = \overline{G_3}\,(\mathbf{D}, \mathbf{d}) \tag{4.10}$$

iii) Substitute (4.7) and (4.10) into (4.5):

$$\mathbf{d} = G_4(\Phi\mathbf{f}) = \overline{G_4}\,(\mathbf{f}) = \overline{G_4}\,(\overline{G_3}\,(\mathbf{D}, \mathbf{d})) \tag{4.11}$$

iv) Substitute (4.9) and (4.10) into (4.2):

$$\mathbf{D} = G_1(\mathbf{P}, \mathbf{f}, \mathbf{d}) = G_1(\overline{G_2}\,(\mathbf{D}, \mathbf{d}), \overline{G_3}\,(\mathbf{D}, \mathbf{d}), \mathbf{d}) = \overline{G_1}\,(\mathbf{D}, \mathbf{d}) \tag{4.12}$$

Let $\mathbf{X} = (\mathbf{D}, \mathbf{d})^\mathrm{T}$. (4.11) and (4.12) form a new fixed-point problem:

$$\mathbf{X} = F(\mathbf{X}) \tag{4.13}$$

where $F(\mathbf{X}) = \begin{pmatrix} \overline{G_1}(\mathbf{D},\mathbf{d}) \\ \overline{G_4}(\overline{G_3}(\mathbf{D},\mathbf{d})) \end{pmatrix}$. Any fixed point for problem (4.8) is clearly a solution for (4.13). Conversely, for any given solution to (4.13), we can always derive corresponding $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{f}$, $\mathbf{v}$ using (4.9), (4.6), (4.10) and (4.7), such that $(\mathbf{D}^\mathrm{T}, \mathbf{P}^\mathrm{T}, \mathbf{Q}^\mathrm{T}, \mathbf{f}^\mathrm{T}, \mathbf{v}^\mathrm{T}, \mathbf{d}^\mathrm{T})^\mathrm{T}$ satisfies (4.8). Therefore fixed problems (4.8) and (4.13) are equivalent.

The existence of equilibrium solution for (4.13) can be shown as follows. In the passenger demand model, the probability for an individual to choose a route is always less than 1. Saturated demand exists on each metropolitan O-D pair, because each individual has a maximum number of potential trips and the population is finite. One upper bound for route demand $\overline{G_1}(\mathbf{D}, \mathbf{d})$ is the saturated demand on the corresponding metropolitan O-D pairs. Let $\mathbf{D_{sat}}$ denote a $K \times 1$ vector of such upper bounds. Then mapping $\overline{G_1}(\mathbf{D}, \mathbf{d})$ satisfies $\mathbf{0} \leq \overline{G_1}(\mathbf{D}, \mathbf{d}) \leq \mathbf{D_{sat}}$. $\mathbf{f}$ is a non-decreasing function of $\mathbf{Q}$ and therefore $\mathbf{D}$, and a non-increasing function of $\mathbf{d}$. (4.10) implies that an upper bound $\mathbf{f_{sat}}$ governed by $\mathbf{D_{sat}}$ applies: $\mathbf{0} \leq \mathbf{f} \leq \mathbf{f_{sat}}$. Following the second equality in (4.11) and the fact that $\mathbf{d}$ is non-decreasing with $\mathbf{v}$ and therefore $\mathbf{f}$, $\mathbf{d}$ should be bounded as well: $\mathbf{0} \leq \mathbf{d} \leq \mathbf{d_{sat}}$. Let $\Omega$ denote set $\{\mathbf{D}: \mathbf{0} \leq \mathbf{D} \leq \mathbf{D_{sat}}; \mathbf{d}: \mathbf{0} \leq \mathbf{d} \leq \mathbf{d_{sat}}\}$. $\Omega$ is both convex and compact. The above analysis shows that, the continuous mappings of $\mathbf{X}$ on $\Omega$: $\mathbf{F}\Omega \subset \Omega$. Brower's fixed point theorem (Ortega and Rheinboldt, 1970) implies that $\mathbf{F}$ has at least one fixed point in $\Omega$.

The uniqueness of the fixed point problem solution, however, cannot be guaranteed under the model specification. One sufficient condition for the solution uniqueness requires $\mathbf{F}$ be a contractive mapping, i.e. $\exists \varphi < 1$ such that

$$\|\mathbf{F}(\mathbf{X_1}) - \mathbf{F}(\mathbf{X_2})\| \leq \varphi \|\mathbf{X_1} - \mathbf{X_2}\|, \ \forall \mathbf{X_1}, \mathbf{X_2} \in \Omega \tag{4.14}$$

where $\| . \|$ denote the norm of the vector, which measures the distance of the vector space. Using the mean-value theorem, $\|\mathbf{F}(\mathbf{X_1}) - \mathbf{F}(\mathbf{X_2})\| = \nabla \mathbf{F}(\xi)\|\mathbf{X_1} - \mathbf{X_2}\|$, where $\xi$ is some value that lies between $\mathbf{X_1}$ and $\mathbf{X_2}$. If the norm of the Jacobian satisfies $\|\nabla \mathbf{F}(\mathbf{X})\| < 1$, then $\mathbf{F}$ is a contractive mapping and the fixed point problem has a unique solution. Due to highly non-linear nature of $\mathbf{F}(\mathbf{X})$, however, the functional form Jacobian would be extremely complicated for any further investigation. We show in a simplified case that the Jacobian can be greater than one, i.e. in general the necessary conditions is not held (Appendix C).

### 4.2.2.2 ATUE as a VI problem
Under two mild modifications, we can also convert ATUE into a Variational Inequality (VI) formulation, providing a different angle to investigate equilibrium properties and solution procedure. This is the first attempt, to our knowledge, to establish the mathematical equivalence between a VI formulation and a user equilibrium problem with demand governed by a three-level NL structure. We demonstrate that the equivalent VI leads to the same conclusions about the existence and uniqueness of the equilibrium solution.

The two mild modifications are made on the equilibrium component models. First, we assume that flight traffic is independent of airport delays. Considering the rather small estimated delay coefficients in the frequency model, ignoring delay impact on frequency would generate little difference in the equilibrium results. The benefit of doing so is to permit "one-shot" calculation of perceived passenger utilities. Second, the proof of the equivalence is based on a more generic 3NL specification in which—unlike the 3NL

model in Section 4.2.1.1—we do not consider the interaction term between inclusive values and distance. We expect this simplification to only affect demand estimation and subsequent equilibrium outcomes marginally, meanwhile providing more general insight into the equivalence between a multi-level NL-based ATUE and the corresponding VI formulation.

We begin the proof by recalling the nesting structure in Figure X and introducing some additional notations. Let $K$, $Q$, $W$ denote the sets of routes, airport pairs, and metropolitan areas pairs in an air transport network. We define $D_w$ as the passenger demand on metropolitan area pair $w$; $D_{q,w}$ passenger demand on airport pair $q$ within metropolitan area pair $w$; $D_{k,q,w}$ passenger demand route pair $k$ within both airport pair $q$ and metropolitan area pair $w$. Given the three-level nesting structure with no interaction between inclusive value terms and other variables, $D_{k,q,w}$ is calculated as follows:

$$D_{k,q,w} = D_w^{sat} P_{k,q,w} = D_w^{sat} P_w P_{q|w} P_{k|(q,w)}$$

$$= D_w^{sat} \frac{\exp(\frac{c_w + \lambda_P I_w}{\lambda_M})}{1 + \exp(\frac{c_w + \lambda_P I_w}{\lambda_M})} \frac{\exp(\frac{c_{q,w} + \lambda_R I_{q,w}}{\lambda_P})}{\sum_{n \in T_w} \exp(\frac{c_{n,w} + \lambda_R I_{q,w}}{\lambda_P})} \frac{\exp(\frac{c_{k,q,w}}{\lambda_R})}{\sum_{m \in R_{q,w}} \exp(\frac{c_{m,q,w}}{\lambda_R})} \qquad (4.15)$$

where $c_w$, $c_{q,w}$, $c_{k,q,w}$ denote the perceived metropolitan area pair, airport pair, and route specific utilities. $I_{q,w}$ and $I_w$ are the inclusive values of routes serving $q$ within $w$, and of all routes serving $w$. $\lambda_R$, $\lambda_P$, $\lambda_M$ are scale parameters associated with different nest levels. In the 3NL model specification, $c_{k,q,w}$ is the generalized costs composed of route fare, segment frequency, hub airport delay, among other exogenous variable values. Given passenger route flows and that frequency is independent of delay, we compute segment flight traffic, airport delay, and airfare sequentially using a modified version of (4.4) that disregards delay effect on frequency, (4.5) and (4.3), respectively. Their values are then used to calculate $c_{k,q,w}$, and subsequently $I_{q,w}$ and $I_w$. $c_{q,w}$ and $c_w$ do not involve any endogenous variables, and are held constant in system equilibration. If the delay impact on frequency is not ignored, achieving consistent values for segment flight traffic, airport delay, and airfare conditional on passenger route flows would require some additional iterative procedure. However, the delay coefficient in the flight frequency model is so small that results from the above "one-shot" calculation would be only slightly different from those when the delay impact is fully considered.

The ATUE is defined by (4.15), together with the supply side constraints: the modified version of (4.4), (4.3), (4.5), and the following regular network and non-negativity requirements:

$$\sum_{k \in R_{q,w}} D_{k,q,w} - D_{q,w} = 0$$

$$\sum_{q \in R_w} D_{q,w} - D_w = 0$$

$$D_{k,q,w} \geq 0 \qquad \forall k \in R_{q,w}, q \in R_w, w \in W \qquad (4.16)$$

$$D_{q,w} \geq 0$$

$$D_w \geq 0$$

where $R_{q,w}$ denotes the set of feasible routes for airport pair $q$ within metropolitan area pair $w$; $R_w$ the set of feasible airport pairs within metropolitan area pair $w$; $W$ the set of all metropolitan area pairs in the air transportation network.

Now let us define vector $\mathbf{X}$ and its vector function $\mathbf{H(X)}$ as follows:

$$\mathbf{X} = \begin{bmatrix} \vdots \\ D_{k,q,w} \\ \vdots \\ \vdots \\ D_{q,w} \\ \vdots \\ \vdots \\ D_w \\ \vdots \end{bmatrix} \qquad \mathbf{H(X)} = \begin{bmatrix} \vdots \\ c_{k,q,w} - \lambda_R \ln \dfrac{D_{k,q,w}}{D_{q,w}} \\ \vdots \\ \vdots \\ c_{q,w} - \lambda_P \ln \dfrac{D_{q,w}}{D_w} \\ \vdots \\ \vdots \\ c_w - \lambda_M \exp\left[\dfrac{D_w}{D_w^{sat}} \middle/ (1 - \dfrac{D_w}{D_w^{sat}})\right] \\ \vdots \end{bmatrix} \qquad (4.17)$$

where vector $\mathbf{X}$ consists of passenger flows on each route, airport pair, and metropolitan area pair, therefore having $(|K|+|Q|+|W|)$ elements. $\mathbf{H(X)}$ has the same dimension. The calculation of $c_{k,q,w}$, $c_{q,w}$, and $c_w$ in $\mathbf{H(X)}$ reflects the supply-side constraints and follows the procedure elucidated above. With these supply-side constraints, a corresponding VI problem is to find a vector $\mathbf{X}^*$ such that

$$(\mathbf{X} - \mathbf{X}^*)^T \mathbf{H(X^*)} \geq 0 \quad \forall \mathbf{X} \in \Omega \qquad (4.18)$$

where $\Omega$ is defined by (4.16). Having the same supply-side constraints and regular network and non-negativity requirements, it remains to see how (4.18) is equivalent to (4.15).

We express (4.16) in the following matrix form

$$\mathbf{BX} = \mathbf{0}, \mathbf{AX} \geq \mathbf{0} \tag{4.19}$$

where

$$\mathbf{B} = \begin{bmatrix} \Delta_{qk} & -\mathbf{I}_{|Q|} & \mathbf{0} \\ \mathbf{0} & \Delta_{wq} & -\mathbf{I}_{|W|} \end{bmatrix}; \quad \mathbf{A} = \mathbf{I}_{(|K|+|Q|+|W|)}$$

where $\Delta_{qk}$ is the airport pair-route incidence matrix with dimension $|Q| \times |K|$; $\Delta_{wq}$ the O-D pair-airport pair incidence matrix with dimension $|W| \times |Q|$.

A sufficient and necessary condition for the solution $\mathbf{X}^*$ in (4.18) is (Facchinei and Pang, 2003):

$$H(\mathbf{X}) + \mathbf{B}^{\mathbf{T}}\boldsymbol{\mu} - \mathbf{A}^{\mathbf{T}}\boldsymbol{\eta} = \mathbf{0} \tag{4.20.1}$$
$$\mathbf{BX} = \mathbf{0} \tag{4.20.2}$$
$$\boldsymbol{\eta} \geq \mathbf{0} \tag{4.20.3}$$
$$\mathbf{AX} \geq \mathbf{0} \tag{4.20.4}$$
$$(\mathbf{AX})^{\mathbf{T}}\boldsymbol{\eta} = \mathbf{0} \tag{4.20.5}$$

Passenger flow in the air transport network is always positive, i.e. $\mathbf{X} > \mathbf{0}$. This together with $\mathbf{A} = \mathbf{I}_{(|K|+|Q|+|W|)}$ and (4.20.5) implies that $\boldsymbol{\eta} = \mathbf{0}$, under which (4.20.1) can be expressed in scalar form as in (4.21.1)-(4.21.3):

$$c_{k,q,w} - \lambda_R \ln\frac{D_{k,q,w}}{D_{q,w}} + \mu_{q,w} = 0 \quad \forall k, q, w \tag{4.21.1}$$

$$c_{q,w} - \lambda_P \ln\frac{D_{q,w}}{D_w} - \mu_{q,w} + \mu_w = 0 \quad \forall q, w \tag{4.21.2}$$

$$c_w - \lambda_M \exp(\frac{\dfrac{D_w}{D_w^{sat}}}{1 - \dfrac{D_w}{D_w^{sat}}}) - \mu_w = 0 \quad \forall w \tag{4.21.3}$$

It follows from (4.21.1) that

$$D_{k,q,w} = D_{q,w} \exp(\frac{\mu_{q,w}}{\lambda_R}) \exp(\frac{c_{k,q,w}}{\lambda_R}) \quad \forall k, q, w \tag{4.22}$$

Summing all routes within airport-pair $q$ yields

$$D_{q,w} = \sum_{m \in R_{q,w}} D_{m,q,w} = D_{q,w} \exp(\frac{\mu_{q,w}}{\lambda_R}) \sum_{m \in R_{q,w}} \exp(\frac{c_{m,q,w}}{\lambda_R}) \quad \forall q, w \qquad (4.23)$$

After some algebra, $\mu_{q,w}$ can be expressed as

$$\mu_{q,w} = -\lambda_R \log[\sum_{m \in R_{q,w}} \exp(\frac{c_{m,q,w}}{\lambda_R})] = -\lambda_R I_{q,w} \qquad \forall q, w \qquad (4.24)$$

which gives the logsum of alternatives in nest $(q, w)$. Substituting (4.24) into (4.22) yields

$$D_{k,q,w} = D_{q,w} \frac{\exp(\frac{c_{k,q,w}}{\lambda_R})}{\sum_{m \in R_{q,w}} \exp(\frac{c_{m,q,w}}{\lambda_R})} \quad \forall k, q, w \qquad (4.25)$$

Performing similar exercise on (4.21.2) and substituting (4.24) for $\mu_{q,w}$ lead to

$$D_{q,w} = D_w \exp(\frac{\mu_w}{\lambda_P}) \exp[\frac{1}{\lambda_P}(c_{q,w} + \lambda_R I_{q,w})] \quad \forall q, w \qquad (4.26)$$

Sum over airport pairs that belong to market $w$:

$$D_w = \sum_{n \in R_w} D_{n,w} = D_w \exp(\frac{\mu_w}{\lambda_P}) \sum_{n \in R_w} \exp[\frac{1}{\lambda_P}(c_{n,w} + \lambda_R I_{n,w})] \quad \forall w \qquad (4.27)$$

which leads to

$$\mu_w = -\lambda_P \log\{\sum_{n \in R_w} \exp[\frac{1}{\lambda_P}(c_{n,w} + \lambda_R I_{n,w})]\} = -\lambda_P I_w \quad \forall w \qquad (4.28)$$

Substituting (4.28) into (4.26) gives

$$D_{q,w} = D_w \frac{\exp[\frac{1}{\lambda_P}(c_{q,w} + \lambda_R I_{q,w})]}{\sum_{n \in R_w} \exp[\frac{1}{\lambda_P}(c_{n,w} + \lambda_R I_{n,w})]} \quad \forall q, w \qquad (4.29)$$

Now substitute (4.29) into (4.21.3):

$$\frac{1}{\lambda_M} \exp(\frac{\frac{D_w}{D_w^{sat}}}{1 - \frac{D_w}{D_w^{sat}}}) = c_w + \lambda_P I_w \quad \forall w \qquad (4.30)$$

which can be re-written as

$$D_w = D_w^{sat} \frac{\exp[\frac{1}{\lambda_M}(c_w + \lambda_P I_w)]}{1 + \exp[\frac{1}{\lambda_M}(c_w + \lambda_P I_w)]} \quad \forall w \qquad (4.31)$$

Combining (4.25), (4.29) and (4.31) gives

$$D_{k,q,w} = D_w^{sat} \frac{\exp(\frac{c_w + \lambda_P I_w}{\lambda_M})}{1 + \exp(\frac{c_w + \lambda_P I_w}{\lambda_M})} \frac{\exp(\frac{c_{q,w} + \lambda_R I_{q,w}}{\lambda_P})}{\sum_{n \in T_w} \exp(\frac{c_{n,w} + \lambda_R I_{n,w}}{\lambda_P})} \frac{\exp(\frac{c_{k,q,w}}{\lambda_R})}{\sum_{m \in R_{q,w}} \exp(\frac{c_{m,q,w}}{\lambda_R})}$$

which suggests that the VI problem (4.18) is equivalent to the equilibrium condition (4.15).

The feasible set $\Omega$, composed of linear and non-negativity constraints, is non-empty, closed, and convex. In addition, the saturated demand provides upper bounds for metropolitan area pair demand. Therefore, $\Omega$ is compact. It is also obvious that $H(\mathbf{X})$ is continuous. Following Facchinei and Pang (2003), at least one solution exists for the VI problem. In other words, the original ATUE has at least one equilibrium point.

If $H(\mathbf{X})$ is monotone, then the uniqueness of the VI solution will also be guaranteed. Proving the monotonicity is difficult given the highly non-linear form of $H(\mathbf{X})$. The challenge in examining the monotonicity is reminiscent of the difficulty in investigating the Jacobian matrix in the fixed point problem context. The uniqueness issue will be further explored in the following model application section.

### 4.2.3 Differences between ATUE and UTSUE

To obtain further insights into ATUE, it is worthwhile to draw some analogy between ATUE and equilibria in other transportation modes. In the first instance, one may seek to compare ATUE with urban transit user equilibrium, because of the shared nature of being schedule transportation services. Closer scrutiny implies that ATUE may bear more similarities with SUE in the urban traffic context. However, we are not aware of any previous attempts that employ the UTSUE concept to study congestible air transport system equilibrium, which may be attributed to several important departures of ATUE from UTSUE, as discussed below.

First, ATUE combines urban traffic performance-demand equilibrium and classic supply-demand equilibrium. Similar to UTSUE, ATUE contains infrastructure performance functions that define the physical relationship between flight traffic and delay. On the other hand, in contrast to UTSUE, market behavior is captured in ATUE: air carriers compete against each other and maximize profit by providing air service with certain price and quality. The competition effect and profit maximization behavior are implicitly reflected in the empirical fare and frequency models. ATUE also differs from UTSUE in that air travelers' interaction with aviation infrastructure is mediated through airlines. Airlines have the flexibility to adjust frequency and the number of passengers onboard in response to delays. In contrast, car-passenger rate is assumed predetermined and invariant to UTSUE. Travelers, as a consequence, can be viewed to have direct interaction with infrastructure in UTSUE. In effect, the passenger-airline-infrastructure interaction in ATUE is more analogous to the demand-supplier-performance taxonomy proposed in Florian and Gaudry (1980; 1983).

Second, ATUE involves a simpler network representation but more complex travel cost calculation. Treating a flight segment as a link in urban road networks, an air route will consists of only a limited number of links since it is very rare for an air traveler to make more than one stop in a trip. In addition, the number of links is relatively small because of the hub-and-spoke topology of airline networks. As a result of these two factors, the number of feasible routes for air travel between an O-D pair is relatively small compared to the number of potential paths in road networks. Therefore, identifying the set of "feasible" paths is relatively easy in ATUE compared to UTSUE, substantially reducing the computational burden in the passenger assignment process. On the cost side, route travel cost under UTSUE is often modeled as only a function of travel time, which in many instances is assumed additive of travel time on individual links. Air passengers' generalized travel cost, in contrast, involves not only total trip time, which is the sum of scheduled flight time, schedule delay, and congestion delay,[22] but out-of-pocket money to purchase the air ticket. In UTSUE, out-of-pocket money is considered only in the presence of tolls. Among these utility components, only scheduled flight time is additive by segment under our demand model specification. Schedule delay on a one-stop route is collectively determined by the max and min frequencies on the two segments, and fares

---

[22] In principle, airport access and egress time are also integral components in the total trip time. As our focus is airside congestion, they are assumed constant and not explicitly considered in the study.

for routes are not additive across links. Moreover, both frequency and fare are functions of airport delays, which depend upon flight traffic from many other segments.

Focusing on travel time, a further distinction can be made between ATUE and UTSUE. Travel time in UTSUE is positively related to passenger flow on a link, and in some extensions, modeled as an increasing function of traffic volume at a node. In other words, the diseconomies of traffic density manifest itself at both link and node levels. Under ATUE, diseconomies of traffic density only hold at the airport (node) level. On a segment, higher frequency reduces passenger schedule delay, and therefore total travel time. Economies, instead of diseconomies, of traffic density exist at the link level. This important distinction stems from the scheduled nature of air transportation system where the Mohring effect takes effect. Link travel delay, if it ever exists, would be the result of airspace congestion, the extent of which is often much smaller than at airports, at least in the US.

## 4.3 Solution algorithm

The solution algorithm is designed based on the fixed point formulation. To arrive at equilibrium passenger flows, airfare, flight traffic, and airport delay that are consistent with the fixed point system (4.2)-(4.7), the Newton-Raphson method seems the natural path. However, the complex model forms (especially the demand model) and the size of a reasonable air transportation network prevent us from doing so because evaluation and inversion of Jacobian becomes cumbersome and even impossible. With all the departures of ATUE from UTSUE as discussed above, many of the existing algorithms to solve for the equivalent VI will be computationally challenging to implement even with the two mild modifications. In light of these difficulties, we resort to a more conventional, Cobweb style heuristic algorithm to solve for the equilibrium system. The procedure is outlined below.

*Step 0*: Initialization.

> *Step 0.1*: Use mean yield values from the dataset for fare model estimation to generate initial airfare $\mathbf{P}^{(0)}$ over all routes;
> *Step 0.2*: Generate initial flight frequency $\mathbf{f}^{(0)}$.
> *Step 0.3*: Calculate the initial airport traffic $\mathbf{v}^{(0)}$ and delays $\mathbf{d}^{(0)}$ using frequency information in Step 0.2.
> *Step 0.4*: Apply airfare, segment frequency, and airport delay from Steps 0.1-0.3 to the demand model to produce initial passenger flow distribution on the network, $\mathbf{D}^{(0)}$.
> *Step 0.5*: Set iteration counter $n = 1$ and convergence tolerance $g = 0.01$.

*Step 1*: Update segment passenger flows $\mathbf{Q}^{(n)}$ using passenger route demand $\mathbf{D}^{(n-1)}$.

*Step 2*: If $n \geq 2$, then perform convergence check: if $\|(\mathbf{Q}^{(n)} - \mathbf{Q}^{(n-1)})/\mathbf{Q}^{(n-1)}\|_\infty > g$, then stop and report the solution; otherwise go to Step 3.

*Step 3*: Update segment flight frequency $\mathbf{f}^{(n)}$ using $\mathbf{Q}^{(n)}$ and $\mathbf{d}^{(n-1)}$.

*Step 4*: Update airport flight traffic $\mathbf{v}^{(n)}$ and delay $\mathbf{d}^{(n)}$ using $\mathbf{f}^{(n)}$.

*Step 5*: Compute airfare $\mathbf{P}^{(n)}$ from $\mathbf{D}^{(n-1)}$, $\mathbf{Q}^{(n)}$, $\mathbf{d}^{(n)}$.

*Step 6*: Perform 3NL demand generation and network loading based on $\mathbf{P}^{(n)}$, $\mathbf{f}^{(n)}$, $\mathbf{d}^{(n)}$.
Update passenger flows $\mathbf{D}^{(n)}$. Set $n = n + 1$.

## 4.4 Model application

### 4.4.1 Network setup

We consider a hypothetical air transportation network, shown in Figure 4.2, which has a circular form with radius $R = 400$ miles. The network has one hub city located at the center and $n = 50$ identical spoke cities, numbered 1, 2, … , 50, uniformly distributed along the circle. Each city has one airport. The network is fully connected. Each spoke-spoke air travel market has two routing options: non-stop or connecting at the hub. In contrast, only a direct route is available for travel between a spoke city and a hub city. The non-stop distance between two spoke cities, ranging from 50 miles (two spokes cities are neighbors) to 800 miles (two spoke cities are aligned with the hub), is calculated as

$$d_{ij}^n = 2R\sin\{\min(\frac{j-i}{n}\pi, (1-\frac{j-i}{n})\pi)\}, \quad \forall i, j, \ 1 \le i < j \le n \tag{4.32}$$
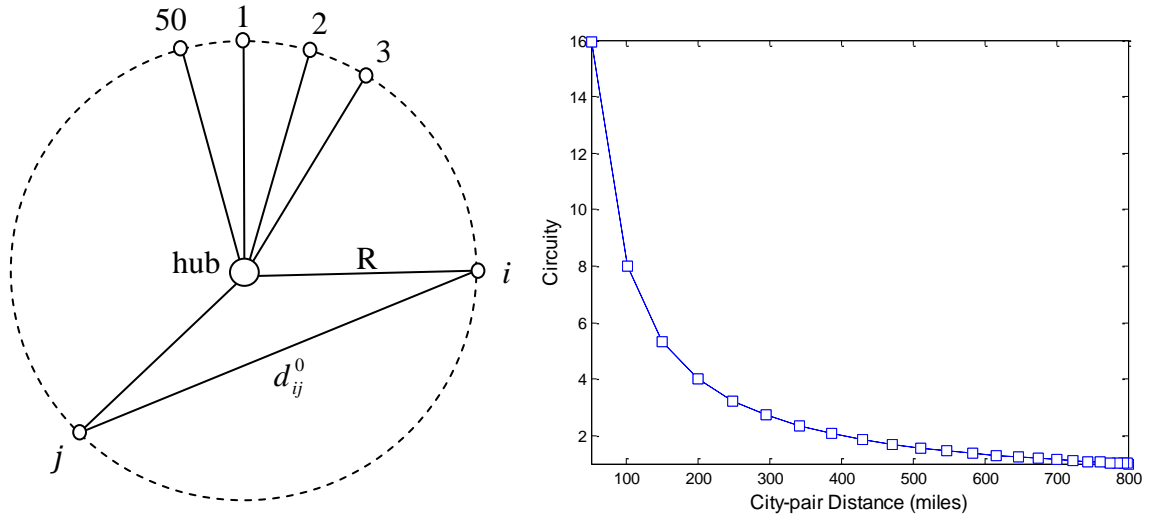


Figure 4.2: Network topology and circuity of one-stop routes

Because of the symmetric setup, all spoke-hub routes are identical and have a distance of 400 miles. Connecting routes in spoke-spoke markets are also identical, but the routing

66

distance is twice the radius (800 miles). The circuity of connecting routes, defined as the ratio between routing and airport-pair distances, depends upon the relative positions of the ending airports on the circle, and therefore the city-pair distance (Figure 4.2).

We assume the population in the hub city and each spoke city to be 10 million and 2 million respectively, roughly corresponding with the size of Chicago and Kansas city metropolitan areas (also note that the non-stop distance between ORD and MCI Airports is 403 miles, very close to the radius). All cities have the same personal income per capita, assumed to be 10,000 dollars each quarter (2004 values).

Values for other variable that are exogenous to the equilibrium process are determined using empirical data. Scheduled flight time on each segment, which is closely related to the segment distance, is calculated based on the following OLS regression results using BTS Airline On-time Performance Database (standard errors in parentheses):

$$(\text{Scheduled Flight Time/min}) = 39.9000 + 0.1205 \times (\text{Distance/miles}) \quad \text{Adj} R^2 = 0.9681$$
$$\phantom{(\text{Scheduled Flight Time/min}) = }\;\; (0.0001) \quad\;\; (0.0950)$$

$$(4.33)$$

Determination of concentration values is largely based on historic averages (2004-2008) in the US air transportation system. We assume spoke-hub and spoke-spoke RouteHHI's to be 0.6 and 0.8. We use the empirical relationship that market-level HHI is about 0.58 times non-stop RouteHHI (regression without constant, $R^2 = 0.87$) to construct spoke-spoke MarketHHI's. Because each spoke-hub market only has one, direct route, MarketHHI's on these routes are equal to the corresponding RouteHHI's. Similarly, as spoke-spoke segments serve exclusively non-stop passengers on the respective spoke-spoke routes, SegmentHHI's on these segments should be the same as the RouteHHI's. Based on results from regressing SegmentHHI on RouteHHI (again without constant, $R^2 = 0.95$), we let SegmentHHI of spoke-hub segments be 0.93 times RouteHHI. Considering that concentration at hub airports is often higher than at spoke airports, we assume concentration at each spoke airport to be 0.2 and 0.6 at the hub airport. For simplicity, any presence of LCC's, airport slot controls, and vacation routes is precluded.

Initial fares are calculated based on the sample means as shown in Tables 3.2 and 3.3 in Chapter 3. Initial flight frequency on spoke-hub segments is set as 1000 flights per quarter (~11 flights per day). As a point of reference, we observe that on ORD-MCI segment about 1100 flights were scheduled in each quarter between 2004 and 2008. Since spoke-spoke segments entail different distances, frequency on these segments is based on a reduced form regression using the same dataset for frequency model estimation. Regression results are presented in Appendix D. We assume airport AAR to be 1,000 operations per day at the hub, which is roughly the size of a medium-sized hub airport like Charlotte (Table 4.4), and 500 at each spoke airport. In the equilibrium analysis, we further assume that peakedness is 5 across all airports, and its change is proportional to airport traffic volume change (i.e. holding coefficient of variation constant). Since neither airport nor monthly dummies will be considered in the hypothetic network, finding reasonable values for the constant in the delay regression model is needed. In light of the

dummy estimates and after having experimented with an array of values, we take 1.5 and 1.2 as the constants for hub and spoke airports. Finally, values for the weather variables in the delay model take their respective means in the data sample.

Table 4.4 Daily AAR's and number of connections in our analysis and real world hub airports (based on 2007 average daily values).

|  | Daily AAR | Number of connections |
| --- | --- | --- |
| ORD | 1,533 | 134 |
| CLT | 1,121 | 62 |
| PHL | 822 | 49 |
| EWR | 708 | 83 |
| DFW | 1,869 | 127 |
| DEN | 1,892 | 105 |
| Hub airport in our analysis | 1,000 | 50 |

## 4.4.2 Initial equilibrium

The convergence to the equilibrium state from initial supply and demand values is achieved within 8 iterations. We have experimented with different starting values of demand which, as implied by the component model estimates, presents the key driver in the system equilibration process. Specifically, we take random draws from a uniform distribution between 0 and 10,000 as the starting values for passenger demand on each route. Such demand is then used to generate initial flight traffic,[23] airport delay, and airfare, whose values replace those from the *Step 0* in the algorithm as inputs for the subsequent steps. This procedure is repeated 500 times. Let vectors $\mathbf{Q}_{[0]}$ denote the equilibrium passenger flow precisely following the algorithm in Section 4.3; $\mathbf{Q}_{[k]}$ the equilibrium passenger flow from the $k$th experiment above ($k = 1, \ldots , 500$). We employ $\|(\mathbf{Q}_{[k]} - \mathbf{Q}_{[0]})/\mathbf{Q}_{[0]}\|_\infty$ to measure the deviation of experimental equilibrium flows from the baseline flow. The distribution of $\|(\mathbf{Q}_{[k]} - \mathbf{Q}_{[0]})/\mathbf{Q}_{[0]}\|_\infty$ over the 500 trials is illustrated in histogram in Figure 4.3. All values are less than 2%, part of which may be attributed to the convergence tolerance chosen ($g$, which equals 0.01). The results suggest that, at least from the segment passenger flow perspective, the equilibrium seems unique and insensitive to the start demand values.

Given the symmetric setup, all spoke-hub markets will share the same set of equilibrium values, shown in Table 4.5. Similarly, we would observe identical supply and demand characteristics in spoke-spoke markets with the same O-D distance. Thus it suffices to examines only the set of routes originating from one city to fully describe the system equilibrium. Figures 4.4-4.6 plot various equilibrium values, as a function of city-pair distance.

---

[23] Initial frequencies are calculated based on the structural rather than reduced form frequency model, in which we assume that delay is almost zero (to avoid meaningless logarithm we take 0.01).
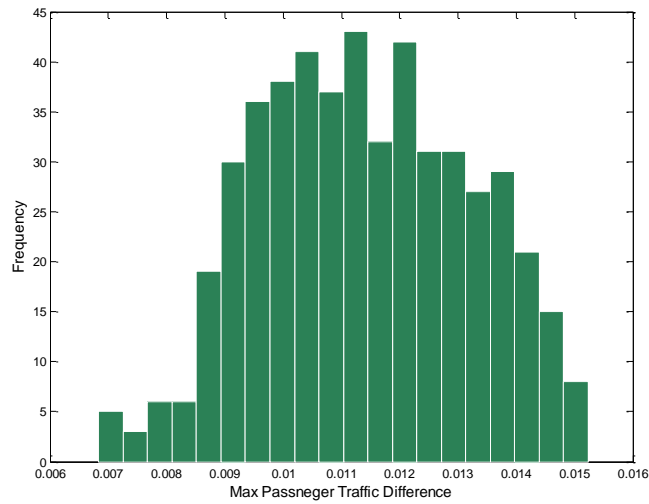
Figure 4.3: Distribution of $\|(\mathbf{Q}_{[k]} - \mathbf{Q}_{[0]})/\mathbf{Q}_{[0]}\|_{\infty}$

Table 4.5 Initial equilibrium values in spoke-hub markets.

| Component | Value |
| --- | --- |
| O-D demand (passengers/quarter) | 57,928 |
| Segment passenger volume (per quarter) | 155,410 |
| Airfare ($) | 154.8 |
| Yield ($/passenger-mile) | 0.387 |
| Segment frequency (flights/day) | 1545 |
| Number of passengers per flight | 101 |
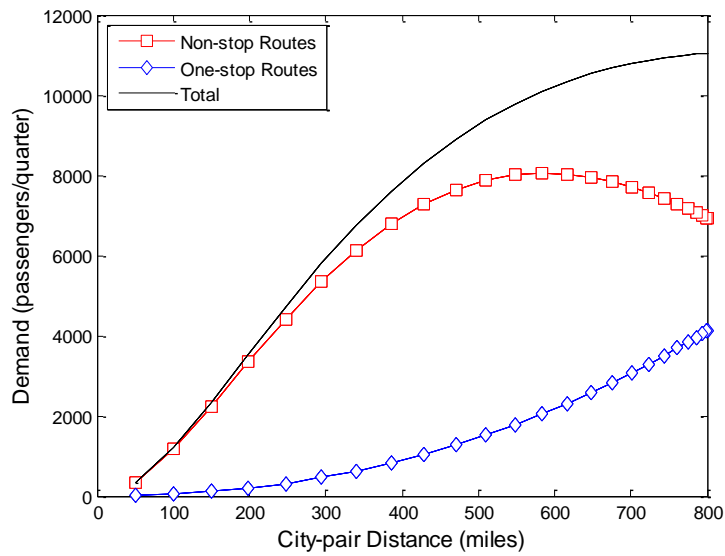


Figure 4.4: Passenger demand in spoke-spoke markets
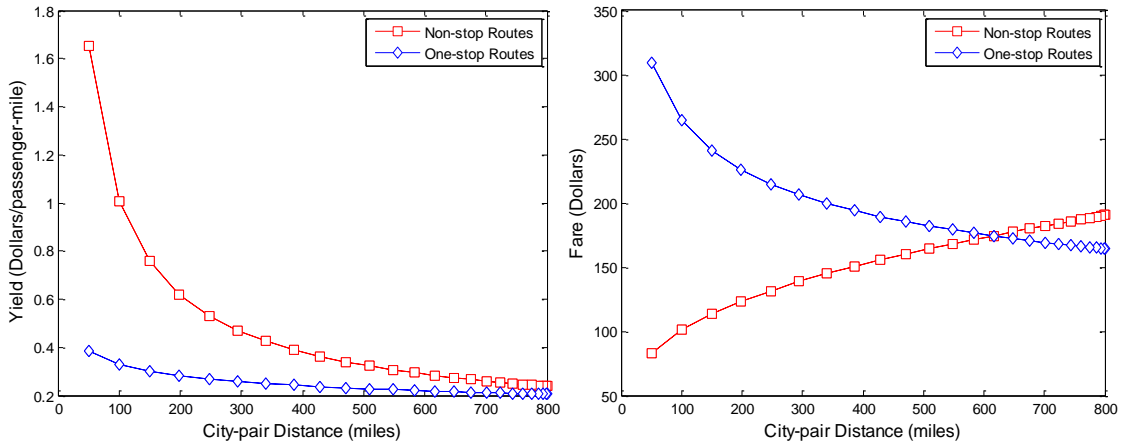
Figure 4.5: Yield and fare in spoke-spoke markets
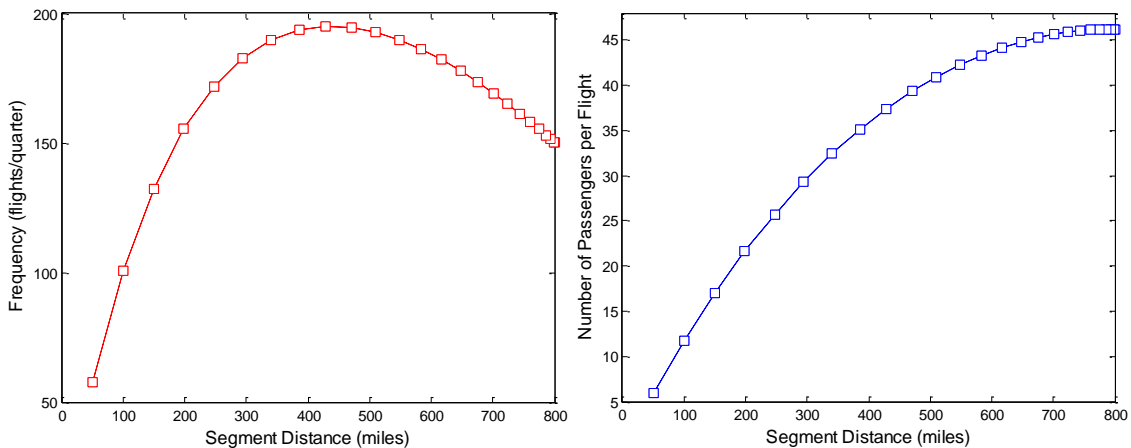


Figure 4.6: Flight frequency and number of passengers per flight in spoke-spoke markets

We first observe that passenger demand in spoke-hub markets, on average 643 travelers per day, is about five times the largest value in spoke-spoke markets, which occurs when the two spoke cities are aligned along one diameter of the circle. This demand discrepancy is largely due to the population difference between a hub and a spoke city (interestingly, the ratio is also five). For spoke-spoke travel, total demand is extremely low on the shortest-haul markets, where competition from surface modes, such as auto and rail, is most intense. As distance increases, air travel would be more advantageous, especially in travel time. Diminishing modal competition results in continuous growth of total passenger demand. However, the growth rate would start to attenuate beyond around 400 miles, where distance starts to create impedance for air travel.

Specific to passenger split in spoke-spoke markets, one-stop routes are very unattractive when the two cities are close, due to the high routing circuity as shown in Figure 4.4. Most travelers on these markets would choose direct service. Both one-stop and direct routing demand increase with distance, until around 600 miles, where much reduced circuity, together with much higher service frequency and comparable ticket price, start

70

to draw passengers from non-stop to connecting routes. In the longest-haul markets, connecting routes transport more than one third in the total number of passengers.

In spoke-spoke markets, changes in yield are primarily shaped by the negative coefficients for distance and—in the case of connecting routes—circuity in the fare models. The effects of route passenger demand and delay are only marginal due to their much smaller coefficients. The drastic declining yield curve for non-stop routes is directly associated with the distance effect; whereas the much moderate change for one-stop yield reflects the net outcome of the distance and circuity effects, the latter of which tends to increase yield with city-pair distance.

Airfare, the product of yield and route distance, manifests itself somewhat differently. Despite the declining yield, non-stop fare increases with city-pair distance, and almost linearly at longer ranges, where only marginal change in yield is observed and distance becomes the major driving force for fare change. On the other hand, as all one-stop routes have the same itinerary distance, the fare curve is essentially a scaling-up of its yield curve. In short-haul markets, because the connecting itinerary distance is much longer than the non-stop distance, remarkable discrepancies exist between fares on the two types of routes. With the continuous decline of one-stop fare and steady increase in non-stop fare, their difference is reduced. The distance of 600 miles marks the crossover point, beyond which connecting routes becomes even cheaper.

Frequency and the average number of passengers per flight vary greatly between spoke-spoke and spoke-hub segments. On the latter ones we observe very high frequency, because these segments serve not only local travelers but connecting passengers on 49 spoke-spoke one-stop routes. 17 flights per day operate on each spoke-hub segment. Each flight has 101 passengers onboard. In contrast, each spoke-spoke segment only serves local non-stop passengers. Not surprisingly, frequency is much lower. The shape of the frequency curve strongly depends upon non-stop passenger volume, with its peak achieved before the point where passenger volume reaches its maximum, because distance dampens the increase in frequency. Consequently, as distance increases, so does the number of passengers per flight. This explains why at longer distances, the number of passengers per flight stays almost constant even with shrinking segment passenger volume.

We compare the above equilibrium values with those observed from a number of U.S. air travel markets with similar population size and distance, some of which are documented in Tables 4.6 and 4.7 (averaged quarterly value between 2004 and 2008). For comparison, we also list equilibrium values for a set of selected segments/markets from the hypothetical network with similar distances (in italics). The equilibrium values obtained from the hypothetical air transportation network are, by and large, in line with the empirical observations. However, the hypothetical network under study has a denser distribution of spoke cities (recall that two neighboring cities are only 50 miles away), passenger consolidation then results in higher hubs-spoke segment passenger traffic. On the other hand, multiple connecting airport choices in the real world contribute to higher total market demand and a greater share of passengers using one-stop routes. The

difference in extent of direct and indirect competition, and the presence of LCC's (which is not assumed in this study), may also explain the differences between the observed and computed fare and yield.

Table 4.6 Supply-demand characteristics of selected segments.

| Segment | Origin population (millions)* | Destination population (millions) | Passengers per quarter | Distance (miles) | Flights per quarter | Passengers per flight |
|---|---|---|---|---|---|---|
| Spoke → Spoke | | | | | | |
| OMA→MCI | 0.86 | 2.06 | 1126 | 152 | 101 | 10 |
| GSO→PIT** | 1.48 | 2.46 | 6460 | 332 | 304 | 19 |
| SDF→RDU | 1.38 | 1.66 | 5784 | 416 | 218 | 27 |
| MCI→PIT | 2.06 | 2.48 | 6941 | 773 | 136 | 49 |
| RDU→MSY | 1.66 | 1.24 | 5682 | 779 | 158 | 36 |
| *Segment 1*** | *2.00* | *2.00* | *2233* | *150* | *132* | *17* |
| *Segment 2* | *2.00* | *2.00* | *6779* | *385* | *194* | *35* |
| *Segment 3* | *2.00* | *2.00* | *7165* | *775* | *155* | *46* |
| Spoke → Hub | | | | | | |
| MSY→ATL | 1.22 | 5.44 | 126087 | 425 | 1169 | 108 |
| RDU→ATL | 1.58 | 5.44 | 145409 | 356 | 1365 | 108 |
| MCI→ORD | 2.06 | 9.63 | 103960 | 403 | 1146 | 91 |
| PIT→ORD | 2.46 | 9.63 | 85605 | 412 | 1250 | 70 |
| OMA→ORD | 0.85 | 9.63 | 69276 | 416 | 1064 | 65 |
| *Segment 4 (spoke→hub)* | *2.00* | *10.00* | *155410* | *400* | *1545* | *101* |

* Population is calculated based on Metropolitan Statistical Areas (MSA).
** We treat PIT as a spoke airport as substantial debubbing of US Airways at the airport started in 2004 (Smith et al., 2006).
*** Italic segments are ones chosen from the hypothetical network.

Besides the segment and route characteristics, the equilibrium is featured by substantial delay at the hub and more moderate delay at the spoke airports, on average about 27.3 min/flight and 11.5 min/flight respectively (first column in Table 4.8). The exorbitant delay at the hub is largely due to connecting traffic, which accounts for two thirds in the total, which is also comparable to observed values in real hub airports in the US (Table 4.9) Compared to real world airport delays, the delay numbers do not seem unreasonable. According to Ball et al. (2010), the system-wide average delay against schedule (schedule padding not included) in 2007 reached 15 min/flight in the US air transportation system, and over 20 min at some of the busiest hub airports (e.g. ORD, JFK, EWR).

The equilibrium results suggest that the hub airport suffers severe capacity constraints and delay. It may seem a sensible decision to increase capacity at the hub. In what follows, we investigate how such capacity investment would mitigate delay, and trigger the system equilibrium shift.

Table 4.7 Supply-demand characteristics of selected markets.

| | O-D distance (miles) | Circuity | Passengers per quarter | Fare (dollars) | Yield ($/ passenger-mile) |
|---|---|---|---|---|---|
| Spoke-spoke markets* | | | | | |
| OMA→MCI | 152 | 1 | 408 | 82 | 0.54 |
| | | 5.39 | 1 | 288 | 0.35 |
| GSO→PIT | 302 | 1 | 2032 | 142 | 0.47 |
| | | 1.60 | 1210 | 163 | 0.34 |
| SDF→RDU | 416 | 1 | 4430 | 118 | 0.28 |
| | | 1.54 | 3968 | 110 | 0.18 |
| MCI→PIT | 773 | 1 | 3033 | 148 | 0.19 |
| | | 1.29 | 3203 | 134 | 0.14 |
| RDU→MSY | 779 | 1 | 4815 | 140 | 0.18 |
| | | 1.19 | 6082 | 108 | 0.12 |
| *Market 1*** | *150* | *1* | *2233* | *114* | *0.76* |
| | | *5.34* | *106* | *241* | *0.30* |
| *Market 2* | *385* | *1* | *6779* | *150* | *0.39* |
| | | *2.08* | *820* | *194* | *0.24* |
| *Market 3* | *775* | *1* | *7165* | *188* | *0.24* |
| | | *1.03* | *3840* | *165* | *0.21* |
| Spoke-hub markets | | | | | |
| MSY→ATL | 425 | 1 | 32384 | 131 | 0.31 |
| RDU→ATL | 356 | 1 | 47988 | 118 | 0.33 |
| MCI→ORD | 403 | 1 | 41499 | 88 | 0.22 |
| PIT→ORD | 412 | 1 | 33414 | 100 | 0.24 |
| OMA→ORD | 416 | 1 | 17357 | 105 | 0.25 |
| *Market 4 (spoke→hub)* | *400* | *1* | *57928* | *155* | *0.39* |

* For each spoke-spoke market, the first line denotes characteristics on the non-stop route; The second on the one-stop route, where circuity, fare, and yield are passenger weighted averages across all connecting routes.
** Italic markets are ones chosen from the hypothetical network.

Table 4.8 Airport delay changes.

| Component | Before | After | Change |
|---|---|---|---|
| Average delay at the hub (min/flight) | 27.3 | 18.4 | -8.6 |
| Average delay at the spoke (min/flight) | 11.5 | 11.3 | -0.2 |

Table 4.9 Proportion of connecting passengers at some major US hub airports (averaged value between 2004 and 2008).

| | ATL | ORD | CLT | DEN | DFW | IAH |
|---|---|---|---|---|---|---|
| Delay (min/flight) | 0.654 | 0.550 | 0.724 | 0.501 | 0.602 | 0.601 |

### 4.4.3 Equilibrium shift in response to capacity expansion

We consider a scenario that increases hub airport arrival capacity by 50%. Engineering practice has shown that the extent of capacity enhancement does not necessarily imply proportional change in the size of infrastructure (e.g. number of runways). Significant increase in airport capacity, in effect, can be achieved through moderate physical infrastructure investment with airspace reconfiguration and new air traffic control procedures.

The equilibrium shift proceeds as follows. We update hub airport delay with the new airport capacity, and subsequently fare and passenger route demand. We then iterate on Steps 1-6 in the equilibration algorithm until the convergence criterion is satisfied. The equilibrium shift is completed in 9 iterations. New airport delay and changes are reported in the 2nd and 3rd column of Table 4.8. Figures 4.7-4.9 illustrate the equilibrium shift in spoke-spoke markets and segments. Changes of supply and demand characteristics in spoke-hub markets and segments are documented in Tables 4.10.
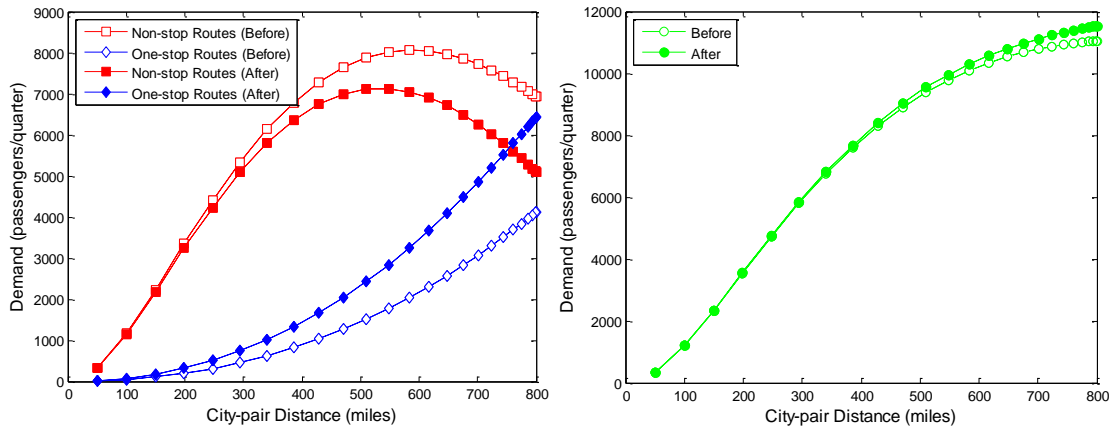


Figure 4.7: Route- (left) and market- (right) specific
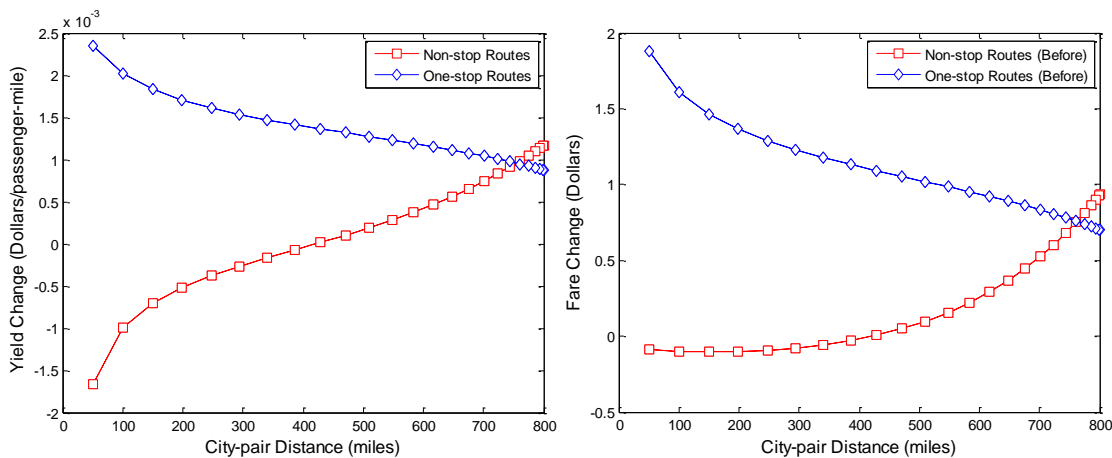demand change in spoke-spoke markets



Figure 4.8: Fare and yield change in spoke-spoke markets
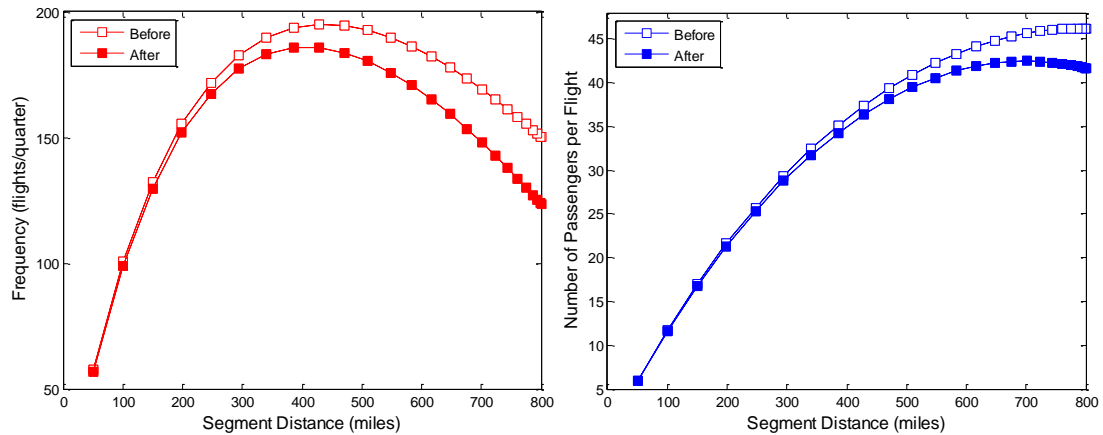
Figure 4.9: Changes in frequency and number of passengers
per flight in spoke-spoke markets

Table 4.10 Changes in spoke-hub markets and segments.

| Component | Before | After | Change |
|---|---|---|---|
| O-D demand (passengers/quarter) | 57,928 | 70,380 | 12,452 |
| Segment passenger volume (per quarter) | 155,410 | 226,190 | 70,780 |
| Airfare ($) | 154.0 | 150.3 | -3.7 |
| Yield ($/passenger-mile) | 0.387 | 0.376 | -0.011 |
| Segment frequency (flights/day) | 1545 | 2004 | 459 |
| Average number of passengers per flight | 101 | 113 | 12 |

As expected, the direct consequence of hub capacity increase is hub delay reduction, by 8.6 min per flight on average, or 32% of the original delay level. Delay at the spoke airports stays almost unchanged. Hub delay reduction improves the service quality of connecting routes in spoke-spoke markets, diverting some passengers from direct routes. This diversion effect is more prominent in longer-haul markets where perceived utilities between the two routing choice are closer, as reflected by the more equal demand split under the initial equilibrium. Passengers' routing choice on these markets are more sensitive to a given utility change. Furthermore, on almost all markets, reduced hub delay makes the option of air travel more attractive, suggesting higher demand would be generated in the air transportation system. One exception is the shortest spoke-spoke markets, where the ratio between perceived utilities from choosing the direct and connecting routes is the highest. After the equilibrium shift, frequency drop on the spoke-spoke segment results in a utility decrease on the direct route, which exceeds the utility increase on the one-stop route. As a consequence, the inclusive value from the bottom level in the 3NL model is slightly smaller after capacity change, suggesting a lower total demand on those markets. While somewhat surprising, this finding is indeed very interesting and reasonable. One can imagine that, under alternative network topologies, such loss in passenger demand could be large for some markets. The market-level demand changes are shown in the 2nd panel of Figure 4.7. In spoke-hub markets, we

observe a more perceivable increase in passenger demand, primarily due to the much higher spoke-hub frequency after hub capacity expansion.

Higher routing demand and reduced delay at the hub lead to the hub-spoke yield and fare which are 1 cent and $3.7 lower than before. Yield change in spoke-spoke markets is much smaller. Note that Figure 4.8 depicts changes in fares rather than their absolute values. On non-stop routes, lower passenger density tends to increase yield, which is offset by the fare reductions resulting from reduced delay. The net yield change, as shown in the left panel of Figure 4.8, can be either positive or negative. Yield change on connecting routes depends further upon increased segment passenger density, which, because of the economies of density, allows airlines to offer lower fare. Furthermore, higher passenger routing demand implies higher yield on one-stop routes. The overall effect of routing demand and segment density increase, combined with delay reduction, is still a positive yield change, especially at short distances where route demand increases the most. Nonetheless, the absolute yield change is less than 0.25 cents even on the most sensitive routes. Similar changing patterns are observed in fare. But the magnitude is dwarfed on short-haul non-stop routes due to much smaller distances.

Consistent with the passenger diversion, airlines would cut frequency in spoke-spoke segments, more significantly on longer-distance segments (18% reduction). The number of passengers on each plane would decrease concurrently, but to a less extent in terms of percentage change, as suggested by the frequency model. Diverted passengers would be accommodated by higher flight traffic and more passengers on the spoke-hub segments. On average, five more flights will operate on each spoke-hub segment per day, with 12 more passengers boarded on each flight. Focusing on each spoke airport, the almost unchanged delay suggests that reduction in flight traffic across all spoke-spoke segments would be almost offset by the added flights on the single segment connecting the hub.

### 4.4.4 Passenger benefits from the expansion

The equilibrium shift brings benefits to travelers in the air transportation system. We use the change in consumer surplus to quantify passenger benefits. Given the Logit demand model structure, the natural choice to measure passenger benefits is logsum (Small and Rosen, 1981; Train, 2003). Specifically for our Three-level Nesting structure, expected consumer surplus for one air travel decision making of an individual[24] on metropolitan area pair market $w$, $E(CS_w)$, is:

$$E(CS_w) = \frac{1}{(\zeta_1^1 + \zeta_2^1 (\text{market distance})_w)\zeta_{\text{logsum}}^2 \zeta_{\text{price}}^3} \log[1 + \exp(\zeta_{w,0}^1 + \zeta_1^1 I_w + \zeta_2^1 I_w (\text{market distance})_w)] + C \qquad (4.34)$$

---

[24] Recall we assume that each individual has a maximum number of potential trips for a given market and quarter. For each of these trips, air travel decision making has to be made as to choose whether to travel by air, and if so, the specific route.

where $\zeta_1^1$ and $\zeta_2^1$ denote the coefficients for the inclusive value at the top level, and the interaction term between that inclusive value and market distance; $\zeta_{w,0}^1$ the sum of perceived utilities other than the two aforementioned terms; $\zeta_{\text{logsum}}^2$ and $\zeta_{\text{price}}^3$ the logsum coefficient at the middle level, and fare coefficient at the bottom level. The division by $(\zeta_1^1 + \zeta_2^1 (\text{market distance})_w) \zeta_{\text{logsum}}^2 \zeta_{\text{price}}^3$ translates utility into dollars based on the fare coefficient and the scaling effect associated with the nesting structure. $C$ is an unknown constant that represents the fact that the absolute level of utility cannot be measured.

The difference in $E(CS_w)$ before and after the hub capacity investment then quantifies an individual's benefit gain from one decision making in market $w$:

$$\Delta E(CS_w) = E(CS_w^{after}) - E(CS_w^{before})$$

$$= \frac{1}{(\zeta_1^1 + \zeta_2^1 (\text{market distacance})_w) \zeta_{\text{logsum}}^2 \zeta_{\text{price}}^3}$$

$$\{ \log[1 + \exp(\zeta_{w,0}^1 + \zeta_1^1 I_w^{after} + \zeta_2^1 I_w^{after} (\text{market distacance})_w)]$$

$$- \log[1 + \exp(\zeta_{w,0}^1 + \zeta_1^1 I_w^{before} + \zeta_2^1 I_w^{before} (\text{market distacance})_w)] \} \qquad (4.35)$$

Since all individuals in market $w$ are considered in the air travel demand modeling with the choice between air travel and outside goods made at the top level, the computed logsum change will capture the full impact of capacity investment on consumer surplus for the entire population, regardless of whether air travel is chosen. Each individual, when making one air travel choice, will receive on average 3.40 cents benefits on the spoke-hub markets. The unit passenger benefits are much smaller in spoke-spoke markets, as illustrated in Figure 4.10. We observe that the unit benefit gains are an increasing convex function of market distance, in which the highest consumer surplus gains, about 0.35 cents per air travel choice, occurs when traveling between any two farthest points in the network. Consumer surplus gains for short-distance travel are much smaller, because the utility discrepancy between non-stop and connecting routes is greater. This discrepancy is reflected and further exaggerated in the sensitivity of the inclusive values because of the exponential form in the logsum term. As a consequence, inclusive values are more sensitive to utility reduction on a non-stop route than to an equal amount of utility increase on the corresponding connecting route. We find that the former effect even dominates the latter on the shortest-haul markets, i.e. the change in inclusive values is negative. This implies that capacity investment leads to not only uneven distribution but even loss of benefits to travelers in part of the air transportation network. The benefit loss is also reflected in the reduction in total demand on these markets, as we have seen in Section 4.4.3. If the reduction in total market demand is large, then the associated loss in traveler benefit could be significant.

The counter-intuitive results of consumer surplus loss are reminiscent of the "Braess Paradox" which suggests that infrastructure investment may worsen travelers' surplus under user equilibrium. However, two differences between the two phenomena are worth

pointing out. First, the Braess Paradox is conventionally elucidated under deterministic user equilibrium with fixed total demand; our ATUE problem assumes demand is stochastic and variable in total. Second, the benefit loss in ATUE derives from reduced traffic on links where there are economies of scale; whereas in Braess Paradox benefit loss comes from increased traffic on links which feature diseconomies of scale.
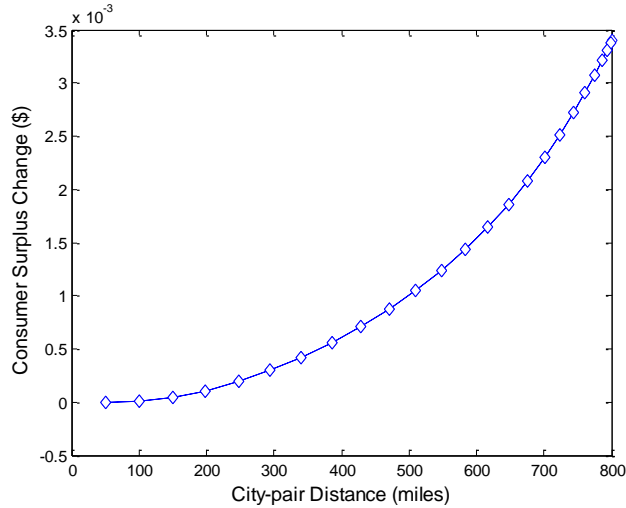


Figure 4.10: Average consumer surplus change per air travel
decision making in spoke-spoke markets

To quantity the overall passenger benefits, the consumer surplus changes computed above are aggregated across markets, and summed over all air travel decision makings and the entire population in the air transportation system. Table 4.11 reports the total traveler welfare gain for one quarter, amounting to 226.6 million dollars.

The aggregate consumer surplus gains can be alternatively obtained using the "rule-of-half" as an approximation. Beginning from the initial equilibrium state, we first compute the equivalent fare change in order to achieve the same market-level demand under the new equilibrium, for each market. This is done by either increasing (or decreasing) the original fare level on the one- and non-stop routes by the same amount, or by the same percentage on each market. After solving for the equivalent fare change and then updating route airfare, the "equivalent" market-level fare after equilibrium shift, which is the average of route fares weighted by route passenger volume, are computed. The consumer surplus gains in market $w$ is given by:

$$(P^0_{market,\,w} - P^1_{market,\,w})(Q^0_{market,\,w} + Q^1_{market,\,w})/2 \tag{4.36}$$

where $P^1_{market,\,w}$ and $P^0_{market,\,w}$ denote the "equivalent" market-level airfare after equilibrium shift and the original average market-level airfare, weighted by original route passenger demand; $Q^1_{market,\,w}$ and $Q^0_{market,\,w}$ are market-level passenger demand before and after the

78

equilibrium shift. These results are then aggregated across markets to obtain system-wide estimates of traveler benefits, which are very close to the passenger welfare gains using the logsum measure (Table 4.11).

Table 4.11 System-wide traveler benefit gains ($million/quarter).

| Logsum | Rule-of-half (same absolute price change) | Rule-of-half (same percentage price change) |
|---|---|---|
| 226.6 | 247.1 | 246.1 |

We also compare hub delay saving and passenger benefit estimates from the ATUE with those under the conventional approach. As discussed in Chapter 2, that approach assumes that, except for airport capacity and delay, everything else would remain unchanged before and after investing in capacity. The calculation of passenger benefit gains under the conventional method is based purely upon the estimate of average flight delay savings, which are converted into dollars by multiplying by passenger value of travel time and the number of passengers on each flight.

The steady state assumption suggests that the results from solving the ATUE represent the long-run equilibrium. Instead of looking at how equilibrium would shift on a year-by-year basis, it may be more appropriate to consider a set of isolated, representative time points. Given the same scenario that hub capacity is increased by 50%, we consider equilibrium states in the base year, and in 5 and 10 years, and compare estimates of delay savings and passenger welfare gains from equilibrium and conventional approaches. We use the US-based forecast (World Bank, 2012; FAA, 2012a) to derive the projected income, population, and flight traffic growth (Table 4.12).[25] The income and population projections will be used to compute ATUE; whereas the flight traffic projection will serve as the input to derive future delay and delay saving estimates.

Table 4.13 shows the hub delay estimates for the three years, with and without capacity investment, using the conventional and equilibrium approaches. All are measured as average delay minutes per flight. Under the conventional approach, base year flight traffic is obtained from the equilibrium results. In 5 and 10 years, hub delay is then computed using the projected flight traffic growth (i.e. 5 and 11%) and the delay model presented in Section 4.2.1.2, with given airport capacity. Absent any bounding forces as in the equilibrium process, the conventional method produces increasingly higher levels of future delay than the equilibrium approach if no capacity investment is made (1st vs. 4th lines). With capacity investment, hub delay will be reduced more drastically (3rd vs. 6th lines). This is because the conventional method only considers the physical relationship among delay, flight traffic, and airport capacity, and does not account for passenger diversion between different routes and induced demand, both of which counteract the hub delay reduction.

---

[25] We use historic data from World Bank (2012) as the assumed population growth rate. Real GDP and flight operation projections are provided by FAA (2012a). Using the population and GDP information we derive the growth rate of GDP per capita, which we assume is proportional to income per capita increase.

Table 4.12 Income, population and flight traffic
growth compared to the base year.

|         | Income per capita | Population | Flight Traffic |
|---------|-------------------|------------|----------------|
| Year 5  | 8%                | 5%         | 5%             |
| Year 10 | 16%               | 10%        | 11%            |

Table 4.13 Hub delay and delay saving estimates under the
conventional and equilibrium approaches.

|              |                    | Base year | In 5 years | In 10 years |
|--------------|--------------------|-----------|------------|-------------|
| Conventional | Without investment | 27.3      | 33.5       | 43.1        |
|              | With investment    | 12.4      | 14.3       | 16.9        |
|              | Savings            | 14.9      | 19.2       | 26.2        |
| Equilibrium  | Without investment | 27.3      | 30.2       | 33.2        |
|              | With investment    | 18.4      | 20.3       | 22.4        |
|              | Savings            | 8.9       | 9.9        | 10.8        |

Figure 4.11 shows the system-level passenger benefit estimates. We again employ logsum to measure the benefit under the equilibrium approach. If the conventional method is considered, passenger benefits would only cover those from the overestimated hub delay savings. We use the US Department of Transportation recommended value of passenger travel time, updated to 2004 values, to convert the delay savings into dollars. Despite the much greater delay saving estimate (even increasingly so in future years), the conventional method only produces about one third to half of total benefits from the equilibrium approach. This implies that the bulk of passenger benefits are reaped through other dimensions in the equilibrium shift than hub delay reduction. Schedule delay reduction on the spoke-hub segments, fare decrease in spoke-hub and certain non-stop spoke-spoke routes, and more passengers now using the air transportation system all contribute to the total benefits to passengers. Such results are consistent with those obtained in Chapter 2, although the air transport system equilibrium is modeled from quite distinct perspectives and with different network setups. The findings again highlight the importance of accounting for the systematic response in conducting investment benefit analysis. Failure to do so, as is likely the case in current engineering practice, will seriously bias the results and risk leading to wrong investment decisions.
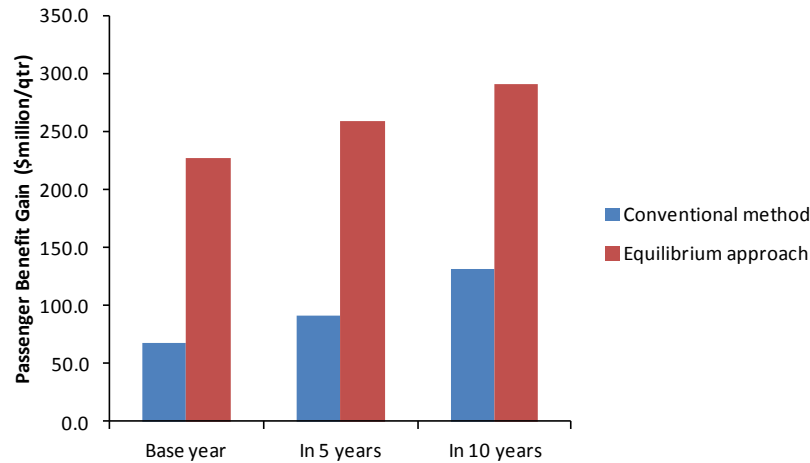
Figure 4.11: Passenger benefit estimates (in $million/quarter)
under the two approaches

## 4.4.5  Sensitivity analysis of capacity investment

So far the discussion of capacity investment has been focused on one scenario, increasing capacity by 50% at the hub airport. In the real world, decision makers are often faced with the question of how much capacity should be invested. While sometimes the question of where to invest may also be relevant, the answer is often the capacity constrained airport(s), which can be easily identified by looking at airport delay levels. In the following analysis, attentions are focused on how passenger benefits will be affected by the baseline capacity (and therefore the delay level) and the amount of investment.

Figure 4.12 shows the benefit gains (using logsum) under four baseline hub capacity levels (1000, 1300, 1600, 1900 operations per day), each of which subject to five capacity investment scenarios, escalating in a 200 increment from 200 to 1000 additional operations per day. Each point on the graph represents the incremental, rather than cumulative benefits. For instance, the point 600 on the 1300 curve means the difference in consumer surplus gains between investing to increase the capacity by 400 operations per day (i.e. from 1300 to 1700) and investing to increase the capacity by 600 operations per day (from 1300 to 1900). These curves reveal the inverse relationship between baseline capacity and investment benefits: greater capacity constraints at the airport, i.e. smaller baseline capacity, imply higher returns from the same amount of capacity investment. Focusing on each individual curves, the downward slopes suggest diminishing returns to capacity investment.

Note that the incremental benefits, albeit diminishing, are still significant in magnitude. One may be interested in the ultimate benefits from expanding the hub capacity to infinity. The consequent consumer surplus gain is $1141 million, about five times the benefits from increasing capacity by 50%. This somewhat surprising result, indeed, is associated with drastic network flow change. As shown in Figure 4.13, under the infinite-capacity equilibrium, higher total demand will appear, especially on long-haul markets, where the vast majority of travelers will choose connecting routes. Due to passenger diversion and consolidation, eliminating hub delay would increase frequency on the

81

spoke-hub segments by more than two-fold, with 37 daily flights—each now having 140 passengers onboard—operating on each segment. The drastic frequency increase represents an important contributor in the overall benefits. Airfare on the spoke-hub route would be reduced most substantially, $17 less per traveler, mainly due to the economies of density; in spoke-spoke markets, price change is much less significant. Some 4-min average delay will persist at the hub, which should be interpreted as the components of flight delay that would persist even with infinite airport capacity. The counterfactual analysis implies that, in the current air transportation network, continuous capacity investment could eventually lead to substantial change in air transport service supply and traffic patterns.
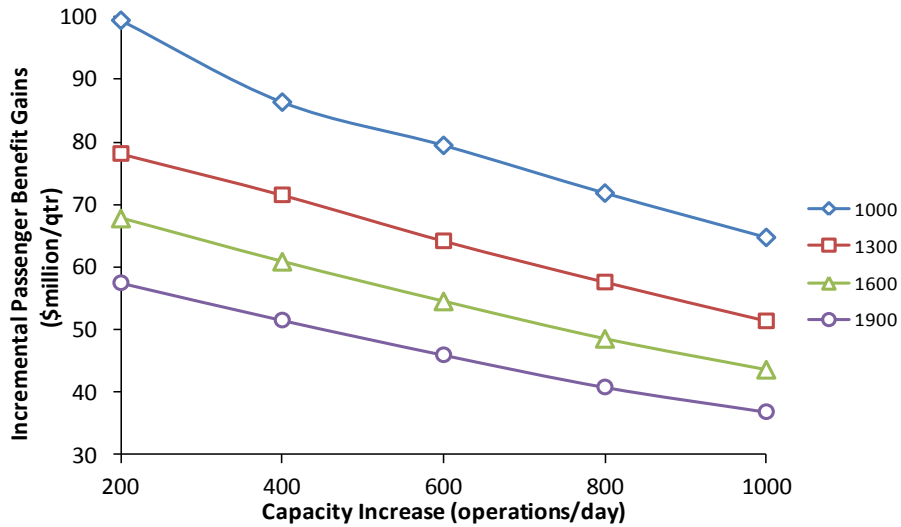


Figure 4.12: Sensitivity of passenger benefit estimates to baseline capacity and capacity increment (in $million/quarter)
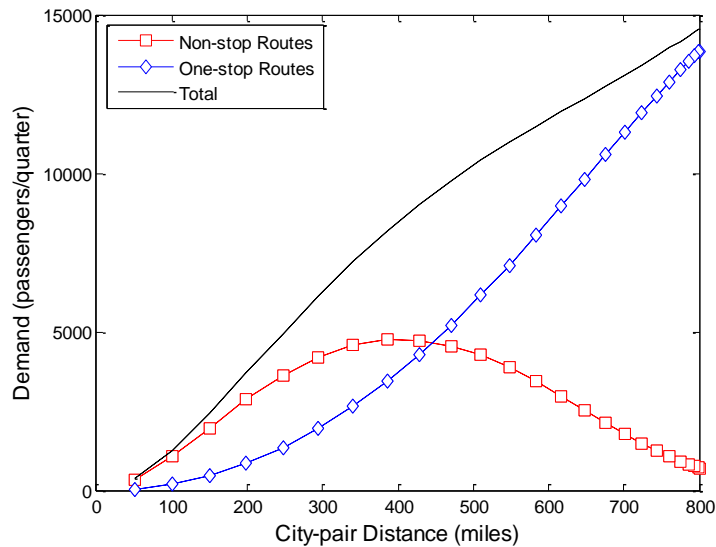


Figure 4.13: Passenger demand in spoke-spoke routes with infinite hub capacity

## 4.5 Summary

This chapter approaches the air transportation system equilibrium from a traveler-centric perspective while taking into consideration constraints and feedback from system supply. Each individual makes air travel decisions to maximize her/his utility, taking into consideration the level of supply, characterized by airfare, frequency, and airport delay, in the system. The passenger demand generation and route assignment process is modeled using a three-level Nested Logit decision structure. Taking together demand and supply-side constraints, the ATUE is formulated as both a fixed point and a variational inequality problem, with existence of the equilibrium solution guaranteed but not the uniqueness in general. Nevertheless, the equilibrium seems unique and insensitive to the starting values under the network examined in this chapter. The equilibrium values, while generated in a hypothetical network setting, are comparable to supply-demand characteristics observed in the real air transportation system.

We have investigated in detail the response of equilibrium to airport capacity investment. Reduced delay improves the hub connection quality, therefore diverting passengers from non-stop to one-stop routes, and attracting more travelers using the air transportation system. Airlines concurrently increase frequency and aircraft size in spoke-hub segments, while cutting back schedules and experiencing a drop in the number of passengers per flight in spoke-spoke segments. The magnitude of adjustment in flight frequency and aircraft size is consistent with the findings in the previous chapter that airlines prefer to adjust frequency in response passenger demand change. Passenger diversion and consolidation create further density on the spoke-hub segments, allowing local passengers to enjoy much higher service frequency. Compared to frequency increase, fare change in the non-stop markets is much smaller, and even more so in spoke-spoke markets.

The results suggest that while hub delay reduction triggers an equilibrium shift, the major underlying driver is frequency, or the Mohring effect. In the air transportation network, the Mohring effect is further exploited with each hub-spoke segment serving both local and connecting passengers on many routes. Passengers enjoy higher frequency not only from demand increase on the route they choose, but many other routes sharing the same segment. As only one routing option exists on each spoke-hub market where passenger schedule delay will be reduced most substantially, unit consumer surplus gain from hub capacity investment is much higher than for spoke-spoke O-D travel. Although one-stop, spoke-spoke routes also enjoy the same high frequencies, the market level unit consumer surplus gain is compromised by frequency decrease and utility reduction for non-stop routes. We find that when the non-stop route is extremely favorable in the baseline, adding hub capacity will slightly reduce the overall welfare for passengers on that market.

Nonetheless, the system-wide benefits to travelers from hub investment is considerable, and much greater than would be obtained under the conventional approach that assumed benefits come entirely from delay reduction. The conventional approach tends to overestimate delay savings, but some of delay reduction will be offset and transformed into higher flight traffic and more passengers using the airport. On the other hand, the

conventional approach would seriously underestimate the total passenger benefits, as the bulk of them comes from impacts other than delay reduction through the equilibrium shift. Such insights echoes the findings in Chapter 2 where the equilibrium is approached from the airline competition standpoint. Finally, our sensitivity analysis shows the decreasing returns to baseline capacity, and diminishing returns to capacity investment. The potential benefit from providing infinite capacity at the hub is much greater than those from any conceivable amount of capacity investment. These benefits are accompanied by significant network flow change, in particular, drastically higher frequency on spoke-hub segments.

The ATUE framework presented in this chapter provides an alternative avenue to the more classic airline competition modeling approach in understanding the air transport system equilibrium. Instead of looking at airline specific behavior, we model interactions between travelers and airlines at more aggregate—route, segment, and airport—levels. This approach therefore reduces computational burden and would be especially suitable for aviation analysts and decision makers whose interests focus on system performance as a whole rather than individual carriers. To improve this ATUE approach, future attempts may be directed in several areas. First, the equilibrium only implies the number of passengers onboard on each flight. It will be interesting to explicitly characterize load factor and aircraft size, which may need to take account of practical supply-side constraints, such as indivisibility of flights and limited aircraft types. For example, airlines may not be able to deploy aircraft with the most appropriate size on a segment given the passenger density but use the closest type available in the fleet. The break-even load factor requirement may also apply in choosing the aircraft type. Second, an important assumption made in the analysis is that the extent of competition is exogenous to the equilibrium process. Empirical investigation is clearly warranted in this area to explore any potential impact of capacity expansion on competition. Furthermore, it may be interesting to examine equilibrium under alternative network forms, such as adding asymmetry and allowing multiple hubs, and test the robustness of the equilibration algorithm and uniqueness of the solution. One particularly interesting modification could be to consider a network with partial connection, such that the possibility of introducing new service in the system can be further allowed.

# 5. Conclusions and Future Research

## 5.1 Conclusions

This research proposes equilibrium-based approaches to investigate benefit gains from aviation infrastructure capacity investment. It contributes to the existing literature by explicitly considering the flight delay effects on passenger demand, airline cost, airfare, and flight frequency, as well as the interplays among these system components in air transport system equilibrium and equilibrium shift in response to infrastructure capacity investment. In particular, we explicitly recognize that the change in service quantity comprises an integral part in total traveler benefits. At the flight segment level, service quantity is measured by flight frequency. As in the urban transit system, flight frequency is associated with a positive feedback effect (i.e. Mohring effect) on system demand. Another positive feedback is due to economies of density, i.e. higher density in a given network results in lower unit operating cost and therefore even higher density. Both feedback effects are strengthened through passenger consolidation in hub-and-spoke networks, but will be disrupted by the occurrence of flight delay at airports, i.e. diseconomies of node density in the network. Capacity-constrained system equilibrium derives from the competing forces of aforementioned two positive feedback loops and the diseconomies of node density—although a different equilibrium still exists absent the latter force.

Given the dual relationship between travelers and airlines in the air transportation system, we model the system equilibrium from two alternative perspectives. First, system equilibrium can be viewed as the result of airline competition. Each airline maximizes its profit by choosing the best pricing and scheduling strategies in a competitive environment, taking into consideration travelers' utility structure. Delay enters both traveler utility and airline cost functions. We find that the occurrence of delay leads to lower demand and higher passenger generalized cost, in spite of reduced airfare due to service degradation. To avoid excessive delays, airlines tend to use larger aircraft. The resultant cost savings from economies of aircraft size partially offset the delay-induced operating cost increase. As expected, the returns to economic welfare are diminishing to baseline capacity. Compared to the conventional benefit assessment method, the airline competition equilibrium model generates a lower estimate of airport delay savings because of induced demand. The delay saving and system benefit estimates are both biased under the conventional approach, because of its failure to recognize delay as a constraint on demand, and to capture welfare gains from schedule delay reduction and induced demand.

The second view of the air transport system equilibrium is user (i.e. traveler)-centric. Each traveler in the system maximizes her/his utility when making air travel decisions, with full knowledge about market supply and performance characteristics, such as airfare,

85

flight frequency, and flight delay. In this study, determination of these characteristics is based on the empirical models. In Chapter 3, we estimate econometric models for airfare and frequency, at route and segment levels respectively. Results show that high airport delay—either at origin, destination, or hub—leads to higher airfare. The fare increase should be interpreted as the net effect of airlines' tendency to pass delay cost to passengers while also compensating for service quality degradation. Higher airport delay discourages airlines from scheduling more flights on relevant segments. However, after controlling for passenger demand and market, route, airport, and segment structures, flight delay effects on airfare and flight traffic are fairly small. Two other component models are presented in the first part of Chapter 4. The first one is the airport delay model, in which we find that airport delay is strongly dependent on the ratio of flight traffic volume and airport capacity. A previously estimated three-level Nested Logit model is used to characterize air travel demand generation and passenger assignment process. In particular, high delay at a given hub is found to reduce the probability of one-stop routes through that hub being chosen, and the total market demand altogether. These component models form a comprehensive empirical basis for delay and congestion analysis in the air transport system.

With these component models, air transport user equilibrium is formulated as a fixed point problem in the second part of Chapter 4. The equilibrium has at least one solution based on the Brouwer's fixed point theorem; whereas the sufficient condition for the uniqueness is in general not guaranteed. With mild assumptions, we further show an equivalent variational inequality formulation, under which similar conclusions about solution existence and uniqueness are obtained. We then apply the air transport user equilibrium to a hypothetical network and solve for the equilibrium with a simple heuristic algorithm. The equilibrium convergence is robust to initial demand values, suggesting that there is a unique equilibrium for this particular model instance. Hub capacity investment attracts passengers from non-stop routes, and generates new travel demand on routes involving the hub airport. Total demand will increase in most spoke-spoke markets, except for the ones with the highest one-stop route circuity, where passengers on non-stop routes dominate in total market demand. This counter-intuitive result carries the important implication that capacity investment does not necessarily generate benefits across all markets. With changes in flight delay, schedule delay, airfare, and total demand, the user equilibrium model predicts much higher passenger benefits from capacity investment than the conventional method, despite a smaller estimate of hub delay saving. These findings are largely consistent with those under the airline competitive equilibrium.

## 5.2   Comparison of the two equilibrium approaches

One may raise the question of which of the two above avenues (airline competitive and user equilibrium) should be chosen in practice to analyze capacity-induced benefits in real air transportation networks. While in principle the two approaches should yield similar results, each has its advantages and disadvantages.

Compared to the user equilibrium model which primarily focuses on travelers' welfare gains, the airline competitive equilibrium provides explicit benefit estimates for both passengers and airlines. To apply the airline competitive equilibrium to real air transportation networks, the airline gaming behavior should be modeled on the scale of the entire network rather than individual routes. In real air transportation networks, a city pair may be served by multiple routes (direct and connecting via different hubs), each of which by multiple airlines. Therefore, modeling individuals' air travel choice will be more complicated than described in Chapter 2, necessitating the use of random utility framework—as in the user equilibrium model—but with a much larger set of route-airline choices. On the other hand, the airline competitive equilibrium model will yield detailed equilibrium information about passenger demand pattern, market supply characteristics, and flight delay, making possible airline- and route-specific analysis for capacity investment. The capability of analyzing the interactions among airlines will be especially useful when the consequence on competition of capacity investment is a concern. However, this capability has a price. Algorithmically, modeling the network-wide gaming behavior may require solving an array of large-scale profit maximization sub-problems, which substantially increases computation time for equilibration. In addition, existence and uniqueness of the equilibrium solution is often not guaranteed. Practical remedies to circumvent this potential issue includes introducing quasi-equilibrium, and starting the equilibration process from different strategy profiles (Hansen, 1990; Adler, 2001, 2005; Li et al., 2010).

Interactions among air carriers, while useful, are not of the primary concern of federal agencies, such as the FAA, who are the likely performer of system-wide aviation infrastructure investment analysis. In this regard, the user equilibrium approach may serve the benefit assessment purpose more efficiently, because it does not necessitate the direct modeling of airline gaming dynamics. Competition effects are reflected in the determination of airfare and flight frequency. Compared to the airline competition equilibrium, the existence of user equilibrium has been theoretically proven in the present study. The equilibration process only involves simple heuristics, therefore exhibiting considerable computational advantages. The improvement in computational efficiency will be especially attractive in simulating system response in large-scale networks. The user equilibrium approach also bears a higher compatibility with the existing NAS-wide simulation tools, and therefore integration will be more convenient. Nonetheless, the user equilibrium model also faces some potential issues. It assumes unchanged competition structure before and after capacity investment, to validate which requires further empirical evidence. Furthermore, airline profit change due to capacity investment has not been explicitly modeled in the current version of user equilibrium model. This might be resolved by introducing some aircraft specific cost components at the flight segment level. However, from a macroscopic vantage point, the airline industry is largely break even over the long run (Jiang and Hansman, 2006). We have conducted some preliminary time-series analysis using US quarterly aggregate data from 1991 to 2009, and found no significant effects of flight delay on industry profit. We speculate that, since delay has no effect on profit, neither does any attempt to reduce delay. While in the short term capacity investment reduces delay cost and increases airline profit, congestion will resume even faster as demand is spurred, causing airline new delay cost and profit loss,

the latter of which neutralizes the airline benefit gains right after the capacity investment. As a consequence, overall capacity may have no effect on producer surplus. Certainly, further investigation is deserved to confirm such speculation. Finally, results from both airline competitive and user equilibrium models need to be compared with reality, in order to assess the predictive capabilities of the two approaches. Such attempts, however, has to be left for future research.

## 5.3   Further research recommendations

The research presented in this thesis can be extended to several other areas. First, in the present research we assume a fixed network. In particular, the network is assumed fully connected in the user equilibrium model. One direction for future research is to consider endogenous network adaptation. In real air transportation networks, non-stop service may not serve some city pairs in the network due to low demand density; travel on these pairs has to pass through the hub airport. High hub delay may result in airlines deliberately avoiding the hub and introducing direct links between the cities. In a reverse manner, hub capacity investment reduces generalized cost of traveling on the one-stop routes. Passenger diversion may render the existing non-stop routes financially unviable and result in discontinuities in the network. In a multi-hub system, establishing new service could also mean opening up a flight link between a spoke city and a hub, which creates non-stop spoke-hub services as well as many one-stop spoke-spoke routing options related to the spoke city. The endogenous consideration of network structure requires introducing a more inclusive set of existing and potential routing choices, and threshold constraints on feasible services. Since the network is no longer predetermined, the existence of new system equilibrium also warrants re-examination.

Second, while the air transport equilibrium models in the our study are assumed in steady state in order to capture the long-run equilibrium behavior, with the lagged structure of the delay variables the equilibrium framework can be extended for investment decision making on both spatial and temporal dimensions. The spatial dimension relates to the questions of where and how much capacity to be invested; the temporal dimension determines the optimal timing for investment. Decision making tends to maximize total net benefit over a certain planning horizon, taking into consideration airline and passenger responses to capacity increase. A multi-stage approach may be appropriate for this modeling purpose. Since capacity investment often requires cost recovery, decision making can be further extended to determining the landing fees, and exploring how different landing fees would affect the system equilibrium.

Third, the timing issue further leads us to consider the general approaches to mitigate congestion. As mentioned in the outset of Chapter 1, besides capacity investment, demand management schemes, such as congestion pricing and slot control, represent another alternative. Our research suggests that capacity investment brings double dividend, i.e. reduced flight delay against schedule and schedule delay; whereas demand management often forces airlines to move flights from peak to off-peak hours, or cut flights from the original schedule. Therefore, delay reduction benefits from demand

management are realized at the expense of increased schedule inconvenience, i.e. higher schedule delay (Swaroop et al., 2012). On the other hand, due to its lumpy and expensive nature, major infrastructure capacity investment is performed on a decade time scale. Demand management, in contrast, incurs almost no implementation cost and is implemented on a daily basis. The best delay mitigation strategy may be a combination of both capacity- and demand-side solutions. Appropriate demand management strategies hinges upon the level of infrastructure capacity. On the other hand, continuous demand growth should prompt pricing signals (e.g. congestion fee, slot values) for capacity investment. The development of an integrated approach encompassing both capacity investment and demand management choices will be another very interesting area for future research.

Fourth, the investment decision making could be approached in a multimodal context. Intermodalism has been recognized as a promising means to effectively reduce airport congestion by shifting passengers away from airplanes (Resource Systems Group et al., 2010). To promote the development of an efficient intercity multi-modal transportation system, the provision of adequate infrastructure capacity for each modes is critical. The equilibrium framework in the thesis could be expanded to encompass other non-air modes, such as auto, intercity bus, and regular rail, and study the multimodal system equilibrium. Toward this end, the passenger demand model will need to be further specified to incorporate the characteristics of competing modes. Also indispensible are supply-side models for the non-air modes, which capture the different components in generalized travel cost in choosing each mode. The intermodal investment decision making will be especially relevant to the current debate and discussion on choosing between new high-speed rail lines and greater capacity at existing airport, and help inform future decision making.

Finally, decisions on capacity investment, or the design of integrated congestion mitigation strategies, should further include the negative environmental externalities, especially climate change impact, from aviation operations. Although capacity expansion reduces flight delay and associated $CO_2$ and other emissions, it also invites more traffic, creating much greater additional emissions. In contrast, demand management does not involve induced traffic. Therefore, total emissions will be reduced. Adding environmental externalities in the benefit analysis will make capacity expansion less favorable than demand management and defer the action of investment. Facing emission penalties, airlines may choose to use slower, more fuel efficient turboprops than jets, which directly affects scheduled travel time and airline operating cost. Exploring the extent to the environmental impact on system equilibrium will be useful in shaping future policies for both aviation infrastructure investment and environment protection.

# References

Abrahams, M., 1983. A service quality model of air travel demand: an empirical study. Transportation Research Part A: Policy and Practice 17 (5), 385-393.

Adler, N., 2001. Competition in a deregulated air transportation market. European Journal of Operations Research 129, 337-345.

Adler, N., 2005. Hub-spoke network choice under competition with an application to western Europe. Transportation Science 39 (1), 58-72.

Bailey, E., Graham, D., Kaplan, D., 1985. Deregulating the airlines. The MIT Press, Cambridge, Massachusetts and London, England.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B., 2010. Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. NEXTOR Report prepared for the Federal Aviation Administration.

Bamberger, G., Carlton, D., 2003. Airline networks and fares. In: Jenkins, D. (eds.), Handbook of Airline Economics, 2nd edition, McGraw Hill.

Berry, S., Carnall, M., Spiller, P., 1996. Airline hubs: costs, markups and the implications of customer heteroskedasticity. NBER Working Paper No. 5561.

Belobaba, P., 2009a. Fundamentals of pricing and revenue management. In: Belobaba, P., Odoni, A. and Barnhart, C. (eds.), The global airline industry, John Wiley & Sons, Inc.

Belobaba, P., 2009b. The airline planning process. In: Belobaba, P., Odoni, A. and Barnhart, C. (eds.), The global airline industry, John Wiley & Sons, Inc.

Boeing, 2011. Current market outlook: 2011-2030. <http://www.boeing.com/commercial/cmo/index.html> (accessed 03.12.12).

Borenstein, S., 1989. Hubs and high fares: dominance and market power in the U.S. airline industry. The Rand Journal of Economics 20 (3), 344-365.

Britto, R., Dresner, M., Voltes, A., The impact of flight delays on passenger demand and societal welfare. Transportation Research Part E: Logistics and Transportation Review 48 (2), 460-469.

Brueckner, J., Spiller, P., 1994. Economies of traffic density in the deregulated airline industry. Journal of Law and Economics 37 (2), 379-415.

Brueckner, J., Zhang, Y., 2001. A model of scheduling in airline networks: how a hub-and-spoke system affects flight frequency, fares and welfare. Journal of Transport Economics and Policy 35 (2), 195-222.

90

Brueckner, J., Flores-Fillol, R., 2007. Airline schedule competition. Review of Industry Organization 20, 161-177.

Brueckner, J., Girvin, R., 2008. Airport noise regulation, airline service quality, and social welfare. Transportation Research Part B: Methodological 42 (1), 19-37.

Bruckner, J., Zhang, A., 2010. Airline emission charges: effects on airfares, service quality, and aircraft design. Transportation Research Part B: Methodological 44 (8-9), 960-971.

Brueckner, J., Lee, D., Singer, E., 2011. Airline competition and domestic U.S. airfares: A comprehensive reappraisal. <http://www.socsci.uci.edu/~jkbrueck/price%20effects.pdf> (accessed 04.23.12).

Bureau of Economic Analysis (BEA), 2011. Statistical Areas—Metropolitan Statistical Areas (MSAs), Micropolitan Statistical Areas, Metropolitan Divisions, Combined Statistical Areas (CSAs). <http://bea.gov/regional/docs/msalist.cfm> (accessed 11.15.11).

Bureau of Transportation Statistics (BTS), 2012.Airline On-time Performance. <http://www.transtats.bts.gov/Oneway.asp?Display_Flag=0&Percent_Flag=0> (accessed 04.11.12).

Caves, D. W., Christensen, L. R., Tretheway, M. W., 1984. Economies of density versus economies of scale: why trunk and local service airline costs differ. Rand Journal of Economics 15 (4), 471-489.

Chi, J., Koo, W., 2009. Carriers' pricing behaviors in the United States airline industry. Transportation Research Part E: Logistics and Transportation Review 45 (5), 710-724.

Resource Systems Group, Coogan, M., Hansen, M., Ryerson, M., Kiernan, L., Last, J., Marchi, R., Yatzeck, R., 2010. Innovative approaches to addressing aviation capacity issues in coastal mega-regions. Airport Cooperative Research Program Report 31, Transportation Research Board, Washington D.C.

De Neufville, R., Odoni, A., 2003. Airport systems: planning, design and management. McGraw-Hill Companies, Inc., New York.

Douglas, G., Miller, C., 1974. Economic regulation of domestic air transport: theory and policy. The Brookings Institution.

Dresner, M. Trethway, M., 1992. Modelling and testing the effect of market structure on price: the case of international air transport. Journal of Transport Economics and Policy 26 (2), 171-184.

Dresner, M., Lin, J., Windle, R., 1996. The impact of low-cost carriers on airport and route competition. Journal of Transport Economics and Policy 30 (3), 309-328.

Else, P., 1985. Optimal pricing and subsidies for scheduled transport services. Journal of Transport Economics and Policy 19 (3), 263-279.

Eriksen, S.E., 1978. Demand models for U.S. domestic air passenger markets. Department of Aeronautics and Astronautics, MIT report No. FTL-R78-2, Cambridge, Massachusetts.

Evans, A., 2010. Simulating airline operational responses to environmental constraints. Ph.D. Dissertation, University of Cambridge.

Facchinei, F., Pang, J., 2003. Finite-dimensional variational inequalities and complementarity problems, vol. 1. Springer, New York.

Federal Aviation Administration (FAA), 1999. FAA Airport Benefit-Cost Analysis Guidance. Report to the Office of Aviation Policy and Plans. <http://www.faa.gov/ regulations_policies/policy_guidance/benefit_cost/media/1999_FAA_Airport_Benefit_C ost_Analysis_Guidance.pdf> (accessed 04.23.12).

Federal Aviation Administration (FAA), 2007a. Capacity needs in the National Airspace System: an analysis of airports and metropolitan area demand and operational capacity in the future. <http://www.faa.gov/airports/resources/publications/reports/media/fact_2.pdf> (accessed 04.11.12).

Federal Aviation Administration (FAA), 2007b. FAA long-range aerospace forecasts fiscal years 2020, 2025 and 2030. <http://www.faa.gov/data_research/aviation/long-range_ forecasts/media/long07.pdf> (accessed 04.07.12).

Federal Aviation Administration (FAA), 2008. Congestion Management Rule for John F. Kennedy International Airport and Newark Liberty International Airport; Final Rule. Federal Register 73 (198), 60544-60571.

Federal Aviation Administration (FAA), 2009a. Congestion Management Rule for LaGuardia Airport; Final Rule. Federal Register 74 (195), 52132-52134.

Federal Aviation Administration (FAA), 2009b. OEP 35. <http://aspmhelp.faa.gov/index. php/OEP_35> (accessed 04.10.12).

Federal Aviation Administration (FAA), 2012a. Fact Sheet—FAA Forecast. <http://www. faa.gov/news/fact_sheets/news_story.cfm?newsId=13395> (accessed 04.11.12).

Federal Aviation Administration (FAA), 2012b. The NextGen nucleus: transformational and supporting programs. <http://www.faa.gov/nextgen/slides/?slide=4> (accessed 04.23. 12).

Flores-Fillol, R., 2010. Congested hubs. Transportation Research Part B 44 (3), 358-370.

Florian, M. and Gaudry, M. 1980. A conceptual framework for the supply side in transportation systems. Transportation Research Part B 14 (1), 1-8.

Florian, M. and Gaudry, M. 1983. Transportation systems analysis: illustrations and extensions of a conceptual framework. Transportation Research Part B 17 (2), 147-153.

Forbes, S., 2008. The effect of air traffic delays on airline prices. International Journal of Industrial Organization 26, 1218-1232.

Frankena, M., 1983. The efficiency of public transport objectives and subsidy formulas. Journal of Transport Economics and Policy 17 (1), 67-76.

Ghobrial, A., Kanafani, A., 1995. Future of airline rubbed networks: some policy implications. Journal of Transportation Engineering 121 (2), 124-134.

Gillen, D., Oum, T., Tretheway, M., 1985. Airline cost and performance: implications for public and industry policies. Center for Transportation Studies, University of British Columbia, Vancouver, Canada.

Gillen, D., Oum, T., Tretheway, M., 1990. Airline cost structure and policy implications: a multi-product approach for Canadian airlines. Journal of Transport Economics and Policy 24 (1), 9-34.

Gillen, D., Morrison, W., Stewart, C. 2002. Air travel demand elasticities: concepts, issues and measurement. Report prepared for Department of Finance, Canada.

Gillen, D., Hazledine, T., 2011. The new price discrimination and pricing in airline markets: implications for competition and antitrust. Working Paper. University of British Columbia.

Givoni, M., Rietveld, P., 2009. Airline's choice of aircraft size – Explanations and implications. Transportation Research Part A: Policy and Practice 43 (5), 500-510.

Goolsbee, A., Syverson, C. 2008. How do incumbents respond to the threat of entry? Evidence from major airlines. Quarterly Journal of Economics, 123, 1611-1633.

Government Accountability Office (GAO). 2010. Integration of current implementation efforts with long-term planning for the next generation air transportation system. Report No. GAO-11-132R.

GRA, Inc., 2004. Economic Values for FAA Investment and Regulatory Decisions, a Guide. Report prepared for the FAA Office of Aviation Policy and Plans, Washington D.C.

Graham, D., Kaplan, D., Sibley, D., 1983. Efficiency and competition in the airline industry. The Bell Journal of Economics 14 (1), 118-138.

Hansen, M., Kanafani, A., 1989. Hubbing and airline costs. Journal of Transportation Engineering 116, 581–590.

Hansen, M., 1990. Airline competition in a hub-dominated environment: an application of non-cooperative game theory. Transportation Research Part B: Methodological 24 (1), 27-43.

Hansen, M., 1995. Positive feedback model of multiple-airport systems. Journal of Transportation Engineering 121 (6), 453-460.

Hansen, M., Weidner, T., 1995. Multiple airport systems in the United States: Current status and future prospects. Transportation Research Record: Journal of the Transportation Research Board, No. 1506, Transportation Research Board of the National Academies, Washington, D.C., 8–17.

Hansen, M., Gillen, D., Djafarian-Tehrani, R., 2001. Aviation infrastructure performance and airline cost: a statistical cost estimation approach. Transportation Research Part E: Logistics and Transportation Review 37 (1), 1-23.

Hansen, M., 2002. Micro-level analysis of airport delay externalities using deterministic queuing models: a case study. Journal of Air Transport Management 8 (2), 73-87.

Hansen, M., Hsiao, C., 2005. Going south? Econometric analysis of U.S. airline flight delays from 2000 to 2004. Transportation Research Record: Journal of the Transportation Research Board, No. 1915, Transportation Research Board of the National Academies, Washington, D.C., 2005, 85–94.

Hansen, M., Wei, W., 2006. Multivariate analysis of the impacts of NAS investment: a case study of a capacity expansion at Dallas-Fort Worth airport. Journal of Air Transport Management 12 (5), 227-235.

Hansen, M., 2010. Costing Delay in the National Airspace System. Presented at the National Aeronautics and Space Administration (NASA) Ames Research Center, Mountain View, California.

Hansen, M., Kwan, I., 2010. US flight delay in the 2000's: an econometric analysis. Proceedings of the 90th Transportation Research Board Annual Meeting.

Hofer, C., Windle, R., Dresner, M., 2008. Price premiums and low cost carrier competition. Transportation Research Part E: Logistics and Transportation Review 44 (5), 864-882.

Hsiao, C., 2008. Passenger demand for air transportation in a hub-and-spoke network. Ph.D. Thesis, University of California, Berkeley.

Hsiao, C., Hansen, M., 2011. A passenger demand model for air transportation in a hub-and-spoke network. Transportation Research Part E: Logistics and Transportation Review 47 (6), 1112-1125.

Hurdle, G., Johnson, R., Joskow, A., Werden, G., Williams, M., Concentration, potential entry, and performance in the airline industry. The Journal of Industrial Economics 18 (2), 119-139.

Jiang, H., Hansman, J., 2006. An analysis of profit cycles in the airline industry. Proceedings of the 6th AIAA Aviation Technology, Integration and Operations Conference, Wichita, Kansas, September 25-27.

Jorge, J., De Rus, G., 2004. Cost–benefit analysis of investments in airport infrastructure: a practical approach. Journal of Air Transport Management 10 (5), 311-326.

Jorge-Calderón, J. D., 1997. A demand model for scheduled airline services on international European routes. Journal of Air Transport Management 3 (1), 23-35.

Le, L., 2006. Demand management at congested airports: How far are we from utopia? Ph.D. dissertation, George Mason University.

Li, Z., Lam, W., Wong, S., Fu, X., Optimal route allocation in a liberalizing airline market. Transportation Research Part B: Methodological 44 (7), 886-902.

Martin, J., Socorro, M., 2007. A new ear for airport regulators through capacity investment. Transportation Research Part A: Policy and Practice 43 (6), 618-625.

Mayworm, P., Lago, A., McEnroe, J., 1980. Patronage impact of changes in transit fares and services. Report prepared for the Urban Mass Transportation Administration, U.S. Department of Transportation.

Meyer, J., Oster, C., 1984. Deregulation and the new airline entrepreneurs. The MIT Press, Cambridge, Massachusetts.

Miller, B., Clarke, J-P., 2007. The hidden value of air transportation infrastructure. Journal of Technological Forecasting and Social Change 44 (1), 18-35.

Mishan, E., 1977. The plain truth about consumer surplus. Journal of Economics 37 (1-2), 1-24.

Mohring, H., 1972. Optimization and scale economies in urban bus transportation. American Economic Review 62 (4), 591-604.

Morrison, S., Winston, C., 1983. Estimation of long-run prices and investment levels for airport runways. Research in Transportation Economics 1, 103-130.

Morrison, S., Winston, C., 1986. The economic effects of airline deregulation. The Brookings Institution.

Morrison, S., Winston, C., 1989. Enhancing the performance of the deregulated air transportation systems. Brookings Papers on Economic Activity: Microeconomics, pp. 61-112.

Morrison, S., 2001. Actual, adjacent, and potential competition estimating the full eoeect of southwest airlines. Journal of Transport Economics and Policy 35 (2), 239-256.

Morrison, S., Winston, C., 2007. Another look at airport congestion pricing. American Economic Review 97 (5), 1970-1977.

Morrison, S., Winston, C., 2008. The effect of FAA expenditures on air travel delays. Journal of Urban Economics 63 (2), 669-678.

Ortega, J., Rheinboldt, W., 1970. Iterative solution of nonlinear equations in several variables. Academic Press, New York.

Oum, T.H., Zhang, A., Zhang, Y., 1993. Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets. Journal of Transport Economics and Policy 27 (2), 171-192.

Oum, T., Zhang, A., Zhang, Y., 2004. Alternative forms of economic regulation and their efficiency implications for airports. Journal of Transport Economics and Policy 38 (2), 217-246.

Pai, V., 2010. On the factors that affect airline flight frequency and aircraft size. Journal of Air Transport Management 16 (4), 167-177.

Panzar, J., 1979. Equilibrium and welfare in unregulated airline markets. American Economic Review 69(2), 92-95.

Pels, E., Verheof, E., 2002. Airport pricing. Tinbergen Institute Discussion Paper. Department of Spatial Economics, Free University of Amsterdam.

Pels, E., Verheof, E., 2004. The economics of airport congestion pricing. Journal of Urban Economics 55, 257-277.

Post, J., 2006. FAA system-wide modeling: uses, models, and shortfalls. Presented at FAA/Eurocontrol TIM, Madrid, November, 2006.

Post, J. et al., 2008. The modernized national airspace system performance analysis capacity (NASPAC). Proceedings of the 26th International Congress of the Aeronautical Sciences, Septembe, Anchorage, Alaska, U.S.A.

Ramdas, K., Williams, J., 2008. An empirical investigation into the tradeoffs that impact on-time performance in the airline industry. Working paper, University of Virginia, Charlottesville.

Reynolds-Feighan, A., Button, K., 1999. An assessment of the capacity and congestion levels at European airports. Journal of Air Transport Management 5 (3), 113-134.

Richard, O., 2003. Flight frequency and mergers in airline markets. International Journal of Industrial Organization 21 (6), 907-922.

Schipper, Y., Rietveld, P., Nijkamp, P., 2002. European airline reform: An empirical welfare analysis. Journal of Transport Economics and Policy 36 (2), 189-209.

Schipper, Y., Rietveld, P., Nijkamp, P., 2003. Airline deregulation and external costs: a welfare analysis. Transportation Research Part B: Methodological 37 (8), 699-718.

Small, K., Rosen, H., 1981. Applied welfare economics with discrete choice models. Econometrica 49 (1), 105-130.

Smith, B., Hoel, L., Repaka, V., 2006. A comparative evaluation of air and auto travel between urban regions. Research Report No. UVACTS-14-5-97. Center for Transportation Studies, University of Virginia.

Steinbach, M., Giles, S., 2005. Developing a model for joint infrastructure investment. Technical paper. The MITRE Corporation.

Swaroop, P., Zou, B., Ball, M., Hansen, M., 2012. Do more US airports need slot controls? A welfare based approach to determine slot levels. Transportation Research Part B: Methodological, forthcoming.

Train, K., 2003. Discrete choice methods with simulation. Cambridge University Press, New York.

Turnovsky, S., Shalit, H., Schmitz, A., 1980. Consumer's surplus, price instability, and consumer welfare. Econometrica 48 (1), 135-152.

Varian, H., 1992. Microeconomic analysis, 3rd Edition. W.W. Norton & Company, New York.

Vaze, V., Barnhart, C., 2012. Modeling airline frequency competition for airport congestion mitigation. Transportation Science, forthcoming.

Wang, X., Zou, L., Dresner, M., 2009. Assessing the price effects of airline alliances on parallel routes. Transportation Research Part E 45 (4), 627-641.

Wei, W., Hansen, M., 2003. Cost economics of aircraft size. Journal of Transport Economics and Policy 37 (2), 279-296.

Wei, W., Hansen, M., 2005. 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.

Wei, W., Hansen, M., 2006. Multivariate analysis of the impacts of NAS investments: A case study of a capacity expansion at Dallas-Fort Worth Airport. Journal of Air Transport Management 12 (5), 227-235.

World Bank, 2012. Population projection tables by country and group. <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTDATASTATISTICSHNP/EXTHNPSTATS/0,,contentMDK:21737699~menuPK:3385623~pagePK:64168445~piPK:64168309~theSitePK:3237118~isCURL:Y,00.html> (accessed 04.04.12)

Zhang, A., Zhang, Y., 2006. Airport capacity and congestion when carriers have market power. Journal of Urban Economics 60, 229-247.

Zhang, Y., 2010. Network structure and capacity requirement: The case of China. Transportation Research Part E: Logistics and Transportation Review 46 (2), 189-197.

Zou, B., Hansen, M., 2012. Impact of operational performance on air carrier cost structure: Evidence from US airlines. Transportation Research Part E: Logistics and Transportation Review, forthcoming.

Zou, L, Dresner, M., Windle, R., 2011. Many fields of battle how cost structure affects competition across multiple markets. Journal of Transport Economics and Policy 45 (1), 21-40.

# Appendix A: A Proof of (2.15) based on Empirical Data

Using the demand function (2.6) and considering the symmetry of the two airlines, the aggregate demand function in the market is

$$Q = Q_1 + Q_2 = 2\alpha_0 - 2(\alpha_1 - \alpha_2)P_m - \frac{4(\alpha_1 - \alpha_2)\gamma}{f_m} \tag{A.1}$$

where $P_m = P_1 = P_2 = P, f_m = f_1 + f_2 = 2f$. Empirical studies have shown that the market level frequency elasticity $\varepsilon_f^0$ is less than 1 (Jorge-Calderón, 1997; Hsiao, 2008). In our model, the corresponding elasticities are expressed as

$$\varepsilon_f^0 = \frac{\partial Q}{\partial f_m}\frac{f_m}{Q} = \frac{2(\alpha_1 - \alpha_2)\gamma}{fQ} \tag{A.2}$$

If (2.15) holds, then the LHS in (2.13) is monotonically decreasing. Rearranging the LHS term to the RHS and multiplying both sides by 3/2, (2.14) becomes

$$[\alpha_0 - (\alpha_1 - \alpha_2)\tau] - \frac{3}{2}\frac{(\alpha_1 - \alpha_2)\gamma}{f} > 0 \tag{A.3}$$

which we want to show to be plausible in the real world. Note

$$[\alpha_0 - (\alpha_1 - \alpha_2)\tau] - \frac{3}{2}\frac{(\alpha_1 - \alpha_2)\gamma}{f} > [\alpha_0 - (\alpha_1 - \alpha_2)P] - \frac{(\alpha_1 - \alpha_2)\gamma}{f} - \frac{1}{2}\frac{(\alpha_1 - \alpha_2)\gamma}{f}$$

$$= Q_1 - \frac{1}{2}\frac{(\alpha_1 - \alpha_2)\gamma}{f} = \frac{Q}{2} - \varepsilon_f^0\frac{Q}{4} = (1 - \frac{\varepsilon_f^0}{2})\frac{Q}{2} \tag{A.4}$$

The first inequality stems from the fact that price is set to be higher than the marginal cost per seat. The fact that frequency elasticity is often less than one suggest that the last term be positive, i.e. (2.15) holds true.

# Appendix B: Correspondence between MAS and Metropolitan Areas

| Area | Airport | Airport Code |
|------|---------|:------------:|
| Chicago, IL | O'Hare International | ORD |
| | Chicago Midway International | MDW |
| New York City, NY | Newark International | EWR |
| | LaGuardia | LGA |
| | John F. Kennedy International | JFK |
| | Long Island MacArthur | ISP |
| Los Angeles, CA | Los Angeles International | LAX |
| | Ontario International | ONT |
| | Orange County/John Wayne | SNA |
| | Burbank Bob Hope | BUR |
| | Long Beach | LGB |
| | Palm Springs International | PSP |
| Dallas/Ft. Worth, TX | Dallas/Ft. Worth International | DFW |
| | Love Field | DAL |
| San Francisco, CA | San Francisco International | SFO |
| | Oakland International | OAK |
| | San Jose International | SJC |
| Washington D.C. | Washington Reagan National | DCA |
| | Dulles International | IAD |
| | Baltimore International | BWI |
| Miami, FL | Miami International | MIA |
| | Ft. Lauderdale-Hollywood International | FLL |
| Houston, TX | Houston International | IAH |
| | William P. Hobby | HOU |

# Appendix C: Investigation of the Jacobian Matrix $\nabla F(X)$

To investigate the properties of $\nabla F(\mathbf{X})$, first recall the estimation results of the fare and frequency models. The coefficients for airport delay are—albeit significant—very small. The same occurs to the coefficient for route passenger volume in the fare model. To maintain the analytical tractability, we ignore these effects in the following analysis. We further assume fixed O-D demand which is assigned to routes following a multinomial Logit model form, and do not differentiate marginal utility of flight frequency, i.e. the coefficients for $f_{\text{non-stop}}$, $f_{\text{min, one-stop}}$, $f_{\text{max, one-stop}}$ are identical. These simplifications enable analytical insights into the properties of the Jacobian matrix while maintaining the fundamental structure of the more complete equilibrium under study.

Given an O-D pair $w$, the observed utility of a representative traveler consists of three parts:

$$V_r = \beta_0 + \beta_1 \sum_a [\ln(f_a) \cdot \delta_{ar}] + \beta_2 \sum_k \eta_{kr} d_k \tag{C.1}$$

where $f_a$ is flight frequency on segment $a$; $\delta_{ar}$ the link-route indicator defined in Section 4.2.2.1; $\eta_{kr}$ equals 1 if route $r$ uses airport $k$ as the connecting hub, and 0 otherwise; $d_k$ the delay at airport $k$. $\beta_0$, $\beta_1$, $\beta_2$ are coefficients. The second term captures the frequency effect on traveler utility; the third term the effect of hub airport delay; all other effects, assumed constant, are embedded in the constant $\beta_0$. Passenger demand on route $r$ can be expressed as

$$D_r = D_{tot} \cdot P_r = D_{tot} \cdot \frac{\exp(V_r)}{\sum_{l \in R_w} \exp(V_l)} \tag{C.2}$$

We express airport delay as only a function of the VC ratio—everything else is treated as constant. In effect, this corresponds to a simpler version of the delay model in Chapter 4. Delay at airport $k$ equals

$$d_k = B_k \exp(\gamma_k \frac{\sum_a \zeta_{ka} f_a}{C_k}) = B_k \exp[\gamma_k \frac{\sum_a \zeta_{ka} G_a(s_a)^\alpha}{C_k}] = B_k \exp[\gamma_k \frac{\sum_a \zeta_{ka} G_a(\sum_j \delta_{aj} D_j)^\alpha}{C_k}] \tag{C.3}$$

where $\zeta_{ka}$ equal to 1 if the end of (directional) segment $a$ is airport $k$, and 0 otherwise; $s_a$ denotes passenger volume on segment $a$. $f_a = G_a(s_a)^\alpha = G_a(\sum_j \delta_{aj} D_j)^\alpha$ stems from the log-log frequency equation. Constants $B_k$ and $G_a$ capture all other factors for airport $k$ and segment $a$ that remain constant in the equilibration process. Under such simplification,

the analytical expression of airport delay as a function of route passenger demand allows reduced dimensions of the fixed point problem for which only $\boldsymbol{D}$ will appear in the formulation. More specifically, we attempt to solve the following fixed point problem

$$D_r = D_{tot} \cdot \frac{\exp(V_r(\boldsymbol{D}))}{\sum_{l \in R_w} \exp(V_l(\boldsymbol{D}))} \quad \forall r \in R_w, \, w \in W \tag{C.4}$$

The element on the $r$th row and $j$th column of the Jacobian matrix, $J_{rm}$, equals

$$J_{rm} = \frac{\partial D_r}{\partial D_m} = D_r(1 - P_r)\frac{\partial V_r}{\partial D_m} - D_r \sum_{\substack{l \in R_w \\ l \neq r}} P_l \frac{\partial V_l}{\partial D_m} \tag{C.5}$$

where

$$\frac{\partial V_r}{\partial D_m} = \beta_1 (\sum_a \delta_{ar} \frac{1}{f_a} \frac{\partial f_a}{\partial D_m}) + \beta_2 \sum_k \eta_{kr} \frac{\partial d_k}{\partial D_m} \tag{C.6}$$

Substituting $f_a = G_a(\sum_j \delta_{aj} D_j)^\alpha$ into $\dfrac{\partial f_a}{\partial D_m}$ yields

$$\frac{\partial f_a}{\partial D_m} = G_a \alpha (\sum_j \delta_{aj} D_j)^{\alpha-1} \sum_j \delta_{aj} \frac{\partial D_j}{\partial D_m} = G_a \alpha (\sum_j \delta_{aj} D_j)^{\alpha-1} \delta_{am} = \alpha \frac{f_a}{\sum_j \delta_{aj} D_j} \delta_{am} \tag{C.7}$$

Similarly, replacing $d_k$ by (C.3) and substituting $\dfrac{\partial f_a}{\partial D_m}$ by (C.7) yield

$$\begin{aligned}
\frac{\partial d_k}{\partial D_m} &= B_k \exp[\frac{\gamma_k}{C_k} \sum_a \zeta_{ka} f_a] \frac{\gamma_k}{C_k} \sum_a \zeta_{ka} \frac{\partial f_a}{\partial D_m} = d_k \frac{\gamma_k}{C_k} \sum_a \zeta_{ka} \frac{\partial f_a}{\partial D_m} \\
&= d_k \frac{\gamma_k}{C_k} \sum_a \zeta_{ka} \alpha \frac{f_a}{\sum_j \delta_{aj} D_j} \delta_{am}
\end{aligned} \tag{C.8}$$

Combining (C.6) with (C.7) and (C.8) leads to

$$\frac{\partial V_r}{\partial D_m} = \beta_1 (\sum_a \delta_{ar} \delta_{am} \alpha \frac{1}{\sum_j \delta_{aj} D_j}) + \beta_2 \sum_k \sum_a \eta_{kr} \zeta_{ka} \delta_{am} d_k \alpha \frac{\gamma_k}{C_k} \frac{f_a}{\sum_j \delta_{aj} D_j} \tag{C.9}$$

Focusing on the first term in (C.9)

$$\frac{\partial V_r}{\partial D_m} = \begin{cases} \beta_1 \sum_a (\alpha \frac{1}{\sum_j \delta_{\bar{a}j} D_j}) + \beta_2 \sum_k \sum_a \eta_{kr} \zeta_{ka} \delta_{am} d_k \alpha \frac{\gamma_k}{C_k} \frac{f_a}{\sum_j \delta_{aj} D_j} & r,m \text{ share segment(s) } \bar{a} \\ \\ \beta_2 \sum_k \sum_a \eta_{kr} \zeta_{ka} \delta_{am} d_k \alpha \frac{\gamma_k}{C_k} \frac{f_a}{\sum_j \delta_{aj} D_j} & r,m \text{ do not share any common segment} \end{cases}$$ (C.10)

In order for the second term in (C.9) to be non-zero, $\eta_{kr} = \zeta_{\bar{ka}} = \delta_{\bar{am}} = 1$ must hold, which means: 1) route $r$ is a one-stop route; 2) route $m$ contains the segment $\bar{\bar{a}}$ whose ending point is the hub airport used by route $r$. Graphically, this can only occur in one of the three situations below:
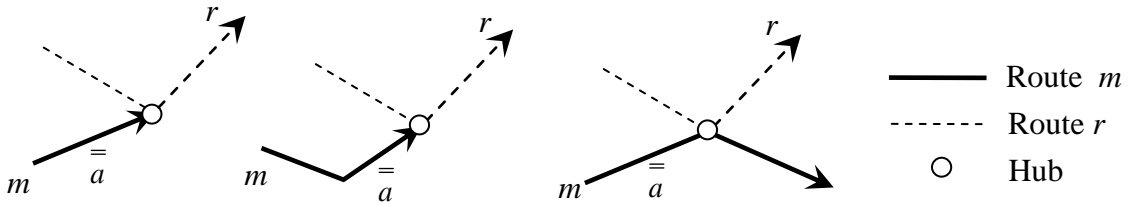


Figure C.1: Three possibilities for the second terms in (C.9) to be non-zero

It should be noted that partial and full overlapping of routes can occur in the last situation.

In order to see $\| J \|$ depends upon D values, we consider a simple network as in Figure C.2. The network has a symmetric shape with three spoke cities equal-distantly distributed surrounding a hub city. We assume each city has one airport. This network consists of 6 spoke→spoke non-stop routes, 6 spoke→hub→spoke connecting routes, 3 spoke→hub non-stop routes, and 3 hub→spoke non-stop routes. Consider non-stop route $r$: $E_1 \rightarrow E_2$ and choose $l_1$-norm,

$$\| J \|_1 = \max_r \sum_m | J_{rm} | = \max_r \sum_m | \frac{\partial D_r}{\partial D_m} | = \max_r \sum_m | D_r(1 - P_r) \frac{\partial V_r}{\partial D_m} - D_r \frac{\partial V_{\tilde{r}}}{\partial D_m} |$$ (C.11)

Since only one connecting route exists for O-D pair $E_1 E_2$, $\tilde{r}$ denotes the connecting route $E_1 H E_2$.
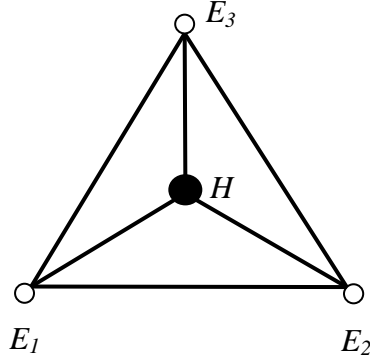
Figure C.2: A simple network

To better illustrate the intuition, in the following we further assume a "congestion free" world, i.e. the second term in (C.9) is ignored. Recall that the essential of the first term in (C.9) is the marginal impact of demand on route $m$ on route $r$. In total six different routes will lead to non-zero values for $|D_r(1-P_r)\frac{\partial V_r}{\partial D_m} - D_r\frac{\partial V_{\tilde{r}}}{\partial D_m}|$. The following discusses each of the situations, which

1) route $m$ is $E_1E_2$, identical to $r$:

$$|D_r(1-P_r)\frac{\partial V_r}{\partial D_m} - D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m}| = D_r(1-P_r)\frac{\partial V_r}{\partial D_m} = D_r(1-P_r)\beta_1(\alpha\frac{1}{D_r}) = (1-P_r)\beta_1\alpha$$

2) route $m$ is $E_1HE_2$, identical to $\tilde{r}$ :

$$|D_r(1-P_r)\frac{\partial V_r}{\partial D_m} - D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m}| = D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m} = D_rP_{\tilde{r}}\beta_1\alpha(\frac{1}{s_{E_1H}} + \frac{1}{s_{HE_2}})$$

where $s_{E_1H}$ and $s_{HE_2}$ denote, respectively, segment passenger traffic on segments $E_1H$ and $HE_2$.

3) route $m$ is $E_1H$, which shares segment $E_1H$ with $\tilde{r}$ :

$$|D_r(1-P_r)\frac{\partial V_r}{\partial D_m} - D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m}| = D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m} = D_rP_{\tilde{r}}\beta_1\alpha\frac{1}{s_{E_1H}}$$

4) route $m$ is $E_1HE_3$, which shares segment $E_1H$ with $\tilde{r}$ :

$$|D_r(1-P_r)\frac{\partial V_r}{\partial D_m} - D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m}| = D_rP_{\tilde{r}}\frac{\partial V_{\tilde{r}}}{\partial D_m} = D_rP_{\tilde{r}}\beta_1\alpha\frac{1}{s_{E_1H}}$$

5) route $m$ is $HE_2$, which shares segment $HE_2$ with $\tilde{r}$ :

$$| D_r (1-P_r) \frac{\partial V_r}{\partial D_m} - D_r P_{\tilde{r}} \frac{\partial V_{\tilde{r}}}{\partial D_m} | = D_r P_{\tilde{r}} \frac{\partial V_{\tilde{r}}}{\partial D_m} = D_r P_{\tilde{r}} \beta_1 \alpha \frac{1}{s_{HE_2}}$$

6) route $m$ is $E_3HE_2$, which shares segment $HE_2$ with $\tilde{r}$ :

$$| D_r (1-P_r) \frac{\partial V_r}{\partial D_m} - D_r P_{\tilde{r}} \frac{\partial V_{\tilde{r}}}{\partial D_m} | = D_r P_{\tilde{r}} \frac{\partial V_{\tilde{r}}}{\partial D_m} = D_r P_{\tilde{r}} \beta_1 \alpha \frac{1}{s_{HE_2}}$$

Summing up the above six terms yields

$$\sum_m | D_r (1-P_r) \frac{\partial V_r}{\partial D_m} - D_r \frac{\partial V_{\tilde{r}}}{\partial D_m} | = (1-P_r)\beta_1\alpha + 3 D_r P_{\tilde{r}} \beta_1 \alpha (\frac{1}{s_{E_1H}} + \frac{1}{s_{HE_2}})$$

$$= (1-P_r)\beta_1\alpha + 3 D_r (1-P_r) \beta_1 \alpha (\frac{1}{s_{E_1H}} + \frac{1}{s_{HE_2}}) \qquad (C.12)$$

$$= (1-P_r)\beta_1\alpha (1 + \frac{3D_r}{s_{E_1H}} + \frac{3D_r}{s_{HE_2}})$$

If $\| \boldsymbol{J} \|$ is less than 1, then mapping $\boldsymbol{D}$ becomes contractive. The Norm-Equivalence Theorem suggests that all norms on $R^n$ are equivalent. Under the $l_1$-norm, the above sufficient condition requires

$$\sum_m | D_r (1-P_r) \frac{\partial V_r}{\partial D_m} - D_r \frac{\partial V_{\tilde{r}}}{\partial D_m} | < 1 \qquad \forall m \qquad (C.13)$$

Note that $\beta_1$ is the coefficient for the frequency variable. Under the multinomial logit form, the demand elasticity for route $r$ with respect to frequency is

$$\varepsilon_f = \frac{\partial D_r}{\partial f_r} \frac{f_r}{D_r} = \frac{\partial (D_{tot} P_r)}{\partial f_r} \frac{f_r}{D_{tot} P_r} = \frac{\partial P_r}{\partial f_r} \frac{f_r}{P_r} = \frac{\partial V_r}{\partial f_r} P_r (1-P_r) \frac{f_r}{P_r} = \frac{\beta_1}{f_r} P_r (1-P_r) \frac{f_r}{P_r}$$

$$= \beta_1 (1-P_r) \qquad (C.14)$$

Replacing $\beta_1$ by $\varepsilon_f /(1-P_r)$ in (C.12) gives

$$\sum_m | D_r (1-P_r) \frac{\partial V_r}{\partial D_m} - D_r \frac{\partial V_{\tilde{r}}}{\partial D_m} | = \varepsilon_f \alpha (1 + \frac{3D_r}{s_{E_1H}} + \frac{3D_r}{s_{HE_2}}) \qquad (C.15)$$

Hsiao (2008) shows that, in the US, route demand elasticity with respect to non-stop route frequency is about 1.2 under a multinomial logit specification (and ~0.8 if a Nested Logit is used). $\alpha$ comes from the frequency model. According to our estimate, $\alpha = 0.65$. In this particular case, the feasible set $\boldsymbol{\Omega}$ is defined as $\{D_r: 0 \le D_r \le D_{tot,w}, \ \forall r \in R_w, w \in W\}$. Assuming identical spoke

cities and the same $D_{tot,w}$ across all O-D pairs, clearly the following set of demand $\boldsymbol{D_0} = \{D_r = 0.5D_{tot,w}, \ \forall \, r \in R_w, \ w \in W\} \in \boldsymbol{\Omega}$. Under this set of demand, $s_{E_1H} = s_{E_2H} = 2D_{tot,w}$. (C.12) then equals

$$1.2 \times 0.65 \times (1 + \frac{3 \times 0.5D_{tot,w}}{2D_{tot,w}} + \frac{3 \times 0.5D_{tot,w}}{2D_{tot,w}}) > 1.$$ Therefore, $\| \boldsymbol{J} \| < 1$ does not always hold, i.e.

mapping $\boldsymbol{F}$ is not necessarily contractive.

# Appendix D: Estimation Results of the Reduced Form Frequency Model

Using the same dataset, Table D.1 reports OLS estimation results of a reduced form of frequency model (3.3). Standard errors are similarly clustered by metropolitan area pair. OriginIncome and DestIncome denote the total income in the original and destination metropolitan areas; OriginConnRatio and DestConnRatio the ratio of connecting to local O-D passengers (based on outbound traffic) at the origin and destination airports. The remaining variables follow the same definition as in Table 3.6. All continuous variables except for OriginConnRatio and DestConnRatio take logarithmic values.

Table D.1 Estimation results for the reduced form frequency model.

| Variable | Est. | Std. Err. |
|---|---|---|
| OriginIncome | 0.2375*** | 0.0179 |
| DestIncome | 0.2364*** | 0.0178 |
| OriginConnRatio | 0.2394*** | 0.0188 |
| DestConnRatio | 0.2370*** | 0.0187 |
| Dist | -0.4180*** | 0.0201 |
| SegmentHHI | -0.7705*** | 0.0205 |
| OriginHHI | 0.0414* | 0.0224 |
| DestHHI | 0.0424* | 0.0223 |
| OriginL4Delay | -0.0714** | 0.0293 |
| DestL4Delay | -0.0800*** | 0.0296 |
| Vacation | 0.1032*** | 0.0268 |
| SlotControl | -0.1431*** | 0.0336 |
| OriginMAS | 0.0262 | 0.0395 |
| DestMAS | 0.0273 | 0.0389 |
| LCC | 0.4794*** | 0.0280 |
| PortionLCC | -0.3234*** | 0.0585 |
| MASPair | -0.0407 | 0.0340 |
| Constant | 0.2987 | 0.4299 |
| Number of Observations | 65,033 | |
| $R^2$ | 0.5555 | |

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

The above estimated model is then used to generate the initial segment flight frequencies, as in *Step 0.2*, with zero values assumed for OriginConnRatio, DestConnRatio, OriginL4Delay, and DestL4Delay on all segments.