Flight Delays, Capacity Investment and Social Welfare under Air Transport Supply-Demand Equilibrium

Bo Zou, Mark Hansen
National Center of Excellence for Aviation Operations Research, Institute of Transportation Studies, University of California, Berkeley, CA 94720, USA

Abstract
This paper analyzes benefits from aviation infrastructure investment under competitive supply-demand equilibrium. The analysis recognizes that, in the air transportation system where economies of density is an inherent characteristic, capacity change would trigger a complicated set of adjustment of and interplay among passenger demand, air fare, flight frequency, aircraft size, and flight delays, leading to an equilibrium shift. An analytical model that incorporates these elements is developed. The results from comparative static analysis show that capacity constraint suppresses demand, reduces flight frequency, and increases passenger generalized cost. Our numerical analysis further reveals that, by switching to larger aircraft size, airlines manage to offset part of the delay effect on unit operating cost, and charge passengers lower fare. With higher capacity, airlines tend to raise both fare and frequency while decreasing aircraft size. More demand emerges in the market, with reduced generalized cost for each traveler. The marginal benefit brought by capacity expansion diminishes as the capacity-demand imbalance becomes less severe. Existing passengers in the market receive most of the benefit, followed by airlines. The welfare gains from induced demand are much smaller. The equilibrium approach yields more plausible investment benefit estimates than does the conventional method. In particular, when forecasting future demand the equilibrium approach is capable of preventing the occurrence of excessive high delays.
1 Introduction

Flight delay is a serious and widespread problem in many parts of the world. In the United States, between 2002 and 2007, flights increased by about 22 per cent, but the number of late-arriving flights more than doubled (Ball et al., 2010). Although traffic and delay have declined somewhat recently because of the economic recession, the Federal Aviation Administration (FAA) expects growth to resume, with flight traffic reaching 2007 levels by 2012, and growing an additional 30 per cent by 2025 (Ball et al., 2010).

One of the major causes of flight delays is inadequate capacity in the air transportation system. The Federal Aviation Administration (FAA) has established multi-billion investment plans to enhance the capacity of the system, under the Next Generation Air Transportation System (NextGen) and beyond (Calvin, 2009). Such huge investment must be weighed against the benefits that system users expect to receive, most noticeably through delay savings, which translate into consumer and producer welfare gains. To this end, appropriate benefit assessment methodologies are of critical importance.

Assessing the economic value of investment in aviation infrastructure has attracted attention from both practitioners and academicians. In the practical world, considerable strides have been made in simulation tools such as NASPAC, ACES, and LMINET, which incorporate flight trajectories, weather, en-route and airport capacity constraints, and schedule adjustments to account for capacity constraints in the system (Post, 2006; Post et al., 2008). Benefits as a result of delay reduction are often measured in the form of airline cost savings and shortening of passenger travel time (e.g. Steinbach and Giles, 2005). While intuitive, this sort of assessment is oversimplified because it pays little or no attention to mechanisms through which airlines and air travelers respond to flight delay. While it is recognized that airlines will change flight schedules to avoid exorbitant delays, efforts to account for this in engineering practice have traditionally been arbitrary and simplistic (FAA, 1999). Passenger demand responses to flight delays—either direct or as a result of airline responses—have been studied even less.

In the academic arena, Hansen and Wei (2006) perform a multivariate ex-post analysis to investigate the impact of a major capacity expansion at Dallas-Fort Worth airport. In addition to improved on-time performance, they find that the delay reduction benefit may be offset by flight demand inducement and airline schedule adaptations. In a series of studies, Morrison and Winston explicitly model passenger demand as either a function of delay (Morrison and Winston, 1983), or the full price of a flight that include airline operating cost, passenger time cost, landing fees, and delay cost to airlines and passengers (Morrison and Winston, 1989; 2007). Jorge and de Rus (2004) point out benefits from airport investment include delay savings for existing and diverted traffic. They argue that new capacity would enable increase in departure frequency, and the use of smaller aircraft. The authors further demonstrate an application of their considerations, in a somewhat simplified version based on rules of thumb generally accepted in the aviation industry.

Airlines may also adjust air fares in response to delay changes, which in turn affect passenger demand. A recent work by Miller and Clarke (2008), focusing on maximizing airport net benefit, recognizes that congestion raises airline operating cost, part of which will be passed onto passengers through higher air fare. The high fare then leads to a lower level of air travel demand. Following their path, airlines will further adjust fare according to the new passenger demand. This supply-demand adjustment process will continue until a new equilibrium is reached.

Equilibrium analysis in this setting must account for the fundamental importance of service quality in shaping travel decisions. Delay is certainly one dimension of service quality. Another is the quantity of service provided, whose importance in scheduled transportation services—
particularly urban transit—has long been recognized by researchers (e.g. Mayworm et al., 1980; Frankena, 1983; Else, 1985), but largely overlooked in aviation infrastructure investment analysis. Given an air transport route, researchers often use frequency and/or schedule delay to measure the service quantity provided on that route. Service frequency and delay are interdependent. For a given airport, total traffic consists of all scheduled flights with the airport being either the departure or arrival end. These flights, together with the capacity at the airport, determine the level of airport delays. Facing high delays, airlines may reduce service frequency and resort to larger aircraft. However, most existing studies implicitly assume flight traffic is determined by passenger demand. Ignoring the frequency response of airlines to flight delay could result in inaccurate benefit estimates for capacity investment.

From a broader perspective, the equilibrium analysis in air transport must take into account economies of density in the system. Economies of density—declining average cost from flowing more traffic on the same network—has been identified by many empirical studies such as Caves et al. (1984), Gillen et al. (1985; 1990) at the airline level, and Brueckner and Spiller (1994) on individual route segments. When there is no congestion, the consequences of economies of density brings are two-fold. First, higher density is realized in the form of more plane-miles; more plane-miles translate into higher frequency, improving the service quality to passengers. On the other hand, it may be possible for airlines to operate at a lower cost using larger aircraft and offer cheaper fares to passengers. The overall effect of service quality and fare can be combined into generalized cost, which includes three parts: ticket price, monetized cost of frequency and (potentially) passenger delay. Higher density reduces passengers’ generalized cost, making air travel more attractive. More demand will be generated. The increase in demand results in an even higher density, contributing to a further reduction in passenger generalized cost. Figure 1 illustrates the final outcome of this positive feedback loop. In Figure 1, the demand curve is a function of passenger generalized cost. Accordingly, the supply curve \((S_0)\) reflects the corresponding generalized price airlines would impose on passengers as a function of output. With no congestion, the supply curve \((S_0)\) is downward sloping, and the equilibrium is achieved at point \(G\).

The above picture no longer holds when capacity becomes a constraint. With increased traffic, delay appears due to limited capacity. Lengthening flight time brings extra cost to airlines, diminishing or reversing economies of density. The new supply curve tracks unconstrained downward sloping curve until delays emerge, and then veers higher, as shown by curves \(S_1\) or \(S_2\), with \(S_1\) representing a more severe capacity constraint in the system. Airlines may pass part of their cost increase to passengers through higher ticket prices. They may also choose to cut back service frequency and up-gauge aircraft size, to reduce their own demand for the system capacity. These result in an increase in the generalized cost to passengers. Furthermore, flight delay increases passenger generalized cost, because passengers suffer from the additional time spent in travel, and reduced predictability and reliability of their travel schedules.

The increase in passenger generalized cost is accompanied by suppressed demand, as shown in Figure 1. The equilibrium shifts from \(G\) upward to point \(B\) or \(C\), depending on the extent of the capacity constraints in the system.
A loss of consumer surplus (CS) is directly discernable from the equilibrium shift in Figure 1. If we want to compare an unconstrained capacity case (supply curve represented by $S_0$) with a high capacity constraint case (supply curve represented by $S_1$), the CS loss is represented by area $ABGF$. (The figure does not provide producer surplus (PS) changes. Because the equilibrium shift involves passenger demand, air fare, and airline operating cost, it is difficult to discern graphically the changes in airline profit.)

In infrastructure investment, our interest is in assessing benefits when moving from a more capacity-constrained equilibrium state to a less capacity-constrained one. We use point $B$ in Figure 1 to denote the original equilibrium. After some investment, the equilibrium shifts to $C$. The question is to identify the equilibrium points and use them to assess the associated benefits. Building on the previous qualitative analysis, the present paper contributes to the investment benefit analysis methodology by proposing a new assessment framework that incorporates a shift in equilibrium in the air transportation supply-demand system. We employ an analytical model to investigate how capacity investment will trigger changes in relevant elements (e.g. passenger demand, air fare, flight frequency), and identify the associated welfare gain from both consumer and producer sides. The rest of the paper is organized as follows. A modeling framework is proposed in the next section, based on which we set up the analytical model structure in section 3. Comparative static analysis ensues in section 4. Section 5 presents a set of numerical analyses, which enriches the insights gained in the previous section. We also examine the sensitivity of equilibrium shift to capacity expansions, and compare the welfare estimates by using the equilibrium and conventional assessment methods. Conclusions are offered in Section 6.

2 Model Framework

The relevant elements in the supply-demand equilibrium shift triggered by capacity change include passenger demand, airline cost, air fare, flight service, and flight delay. Figure 2 represents the interactions between these variables, with the arrows denoting causal relationships.
In Figure 2, passenger demand is determined by the generalized cost, consisting of air fare, flight delays, the amount of flight service provided, as well as exogenous factors such as population, income, and characteristic of competing modes. Determining air fare involves airlines’ profit maximizing behavior. At the route level, this requires the knowledge of passenger demand, flight cost structure, and market conditions. Flight service reflects airlines’ scheduling behavior in response to passenger demand, infrastructure capacity, and flight delay. It generates airline production output, which is often measured by flight miles or passenger miles flown. Flight delay appears when the quantity of airline service approaches infrastructure capacity. Based on the production theory, airline cost depends primarily on input prices and output. Empirical evidence finds higher flight delay increases airline operating cost (Hansen et al., 2001; Zou and Hansen, 2010).

The model framework implies that once infrastructure capacity level is changed, the new values of passenger demand, air fare, flight service, airline cost, and flight delay will be endogenously adjusted, leading to a new equilibrium. Certainly, capacity is often affected by weather conditions, but this is not the focus of our study. We assume weather conditions, like other exogenous factors, to be fixed. One may further argue that changes in infrastructure capacity results from investment, the decision-making of which is based upon the level of flight delay in the system. However, due to the lumpy, public, and politically contentious nature of aviation infrastructure investment, the link between investment and delay is tenuous. The link from flight delay to investment is treated as a weak feedback (a dashed line).

Figure 2 The Modeling Framework
In the next section, this proposed framework will be applied to an airline competition model to explore the capacity-related supply-demand equilibrium and how the equilibrium shifts when capacity changes. Despite the existence of a large body of theoretical literature analyzing the economics of airline behavior, relative few efforts have so far been devoted to airline behavior vis-à-vis infrastructure capacity constraints. The following analytical model will provide some useful insights about the interplays among passenger demand, air fare, airline cost, flight traffic and delay, from a microscopic point of view.

3 The Model

3.1 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: \( P_i = P_i + \frac{\gamma}{f_i} \), for \( i = 1, 2 \). We assume the representative consumer has the following utility function:

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

(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:

\[
q_0 + P_1 q_1 + P_2 q_2 = I
\]

(2)

where \( I \) denotes income. The first-order conditions of the corresponding Lagrangian \( L \),

\[
U(q_0, q_1, q_2) - \lambda (q_0 + P_1 q_1 + P_2 q_2 - I)
\]

with \( \lambda \) being the Lagrange multiplier, are

\[
\frac{\partial L}{\partial q_0} = 1 - \lambda = 0
\]

(3.1)
\[
\frac{\partial L}{\partial q_1} = \frac{\alpha_{00}}{\alpha_{01} - \alpha_{02}} - \frac{\alpha_{01}}{\alpha_{01}^2 - \alpha_{02}^2} q_1 - \frac{\alpha_{02}}{\alpha_{01}^2 - \alpha_{02}^2} q_2 - \lambda \overline{P}_1 = 0 \tag{3.2}
\]
\[
\frac{\partial L}{\partial q_2} = \frac{\alpha_{00}}{\alpha_{01} - \alpha_{02}} - \frac{\alpha_{01}}{\alpha_{01}^2 - \alpha_{02}^2} q_1 - \frac{\alpha_{01}}{\alpha_{01}^2 - \alpha_{02}^2} q_2 - \lambda \overline{P}_2 = 0 \tag{3.3}
\]
\[
\frac{\partial L}{\partial \lambda} = -(q_0 + \overline{P}_1 q_1 + \overline{P}_2 q_2 - I) = 0 \tag{3.4}
\]

The second-order conditions are guaranteed since the Hessian is negative semi-definite given the concavity of the utility function. Substituting (3.1) into (3.2) and (3.3) yields the following system of linear inverse demand functions:

\[
\overline{P}_i = \frac{\alpha_{00}}{\alpha_{01} - \alpha_{02}} - \frac{\alpha_{01}}{\alpha_{01}^2 - \alpha_{02}^2} q_i - \frac{\alpha_{02}}{\alpha_{01}^2 - \alpha_{02}^2} q_{-i}, \ i = 1,2 \tag{4}
\]

where the subscript \(-i\) denotes the competing airline. Incorporating the generalized cost expression and solving (4) for \(i=1,2\) lead to the following “symmetric” demand function

\[
q_i = \alpha_{00} - \alpha_{01} P_i + \alpha_{02} P_{-i} - \frac{\alpha_{01}'}{f_i} + \frac{\alpha_{02}'}{f_{-i}}, \ i = 1,2 \tag{5}
\]

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

\[
Q_i = \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1'}{f_i} + \frac{\alpha_2'}{f_{-i}}, \ i = 1,2 \tag{6}
\]

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

\[
Q_i = \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1'}{f_i} + \frac{\alpha_2'}{f_{-i}} - \mu L, \ i = 1,2 \tag{7}
\]

\[
\sum_{i=1}^{2} Q_i = \sum_{i=1}^{2} \left( \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1'}{f_i} + \frac{\alpha_2'}{f_{-i}} - \mu L \right)
\]

\[
\sum_{i=1}^{2} Q_i = \sum_{i=1}^{2} \left( \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1'}{f_i} + \frac{\alpha_2'}{f_{-i}} - \mu L \right)
\]

\[
\sum_{i=1}^{2} Q_i = \sum_{i=1}^{2} \left( \alpha_0 - \alpha_1 P_i + \alpha_2 P_{-i} - \frac{\alpha_1'}{f_i} + \frac{\alpha_2'}{f_{-i}} - \mu L \right)
\]
where $\mu = k \cdot n \cdot (\alpha_0 - \alpha_{0i}) = k \cdot (\alpha_i - \alpha_s)$ 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 $N(f_1 + f_2)/K$, 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. $L = L(f_1 + f_2)$.

### 3.2 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:

$$C_i = c_0 + \tau s_i + \eta s_i L$$

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.

### 3.3 Competition and Equilibrium

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

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

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

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

---

$^1$ 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 paper refers specifically to traffic volume of arrivals.

$^2$ 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 paper as load factor is assumed to be 1, economies of density on the supply side are primarily embodied in the first aspect.
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:

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

(10.1)

\[ \frac{\partial \pi_1}{\partial f_1} = (P_1 - \tau - \eta L)\left(\frac{\alpha_1 \gamma}{f_1^2} - \frac{\partial L}{\partial f_1} - \eta \frac{\partial L}{\partial f_1}\right) - c_0 - \eta \frac{\partial L}{\partial f_1} (\alpha_0 - \alpha_1 P_1 + \alpha_2 P_2 - \frac{\alpha_1 \gamma}{f_1} + \frac{\alpha_2 \gamma}{f_2} - \mu L) = 0 \]  

(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, \( \frac{\alpha_1 \gamma}{f_1^2} - \frac{\partial L}{\partial f_1}(c_0 + \eta \frac{\partial L}{\partial f_1} Q_1)/(P_1 - \tau - \eta L) > 0 \) according to (10.1) and (10.2). For the delay function \( L \), we further expect marginal delay increase is greater when traffic is at a higher level, i.e. \( \frac{\partial^2 L}{\partial f^2} > 0 \). Then the second-order derivatives

\[ \frac{\partial^2 \pi_1}{\partial f_1^2} = -2\alpha_1 \]  

(11.1)

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

(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 (10.1) and (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

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

(12)

Substituting the above into the FOC frequency equation (10.2) yields

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

(13)

In order to discern potential frequency changes when delay occurs, Equation (13) needs to be simplified. The last term on the right hand side (RHS) of (13) is positive, as \( \alpha_1 > \alpha_2 \). So is the

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.
second-to-last term on the RHS, since substituting (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 (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
\[ F = [\alpha_0 - (\alpha_1 - \alpha_2)\gamma]/f^2 \]
Taking its first order derivative with respect to \(f\), we obtain
\[
\frac{\partial F}{\partial f} = \frac{2(\alpha_1 - \alpha_2)\gamma - 2f[\alpha_0 - (\alpha_1 - \alpha_2)\tau]}{f^4}
\]
(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:
\[
\frac{(\alpha_1 - \alpha_2)\gamma}{f} < \frac{2}{3}[\alpha_0 - (\alpha_1 - \alpha_2)\tau]
\]
(15)

Empirical evidence suggests that it is plausible for (15) to hold. More details are provided in appendix A. Therefore, \(\partial F/\partial f < 0\) and the LHS of (13) is a monotonic decreasing function. When traffic delay occurs, the RHS of (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.

4 Comparative Static Analysis

4.1 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 (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 (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 \(\alpha_2 \rightarrow \alpha_1\)), airlines tend to pass a larger portion of their delay cost to passengers. In normal cases, the portion should be greater than \(1/2\) 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 \(\tilde{P}\).

At equilibrium, demand for each carrier is
\[ Q_i = \alpha_0 - (\alpha_1 - \alpha_2)[P + \frac{\gamma}{f} + \frac{\mu L}{\alpha_1 - \alpha_2}] = \alpha_0 - (\alpha_1 - \alpha_2)\overline{P}, \quad i = 1, 2 \]  

(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 (12) into \( P \) above, the generalized cost under equilibrium, \( \overline{P}_0 \), becomes

\[ \overline{P}_0 = \frac{\alpha_0 + \alpha_1(\tau + \eta L)}{2\alpha_1 - \alpha_2} + \frac{\gamma}{f_0} \frac{\alpha_1}{(2\alpha_1 - \alpha_2)} + \frac{\alpha_1}{(\alpha_1 - \alpha_2)(2\alpha_1 - \alpha_2)} \mu L \]  

(17)

When there is no delay, generalized cost equals

\[ \overline{P}_0 = \frac{\alpha_0 + \alpha_1\tau}{2\alpha_1 - \alpha_2} + \frac{\gamma}{f_0} \frac{\alpha_1}{(2\alpha_1 - \alpha_2)} \]  

(18)

Comparing (17) with (18), two delay-related terms are added in (17) when congestion occurs: \( \alpha_0 \eta L / (2\alpha_1 - \alpha_2) \) 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 \( \overline{P}_0 > \overline{P}_0 \), i.e. generalized cost will increase.

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 (12) for \( P \) into the demand function (16)

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

(19)

When there is no delay, the corresponding \( \tilde{Q}_{0j} \) equals

\[ \tilde{Q}_{0j} = \frac{\alpha_1}{2\alpha_1 - \alpha_2} [\alpha_0 - (\alpha_1 - \alpha_2)\tau - \frac{(\alpha_1 - \alpha_2)\gamma}{f_0}], \quad i = 1, 2 \]  

(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 (19), demand for each airline becomes less when delay occurs, i.e. \( Q_{0j} < \tilde{Q}_{0j} \), \( i = 1, 2 \).

4.2 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 (19) by \( f_0 \):

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

(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 \(\tau + \frac{c_0}{s_0} + \eta L\) is also left indeterminate as a consequence.

### 4.3 Changes in consumer welfare

The increase in passengers’ generalized cost and the reduction in demand that result from delay are shown in Figures 3 and 4, 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 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).\(^5\) 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 3 and 4.

In the first step, we increase the generalized cost of airline 1 from \(\overline{P}_{0,1}\) to \(\overline{P}_{0,1}'\), with the generalized cost of airline 2 being provisionally unchanged. The corresponding CS loss is the area \(\overline{P}_{0,1}\overline{P}_{0,1}'DB\) in Figure 3. 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 \(\overline{P}_{0,2}\) to \(\overline{P}_{0,2}'\), with the further loss of CS given by the area \(\overline{P}_{0,2}\overline{P}_{0,2}'EH\) in Figure 4. Concurrent with this is the horizontal move of airline 1’s demand curve from D to A (Figure 3). The total CS loss is calculated by adding together the two areas: \(\overline{P}_{0,1}\overline{P}_{0,1}'DB\) and \(\overline{P}_{0,2}\overline{P}_{0,2}'EH\), 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 \(\overline{P}_{0,1}\overline{P}_{0,1}'DB\) and \(\overline{P}_{0,2}\overline{P}_{0,2}'EH\) 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 \(\overline{P}_{0,1}\overline{P}_{0,1}'DB\) and \(\overline{P}_{0,2}\overline{P}_{0,2}'EH\) is equal to twice the area of the trapezoid \(\overline{P}_{0,1}\overline{P}_{0,1}AB\) (or trapezoid \(\overline{P}_{0,2}\overline{P}_{0,2}EF\)), and the two triangles \(DBJ\) and \(EHG\) are of equal size.

\(^4\)\(\frac{\partial}{\partial f}[(\alpha_0 - (\alpha_1 - \alpha_2)\tau - (\alpha_1 - \alpha_2)\gamma/f)/f] = [2(\alpha_0 - (\alpha_1 - \alpha_2)\gamma/f - (\alpha_0 - (\alpha_1 - \alpha_2)\tau)]/f^2.\) 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 \(\epsilon_x\) 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 \(f\).

\(^5\)The authors thank one of the reviewers to point this out.
The welfare changes on the supply side remain analytically indeterministic 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.

5 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 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. 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 = \frac{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 = \frac{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:

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

(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), $\mu = 37.5(\alpha_1 - \alpha_2) / 60 = 37.5 \times (12.5 - 6.25) / 60 = 3.9$ 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 1. In the subsequent analysis, all variables are treated as continuous.$^7$

---

$^6$ 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.

$^7$ 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).
Table 1 Derived Parameter Values

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

5.1 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 (13) become zero. We find the equilibrium solution with the second-order conditions satisfied. The first line in Table 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).\(^8\) Solving Equation (13) yields a new set of equilibrium values (the second line in Table 2). Compared to the ideal case, delay results in smaller frequency, higher passenger generalized cost, and reduced demand, confirming our analytical conclusions.

Table 2 Comparison of Scenarios with and without Capacity Constraint

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Frequency</th>
<th>Air fare</th>
<th>Generalized cost</th>
<th>Airline Demand</th>
<th>AC size</th>
<th>Unit operating cost</th>
<th>Traffic/capacity ratio</th>
<th>Average delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K=\infty$</td>
<td>7.6</td>
<td>98.9</td>
<td>130.3</td>
<td>485.7</td>
<td>63.6</td>
<td>91.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$K=720$</td>
<td>5.6</td>
<td>96.0</td>
<td>143.2</td>
<td>405.0</td>
<td>71.9</td>
<td>91.5</td>
<td>0.94</td>
<td>7.3</td>
</tr>
</tbody>
</table>

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 (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.

\(^8\)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).
5.2 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.

5.2.1 Changes in the supply-demand characteristics
Holding the market potential constant, capacity increase reduces the traffic volume/capacity ratio and delay. Figure 5 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 6. 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.

With increasing airport capacity and continuing rise in passenger demand, airlines tend to schedule more flights (Figure 6). Frequency seems more sensitive to capacity level than does passenger demand, because airlines also decrease aircraft size as traffic increases. Figure 6 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 1st term on the RHS of (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 7).

![Figure 5 Delay and Volume/capacity Ratio vs. Airport Capacity](image)
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 7).

Figure 8 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.
5.2.2 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 4.3, at each baseline capacity, we calculate total CS change as

\[ \frac{1}{2} (\bar{P}_1^0 - \bar{P}_1^1)(Q_1^0 + Q_1^1) + \frac{1}{2} (\bar{P}_2^0 - \bar{P}_2^1)(Q_2^0 + Q_2^1) \],

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 \((\bar{P}_1^0 - \bar{P}_1^1)(Q_1^0 + Q_1^1)\) is needed. By the same token, CS gain for induced demand is obtained as \((\bar{P}_1^0 - \bar{P}_1^1)(Q_1^1 - Q_1^2)\), in which \(Q_2^2\) 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 \((\bar{P}_1^0 - \bar{P}_1^1)(Q_1^1 - Q_1^2)\) equals twice the illustrative area \(DBJ\) in Figure 3. 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 9 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 4. 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 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 9 Welfare Gain under Different Baseline Capacity Levels for a Fixed Capacity Increment

5.3 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 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\(^9\) 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

\(^9\) We use the same passenger value of time ($\$37.5/hr$), the one used to determine the parameter $\mu$ before.
demand, as well as taste variation, whose contribution is $\alpha_\delta (1 + \delta) \Delta$.\(^{10}\) 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 5.2.2, 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 benefit using the conventional method, the temporal patterns are very different. As shown in Figure 9, 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 11. 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.

\(^{10}\) Suppose demand for airline 1 in year $k$ equals $Q_{1,k} = \alpha_0 - \alpha_1 P_{1,k} + \alpha_2 P_{2,k}$. According to our treatment of socio-economic impact on parameters, airline 1’s demand in the following year becomes $Q_{1,k+1} = \alpha_0 (1 + \delta)(1 + \Delta) - \alpha_1 (1 + \delta)P_{1,k+1} + \alpha_2 (1 + \delta)P_{2,k+1}$. Since response from the supply side is not considered, $P_{1,k+1} = P_{1,k}, P_{2,k+1} = P_{2,k}$ 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.
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.

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 10). As illustrated...
before, the equilibrium permits demand to self-adjust so that exceedingly high delay can be prevented.

6 Conclusion

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 paper 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 paper 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.

Acknowledgement

This research was funded by the Federal Aviation Administration through a grant to the National Center of Excellence for Aviation Operations Research (NEXTOR) for “Air Transport Supply-Demand Equilibrium Models that are Sensitive to NAS Investment Levels”. The enthusiastic support of Joseph Post for this project is gratefully acknowledged. An earlier version of this paper was presented at the Kuhmo-Nectar Conference on Transport Economics 2010, in Valencia, Spain. The first author would like to thank Leonardo J. Basso and other seminar participants for helpful comments and particularly Jan Brueckner for his early presentation at UC Berkeley which inspired part of this work, as well as his very helpful suggestions on this paper. Additional gratitude extends to the anonymous referee and Mogens Fosgerau, the co-editor of the issue, for their valuable suggestions.

Reference


Calvin L., Scovel III., 2009. Actions Needed to Meet Expectations for the Next Generation Air Transportation System in the Mid-Term. Testimony before the Committee on Transportation and Infrastructure, Subcommittee on Aviation, United States House of Representatives.


Appendix A: A proof of Equation (12) based on empirical data

Using the demand function (6) and considering the symmetry of the two airlines, the aggregate demand function in the market is

\[ Q = Q_1 + Q_2 = 2\alpha_0 - 2(\alpha_1 - \alpha_2)P_m - \frac{4(\alpha_1 - \alpha_2)\gamma}{f_m} \]  \hspace{1cm} (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

\[ \varepsilon_f^0 = \frac{\partial Q}{\partial f_m} \frac{f_m}{Q} = \frac{2(\alpha_1 - \alpha_2)\gamma}{fQ} \] \hspace{1cm} (A-2)

If (15) holds, then the LHS in (13) is monotonically decreasing. Rearranging the LHS term to the RHS and multiplying both sides by \( 3/2 \), (14) becomes

\[ [\alpha_0 - (\alpha_1 - \alpha_2)\tau] - \frac{3}{2} \frac{(\alpha_1 - \alpha_2)\gamma}{f} > 0 \] \hspace{1cm} (A-3)

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

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

\[ = Q_1 - \frac{1}{2} \frac{(\alpha_1 - \alpha_2)\gamma}{f} = \frac{Q - \varepsilon_f^0 Q}{2} = (1 - \varepsilon_f^0) \frac{Q}{2} \] \hspace{1cm} (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. (15) holds true.