Market-based Airport Demand Management
– Theory, Model and Applications

by

Terence Ping-Ching Fan

B.A.Sc., Mechanical Engineering, University of British Columbia, Canada (1997)
SM, Aeronautics and Astronautics, Massachusetts Institute of Technology, USA (1999)
SM, Technology and Policy, Massachusetts Institute of Technology, USA (1999)

Submitted to the Engineering Systems Division
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Transportation Systems and Policy Analysis
at Massachusetts Institute of Technology

February 2004
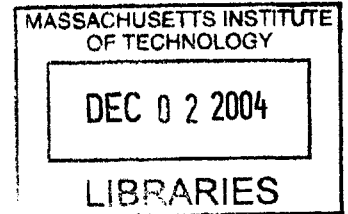
Author ....
Center for Transportation and Logistics, Engineering Systems Division
22 September 2003

Certified by ....
Amedeo R. Odoni
T. Wilson Professor of Aeronautics and Astronautics and of
Civil and Environmental Engineering, Thesis Committee Chair and Supervisor

Certified by ....
Nancy Rose
Professor of Economics, Thesis Supervisor

Certified by ....
Cynthia Barnhart
Professor of Civil and Environmental Engineering, Committee Member

Certified by ....
Peter Belobaba
Principal Research Scientist, Committee Member

Accepted by ....
Richard de Neufville
Professor of Engineering Systems and of Civil and Environmental Engineering
Chair of Education Committee, Engineering Systems Division

BARKER

Market-based Airport Demand Management – Theory, Model and Applications

by

Terence Ping-Ching Fan

Abstract

The ever-increasing demand for access to the world's major commercial airports combined with capacity constraints at many of these airports have led to increasing air traffic congestion. In particular, the scarcity of airside (take-off and landing) capacity at these airports has not been appropriately priced, leading to excessive demand as in the Tragedy of the Commons. Congestion pricing, as a classical economic approach to the efficient allocation of constrained transportation infrastructure capacity, has a long history of theoretical development. However, its application in the airport setting must deal with a set of important differences from the classical urban roadway setting. These differences have eluded the attention of researchers until very recently. They stem from the following set of complications: i) the peak and off-peak periods at congested airports are often less distinguishable than in the urban transport context; ii) airlines are a dominant intermediary between an airport's capacity and passengers as the end-users of that capacity; and iii) airlines operate groups of flights, as distinct from the atomistic behaviour of individual commuters.

To address these complications, an analytical model is developed to explore the impact of congestion pricing at airports and understand potential airline responses under a range of assumptions about the market's structure. Through a set of numerical experiments, carried out with the help of a probabilistic queuing model, we compare the economic benefits resulting from adopting fine versus coarse congestion tolls for the cases of markets with symmetric and asymmetric carriers. Given sustained demand for access to an airport and reasonably elastic responses in terms of frequency adjustments, the benefits to carriers of instituting congestion pricing generally exceed the amount of tolls collected. While a system of fine or graduated tolls is suited for all airports, systems of coarse or uniform tolls, which can be implemented more easily, are applicable only at airports with fairly symmetric carriers that hold approximately equal frequency shares.

In addition to congestion pricing, slot lease auctions can also be an effective means for promoting an economically efficient use of scarce airport capacity. In practice, the impact of slot lease auctions is similar to that of coarse tolling. Slot auctions are therefore applicable, in pure form, at airports with symmetric carriers. At these airports, a market-based demand management policy can comprise both congestion pricing and slot lease auctions. With respect to implementation, simultaneously ascending auctions recently used in the context of allocating electromagnetic spectra can be appropriately adopted to airports. A lump-sum subsidy can be used to promote specific socially desirable goals in the allocation of scarce airport capacity.

Several airport authorities around the world, currently using purely administrative or hybrid forms of demand management, have developed sophisticated techniques for defining and managing their constrained airport capacity. Some of these techniques can be useful in

developing market-based demand management policies. As an interesting case study, the experience of New York's LaGuardia Airport (LGA) in coping with a sudden increase in demand subsequent to the passage of the Wendell-Ford Aviation Act for the Twenty-first Century in 2000 is examined. The estimated impact of the temporary "slot lottery" at LGA demonstrates how even small reductions in the number of flights operated at a busy airport can bring about dramatic reductions in congestion delay. It also provides clear evidence of the extent of under-pricing access to many congested airports in the United States. The experience at LGA is contrasted with two other representative airports in the US to demonstrate the different policy needs depending on the specific airport characteristics.

Thesis Co-Supervisor: Amedeo Odoni
Title: T. Wilson Professor of Aeronautics and Astronautics and of
                        Civil and Environmental Engineering, MIT

Thesis Co-Supervisor: Nancy Rose
Title: Professor of Economics, MIT

To my parents

Acknowledgement

Content

List of Figures

8

List of Tables

Chapter 1

Introduction

Airport congestion in the US has been the focus of much public attention in the late 1990's. At least 58 articles were published on this topic in the *New York Times*, *USA Today*, and *Wall Street Journal* in 2000 alone (Mayer and Sinai, 2002). According to the Airline Service Quality Performance (ASQP) database[1], compiled by the Bureau of Transportation Statistics in the Department of Transportation, only 63% of all reported flights in 2000 arrived within 15 minutes of the scheduled arrival time, compared with 78% the year before. In particular, among the 25 carrier-route combinations that experienced at least 90% late arrivals in 2000 by this definition, 18 had New York LaGuardia airport as either the origin or destination (US Department of Transportation, 2000/01). This occurred against a backdrop of ever-longer scheduled flight times for the same routes (Mayer and Sinai, 2002 and Bratu, 2003). While air traffic in the US experienced an abrupt drop following the terrorist attack of September 11, 2001, it is expected to revert to a long-term growth trend. Congestion may thus become a major policy issue once again.

The problem of airport congestion belongs to a broader class of problems known among economists as the "Tragedy of the Commons", in which the under-priced public utility attracts excessive demand. In microeconomic theory, the general problem of externality arises when the cost of a service or product to an individual or company (the private or *"internal"* cost) is not equal to the total cost to society (i.e., full social cost). Negative externalities occur when in the absence of regulation the *internalized* cost is less than the full social cost, i.e., the action of one individual imposes a net cost on others.

---

[1] This includes U.S. domestic jet services operated by eleven major carriers to/from 29 major airports.

A classic example of negative externalities is an industrial outfit discharging harmful pollutants to the environment for which the firm is not appropriately charged. In this example, the cost of producing an incremental unit of product to the firm does not account for the cost that society incurs in cleaning up the pollutants discharged. In other words, the marginal private cost, or MPC, is less than the marginal (total) social cost, or MSC. The difference between MSC and MPC is the *external* cost imposed on others by this firm.

When the MSC exceeds the MPC, appropriate taxation can induce a profit-maximizing firm to produce at a level that is socially optimal. The development of this reasoning can be traced to Pigou (1920) and Knight (1924). A tax designed to induce firms to produce (or consumers to consume) at the socially optimal quantity has subsequently been referred to as a "Pigouvian tax". This tax is set such that each firm faces a total marginal production cost equal to the MSC (this is also known as "first-best" pricing when all prices are at the socially optimal level), i.e., is equal to the difference between MSC and MPC at the socially optimal level of production.

Baumol and Bradford (1970) provide an overview of many research papers on the broader topic of optimal taxation. Among many papers in this area, Diamond (1973) noted the possibility of imperfect corrective pricing when different participants give rise to different externalities. In transportation, aircraft operators, as the main contributors to airport airside congestion in the US, tend to exhibit characteristics similar to the consumers described in Diamond (1973). This similarity will provide the starting point for the analytical framework developed in Chapter 4 of this thesis.

Side-stepping the recommendation of Pigou and Knight, Coase (1960) demonstrated that, given well-defined and easily tradable property rights, externalities could be internalized through transaction cost-free negotiations. These can be conducted in the absence of government intervention. In other words, for the polluting firm example above, instead of paying a Pigouvian tax to the government, the firm could compensate

those whose right to a clean environment are harmed by the pollution. Easily distinguishable and enforceable property rights are therefore two key implicit assumptions. This reasoning is also the basis of the "buy-and-sell" feature of the High Density Rule (HDR) at several US airports, a hybrid demand management system that will be discussed in greater detail in Chapter 2.

In the context of transportation, Vickrey (1963, 1969) applies the concept of tolls to transport infrastructure. Assuming that the congested transport infrastructure can be physically expanded to relieve congestion, an optimal investment policy would be to expand the facility when the congestion cost exceeds the capacity expansion cost. Users of transport infrastructure should therefore be charged either the congestion cost (the additional cost one user imposes on other users), or the capacity expansion cost (should expansion be warranted).

Mohring and Harwitz (1962) demonstrated that the revenue from congestion tolls would fall short of construction costs if there were economies of scale in providing the infrastructure capacity, and would exceed the cost if there were diseconomies of scale in capacity expansion. In cases where simple marginal cost pricing is inadequate to fund the transport infrastructure, "Ramsey" pricing (1927), where consumers' charges are inversely proportional to their demand elasticities, is advocated. Oum and Zhang (1990) discuss the case where discontinuities in expanding transport capacity may necessitate some operating budget deficits early and surpluses later on in the life cycle of the capacity.

Other notable work on the economics of congestion on urban roadways and highways includes Arnott et al (1993), which incorporates certain dynamic elements (e.g., inter-temporal substitution) of deterministic queues in the analysis of congestion bottlenecks, Arnott and Kraus (1995), which shows that marginal cost pricing is still possible in cases where commuters behave in the same way in travel but differ from one another in shadow values of time and work, and McDonald (1995), which examines the use of second-best pricing where congestion pricing is possible only on a small part of

the urban road system. Small (1992) provides a survey of the literature on the analysis of congested urban roads.

In the context of airports, Levine (1969) and Carlin and Park (1970) provide some of the earliest discussion and numerical illustrations on airside congestion pricing. The former argued for the use of proper pricing in solving airport congestion, while the latter empirically estimated the marginal cost of airport airside congestion. Walters (1978) provides a survey on the early literature on airport economics. Morrison (1983, 1987) relied on regression analyses to estimate airport airside congestion delay and the appropriate tolls, while Daniel (1995), and Daniel and Pahwa (2000) demonstrate via stochastic queues the different ways of implementing congestion tolls at hub airports. For a comprehensive view on recent congestion pricing plans considered at several major airports, as well as some of the political cases for and against it, de Neufville and Odoni (2003) provide an extensive exposition.

Until this point, the analysis of airport congestion relies on the same principles as have been developed in the context of urban (highway) transportation. However, airport airside congestion is more complicated than urban transportation in at least the following ways:

i) The latter often has clearly defined, relatively short peak periods in the midst of non-peak hours. Thus, the inter-temporal analysis can be conveniently divided into a static analysis of the peak period, a static analysis of the non-peak period, and a peak-versus-non-peak substitution analysis. The distinction between peak versus non-peak periods can be much less clear in airports, i.e., congestion can be present all day (thus necessitating the dynamic modelling of queues – explored in more detail in Chapter 3).

ii) At airports, airlines act as an intermediary between the transportation facility's capacity (runway systems) and the end-users of this capacity (i.e., passengers), while there is no such equivalent intermediary in the urban or highway

context. The complete equilibrium analysis of the former can therefore be very complicated, although some simplifying assumptions can render the analysis sufficiently tractable (as will be discussed in Chapter 4).

iii)   In the urban/highway context, individual commuters (automobile drivers) act in an atomistic fashion, with each individual among many making his or her own decisions whether or not (or when) to drive. At airports, the airline intermediary decides whether or not to operate a *group* of flights, which altogether can account for a significant portion of all the flights operated in any time period; the ramifications of this will be discussed in Chapter 4.

The last point is recognized by Daniel (1995), Hansen (2002), and Brueckner (2002, 2003). In particular, a deterministic queuing model based on the notion of cumulative flow diagrams was used in Hansen (2002) to estimate congestion delays and external delay costs at Los Angeles International Airport (LAX). The analysis led to conclusions strikingly similar to those presented in Chapter 3. Brueckner (2002, 2003) shows that the traditional principle of congestion pricing requires modifications at airports where certain carriers have considerable frequency shares (e.g. hub carriers at their hubs). This thesis takes all of the points i) through iii) into consideration.

Instead of treating congestion pricing as the only viable policy instrument, this thesis broadens the discussion to include other forms of demand management that are also aimed at constraining or reducing the number of flight operations within a specific time period at busy airports. A notable economic approach orthogonal to the concept of congestion pricing (where the airport authority sets a socially desirable price) is slot auction (where airport users name their prices for slots and only the highest bidders will be allocated the capacity). This, together with Hansen's proposal (2002) on how to determine which specific flights should be denied access to scarce airport capacity under demand management, can be drawn upon to transform market-based remedies for airport congestion into public policy.

Different demand management strategies have in fact been practiced at most airports worldwide. While none of these conform exactly to the economic approach described earlier, many still offer important insights as to how the limited take-off and landing capacity at a congested airport can be effectively managed. These experiences, including those in the US, will be examined briefly in Chapter 2.

In particular, the experience at New York's LaGuardia airport offers a glimpse of the enormous impact demand management can potentially have on relieving airport congestion. This, together with the method through which congestion delays are modeled, will be discussed in Chapter 3.

An analytical framework designed to investigate the impact of different market structures on the efficacy of demand management will be developed in Chapter 4. The basic framework developed will then be followed by numerical experiments. Notable insights based on the findings of Chapter 4 will be recast in the public policy context in Chapter 5. In particular, an alternative mechanism to congestion pricing, namely slot lease auction, will also be introduced in Chapter 5.

Chapter 2

Brief Survey of Airport Demand Management Practices

No airport has to date practiced a pure form of congestion pricing, although one such proposal was close to being formally adopted at Boston Logan Airport (see de Neufville and Odoni, 2003). This is perhaps partly a consequence of the complex issues mentioned at the end of Chapter 1. Instead, past and current industry practices toward relieving airport congestion have ranged from a totally non-interventionist approach in the overwhelming majority of airports in the US (Section 2.1), to purely administrative procedures adopted at most major airports outside of the US (Section 2.2). The worldwide experience with these practices is briefly surveyed in Fan and Odoni (2002) and in this Chapter.

Occasionally, one can find in practice policy elements that combine economic principles with administrative procedures. These so-called hybrid approaches will also be examined in this Chapter (Section 2.3). While none of the congestion relief practices, either non-interventionist or purely administrative, exactly match the kind of market-based demand management policies advocated in Chapter 5, specific elements of the practices described in this Chapter, as well as the lessons learned offer important public policy insights.

2.1   Experience of a Largely Non-interventionist Approach – the Case of the US

The deregulation of the airline industry in the US has made possible free entry and withdrawal of services, freedom in scheduling frequency and equipment, as well as pricing among domestic carriers. In response, the hub-and-spoke route system has been

widely adopted by major carriers. The economics and policy implications of "hubbing" have been extensively studied, notably by Borenstein (1989), McShan and Windle (1989), Kahn (1993), Borenstein and Rose (1994), Levine (1994), Zhang (1996), and Borenstein and Netz (1999). The concentration of operations at hub airports is an important cause of congestion delays, as demonstrated by Mayer and Sinai (2002), and an effective demand management policy must take this into account.

No form of demand management is currently in use as a congestion relief measure at any commercial airport in the U.S., except for the four operating under the High Density Rule, or HDR. The experience of HDR at these four airports will be discussed in Section 2.3. Apart from these four airports, however, the U.S. environment offers important insights into the pattern of airport usage resulting from a largely non-interventionist policy regarding airport access.[2]

Throughout the past few decades, demand for passenger air transport in the U.S. has been rising steadily. Figure 2-1 shows that, except for 1991 and 2001/2002, the revenue passenger-miles (RPM, a standard measure of passenger air transport demand) performed by U.S. carriers in the scheduled domestic market (including regional carriers that report traffic statistics to the US Department of Transportation) have increased in every year since deregulation. Included in this graph are operational statistics from all major US trunk carriers and their predecessors during the period sampled (including Alaska and Southwest), as well as most other national and regional carriers that reported operational statistics in the same period (including JetBlue, AirTrans, American Eagle, Atlantic Southeast Express, Continental Express, Executive Airlines, Horizon Air, Mesaba, Mesa, Midway, Midwest Express, TransStates, and Air Wisconsin). The RPM in 2000 were roughly three times the number in 1982, resulting from a remarkable average annual growth rate in excess of 6% over an 18-year span. While there has since been a small decline in demand in the aftermath of the terrorist attacks, the longer-term projection is still for continued traffic growth.

---

[2] At some airports, notably Santa Ana/Orange County, the number of aircraft movements is constrained due to noise concerns but not congestion. These instances fall outside of the scope of this thesis.

Figure 2-1. Long-run Passenger Air Travel Demand in the U.S.
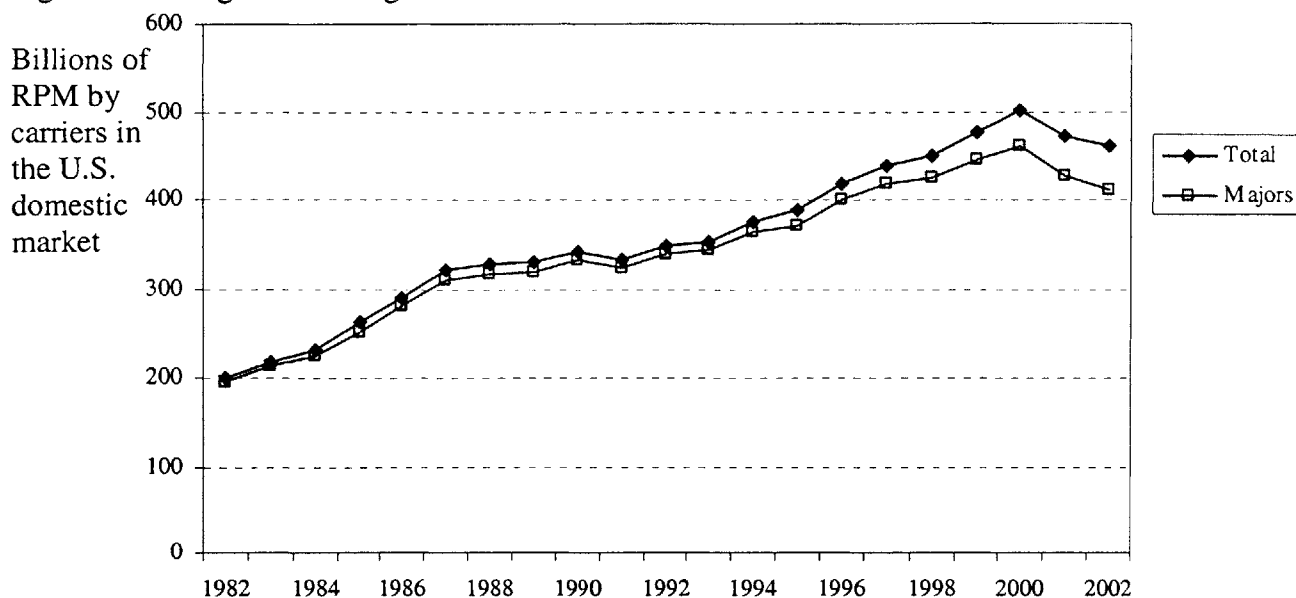
Source: U.S. Department of Transportation (DoT) Form 41

To accommodate long-term passenger growth on any given market, air carriers can pursue any combination of four options: i) increase the number of existing flights offered (many of which are designed to facilitate connections at hub airports); ii) increase the number of direct, hub-bypassing flights; iii) increase the size of the aircraft they fly; or iv) operate with higher load factors (i.e., the percent of seats occupied by revenue-paying passengers). There are two important points worth noting from these options. First, note that i) and ii) differ in that while the former has little impact on the average stage-length flown and *exacerbates* congestion delays at hub airports, the latter generally increases the average stage-length flown and tends to *relieve* congestion at hubs. Second, if carriers prefer to pursue either (i) or (ii), but not (iii), not only will the access to under-priced airports be over-subscribed, but also the manner in which scarce airport airside capacity will be utilized will be an unnecessarily inefficient one.

Based on the Form 41 data reported from the US Department of Transportation, Figure 2-2 decomposes the RPM growth into its component trends, and demonstrates that

the strong RPM growth of U.S. carriers during the 1990's was achieved through a combination of longer stage lengths, higher load factors and more flights.

Figure 2-2. Component Trends for Passenger Demand in U.S.
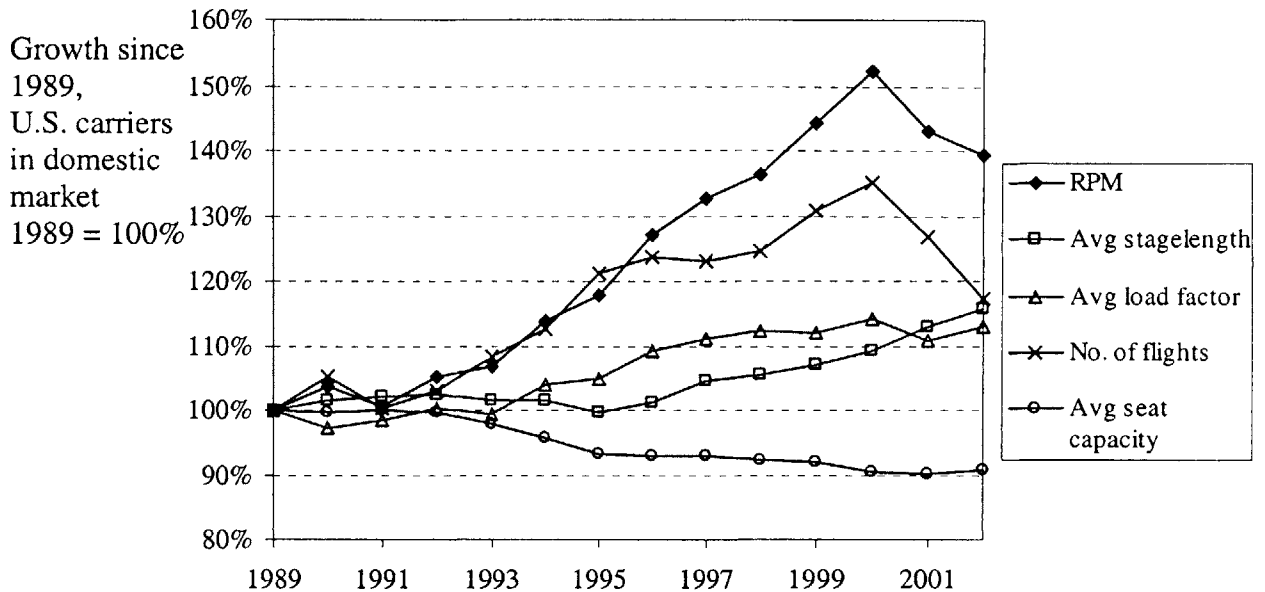
Source: US Dept. of Transportation, Form 41.

The dramatic 15% rise in average stage-length from 1995 to 2002 is evidence of the growing prevalence of longer, hub-bypassing flights. As reasoned above, this hints at the worsening congestion at non-hub airports compared with hub airports. Meanwhile, the average aircraft size operated by trunk carriers, as measured by available seats per flight, has declined noticeably by about 10% since 1992. Since the decrease in aircraft size occurred over a full decade, it is reasonable to conjecture that this preference for small aircraft to accommodate growth is a sustained and rational economic response, *largely in the absence of airport demand management*, by profit-maximizing carriers. This observation is consistent with a widely accepted tenet of the airline industry: the benefits, in terms of increased market-share, which can be obtained from additional flight frequency on competitive markets, often outweigh the per-seat cost advantages provided by larger aircraft. In addition, Fan (2002) shows that, as long as the willingness to pay among passengers spans a broad range and an effective revenue management system is in

place, smaller aircraft may produce higher operating profits than larger ones, even with the same frequencies and market shares.

The preference of carriers for small aircraft and high frequencies need not be a problem, as long as there is adequate airport capacity. This, unfortunately, is not the case at many major U.S. airports. There is abundant evidence of worsening air traffic congestion in the U.S. before September 11, 2001. Partly as a response, airlines in the U.S. have increased their scheduled block times (or gate-to-gate times) between city pairs over the years. Shumsky (1993) found that the average scheduled block time of 216 randomly selected flights, which reflected the overall mix of flight-length characteristics, increased from 107 minutes to 117 minutes between 1987 and 1991. Bratu (2001) similarly estimated that the average scheduled block time on the 1,000 most frequently travelled routes in the U.S. increased from 121.9 to 125.7 minutes between 1995 and 1999. Taking into account all reported delays from primarily major US domestic carriers, Mayer and Sinai (2002) estimate that flight delays are increasing in hubbing and decreasing in airport market concentration, but the hubbing effect dominates empirically.

In the absence of major capacity increases in at least the short- to medium-term, as demonstrated by Table 2-1, the domination of strategies emphasizing increased flight frequencies may portend the eventual necessity of demand management measures for airport airside capacity. In particular, Figure 2-2 suggests that the perceived or internalized costs associated with growing congestion at many of the busiest airports in the U.S. in the past decade were generally insufficient to discourage carriers from pursuing the "higher frequency, smaller aircraft" strategies.

Under very specific circumstances, however, aircraft operators may internalize much of the *MSC*. Brueckner (2002) demonstrates this for the theoretical case of a perfectly price-discriminating monopolist. One realistic case is the operations of the cargo carrier Federal Express at an airport where it is essentially the only operator during certain parts of the day (e.g. Memphis at late night and early morning hours). Since the congestion costs generated by Federal Express aircraft movements fall almost exclusively

on other Federal Express movements, the company internalizes practically all the *MSC* when it comes to delay due to congestion. Stated differently, putting aside noise and pollution concerns, the *MPC* due to delays incurred by any night-time Federal Express flight at Memphis Airport is roughly equal to the *MSC*. It is to be expected then that Federal Express will operate in Memphis exactly the number of flights that maximizes its total economic welfare, taking into account the "external" (to other flights but internalized by the same carrier) delay costs associated with each flight. Moreover, Federal Express has significant flexibility in scheduling departure and arrival times: its customers do not care about the exact time their packages arrive at and depart from the Memphis transfer hub, as long as they arrive by, say, 9 a.m. the next morning, or by another specified time within the next business day at their final destinations. In the case of Memphis Airport, no demand management measure would therefore be necessary, with or without congestion. The aircraft operator can be relied upon to manage demand and congestion in an economically efficient way.

Table 2-1. Plans for Capacity Expansion at U.S. Airports with the Most Delays

Airport	Plans[a] for capacity expansion?
New York LaGuardia	No
Newark	No
New York Kennedy	No
Chicago O'Hare	Probable[b]
San Francisco	No
Philadelphia	No
Atlanta	Yes
Boston	No

Notes:
a. Short- to medium-term plans that will increase current capacity by 5% or more.
b. Probable, according to recent news clips.
Source: FAA Airport Capacity Benchmark Report 2001; FAA's OPSNET.

A parallel, although more complicated argument can be made about airports that are dominated by a single passenger airline. Figure 2-3 shows the peaks and valleys of scheduled flight movements (estimated from *Official Airline Guide*, 2000) at Atlanta Hartsfield Airport, where Delta Airlines has its main hub and carries more than 75% of

all passengers. An airline wishing to set up an efficient connecting complex (with banks of arriving flights alternating with banks of departing flights) with a short average connecting time, has no choice other than to distribute unevenly scheduled departure and arrival times. However, as shown in Figure 2-3, the airline can arrange so that the peaks of flight movements – that may often exceed airport capacity during short time intervals – are interspersed with "valleys" that provide sufficient "recovery time", at least on good-weather days, for delays incurred during the previous peaks. Since the dominant airline again internalizes much of the *MSC* in such cases, there is little role for *external* intervention in the form of demand management measures at such airports. The logical area for policy intervention would be to guard against monopolistic fare pricing.

Figure 2-3. Scheduled Flight Operations at Atlanta Hartsfield Airport (ATL)

Time of day (5 = 05:00 am – 05:14 am, etc.)
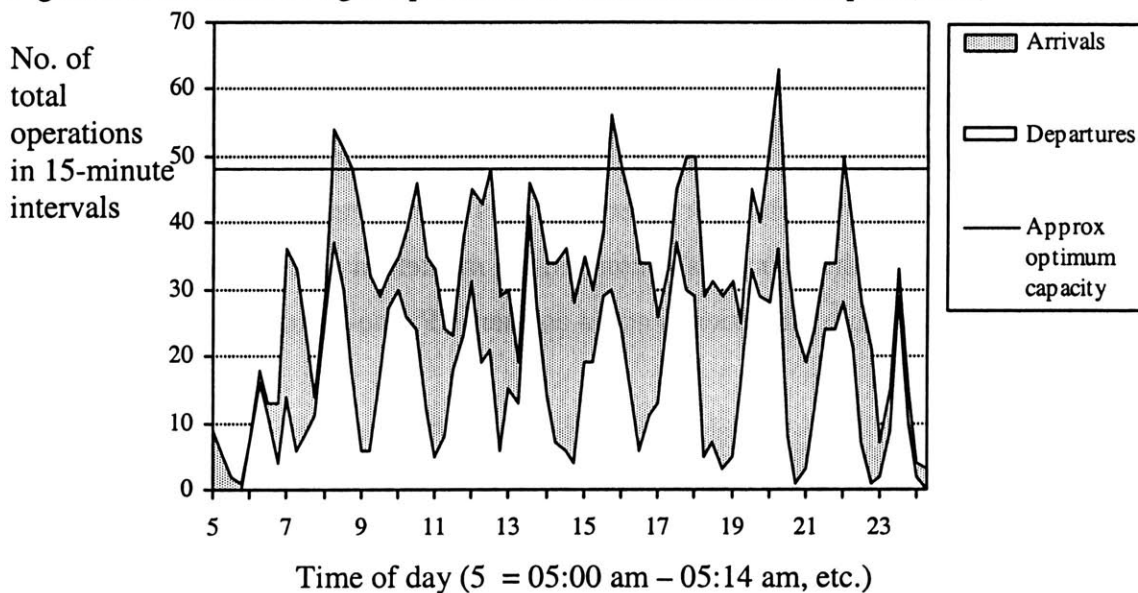
In contrast, at congested airports with a large number of competing carriers and with no dominant operator(s), little internalization of *MSC* takes place. In these cases, the *MPC* in general will be equal to only a fraction of the associated *MSC*. As a result, severe congestion may result in the absence of demand management. The experience at New York's LaGuardia airport during 2000 and 2001 constitutes a prime example. The insights from that experience will be presented in Chapter 3. In Chapter 4, the economics

and attractiveness of demand management in general will be examined in a more rigorous theoretical context.

In summary, the U.S. experience shows that in the absence of demand management in a deregulated environment, airlines tend to pursue a strategy of high frequency with small aircraft. At capacity-constrained airports, this heightens the need for demand management measures.

2.2   Worldwide Experience in Purely Administrative Schedule Coordination

In stark contrast to the non-interventionist approach in most of the US, many major airports around the world have long relied on purely administrative measures to manage the demand for airport access. These measures, commonly referred to as "schedule coordination", are introduced in this Section. It is important to note that while these measures have been in place for several decades at most major airports worldwide, they do not incorporate the principles of competitive markets in their operation (and, thus, the resulting capacity allocation is not in any sense Pareto-optimal in economic terms).

The International Air Transport Association (IATA) organizes Schedule Coordination Conferences (SCC) every November and June, where representatives of numerous airports, civil aviation organizations, and nearly 300 airlines from around the world meet to coordinate scheduled departure and arrival times in view of airport operational constraints. For the purposes of schedule coordination, airports are classified into three categories: *Level 1* (or *"non-coordinated"*) airports are those whose capacities are adequate to meet demand. *Level 2* (or *"coordinated"* or *"schedule facilitated"*) airports are those where demand is approaching capacity and "some cooperation among potential users is required to avoid reaching an over-capacity situation" (IATA, 2000). To this purpose, a "schedules facilitator" is appointed who seeks voluntary cooperation on schedule changes by the airlines to avoid congestion. *Level 3* (or *"fully coordinated"*)

airports are those deemed sufficiently congested to require the appointment of a schedule

coordinator whose task is to resolve schedule conflicts and allocate available slots. *All*

requests for slots at Level 3 airports must be reviewed and cleared by the schedule

coordinator. About 140 airports worldwide are currently designated as fully coordinated

and use IATA's slot coordination approach. For scheduling international flights, several

US airports (New York Kennedy, Chicago O'Hare, Los Angeles International, San

Francisco and Miami) are considered fully coordinated. Special provisions are made to

give priority to slot requests by international carriers at these US airports. Table 2-2

shows the level of schedule coordination at some of the world's busiest airports outside

the United States.

Table 2-2. Level of Schedule Coordination at Major Airports

City and Airports	Level 3	Level 2	City and Airports	Level 3	Level 2
Amsterdam	Yes		Osaka – Kansai		Yes
Brussels*	Yes		Paris – Charles		
Frankfurt-am-Main	Yes		de Gaulle	Yes	
Hong Kong	Yes		Paris – Orly	Yes	
Johannesburg	Yes		Rome – Fiumicino	Yes	
London – City		Yes	Seoul – Incheon	Yes	
London – Gatwick*	Yes		Sydney	Yes	
London – Heathrow*	Yes		Tokyo – Haneda	Yes	
London – Stansted*	Yes		Tokyo – Narita	Yes	
Osaka – Itami	Yes		Toronto – Pearson	Yes	

*Airports with a peak-time premium on landing fees in addition to schedule coordination.
Source: IATA (2000)

Each fully coordinated airport must first specify a *declared capacity*, which

indicates the number of aircraft movements per hour (or other unit of time) that the

airport can accommodate. Prospective airport users must submit a formal request for each

and every desired slot. The declared capacity is rationed according to a set of criteria,

among which the principal and overriding one is historical precedent: an aircraft operator

who was assigned a slot in the same previous season ("summer" or "winter") and utilized

that slot for at least 80% of the time is entitled to continued use of that slot. Second

priority is assigned to requests aimed at extending seasonal scheduled service (previous

winter or previous summer) to year-round scheduled service (to the next summer or the

next winter, respectively). Any requests for slots for new services are processed thereafter. Once a slot is awarded in a particular season on the basis of historical precedent, the recipient may use it to serve a different destination from the one served in the previous corresponding season. Slot exchanges between airlines, on a one-for-one (and non-monetary) basis, are allowed. In many cases, short-term leases to code-sharing partners are also allowed.

Slots for scheduled services that have been discontinued, or not used at least 80% of the time, are returned to a "slot pool" for re-allocation. Note that this measure of *utilization* is distinct from the *efficiency* with which each slot is utilized. Any new slots made available through increased airport capacity are also placed in the pool of available slots. At least 50% of the slots in this pool are targeted for new entrants or limited incumbent carriers. However, the definition of a "new entrant" is very restrictive: an aircraft operator qualifies as a new entrant as long as it does not hold more than four slots on the same day, *after* receiving any new slots from the slot pool. Thus a new entrant can be awarded at most two flights per day, hardly sufficient to establish a significant foothold at a major airport.

There are significant variations in the level of sophistication with which these general demand management procedures are applied at different airports. For example, some airports utilize a simple limit on the number of movements that can be scheduled in any single hour of the day, while others may apply a combination of limits that may restrict the number of movements for intervals smaller than one hour. This is shown in Table 2-3 for a number of major international airports.

Note, for example, that Sydney Kingsford-Smith airport controls the number of movements down to the level of 5-minute intervals and utilizes a staggered enforcement period in order to smoothen any intra-hour peaks in the traffic schedule. At Tokyo Narita, the hourly limit is set between 26 and 32 movements per hour depending on the anticipated mix of departing and arriving aircraft, whereas at Frankfurt there are separate

quotas for the maximum number of departures and of arrivals that can be scheduled in any hour (48 and 43 per hour respectively).

Table 2-3. Implementation of Scheduling Capacity at Selected Airports

Airports	Limit of Scheduled Movements Per Interval (2001)							Remarks
	1 day	3 hour	60 min	30 min	15 min	10 min	5 min	
London Heathrow			79-85[a]					As long as average delay < 10 min
Tokyo Narita	367[d]		26-32[a]					Daily limit is a noise constraint
Frankfurt			78	43		16		
Seoul Incheon			37[b]					
Sydney			80[c]		21		8	
Osaka Kansai		81	30					

Notes:
a. Changes based on departure and arrival mix.
b. Enforced in 1-hour periods staggered by 30 minutes, e.g. 1000 – 1059, 1030 – 1129.
c. Enforced in 1-hour periods staggered by 15 minutes, e.g. 1000 – 1059, 1015 – 1109.
d. This includes a limit of 349 international flights and 18 domestic flights per day.

An even more flexible approach is in use at London Heathrow (LHR). First, as in Tokyo Narita, the declared hourly capacity may change by hour of the day and depends on the mix of departures and arrivals in each hour. Moreover, a marked difference exists between the number of slots available in the summer and winter seasons, as shown in Figure 2-4, in order to take into consideration the impact of unfavourable weather conditions in winter. In general, the number of slots at LHR is adjusted from year to year with the objective of maintaining the level of airborne holding to an average of 10 minutes or less per arriving flight.

The declared capacity at an airport may increase gradually over the years as a result of airport infrastructure expansion, air traffic control improvements and airport operator experience with the slot coordination system. The evolution of the number of slots made available at LHR and at Frankfurt over the last decade, as reflected in Table 2-4, illustrates this point.

Figure 2-4. Declared Total Hourly Capacity at London Heathrow, 2000/01

Total hourly
capacity for
allocation
(Departures and
arrivals)

Time of day (e.g. 6 = 6:00 a.m. to 6: 59 a.m. local time)
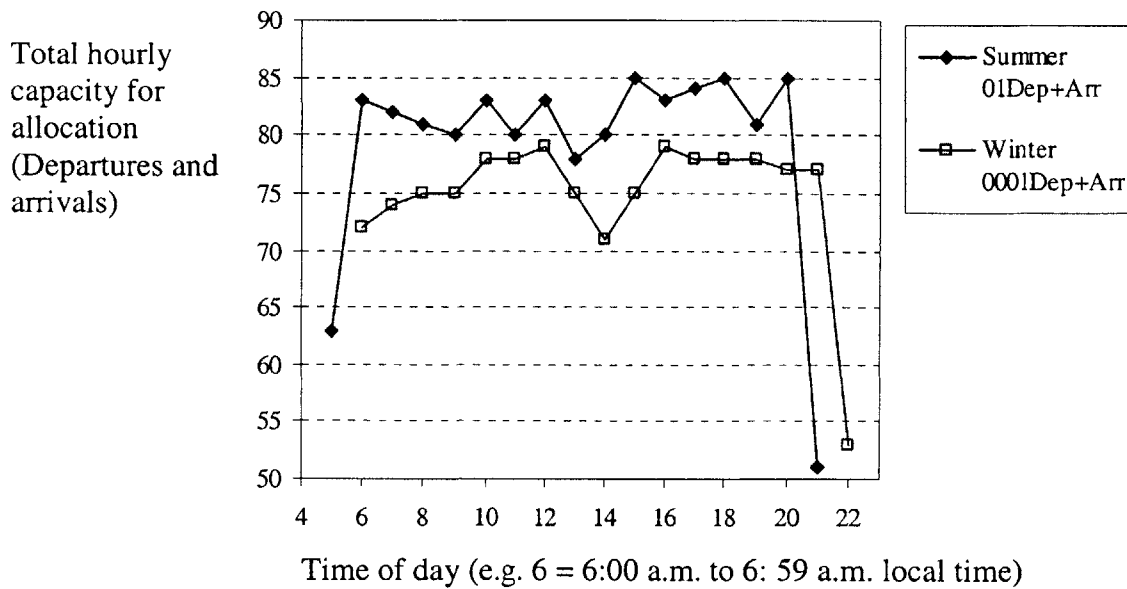
Table 2-4. Change in Total Number of Daily Slots Over the Years

Total daily slots	London Heathrow	Frankfurt-am-Main
1991 summer	1246	1056
1996 summer	1283	1152
2001 summer	1347	1248
Change, 91-01	8%	18%

Source: Slot coordinators for London Heathrow and Frankfurt-am-Main

Local authorities are given considerable leeway under the IATA system in specifying the declared capacity of an airport. Declared capacity need not be determined solely by the capacity of the runway system. Constraints due to the availability of aircraft stands, passenger terminal processing capacity and even aircraft ramp servicing capacity can be taken into consideration. In some cases, local authorities have also gone as far as segmenting airport capacity to serve particular public policy goals. Examples include the designation of blocks of slots reserved for international, domestic and general aviation traffic at Tokyo Narita, and setting aside slots to accommodate regional services within the State of New South Wales at Sydney Airport. At Tokyo Haneda, the Japanese

Ministry of Land, Infrastructure and Transport has even developed a "performance scorecard" which is used for allocating new slots among the three major domestic carriers.

The main appeal of IATA's purely administrative schedule coordination approach is that it "has [maintained] a high degree of coherence and stability in the international air transport system" (IATA, 2000). Indeed, this demand management mechanism has worked well in practice in instances where demand exceeds the supply of airport capacity by a relatively few operations and for only a small number of hours in a day. However, when a significant excess of demand over capacity exists, there is a clear risk that an approach that is entirely detached from economic considerations and incentives may lead to serious distortions of the marketplace. It can be argued that, at some of the most congested airports in the world, the schedule coordination process currently serves as a means for preserving the *status quo*, effectively acting as a regulatory device at the airport level. New competitors may be prevented from entering markets in an effective way, either by being denied slots altogether or by being relegated to slots at inconvenient times of the day (GAO 1990).

In response to such concerns, a number of governments around the world have been taking an increasingly active interest in this matter. In particular, the European Commission (EC) has conducted extensive reviews of the IATA airport capacity allocation procedures. In July 2000, it issued a consultation paper that proposed several fundamental changes in the way the procedures are applied at airports in the European Union (EU). Two key proposals (Baker, 2001; Pagliari, 2001) involved allowing some trading of slots via an auction and a secondary market, as well as a limit of between ten and twenty-five years on the length of time such traded slots can be held. Moreover, the EC proposal included provisions for easing market entry into congested airports by new carriers. Opposition from the Association of European Airlines (AEA) has led to the tabling of these proposed changes. It is probable, however, that the existing purely administrative slot allocation system will be replaced in the EU in the near future with a

hybrid system combining economic elements with administrative procedures. Some examples of hybrid demand management systems are described in the next section.

2.3   Worldwide Experience in Hybrid Approaches to Demand Management

While no airport currently adopts purely economic approaches to demand management as described in Chapter 1, some airports have combined certain features of purely economic and purely administrative approaches to create hybrid demand management mechanisms. In fact, considerable experience already exists with the application of two different types of hybrid demand management systems. The first has been used at some European airports, and the second at HDR airports in the U.S.

In the former case, the purely administrative schedule coordination process of IATA is supplemented with schedules of landing fees designed to discourage airport use during peak traffic periods. Table 2-5 shows the landing fees charged in 2001 at the three London Airports operated by the British Airports Authority. Note that during peak periods a landing fee essentially independent of aircraft weight is charged at Heathrow and Gatwick, creating an incentive for the use of large aircraft[3]. In addition, the BAA imposes a separate per-passenger handling fee, with a higher charge in effect during peak passenger traffic periods.

Unfortunately, the potential effectiveness of the schedule of landing fees shown in Table 2-5 is undermined by the fact that the amounts charged are too low to have much of an impact on most aircraft operators. The reason for this anomaly is that the magnitude of the landing fees is determined through a regulatory process that places a cap on the rate of return on the BAA's assets, taking into account the BAA's *total* revenues from both aeronautical and non-aeronautical (or "commercial") sources. Because the BAA has been highly successful in increasing non-aeronautical revenues over the years, its

---

[3] A surcharge is also imposed on noisy aircraft; this surcharge may range up to 120% of the basic landing fee for those aircraft that do not meet ICAO Annex 16, Chapter 2 requirements.

aeronautical charges have hardly increased (and in many cases have declined in real terms) ever since 1986.

Table 2-5. Basic Aircraft Landing Fee Schedule for London Airports, 2001

Fee per landing	Heathrow		Gatwick		Stansted	
Aircraft weight	Peak[a]	Off-peak	Peak[b]	Off-peak	Peak[c]	Off-peak
MTOW≤ 16 tons	£ 418.50	£ 130	£ 310.50	£ 75	£ 80	£ 70
16< MTOW ≤50	£ 465.00	£ 195	£ 345.00	£ 85	£ 120	£ 85
50 < MTOW	£ 465.00	£ 335	£ 345.00	£ 115	£ 195	£ 95
Special charge for MTOW>250	--	--	--	--	£ 335	£ 335

Source: British Airports Authority (BAA)
Note: MTOW = Maximum take-off weight of the aircraft
Addditional noise surcharges apply; 05:00 – 06:59 is shoulder period in summer.
a. 07:00-09:59, 17:00-18:59 GMT, April 1 to October 31; and around 00:00-04:59 GMT
b. 06:00-11:59 and 17:00-18:59 GMT, April 1 to October 31
c. April 1 to October 31

At Brussels Zaventem Airport, which experienced high rates of traffic growth until the demise of Sabena in 2002, the landing fee paid by an aircraft was computed on the basis of the following formula:

$$Landing\ Fee\ and\ Take\text{-}off\ Fee = U \cdot W \cdot E \cdot D$$

In this formula the landing fee depends on: a basic rate per metric ton, U, which is at €2.98 in 2003; the weight of the aircraft (W, between 25 and 175 metric tonnes); a multiplier (E) that is based on the environmental friendliness of the aircraft type, which ranges between 0.90 and 1.70[4]; and a "day-night" multiplier (D), which is set at 2 for flights taking off or landing between 11 pm and 05:59 am local time[5]. (Brussels International Airport Company, 2003). The day-night multiplier D was also used as a peak-time multiplier for several peak hours during the day prior to the demise of Sabena, and this can be re-instated if necessary.

---

[4] The use of the value of E=0.9 offers a discount or rebate to the least noisy aircraft.
[5] The use of D=2 for 23:00 to 05:59 is to discourage the nighttime use of the airport for noise alleviation.

A second, very different type of hybrid system is the one used at the four HDR airports in the United States: New York LaGuardia, New York J. F. Kennedy (only during certain hours of the day), Washington National (now Ronald Reagan) and Chicago O'Hare. When inaugurated in 1968, HDR operated as a purely administrative procedure. However, soon after the deregulation of the US airline industry, the Schedule Coordination Committees were unable to satisfy the demand for slots at these airports and remained in a deadlock for several years. The Department of Transportation abolished these Committees on December 16, 1985, and those carriers who held slots on this date became the effective "owners" of these slots (see, for example, Gleimer, 1996). Since that date, a "buy-and-sell" environment has existed at these airports. Slots can be traded in a market in which not only aircraft operators, but also other commercial entities can participate. Current holders of slots allocated to domestic operations[6] under HDR may sell or lease them, just like any other commodity, subject to certain restrictions. The U.S. Department of Transportation, however, has the right to withdraw slots at any time to satisfy any shortfall in the number of slots available for international flights and for Essential Air Service flights. In addition, a use-it-or-lose-it provision returns a slot to a pool of unused slots for re-allocation, if its current holder utilizes it for less than 80% of the time.

Gleimer (1996) reports that under the buy-sell rule the number of slot sales has declined over the years up to 1996. While there has not been a detailed study on this subject after 1996, the pattern of flight frequencies by different carriers at the four HDR airports in the US appears to suggest that this trend has continued. This decline appears to be related to the reluctance of carriers to allow scarce airport capacity resources to be transferred to their competitors. Some carriers have opted for leasing out slots, instead of selling them, even after reducing their own operations at the HDR airports. In this way, the slots meet the minimum use requirement, while continuing to be the property of the original owner. However, since these leases are often of short duration, the carriers holding the leases may be discouraged from making such long-term commitments as in

---

[6] Since 1998, Canadian carriers have also been able to freely monetize slot holdings through buying, selling and trading slots under HDR. See the 1995 U.S. – Canada bilateral, Annex II, § 1, paragraph 5.

building terminal buildings, intra-airport transport links and regional transport links. This is one potential drawback of granting carriers full ownership of slots in perpetuity. Meanwhile, the concentration of slot ownership at high-density airports has increased over the years.[7] For example, Table 2-6 shows that the share of daily departures at Chicago O'Hare held by the two dominant carriers has increased over time. It is not clear from Table 2-6, whether the buy-and-sell rule has encouraged competition in the case of Chicago O'Hare.

Table 2-6. Percent of Daily Departures at Chicago O'Hare

Airline	1985[1]	1987[1]	2000[2]
United	41.4	45.8	45.9
American	27.6	31.4	38.8
Northwest	8.3	3.9	2.4
Continental	5.8	5.0	1.8
Delta	5.5	4.8	2.1
US Airways	4.5	3.9	1.7
TWA	4.1	3.1	0.8
Others	2.8	2.1	6.5

Sources: 1) Borenstein (1988); 2) *Official Airline Guide*

Under the buy-and-sell provision of the HDR, slots can be traded only with the consent of the owner. For strategic reasons, the highest bidder for these slots may still be denied the right to purchase these slots from its competitors. In a more pro-competitive approach, the airport authority can administer slot auctions such that the slot tenure is allocated to the highest bidder. While slot auction in the airport setting has not been put into practice, experiences in auctioning electromagnetic spectra offer interesting insights on the potential usefulness of this approach. This will be discussed in greater detail in Chapter 5.

In short, hybrid approaches to demand management, which combine elements of both purely administrative and economic measures, hold interesting prospects in a comprehensive, market-based demand management policy. However, each of these

---

[7] See, for example, General Accounting Office (1990).

policy approaches seems to have some drawbacks in one way or another. Thoughtful planning is needed to put together a demand management policy that would be both socially desirable and acceptable to all stakeholders. In Chapter 3, the recent experience at New York LaGuardia Airport, one of the most congested in the U.S., will be examined in detail. Chapter 4 then develops a theoretical framework to analyze the appropriateness and attractiveness of congestion pricing at different types of airports and market environments.

Chapter 3

Potential Impact of Demand Management on Congestion Delay

While virtually none of the demand management approaches examined in Chapter 2 are market-based, one of the prime reasons behind their continued existence is the enormous congestion delay that the absence of these measures would cause. The primary objective of this Chapter is to demonstrate precisely the magnitude of congestion relief that demand management can bring about, thereby motivating the pursuit of a market-based demand management policy.

Section 3.1 describes a queuing model called DELAYS that is used to simulate airport airside congestion delays. Section 3.2 applies DELAYS to New York's LaGuardia airport, where a crucial change in the decades-old High Density Rule (HDR) was instituted in early 2000 and followed by a temporary demand management policy in early 2001. The experience highlights how current landing charges contribute to the over-utilization, i.e., excessive congestion at that airport. Section 3.3 compares the level of congestion at a few representative airports in the US, and draws insights on the important differences that must be considered in any nation-wide implementation of any market-based demand management policy.

3.1   The DELAYS Model

Carlin and Park (1970) is the first paper to our knowledge that gave an empirical estimate of the marginal delay cost of congestion caused by operating an excess of flights. It proceeds by identifying periods of time when the runway at New York

LaGuardia was "busy", i.e., with airplanes in a queue to depart. Any additional flight which is added to this aircraft queue must delay all subsequent flights in the same queue by the amount of time it occupies the runway, i.e., its "service time". This delay which is caused by any additional flight scheduled to depart at some specific time is thus a close approximation to its marginal delay that can be attributed to that flight. Multiplying this delay by the appropriate value of time gives the marginal delay cost.

Over the years, models that simulate the dynamic evolution of queues have grown in sophistication. This Section describes a probabilistic queuing model, DELAYS, that has been developed over the years at MIT. This model is well suited to estimate the impact of demand management measures in an airport setting. The reason behind the preference of this model to an empirical approach, such as that of Carlin and Park described above, has to do with the difficulties associated with in using the latter approach:

i) each airline has its own way of estimating the scheduled flight time of a particular flight, and it is thus difficult to measure "delay" against a single standard time;

ii) even if a standard metric of flight delay were available, delays can result from a variety of reasons apart from congestion at a given airport, including delays due to congestion in en route airspace or at other airports, delays related to aircraft crew availability and gate availability during peak hours. It is thus extremely hard to ascribe delays to local capacity constraints.

Instead a model DELAYS can be used to investigate how congestion is generated at busy airports, to explore the sensitivity of delays to changes that may result from demand management measures and to quantify the external delays caused by additional runway movements. While approximations must be made in constructing such a model, this approach is extremely useful as a means of "isolating" the effect of the local demand

vs. capacity relationship from other operational concerns, and hence of providing important policy insights.

The DELAYS model used in this thesis is one based on queuing theory. It is based on the following observations (Barnhart et al., 2003). Estimating delays for any set of actual or hypothetical conditions at a system of runways poses a challenge to operations researchers: the closed-form results developed in the voluminous literature on classical steady-state queuing theory are largely irrelevant – at least, when it comes to the really interesting cases. The reason is that the assumptions that lead to these closed-form results are generally not valid at busy airports. First, service rates and, especially, arrival rates typically vary considerably through the course of a day of airport operations – unless the scheduling of flights during peak times is artificially restricted by the imposition of "slot systems". Second, demand for use of the runway system may exceed capacity ($\rho > 1$), possibly for extended periods of time, most often when weather conditions are less than optimal. Any examination of airfield delays must therefore take explicitly into account the non-stationary nature of airport queues.

This has motivated the development of numerical approaches to the problem of computing airport delays. In his pioneering paper on the subject, Koopman (1972) made the following argument:

1. The non-homogeneous Poisson process provides a reasonable approximation of demand (arrivals and departures) for access to the runway systems of major airports.

2. The random variables that describe the duration of service times for arrivals and for departures at an airport have probability distributions that are "less random" than the negative exponential and "more random" than deterministic (constant).

3. Hence, the queuing characteristics of an airport with k runways ("servers") are bounded by the characteristics of the $M(t)/M(t)/k$ and the $M(t)/D(t)/k$ queuing models, each providing "worst case" and "best case" estimates, respectively. Note

that this allows for dynamic changes in the service rates, as well as the demand rates.

Based on these premises, Koopman solved the differential equations and the difference equations describing, respectively, $M(t)/M(t)/k$ and $M(t)/D(t)/k$ queuing models, using the typical daily demand profiles for the runway systems of the John F. Kennedy International (JFK) and the LaGuardia (LGA) airports in New York. Plots of the expected queue lengths as a function of time over a 24-hour period showed that the results obtained from the two non-stationary queuing models were quite close to each other. Koopman thus argued that queuing characteristics at congested airports are rather insensitive to the precise form of the probability distribution of service times and that a reasonable interpolation of the results from the two models can offer a good approximation to the true queuing statistics.

Extending the work of Koopman, the $M(t)/E_k(t)/k$ system was proposed by Kivestu (1976) as a model that could be used to compute directly approximate queuing statistics for airports – rather than solving separately the $M(t)/M(t)/k$ and $M(t)/D(t)/k$ models and then somehow interpolating their results. Note that negative exponential service times (M) and constant service times (D) are simply special cases of the Erlang (E_k) family, with $k = 1$ and $k = \infty$, respectively. Kivestu suggested that k should be determined from this relationship:

$$\frac{E[S]}{\sigma_S} = \sqrt{k} \qquad \qquad ...[3\text{-}1]$$

where $E[S]$ denotes the expected value of the service time (runway occupancy time), and σ_S denotes and the standard deviation of the service time.

Both $E[S]$ and σ_S can be measured from field data or can be computed through one of the many existing airport capacity models. Kivestu also developed a powerful numerical approximation scheme that solves the Chapman-Kolmogorov equations for the

$M(t)/E_k(t)/k$ system efficiently. Malone (1995) has demonstrated the accuracy and practicality of Kivestu's approach and developed additional efficient approximation methods well suited to the analysis of dynamic airfield queues.

An implementation of Kivestu's model called DELAYS (Stamatopoulos, 2000) was used extensively in this thesis to model airport queues in this chapter and to perform the numerical experiments described in Chapter 4. In the DELAYS model, the runway system of an airport and their ancillary airspace are represented as a single server where aircraft queue up to depart from or arrive at. As depicted in Figure 3-1, the airplanes waiting to take-off and to land at the congested airport are modeled as one queue, with the same distribution of service times on the runway system. DELAYS takes as inputs the sustainable capacity of an airport's runway system in different time periods, and the number of flight operations demanded per period, and returns the probability distribution of the evolving aircraft queue. The model estimates the evolution over time of the combined queue of aircraft waiting to take-off from or land at the runway system, and does not take into account any additional delays due to surface movements, airspace congestion, congestion at other airports, etc. For convenience, the unit of time used for air traffic demand and for airport capacity is the hour, i.e, the demand and the capacity inputs are expressed in terms of movements per hour.

As already indicated, DELAYS models the service times of individual airplanes as an Erlang random variable. The value $k = 9$ was used reflecting the fact that the standard deviation of runway service times at airports are typically in the order of one-third of the expected service times (see [3-1]). The value of the hourly sustainable capacity, μ, is the inverse of the expected service time. The demand for this queuing system, i.e., the arrival of departing or landing aircraft at the end of this queue, is modelled as a non-homogeneous Poisson process, with a dynamic arrival rate of $\Lambda(t)$ flights in hour t, equal to the sum of the expected arrival rates of landing and departing aircraft in hour t. (The capital form of λ denotes the aggregate number of flights operated by all carriers.) The aircraft are served on a first-come, first-served basis, with no allowance for different service priorities.

39

Figure 3-1. Modelling LGA as a Single-Server, Infinite-Capacity System

Queue of departing and arriving aircraft   Departed or arrived aircraft

Legend: Departing aircraft; Arriving aircraft.
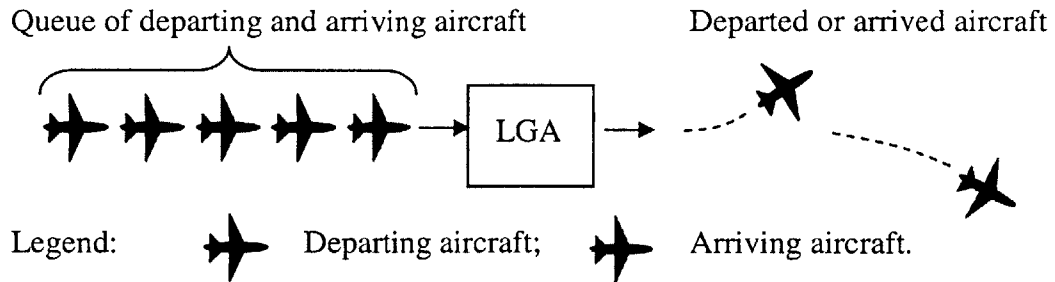
To check that the output of DELAYS is indeed realistic, the expected aircraft waiting times generated by DELAYS were compared with the observed flight delays at LaGuardia Airport (LGA) for selected days of November 2000 and August 2001. The field data were obtained from the Airline Service Quality Performance (ASQP) database of the U.S. Department of Transportation. U.S. carriers with more than USD $1 billion in revenues (last adjusted in 1999) are required to report for each of their flights the gate pushback time, the wheels-off time, the wheels-on time and the gate arrival time for at least 29 large airports to the ASQP database. In practice, these carriers report the statistics on many more airports than required. These statistics are published on a monthly basis.

In the US, most regional jets and virtually all turboprops are operated by separate subsidiaries of mainline carriers. Thus, statistics on these flights are not reported as part of the ASQP database. Assuming that these flights are randomly interspersed with other flights operated by mainline carrier jets and therefore experience similar delays at major airports, the average flight delays generated from the ASQP database should still be indicative of the level of congestion experienced by all the flights.

In DELAYS, the departing and arrival aircraft queues are considered as one queue, and the expected waiting time that the model computes refers to the time that the airplanes have to wait for their turn to take off or land. A comparable measure of delay can be estimated from the ASQP data for departing aircraft by subtracting an estimated taxi time from the "taxi-out" time, i.e., the actual amount of time that elapses between

40

pushback from the passenger gate and the moment of "wheels-off" (when the airplane takes off from the runway). For arriving aircraft, such a demarcation is much less definitive. As such, only the average delays for departing aircraft in the ASQP database were used for comparison with the delays estimated by the DELAYS model. A standard taxiing time of 15 minutes between the runway and the passenger terminal is assumed, i.e., any extra time beyond 15 minutes between "pushback" and "wheels-off" is considered congestion delay.

To compare actual delay statistics from the ASQP database with those predicted from DELAYS, several days in November, 2000 and in August, 2001 were used to represent the pre- and post-slot-lottery situation at LGA (see Section 3.2). Schedule and fleet assignment changes after the slot lottery had stabilized by August 2001. This was also the last full month before demand at LGA declined dramatically due to the September 11, 2001 events. To minimize adverse influences on airport capacity from poor weather conditions, only those weekdays with good visual flight rule (VFR) conditions have been chosen (e.g., visibility at 10 miles or more, clear or broken clouds above 2,500 ft, wind speed less than 10 knots for the entire day, and similar conditions for at least one day before). These requirements resulted in the selection of the dates of November 3, 8 and 13 in 2000 and August 15 and 22 in 2001. It is important to note that fine weather persisted for a few consecutive days around those selected dates in November, assuring near-perfect flying weather on the dates selected. In August 2001, weather conditions were more variable than in November, with intra-day cloud and wind variations as well as no consecutive days with near-perfect weather conditions.

To add more realism to the DELAYS predictions, the published hourly flight schedules (from the *Official Airline Guide*) were reduced by the same cancellation rate observed in the ASQP database in the corresponding hour on the same dates. For example, if, on 13 November 2000, 2% of the departing and arriving flights between 08:00 and 08:59 were cancelled according to the ASQP database, then the number of flights published in the OAG schedules for this hour (which is assumed to be the demand for the hour) is also reduced by 2%. This implicitly assumes that the hourly cancellation

rate for turboprop and regional jets (not reported in ASQP) is the same as that observed for jets operated by major carriers (reported in ASQP). Further, four general aviation flights per hour are added to the modified schedules between 06:00 and 23:59, and a capacity of 81 flights per hour, corresponding to the reported sustainable VFR capacity for LGA, is used. To account for the potential variations in the cancellation rate among large jet versus regional jet and/or turboprop operations, two DELAYS runs were generated: i) based on the original OAG schedules plus the general aviation flights (OAG + GA), and ii) based on the OAG schedules, with the assumed cancellations subtracted from it, and with the general aviation flights added (OAG – cancellations + GA). At this level of rough approximation, the model should be considered adequate if the actual ASQP data are close to and ideally fall somewhere between the two sets of delay estimates produced by the model.

Figures 3-2, 3-3 and 3-4 compare the predictions from DELAYS with the actual observations from ASQP for the three dates in November 2000 (pre-lottery): 3 November, 8 November and 13 November respectively. For the first two of these three days, the ASQP observations closely resemble the OAG + GA prediction, while for the third day they were above the OAG + GA prediction before late afternoon but between the OAG – cancel + GA and the OAG + GA predictions for the rest of the day. Figures 3-5 and 3-6 show the comparisons for two days in August 2001 (post-lottery): 15 August and 22 August respectively. Except for the morning (and early afternoon for 22 August), the ASQP observations generally fall between the OAG – cancel + GA and the OAG + GA predictions.

Based on the comparisons reported in Figures 3-2 through 3-6, the performance of the DELAYS model was deemed reasonable for the purposes of this thesis. The model seems to capture adequately (for the level of approximation required in our subsequent work) both the general time pattern and the order of magnitude of the delays inferred from the ASQP data. As a more general observation, it is noted that the validation of airport delay models using the ASQP database is, in general, difficult due to the facts that

the ASQP data i) are available for only a subset of all flights and ii) include delays that may be due to a great variety of causes in addition to runway congestion.

Figure 3-2. Comparing DELAYS with ASQP Data at LGA (3 Nov 2000)

Average delay (minutes per flight)
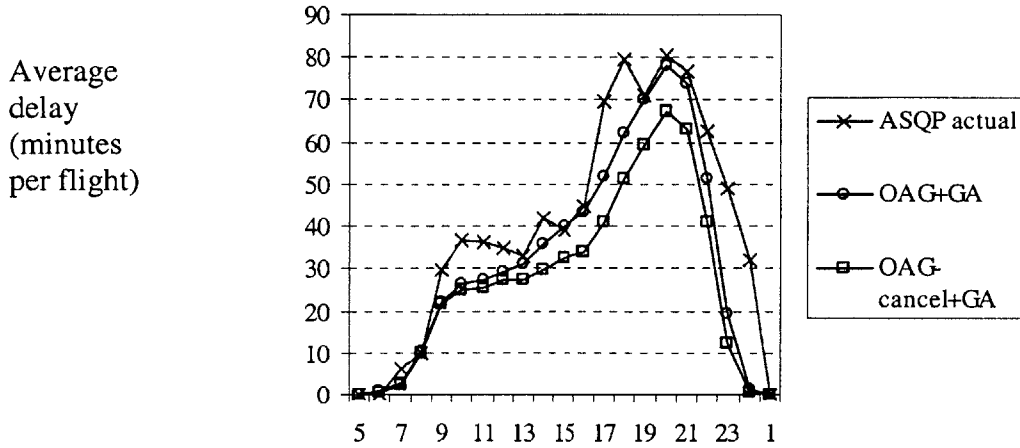
Time of day (e.g. 5 = 05:00 – 05:59 am) of originally scheduled departure/arrival times

Figure 3-3. Comparing DELAYS with ASQP Data at LGA (8 Nov 2000)

Average delay (minutes per flight)
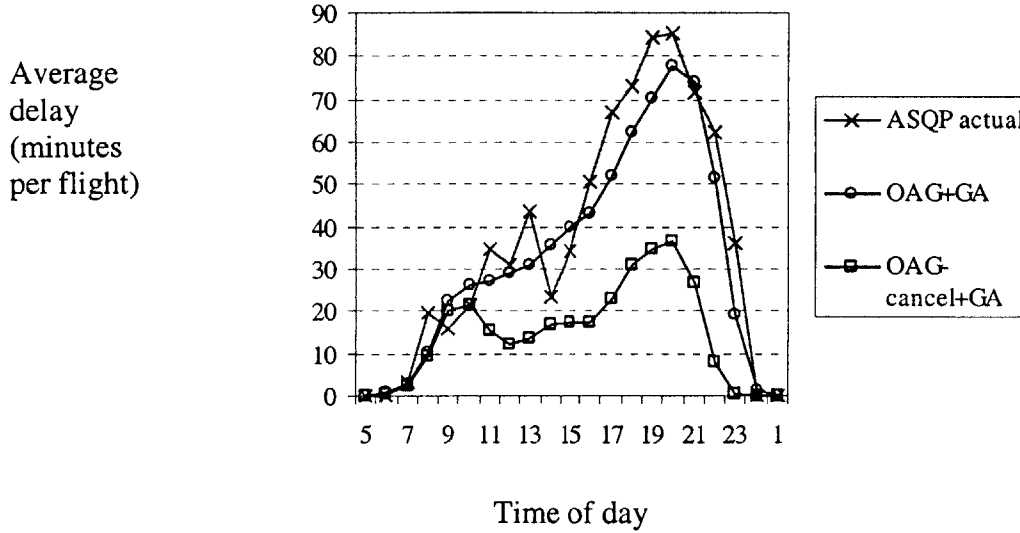
Time of day

Figure 3-4. Comparing DELAYS with ASQP Data at LGA (13 Nov 2000)

Average
delay
(minutes
per flight)
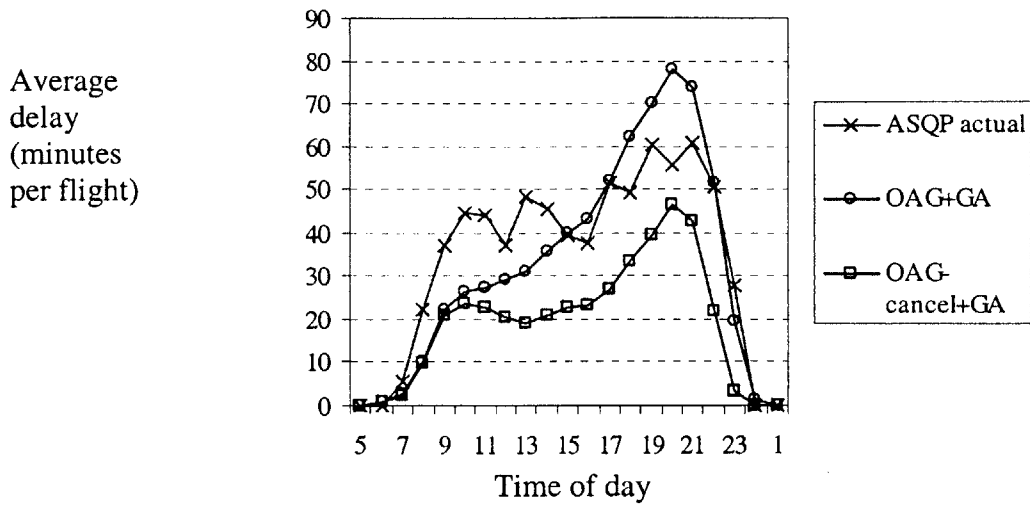
Figure 3-5. Comparing DELAYS with ASQP Data at LGA (15 Aug 2000)

Average
delay
(minutes
per flight)
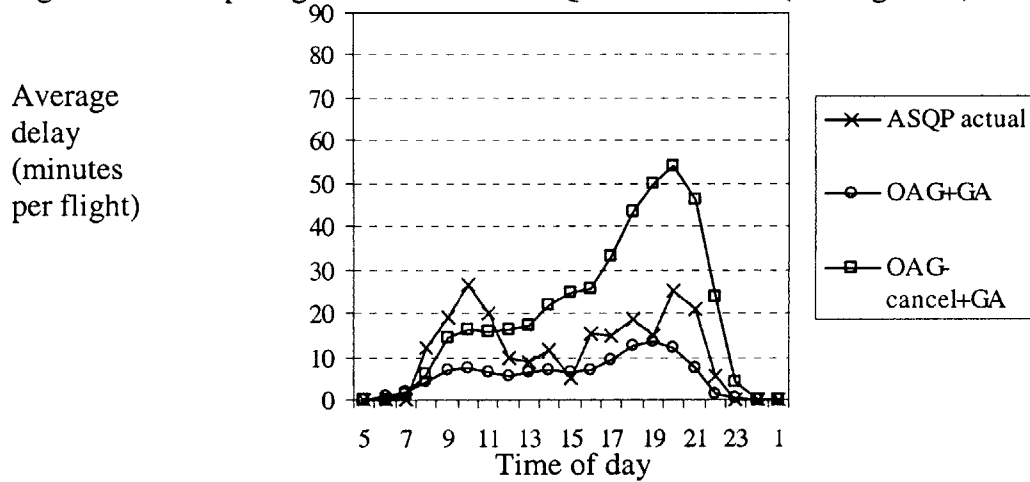
Figure 3-6. Comparing DELAYS with ASQP Data at LGA (22 Aug 2000)

Average
delay
(minutes
per flight)
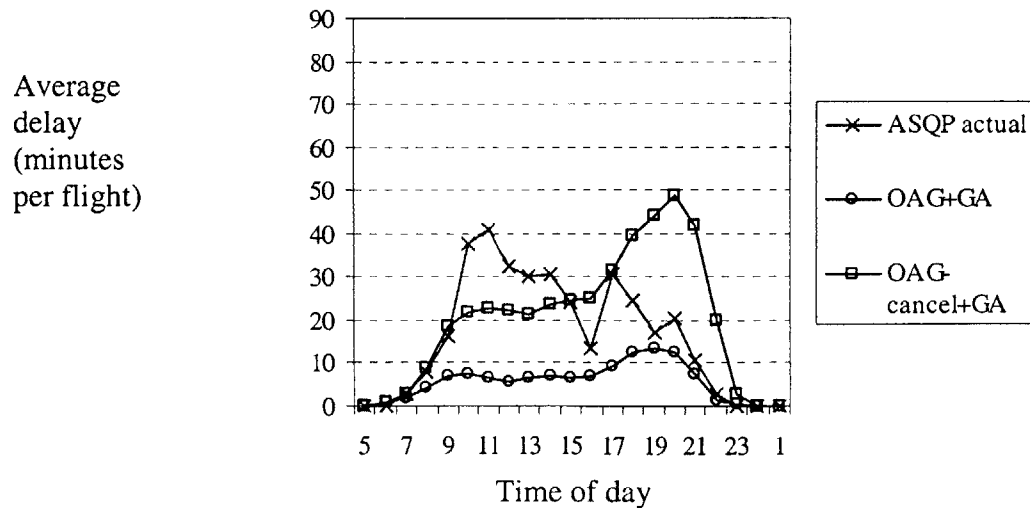
3.2   The Case of LaGuardia

With an adequately realistic model available, congestion delays at major airports can now be examined in greater detail. In particular, DELAYS will now be used to demonstrate how small changes in the number of flights at an already busy airport can have a dramatic impact on congestion delays. This provides a critical motivation for the detailed examination of market-based demand management measures that are the main subject of this research.

Among various major airports in the US, the case of New York's LaGuardia Airport deserves special attention in any discussion of demand management, and will be the focus of this Section. As one of the first airports to be administered under the High Density Rule (HDR) in the late 1960's, LaGuardia (LGA) has long been one of the busiest in the U.S. However, LGA is physically constrained. Sitting on only 680 acres of land, LGA has only two perpendicular and intersecting runways as shown in Figure 3-7. Arriving and departing air traffic must be carefully synchronized. While the current airfield infrastructure can be upgraded to handle somewhat larger aircraft, LGA does not have the physical space to add more runways.

Figure 3-7. Schematic of the Intersecting Runways at LaGuardia Airport (FAA, 2001b)
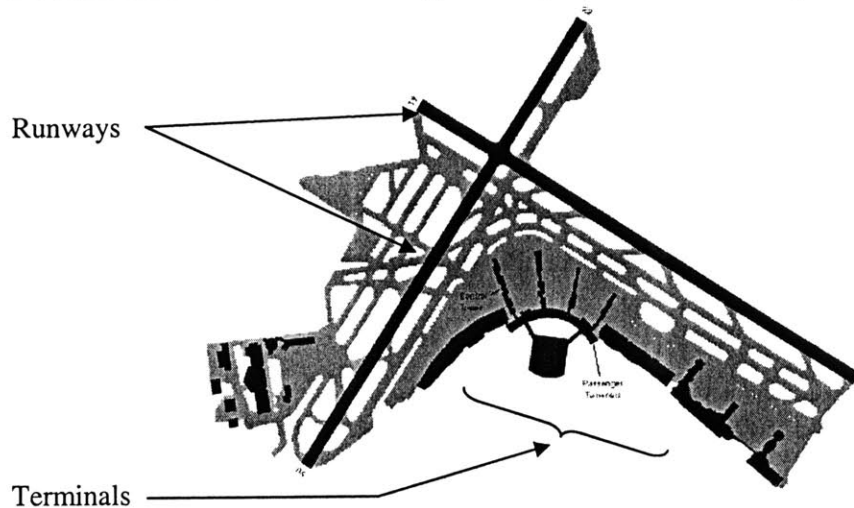
As mentioned in Chapter 2, LGA, together with three other highly congested airports – New York John F. Kennedy, Newark, Washington National (now Ronald Reagan) and Chicago O'Hare – has had its allowable number of aircraft operations restricted since 1968 under the High Density Rule (HDR). The purpose of HDR was to contain delays at these airports. Newark was removed from HDR in October 1970. In addition to HDR, a perimeter rule has also been in place at LGA, limiting new scheduled flights to and from LGA to no more than 1,500 miles. In the spring of 2000, the allowable number of flight operations per hour at LGA was constrained to an average of about 1,050 scheduled flight operations (arrivals and departures) on weekdays.

On April 5, 2000, the Wendell H. Ford Aviation Investment and Reform Act for the 21st Century, or AIR-21, was enacted, exempting from the HDR slot limitations certain flight operations, namely those performed by aircraft with a capacity of 70 seats or fewer and operating between LGA and small airports (referred to as "small hubs and non-hubs"). In the first seven months after AIR-21 was enacted, airlines sought to schedule more than 600 new operations a day at LGA. As of November, 300 of these new movements had been inaugurated, bringing the average total on a typical weekday to about 1,350. Another 200 movements were scheduled to start by the end of January 2001.

The result of the new schedules at LGA was an unprecedented level of flight delay and numerous flight cancellations on a daily basis. According to the FAA's OPSNET (Air Traffic Operations Network Database) statistics, LGA alone accounted for more than 25% of the serious flight delays (more than 15 minutes) experienced at *all* commercial U.S. airports in the fall of 2000, compared with 10% for the previous year. Nonetheless, carriers did not indicate any intention to reduce their level of flight operations, or to even scale down some of the pre-announced schedule expansions (FAA, 2001a).

Based on the discussion in Chapter 2, the carriers' reluctance to reduce their planned level of operation is quite understandable. LGA has never been a "hub" for any particular airline. Figure 3-8 shows the frequency shares of the largest, second largest,

third largest, etc., operators at selected major airports in the U.S. The frequency shares of the principal carriers are smaller at LGA than at most other major airports in the U.S. In other words, the degree to which carriers internalize the cost of congestion delay at LGA is smaller than at other airports surveyed. As a result, carriers have little incentive to reduce by their own choice their level of operations to reduce congestion.

Figure 3-8. Frequency Shares at Selected Major U.S. Airports
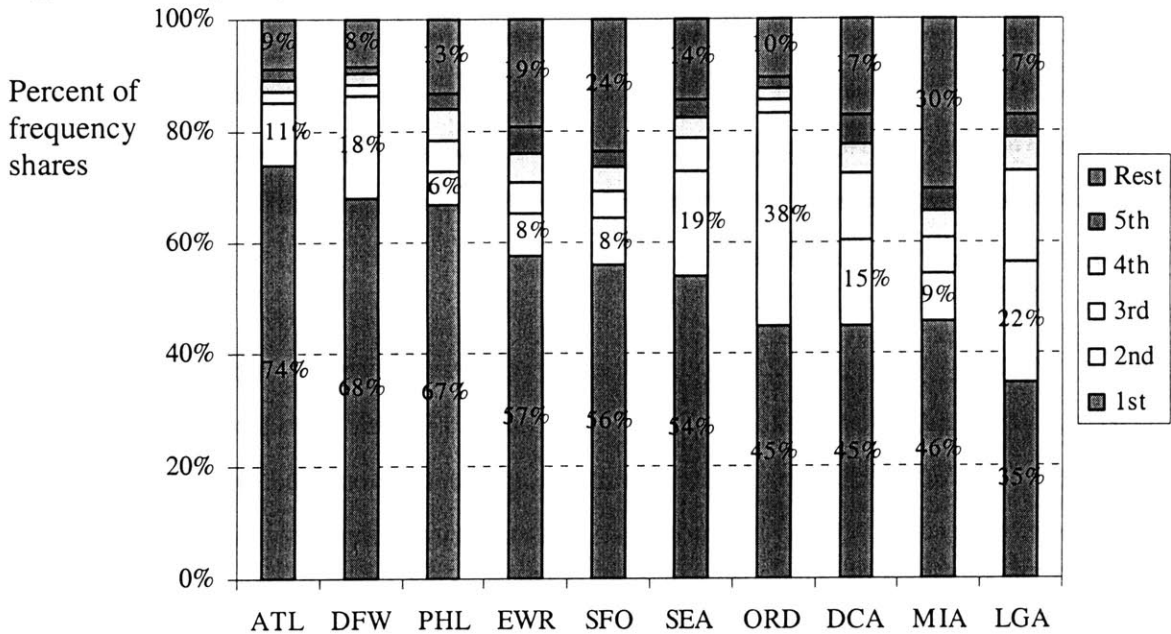
Legend: 1st – Carrier with highest frequency,
             2nd – Carrier with second highest frequency, etc.

Airports: ATL – Atlanta Hartsfield; DFW – Dallas/Fort Worth;
             PHL – Philadelphia;   EWR – Newark;
             SFO – San Francisco;   SEA – Seattle/Tacoma,
             ORD – Chicago O'Hare;   DCA – Washington National (Reagan);
             MIA – Miami;   LGA – New York LaGuardia.

Source: Schedule as of Monday, 13 November, 2000; *Official Airline Guide MAX.*

As an interim solution to the unprecedented congestion at LGA, the FAA, with strong support from the operator of the airport, the Port Authority of New York and New Jersey (PANYNJ), imposed a limit on the number of slot exemptions granted under AIR-21. These "AIR-21 slots" were allocated among eligible flights through a lottery that was conducted in December 2000 and took effect on January 31, 2001 (FAA, 2000). The

lottery was designed so that scheduled movements in any hour would not exceed 75, which, when added to 6 hourly operations from general aviation, represents a level that the airport was deemed adequate to accommodate at reasonable levels of delay in good weather conditions (FAA, 2000).

While the slot lottery took effect by February 2001, there was a considerable lag until the new schedules and requisite operational adjustments were fine-tuned by the carriers. The net result of the "slot lottery" was a reduction in the total number of scheduled flight operations at LGA, from about 1,350 per day in November 2000 to about 1,200 per weekday in August 2001, i.e., about a 10% drop in traffic. Indeed, the severity of delays and the number of cancellations at LGA declined significantly after January 2001, compared to the fall of 2000.

Figure 3-9 shows how the hourly level of scheduled flight operations changed from the pre-lottery (Nov 00) to the post-lottery (Aug 01) level. To account for general aviation traffic, four additional flight movements per hour were added to the scheduled operations between 06:00 am and midnight. This is because airport statistics indicate an actual average of 3 to 4 general aviation operations per hour at LGA during the period in question, despite the fact that the FAA had allowed for 6 general aviation operations per hour when it set the target level of scheduled operations to 75 movements. For reference, the VFR capacity of 81 flight movements per hour is also shown in the Figure. In contrast with the "peaks-and-valleys" profile of many hub airports (see, for instance, Atlanta, Figure 2-3 in Chapter 2) a consistently high and steady level of demand persists throughout the day at LGA. As such, even if the flight schedules at LGA could be distributed perfectly evenly throughout the day, the extent of flight delays would still not be significantly reduced. At high overall levels of demand, small delays early in the day can easily propagate to flights later in the day, even in the absence of network effects (i.e., a late outbound flight from LGA may automatically delay the arrival of an inbound flight, if these two are operated with the same aircraft).

Figure 3-9. Time-of-day Profile of Scheduled* Operations at LGA
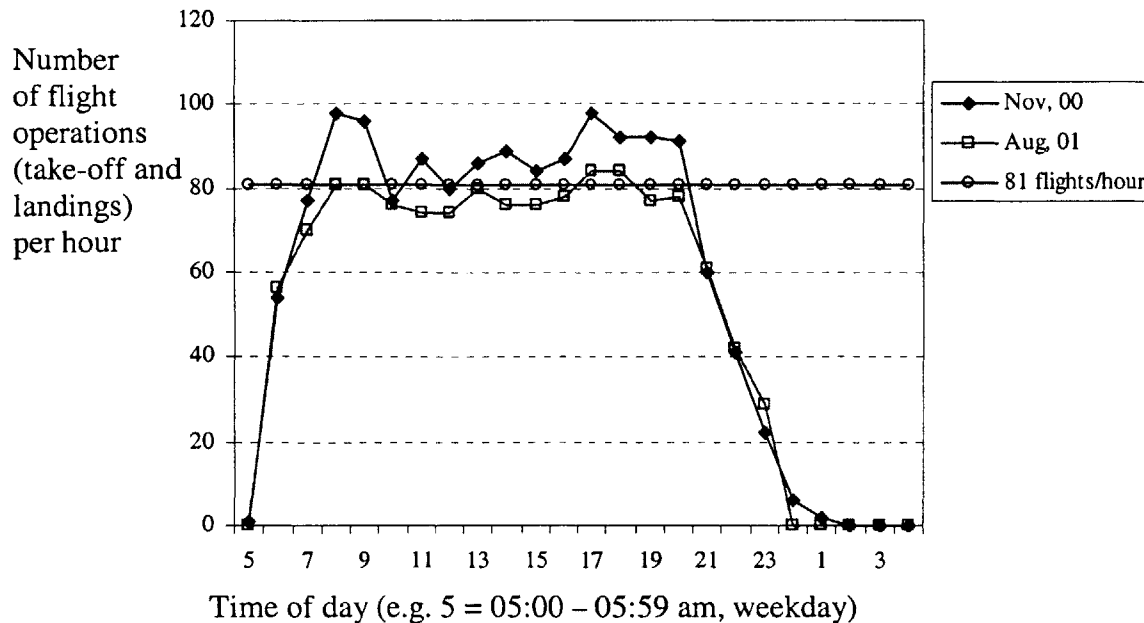
* Including 4 general aviation operations per hour between 06:00 am and midnight.
Source: *Official Airline Guide*

Using DELAYS, the extent of airside congestion at LGA can be modelled. Using the hourly flight profiles shown in Figure 3-9, Figure 3-10 compares the average flight delays as predicted by the DELAYS model both before and after the slot lottery. As shown in Figure 3-10, the average delay per operation prior to the slot lottery increases monotonically from early morning until about 8 p.m. in the evening. At the evening peak hour in November 2000, flights to and from LGA can expect an average of an hour of delays due solely to congestion. In comparison, the *10%* drop in traffic as a result of the slot lottery reduced the average delay at the peak evening hours by about *80%*. In total, about 910 aircraft-hours of congestion delay were expected per weekday in November at LGA, compared with "only" about 150 aircraft-hours in August.

It is possible to estimate how much additional delay an incremental flight at different times of the day would cause (or in how much delay reduction one less flight would result). This can be estimated by computing the difference between the total daily delay estimates provided by two DELAYS runs, one with one additional flight added to

the original schedule in the hour concerned, and the other with the original schedule minus one flight in the same hour. The difference in total aircraft-hours of delay between these two runs is then divided by two to obtain the "marginal" delay caused by one additional or one fewer flight in that hour. This process was repeated for each one of the 24 hours of the day.

Figure 3-10. Estimated Average Delay Before and After Slot Lottery at LGA
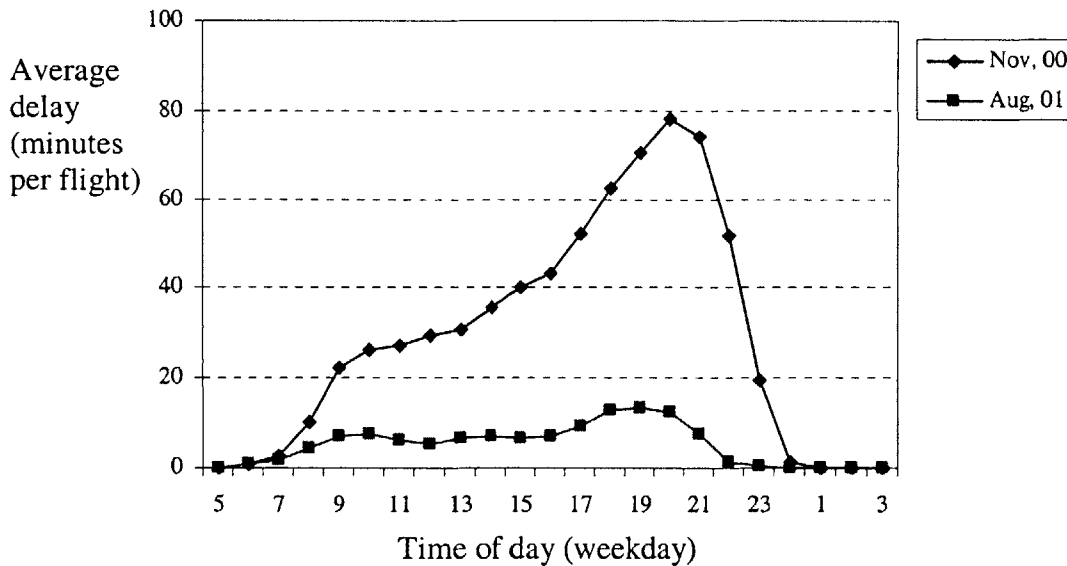
Figure 3-11 shows the estimates of marginal delay at LGA before and after the lottery. In particular, the almost linear, downward-sloping marginal delay curve for November 2000 suggests that one more flight at 08:00 am would cause an additional delay to practically every one of the more than 1,000 flight movements during the rest of the day, resulting in roughly 13 aircraft-hours of additional total delay. For August 2001, the ripple effect on delay is far smaller in view of the slightly lower number of flights relative to the sustainable capacity.

In Chapter 1, the economic principle of congestion pricing was introduced. It is useful to compare the approximate marginal delay costs associated with an aircraft movement to the access fee that the aircraft operator actually pays. In the case of LGA, operators of scheduled jet flights paid a fee of $5.25 per thousand pounds (lbs) of

maximum gross take-off weight (MGTOW) per landing in 2001 (PANYNJ, 2001). The average MGTOW per aircraft at LGA was about 114,000 lbs (i.e., somewhat smaller than that of a 737-300, due to the presence of a large number of smaller aircraft). Thus, the average fee per flight movement was about $300 (or $600 per departure). Figure 3-12 compares this $300 fee with the estimate of the average marginal delay cost associated with an incremental flight movement at LGA as a function of time of day for the published schedule in August 2001, i.e., *after* the lottery. The estimated marginal delay cost is derived by simply multiplying the estimated marginal delay caused by an additional movement in August 2001 (Figure 3-11) by an estimated average direct operating cost of $1,600 per block hour for aircraft using LGA. It is remarkable that, for the 12-hour period between 08:00 and 20:00, the marginal delay cost is *more than one order of magnitude* greater than the actual average fee per movement. This is consistent with the findings of Carlin and Park (1970), and suggests that, even after the lottery, access to LGA during the busy hours of the day was grossly under-priced.

Figure 3-11. Estimated Marginal Delay Before and After Slot Lottery at LGA
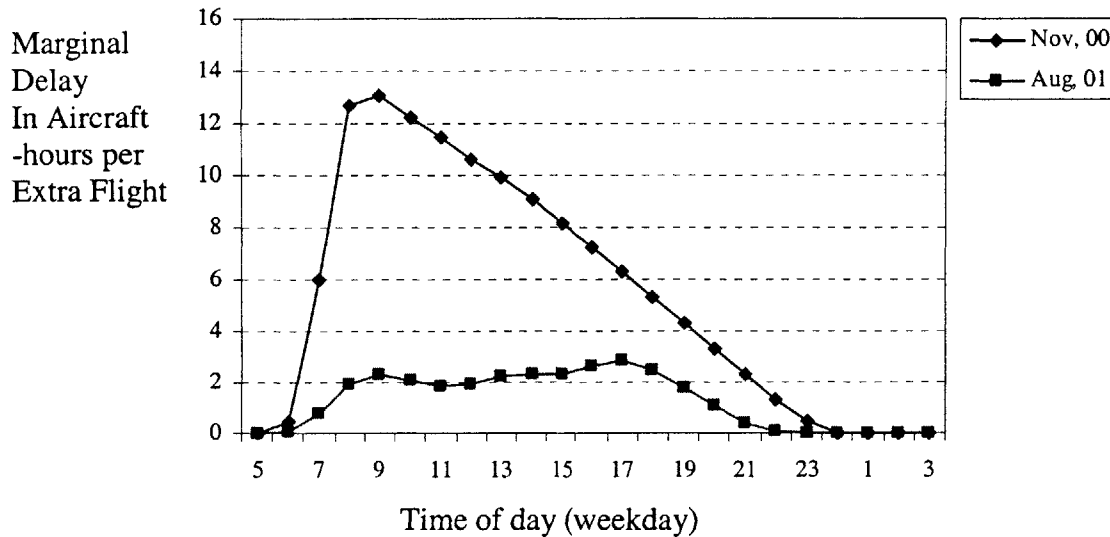
For a separate airport, the Los Angeles International Airport (LAX), Hansen (2002) estimated the congestion delays and external delay costs using a deterministic queuing model based on the notion of cumulative flow diagrams. This analysis led to conclusions strikingly similar to those presented here. Specifically, he found that for

51

many flights, the external delay costs incurred far exceed the reduction in seat-adjusted schedule delay (as a measure of social benefit), often by as much as an order of magnitude, or more.

Figure 3-12. Estimated Marginal Delay Cost and Existing Landing Fees at LGA
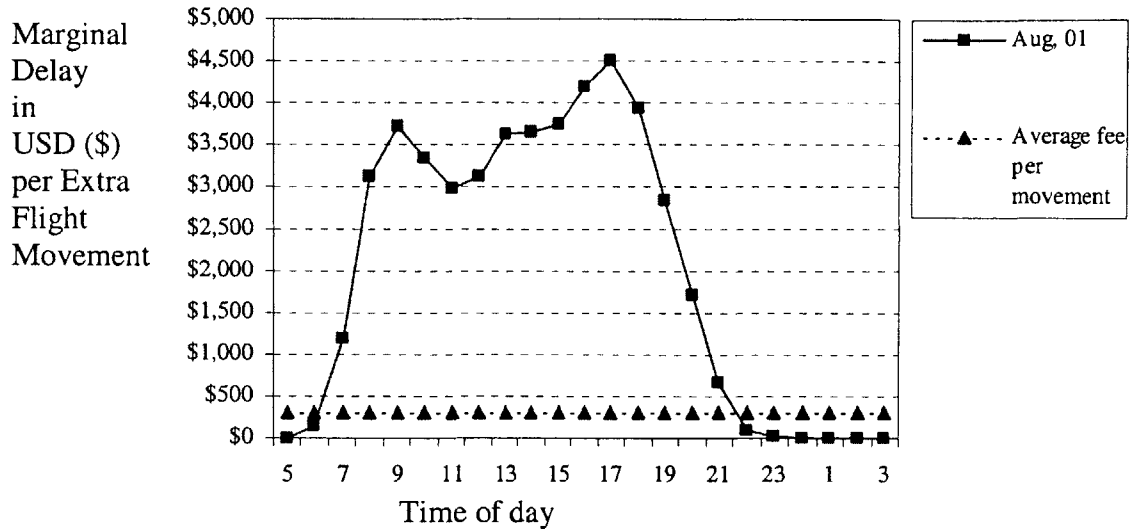
3.3 Beyond LaGuardia

The demand for runway access (as represented by the total scheduled frequency) relative to the sustainable capacity is likely to vary from one airport to another. As such, the level of congestion generated will also be different. This obviously means that, while the underlying economic principles are still valid, the application of congestion pricing may not be necessary in some cases and many require fine-tuning to the local conditions at those airports where application is warranted.

To illustrate these points and to obtain a further understanding of how the observations made in the previous section for LGA may be applicable elsewhere, two other airports, in addition to LGA, were selected as representative examples of the level of airport airside congestion in the US. LGA represents the case where an airport's capacity is fully utilized or overtaxed for most of the day. Boston Logan Airport (BOS) is a typical example of many airports across the US at which demand is roughly equal to or

exceeds capacity for only part of the day. Austin, TX (AUS) is chosen to represent cases where demand is not close to capacity during any significantly long part of the day.

Figure 3-13 compares the demand-to-capacity ratio (hourly demand is shown as a percent of capacity) at these three airports for a typical weekday. The sustainable capacities for these three airports are also shown at the side of the figure. In particular, the capacities listed are those allocated to flights other than military or general aviation. For example, LGA's sustainable capacity is 81 flights per hour, including 6 flights per hour that are allocated to general aviation (although usually fewer than these are actually operated), and hence a capacity of 75 flights-per-hour is shown. The figures for BOS and AUS were obtained from FAA surveys and from the respective airport authorities. In contrast to LGA, BOS only experiences a peak period in the late evening hours, roughly from 4 pm to 9 pm, while AUS does not have any consecutive hours with scheduled flights constituting more than 50% of its capacity for scheduled flights.

Figure 3-13. Demand-to-Capacity Ratios at Representative Airports
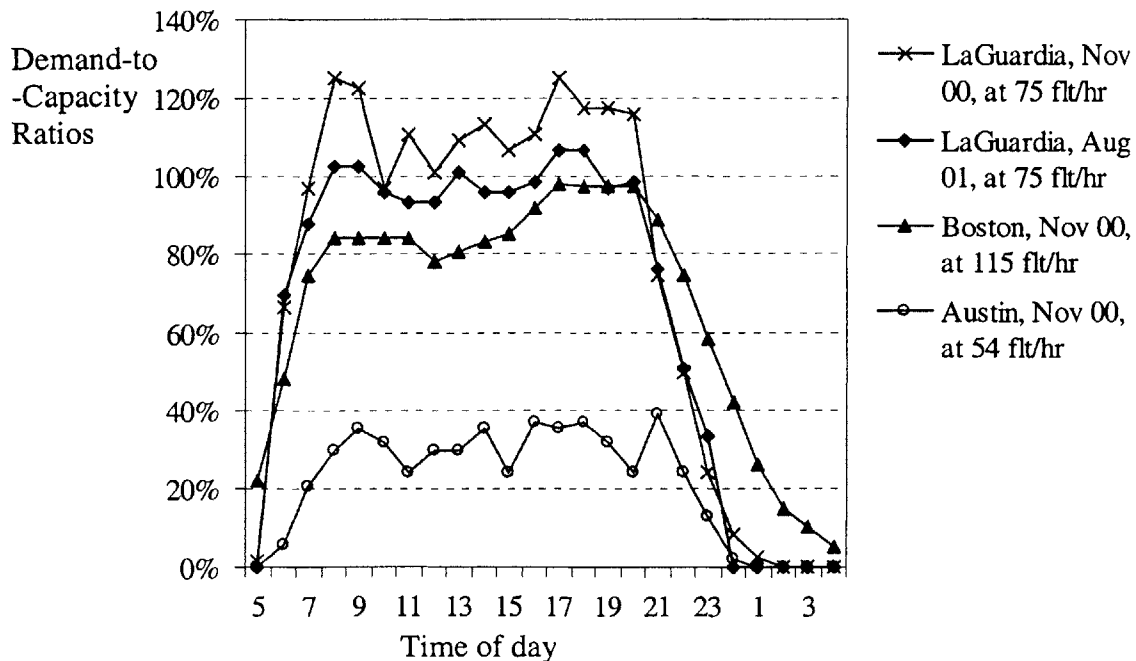
In Figure 3-13, the demand-to-capacity ratio for BOS appears to be only marginally below that for LGA. According to queuing theory, however, this seemingly small gap between demand and capacity can translate into a large difference in the average and total congestion delay. Estimated via DELAYS, Figure 3-14 compares the average delay expected at these three airports, while Figure 3-15 compares the total delay at these airports. Figure 3-14 shows that for flights departing or arriving during the evening peak hours, average delay at pre-lottery LGA can be as much as *one full order of magnitude* more than that at BOS (80 minutes/flight compared with 7 minutes/flight). Even after the lottery, the average delay at LGA during the evening peak hours is still more than twice that at BOS (roughly 15 minutes/flight compared with 7 minutes/flight). Note that these delays are purely due to congestion. In other words, even under the finest weather, and in the absence of airline-related operational anomalies, airplanes departing from or arriving at BOS in the evening are still expected to suffer an average delay of 7 minutes.

Figure 3-14. Estimated Average Delay at Representative Airports
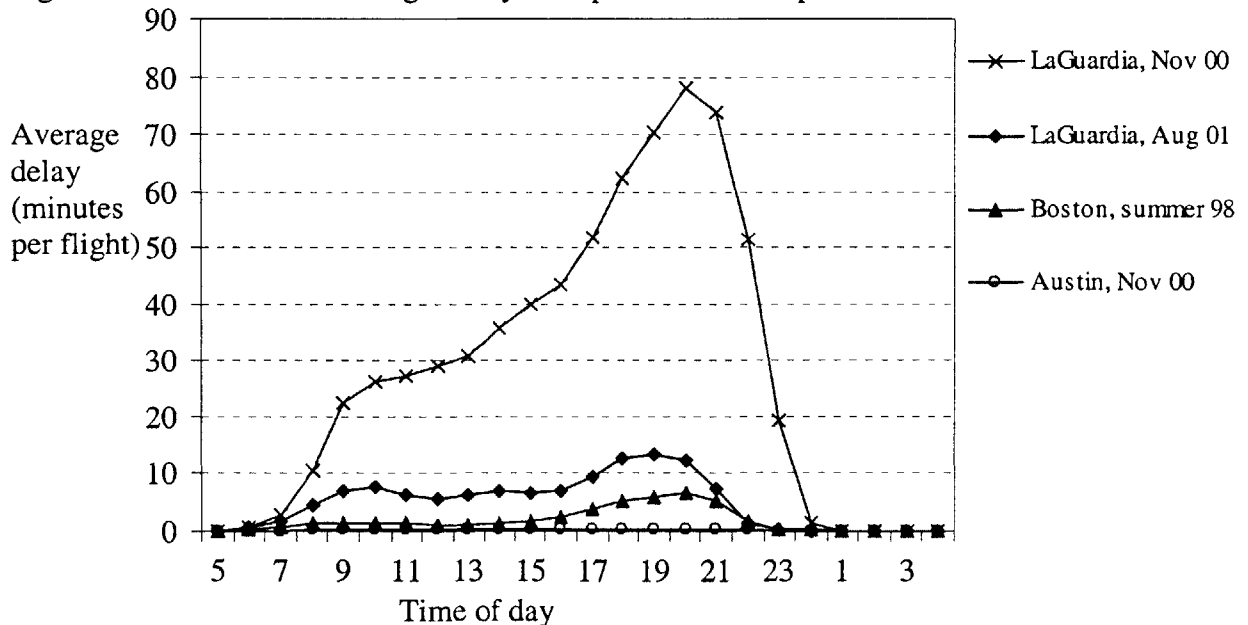
Figure 3-15 shows a similar situation as Figure 3-14. "Total delay" in the context of Figure 3-15 is defined as the total amount of delay that will be incurred by all flights

scheduled to land or to take off during a given hour of the day. The total delay incurred by flights departing or arriving during the evening peak at pre-lottery LGA is one full order of magnitude more than that at BOS (113 aircraft-hours compared with 12 aircraft-hours), and even the post-lottery LGA figure is more than double that at BOS (17 aircraft-hours compared with 12 aircraft-hours). In contrast, AUS experiences virtually no congestion throughout the day.

Figure 3-15. Estimated Total Delay at Representative Airports
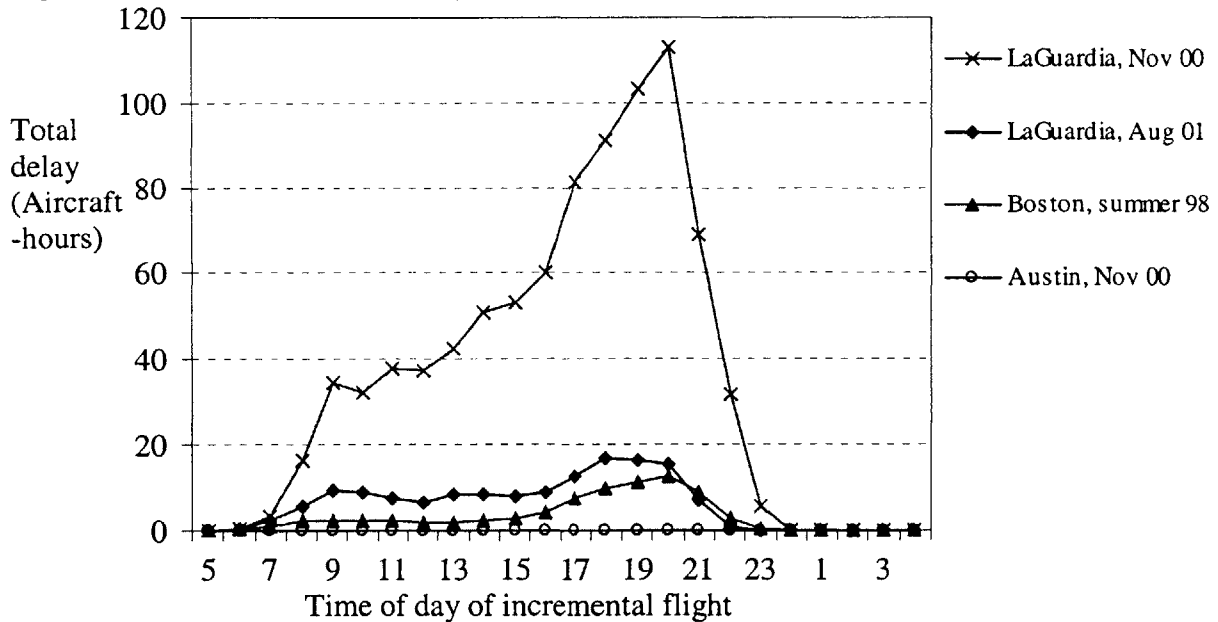
The concept of marginal delay, as defined previously, can also be applied at these airports. Figure 3-16 compares the marginal delay at these three representative airports. Similar to the earlier figures, the marginal delay at BOS is less than half of the post-lottery LGA and roughly one-tenth of the pre-lottery LGA level, during the evening peak hours. At roughly $1,600/aircraft-hour, an incremental flight during the evening peak hour at BOS generates roughly $2,552 in additional operating costs to all other flights. This amount is still much higher than the prevailing landing fee charges, which are of the order of a few hundred dollars per flight. As expected, marginal delay at AUS is negligible.

Figure 3-16. Estimated Marginal Delay at Representative Airports
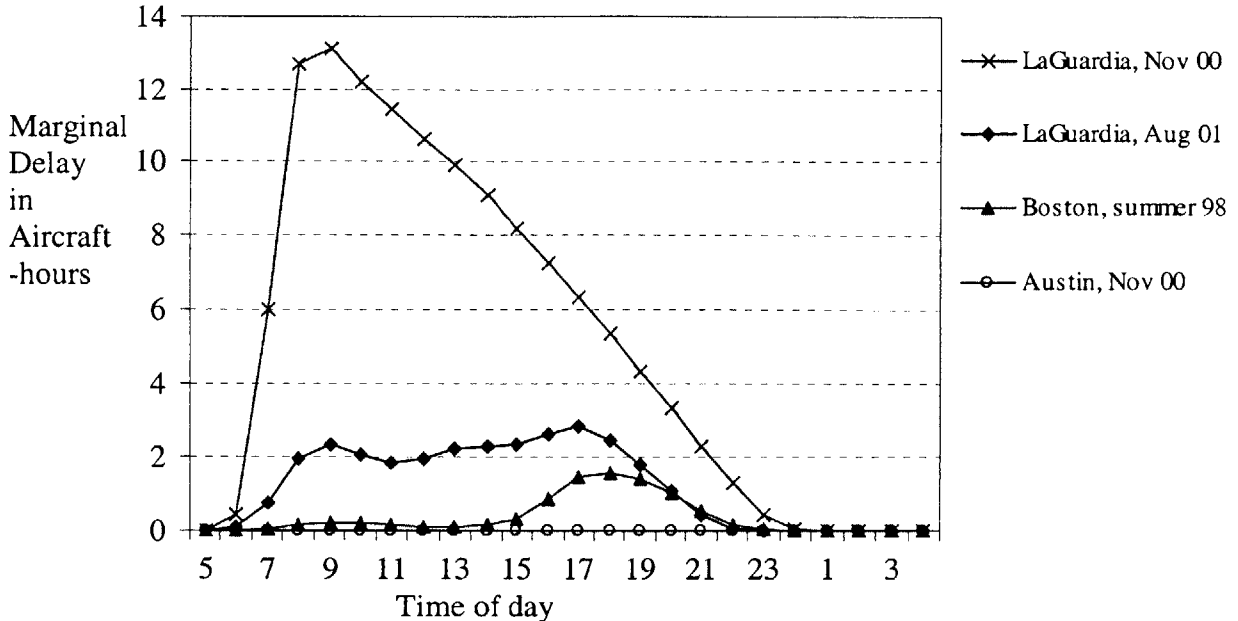
The comparison of the different levels of congestion at these three representative airports offers an important lesson for the formulation of a nation-wide demand management policy: adaptation of a broad policy framework to the local airport conditions is critical to a successful implementation. For example, while LGA clearly warrants a market-based demand management policy that limits the number of flights per hour for practically most of the day on weekdays, BOS may only warrant a policy that aims to encourage carriers to shift peak-hour operations in the evenings to other off-peak periods. The level of the congestion toll that is to be levied under such policies will likely be much lower for BOS than for LGA. In contrast, no demand management measures are needed for non-congested airports such as AUS.

It is also worth emphasizing that the congestion delays computed in this section are based on the assumption of optimal weather conditions at these airports (VFR and little wind). With reduced visibility or high crosswinds, the capacity at these airports can be drastically reduced (although more flights may also be cancelled). In these conditions, delays can be far higher than those estimated here. Even though some airports do not warrant any demand management policy under optimal weather conditions, special

allowances to limit the number of flight operations may be justified under poor weather conditions.

In Chapter 4, analytical tools that will help refine a general demand management policy framework will be introduced. Using the DELAYS model and drawing on the discussion and findings of this Chapter, a number of numerical experiments will be carried out to model the potential impact of demand management under a variety of assumed market structures.

Chapter 4

An Analytical Framework

Chapter 3 has demonstrated that any demand management policy that can reduce excess demand at a capacity-constrained airport, even by a small percentage, has a potentially large economic impact. According to traditional microeconomic theory, congestion pricing, as discussed in Chapter 1, is a useful starting point for a relevant public policy. Based on this principle, Morrison (1983, 1987), for instance, has estimated the necessary congestion tolls for several major airports in the U.S. using regression analyses of the delays observed at these airports. Indeed, congestion pricing in this traditional sense has been in use in downtown Singapore, Central London, and very likely in the near future, in the Netherlands (Honeywill, 2002).

In the context of urban transportation, each driver among many decides whether to use his/her own vehicle or to use public transport, and thus behaves in an "atomistic" fashion. At airports, however, each of the airlines (as the most frequent aircraft operators) decides whether or not to operate a substantial number of flights. In other words, the behaviour of airlines may be far from atomistic as in the case of urban commuters. As a result, there are two major consequences. First, as recognized by Daniel (1995) and Hansen (2002), non-atomistic airlines, or *bulk operators*, may internalize congestion costs to a larger extent than do the atomistic urban commuters. Second, as stated by Brueckner (2003), depending on the pricing structure offered by airlines serving congested airports (and also with simplifying assumptions on passengers' benefit functions), airlines may operate above or below the social optimum in the absence of governmental intervention. In particular, Brueckner concludes that a perfectly price-discriminating monopolist at a congested airport operates at the economic optimum

59

without governmental intervention, while the behaviour of atomistic carriers at certain airports justifies congestion pricing as developed in the urban transport context.

This Chapter extends the advances made by these two papers, among others, in three major directions: i) it demonstrates that even if the consequences of imperfect competition mentioned in Brueckner (2003) are abstracted away, congestion pricing is more complicated than previously thought; ii) it provides a more detailed treatise on more possible market structures between a discriminating monopolist carrier and perfect competition; iii) it suggests a framework for examining the case of asymmetric carriers (which indeed is the most realistic one for most airports).

4.1 Fundamental Model

Sidestepping the issue of imperfect competition, the utilization of scarce airport capacity is the focus of the analytical model developed in this Section. At the heart of the model is the fact that expected congestion delay for carrier i in period t is a function of the number of flights operated by carrier i in all periods up to and including t, as well as the number of flights operated by all other carriers in all periods up to and including t. To improve the tractability of the problem, several assumptions are required. In particular, some of these assumptions are of primary importance to the construction and key conclusions of the model, while others are included to simply improve computational efficiencies and are incidental to the main findings.

The primary assumptions are as follows:

i)   The demand for transportation services faced by a particular carrier is a function of its own frequency and services offered, and independent of such decisions made by another carrier. This assumption significantly improves the tractability of the problem when the effect of different market structures at congested airports is investigated.

ii)   Passengers' demand curve for flights departing from or arriving at a congested airport during a particular time period of the day is assumed to be independent of the demand curve for another such period. In other words, there is no temporal reallocation of the demand. The trade-off that this assumption involves is explained earlier.

iii)   Passengers' demand curves are assumed to be independent of the congestion expected, at least in the short run. The same assumption is required in Brueckner (2001) to improve tractability.

Moreover, secondary assumptions are included to improve computational efficiency but these are not crucial to the findings of the analysis:

iv)   Only one congested airport is considered. All other airports are assumed to be un-congested. Aircraft turnaround times are assumed to be sufficient to avoid delay propagation throughout the airline network.

v)   Only one airplane configuration is assumed for all carriers and hence only one set of operating cost statistic is used. As a result, airlines' operating decision can be fully described by the number of flights operated in each time period, with each flight carrying the same number of passengers. Similarly, the per-hour cost of congestion to the airlines is the same. Once the derivation advances to the use of net marginal revenue per flight, this assumption can be relaxed.

vi)   In the derivation, each flight is assumed to have the same scheduled operating cost (OC) that is independent of congestion. Once the derivation advances to the use of net marginal revenue per flight, this assumption can be relaxed.

vii)   The demand for transportation services for each airline maintains some fixed relationship (e.g. same proportions) to the demand for other airlines. This, together with assumption (vi) greatly simplifies the estimation of any internalized congestion costs.

With these basic assumptions, the model on how the internalization of congestion costs affects different carriers can be quantitatively developed. On the cost side, the total operating cost can be divided into two parts: one that is based primarily on scheduled operating costs (OC) that include any other expected costs not linked to congestion, and all other congestion-related costs (CC). In particular, these costs (OC and CC) are expressed as some function of the carriers' frequencies. Moreover, the part of congestion-related costs (CC) borne by a particular carrier in question is referred to as the private congestion-related cost (PCC), and the part that is borne by other carriers is referred to as the external congestion-related cost (ECC). The sum of PCC and ECC is the social congestion-related cost (SCC).

Specifically, the congestion-related costs are functions of the flight frequency:

$$SCC = \Lambda \cdot c \cdot W(\Lambda) \qquad \qquad ...[4\text{-}1]$$

$$PCC_i = \lambda_i \cdot c \cdot W(\Lambda) \qquad \qquad ...[4\text{-}2]$$

where c is the standard per-hour operating cost of the single airplane type;

W is the expected wait time as a function of the total flight frequency Λ; and

λ_i is the flight frequency operated by carrier i.

Taking the derivative of the PCC with respect to frequency operated by a carrier i (as in Jansson (1998)) yields:

$$MPCC_i = c \cdot W(\Lambda) + c \cdot \lambda_i \cdot \frac{dW}{d\Lambda} \cdot \frac{\partial \Lambda}{\partial \lambda_i} \qquad \qquad ...[4\text{-}3]$$

with $0 \le W(\Lambda); \frac{dW}{d\Lambda} > 0$ and $\frac{\partial \Lambda}{\partial \lambda_i} \equiv 1$.

Taking the derivative of the *SCC* with respect to frequency (λ_i) yields:

$$MSCC_i(\Lambda) = c \cdot W(\Lambda) + c \cdot \Lambda \cdot \frac{dW}{d\Lambda} \cdot \frac{\partial \Lambda}{\partial \lambda_i} \qquad \text{...[4-4]}$$

Or, $\quad MSCC(\Lambda, \lambda_i) = MPCC_i + c \cdot (\Lambda - \lambda_i) \cdot \frac{dW}{d\Lambda} \qquad \text{...[4-5]}$

where the second term in [4-5], $c \cdot (\Lambda - \lambda_i) \cdot \frac{dW}{d\Lambda}$, is the congestion cost *external* to, or not *internalized* by carrier *i*, and each carrier *i* operates λ_i flights in view of each other's operating decisions (leading to an aggregate of Λ flights being operated).

On the demand side, let *MR* denote the marginal revenue for a carrier as a function of its frequency (λ_i). Note that the marginal revenue expected of an incremental flight depends on the competitive conditions of the specific market(s) served by this flight. The profit function (π) becomes:

$$\pi = \int_0^{\lambda_i} MR_i(q) \cdot dq - \int_0^{\lambda_i} MPCC_i \cdot dq - \lambda_i \cdot OC_i \qquad \text{...[4-6]}$$

The first-order necessary condition for profit maximization yields:

$$\frac{\partial \pi}{\partial \lambda_i} = MR_i - MPCC_i - OC_i = 0 \qquad \text{...[4-7]}$$

$$MR_i(\lambda_i) = MPCC_i + OC_i \qquad \text{...[4-8]}$$

$$MR_i(\lambda_i) = MSCC - c \cdot (\Lambda - \lambda_i) \cdot \frac{dW}{d\Lambda} + OC_i \qquad \qquad \ldots [4\text{-}9]$$

The second-order sufficient condition is assumed to be satisfied with:

$$\frac{\partial \pi^2}{\partial^2 \lambda_i} \le 0 \qquad \qquad \ldots [4\text{-}10]$$

As evident in [4-9], as a carrier's frequency (λ_i) increases given Λ constant, the profit-maximizing marginal flight revenue takes into account an increasing share of the *MSCC*. A monopoly carrier, with $\Lambda = \lambda_i$, practically internalizes all of the *MSCC*, and hence does not require public policy intervention from the perspective of congestion (it may, however, justify heightened monitoring for anticompetitive behaviour).

Combining the *MR* and *OC* terms simplifies [4-9] and [4-10] to the profit-maximizing condition using the Net Marginal Revenue function (*NMR*) per flight:

$$NMR_i(\lambda_i) = MSCC - c \cdot (\Lambda - \lambda_i) \cdot \frac{dW}{d\Lambda} \qquad \qquad \ldots [4\text{-}11]$$

where $NMR_i \equiv MR_i - OC_i$ with $\dfrac{dNMR_i}{d\lambda_i} < 0$ in the region of interest $\quad \ldots [4\text{-}12]$

From the welfare perspective, carriers' total profit (Π), i.e., producers' surplus, can be formulated as a function of the aggregate frequency operated:

$$\Pi(\Lambda) = \int_0^\Lambda [NMR_{\Sigma i}(q) - OC(q)] \cdot dq - CC(\Lambda) \qquad \qquad \ldots [4\text{-}13]$$

Note that without fully knowing consumers' preferences, little can be said about consumer surpluses (and hence social welfare). However, maximizing producers' surpluses can contribute to, although not necessarily result in, higher social welfare.

Maximizing Π with respect to the carriers' frequencies yields:

$$\frac{\partial \Pi}{\partial \lambda_i} = NMR_i - MSCC_i(\Lambda) = 0 \qquad \forall \ i \qquad \qquad ...[4\text{-}14]$$

But since $MSCC_i(\Lambda) = MSCC_j(\Lambda)$ $\qquad \forall \ i \neq j$ $\qquad\qquad$ ...[4-15]

$$NMR_i(\lambda_i^S) = NMR_j(\lambda_j^S) = MSCC(\Lambda^S) \qquad \forall \ i \neq j \qquad\qquad ...[4\text{-}16]$$

where the superscript S denotes the surplus-maximizing condition.

Comparing [4-16] with [4-11] shows that if each carrier operates λ_i flights as a profit-maximizing decision in view of other carriers' decisions, the aggregate decision (Λ) will only be surplus-maximizing if each carrier is faced with a tax equal to the amount $c \cdot (\Lambda^S - \lambda_i) \cdot \left. \dfrac{dW}{d\Lambda} \right|_{\Lambda^S}$, i.e., the *external congestion cost* imposed by carrier i's operating decision, where the superscript S denotes the surplus-maximizing condition.

Comparing [4-16] with [4-11] confirms that for a monopoly operator (with $\lambda_i = \Lambda$), the profit-maximizing decision of this operator also maximizes carriers' collective surpluses. As stated in Brueckner (2002, 2003), there is as a result no role for demand management for monopoly carriers.

In general, given any level of toll T_i, carriers would adjust their operating frequency such that:

$$NMR_i(\lambda_i^T) = MPCC_i(\Lambda^{T_i}, \lambda_i^{T_i}) + T_i \qquad\qquad ...[4\text{-}17]$$

where the superscript T denotes carriers' frequency choice in the presence of this toll.

4.2   Revenue Neutrality

The political case for demand management policy can be strengthened if demand management implemented under certain market conditions can be demonstrated to produce net social benefits even if none of the collected toll is put into productive uses. A simple measure, termed the Short-run Benefit-to-cost Ratio (*SRBR*), is devised to gauge just this. The *SRBR* is defined by the gain in producer surplus divided by the congestion toll collected.[8] The short-run nature of this metric limits the potential benefit to the reduction of direct congestion costs, and does not take into account any further benefit the collected toll can generate.

If the total market frequency is reduced through a congestion toll from a no-toll equilibrium Λ^P to some Λ^T while the *NMR* curves remain unchanged, the *change* in producer surplus (Π), $\Delta\Pi$, can be approximated by:

$$\Delta\Pi \approx \Pi\left(\Lambda^T\right) - \Pi\left(\Lambda^P\right) \qquad \ldots [4\text{-}18]$$

Since the sum of congestion toll collected equals $\Sigma_i \lambda_i T_i$, the *SRBR* is defined mathematically as follows:

$$SRBR \equiv \frac{\Delta\Pi}{\sum_i \lambda_i T_i} \qquad \ldots [4\text{-}19]$$

An *SRBR* greater than 1 means that even in the worst-case scenario where the collected toll simply vanishes, carriers will still be better off with a congestion pricing policy with a set of toll T_i than without it. For those cases where the congestion toll can be applied toward some other tangible uses, such as acquiring more land for new runways, the expected benefit can be quantified and added to the *SRBR* metric. For other

---

[8] Note that this definition of revenue neutrality may differ from other such definitions developed for different public policies.

cases where no direct tangible benefits are expected from the congestion toll, and that an *SRBR* is less than 1, the relevant policy administrator may consider redistributing some of the collected tolls back to the carriers (payers of the toll). In particular, the difference between unity and the *SRBR* provides a guide as to how much of the collected tolls should be distributed back to the carriers to achieve revenue neutrality. It is worthwhile to note that in previous proposals for congestion pricing (e.g. see de Neufville and Odoni 2003), all the tolls collected can be redistributed to the toll-payers (although not the same amount to the same operators as the tolls).

In the next few sections, some assumptions on carriers' net marginal revenue functions will be made to generate further insights into public policy. In particular, the discussion will start from the simplest case: one where all aircraft operators are *symmetric* in their operating decisions (Section 4.3). This will then be followed by the more complicated and general case of *asymmetric* operators (Sections 4.4).

4.3   Symmetric Operators

Sections 4.1 and 4.2 described the fundamental model and a performance metric for a market-based demand management policy, using congestion pricing as the policy instrument. Without more knowledge of carriers' *NMR* curves, it is difficult to draw further insights from the model and metric presented. In this and the next section, some incrementally general assumptions will be made with respect to carriers' *NMR* curves. The accompanying analytical tools in support of the fundamental model for market-based demand management policy will be developed in these sections.

To start with the simplest rendition of the fundamental model, consider a congested airport served by carriers with identical net marginal revenue functions ($NMR_{i,t}$ for the same periods). These operators are *symmetric*, meaning that each carrier makes the same profit-maximizing operating decisions as another with or without demand management. Due to the symmetric nature, each carrier maintains a constant and identical

frequency share. The frequency share of each carrier is related to the number of operators, n, in a straightforward manner:

$$\alpha = \frac{1}{n} \qquad \qquad ...[4\text{-}20]$$

Without loss of generality, this frequency share is assumed to be time-invariant over a number of consecutive hours. This can simplify the analysis significantly.

Based on the frequency share information, the number of flights each airline operates in the absence of congestion pricing (i.e., the "privately-optimized" market equilibrium frequency), λ_t^P, for period t, can be expressed as:

$$\lambda_{i,t}^P = \alpha \cdot \Lambda^P = \frac{1}{n} \cdot \Lambda^P \qquad \qquad ...[4\text{-}21]$$

The number of flights each airline operates under a congestion pricing policy that results in a surplus-maximizing level of operations for period t, $\lambda_{i,t}^S$, can be expressed as:

$$\lambda_{i,t}^S = \alpha \cdot \Lambda^S = \frac{1}{n} \cdot \Lambda^S \qquad \qquad ...[4\text{-}22]$$

If a carrier has a frequency share of α, assuming that the carriers' flight times are randomly distributed and that each observes an identical operating cost per unit of time, it "internalizes" exactly a fraction α of the increase in congestion delay. The amount of surplus-maximizing toll levied on a per-flight basis for period t, T_t^S, is therefore the amount of *MSCC* not internalized by the carriers at the desirable level of operation, or:

$$T_t^S = (1 - \alpha) \cdot c \cdot \Lambda_t \cdot \frac{dW}{d\Lambda_t} \qquad \qquad ...[4\text{-}23]$$

Note that the same congestion toll T_i is applicable to all symmetric carriers (hence the subscript i is temporarily dropped). As such, the *SRBR* metric for symmetric carriers (*SRBRsym*) from a no-toll equilibrium to a surplus-maximizing equilibrium can be calculated as follows:

$$SRBRsym_{\Lambda,t} = \frac{\Delta\Pi}{T_t^S \cdot \Lambda_t^S} \qquad \qquad ...[4\text{-}24]$$

By varying the number of symmetric operators, it is therefore possible to examine how the socially optimal congestion toll or the *SRBR* changes. Numerical examples will be presented in Section 4.5.

4.4   Asymmetric Operators

An important assumption in the case of symmetric operators, namely the time-invariance of the frequency share (α) over certain j periods, is also needed in the analysis of asymmetric operators. In the case of asymmetric operators, however, the frequency shares are allowed to differ from one operator to another. In other words, asymmetric carriers, unlike symmetric ones, are allowed to observe different *NMR* curves, e.g., $NMR_{1,t} \neq NMR_{2,t}$, etc.

In particular, consider two ways in which the *NMR* curve of one carrier can differ from those of other carriers:

i) $\dfrac{\partial NMR_i}{\partial \lambda_i}$, and

ii) $NMR_i \ (\lambda_i = 1)$.

Respectively, these refer to the slope and intercept of the *NMR* functions. The impact of these two forms of asymmetry in the *NMR* curves will be explored in the numerical experiments in Section 4.5.

At this point, it is important to distinguish two concepts of congestion tolling: fine versus coarse tolls. Under fine tolling, each carrier is assessed a congestion toll for the hour t, $T_{i,t}{}^S$, equal to the difference between the *MSCC* and the carrier's own *MPCC* at the surplus maximum (i.e., such that each carrier is made to internalize its external cost):

$$T_{i,t}{}^S = (1-\alpha_i{}^S) \cdot c \cdot \Lambda_t \cdot \frac{dW}{d\Lambda_t} \qquad \qquad ...[4\text{-}25]$$

Equation [4-25] is very similar to [4-23], but the subscript i is needed to distinguish the surplus-maximizing toll for one carrier from the one for another. In other words, each asymmetric carrier may be subjected to a different toll under fine tolling, such that each carrier ultimately faces exactly the *MSCC* at the surplus-maximizing frequency. A carrier with a lower frequency share should be subjected to a larger fine toll than a carrier with a higher frequency share. Note in general that each carrier's profit-maximizing frequency share in the presence of congestion but in the absence of tolls is not the same as its congestion-free or surplus-maximizing frequency share:

$$\alpha_i{}^P \equiv \frac{\lambda_{i,t}^P}{\Lambda_t} \neq \frac{\lambda_{i,t}^S}{\Lambda_t} \equiv \alpha_i{}^S \qquad \qquad ...[4\text{-}26]$$

In general, two possibilities occur at the no-toll equilibrium compared with the fine-toll equilibrium: a) $\lambda_i^P \geq \lambda_i^S$ for all i; and b) $\lambda_i^P < \lambda_i^S$ for some i.

The first of these two cases is in fact expected, since a lower toll usually means that more flights are now economically viable. This case is the same as in symmetric carriers described in the last section. Similar to the case of symmetric carriers, the gain in surplus that is reduced by fine tolling can be determined by summing up incremental differences between *MSCC* and the net marginal revenue functions. Mathematically, the surplus gain from a no-toll to fine-toll equilibria (with congestion) is defined by:

$$\Delta\Pi_{no-toll \rightarrow fine-toll} \approx \int_{\Lambda^S}^{\Lambda^P} MSCC \cdot dq - \sum_i [\int_{\lambda^S_i}^{\lambda^P_i} NMR_i(q) \cdot dq] \qquad \ldots[4\text{-}27]$$

With congestion, a carrier's no-toll equilibrium frequency is generally larger than the fine-tolled frequency. This is because congestion tolls increase a carrier's non-congestion operating cost. However, a congestion toll can also reduce a carrier's congestion-related operating cost, and so the net result can be difficult to determine *a priori*. To visualize this, consider the case where the toll is reduced from the surplus-maximizing level to zero, and many carriers rapidly increase their frequency of service. Those carriers with an already large enough degree of cost internalization (in terms of the congestion impact) would choose to stop increasing their frequency of service as their share of the congestion cost increases disproportionately due to actions of these other rapidly expanding carriers. When the fine-toll frequency share of this dominant carrier is large enough compared with others, this carrier may end up reducing its frequency when a congestion toll is lifted (since other carriers are increasing their frequencies much more rapidly), and hence arrive at the second case. Note that this situation may arise when considering the congestion effect alone, and does not involve any assumptions on passengers' travel preferences as required in Brueckner (2003).

Unlike the case of symmetric carriers, it is possible in a market with asymmetric carriers that a complete market withdrawal can be a profit-maximizing decision under tolling. Mathematically, it makes economic sense for carrier i to completely withdraw from a market with the imposition of congestion toll T_i if:

$$NMR_i(1) - MPCC_i - T_i < 0 \qquad \ldots[4\text{-}28]$$

where an integrality in frequency is assumed here (otherwise, the "1" should be "0").

One major concern in implementing fine tolls is that those carriers with low frequency shares will be charged high congestion tolls on a per-flight basis. While this is justified from the perspective of congestion, those carriers with low frequency shares

may be the ones that should be encouraged to operate their services to enhance competition. As well, political realities may dictate the use of a uniform toll for all carriers, or *coarse tolling* to arrive at the same market frequency as fine tolling would achieve.

In a similar fashion as before, the surplus-maximizing coarse toll, T^{C*}, can be defined such that:

$$T^{C*} = \arg\max_{T^C} \Pi(\Lambda^C) \qquad \ldots[4\text{-}29]$$

where the function Π is as defined as in [4-13], and

Λ^C denotes the aggregate frequency operated when each carrier maximizes its profit relative to one another, such that:

$$NMR_i\,(\lambda_i^C) = MPCC_i\,(\Lambda^C, \lambda_i^C) + T^{C*} \quad \text{for each carrier } i \qquad \ldots[4\text{-}30]$$

where the superscript C denotes coarse tolling, and the subscript t has been omitted.

Note that if there are n operators, there are n equations in [4-30]. If the NMR_i's and the $MPCC_i$'s are known and non-degenerate, the λ_i for each carrier has a unique solution. An iterative procedure can be used to determine T^{C*} to satisfy [4-30].

With coarse tolling, certain carriers are bound to face a marginal congestion-related cost above their *MSCC*, while some are bound to face a lower one. As a result, some carriers may have a higher or lower coarse-toll equilibrium frequency than the surplus-maximizing fine-toll frequency.

The welfare gain from the no-toll equilibrium to a coarse-toll equilibrium can therefore be found by:

$$\Delta\Pi_{no-toll \rightarrow coarse-toll} \approx \Delta\Pi_{no-toll \rightarrow fine-toll} - \Delta\Pi_{coarse-toll \rightarrow fine-toll} \qquad \ldots[4\text{-}31]$$

Likewise, the SRBR can be estimated through:

$$\text{SRBR}_{\text{no-toll to coarse-toll}} \approx \frac{\Delta\Pi_{no-toll \rightarrow fine-toll} - \Delta\Pi_{coarse-toll \rightarrow fine-toll}}{\sum_i T_i^C \cdot \lambda_i^C} \qquad \ldots[4\text{-}32]$$

For symmetric carriers, the congestion tolls charged under both fine and coarse tolling are the same. Many theories on congestion pricing implicitly assume both homogeneity and atomicity of the users in calculating congestion tolls and welfare changes. In the context of airport demand management, however, each airline as a decision-making unit decides whether or not to operate a group of flights, which may account for a significant share of traffic at an airport. As a consequence, the imposition of fine tolling versus coarse tolling may result in potentially different responses from each aircraft operator. At this point, the analytical tools have been developed and refined to encompass a broad number of airport scenarios. Numerical experiment is the next logical step, and will be described in the next section.

4.5 Numerical Experiment

Numerical experiments will be used next to provide an order-of-magnitude estimate of the impact of congestion tolling as well as policy insights on demand management in general. To simulate carriers' responses to different congestion tolling methods, further knowledge or assumptions about the carriers' net marginal revenue (*NMR*), marginal social congestion cost (*MSCC*) and marginal private congestion cost (*MPCC*) curves are required.

The DELAYS model described in Chapter 3 provides a means to estimate the *MSCC* curve as a function of total (market) frequency operated within a time period. The queuing model includes assumptions on the probabilistic behaviour of airport capacity,

on the distribution of flight departure and arrival times, and on the distribution of aircraft service time at the congested airport. These assumptions make the delay estimates more realistic than those obtained through deterministic queuing models. However, the use of a deterministic model would not change the fundamental characteristics of the results obtained in this section, as noted by the comparison with Hansen (2002) earlier.

Figure 4-1 shows a few marginal delay curves for changes in total (market) frequency in the first ($t = 1$) of γ consecutive hours of congestion at an airport with an expected capacity $\mu = 80$ flights/hour (combined departures and arrivals), assuming that $\Lambda_t = \mu = 80$ flights/hour in subsequent hours (for $t > 1$). As expected, the marginal delay curve rises more steeply as the number of congested hours (γ) increases. For example, if there are only two consecutive hours of congestion ($\gamma = 2$), the marginal delay due to an additional flight at $t = 1$ reaches 50 aircraft-minutes when there are about 95 flights scheduled in the first hour ($\Lambda_1 \approx 95$). However, when there are four consecutive hours of congestion ($\gamma = 4$), marginal delay reaches the same level when there are only 80 flights in the first hour ($\Lambda_1 \approx 80$). To arrive at the marginal social congestion cost, we simply multiply this marginal delay by the average cost of delay (c), and then add to the average (current) cost of delay ($c \cdot W$).

Figure 4-1.   Marginal delay at different consecutive hours of peak demand

Capacity (μ) = 80 flights/hour, for different γ, with $\Lambda = \mu$ for $t = 2, 3, ..., \gamma$
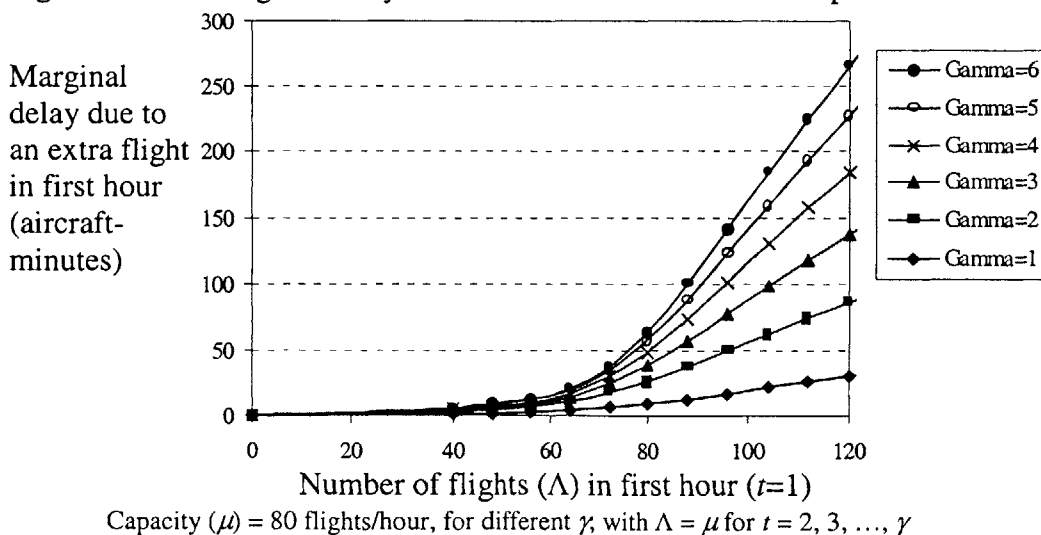
Increasing the aggregate flight frequency to $\Lambda_t = 1.1\cdot\mu = 88$ flights/hour for $t > 1$ while keeping everything else constant as above results in an increase to the marginal delay at different flight frequency at $t = 1$. This is demonstrated in Figure 4-2, with $\gamma = 5$. This is expected because the higher the Λ_t for $t > 1$, the more airplanes there are behind the queue and hence a net increase in the total marginal delay is expected.

Figure 4-2. Effect of increasing the aggregate frequency for all $t > 1$

Marginal delay due to an extra flight in first hour (aircraft-minutes)

Number of flights (Λ) in first hour (t=1)

Capacity (μ) = 80 flights/hour, Λ_t = 88/hr or 80/hr for t = 2, 3, ..., γ for γ = 5
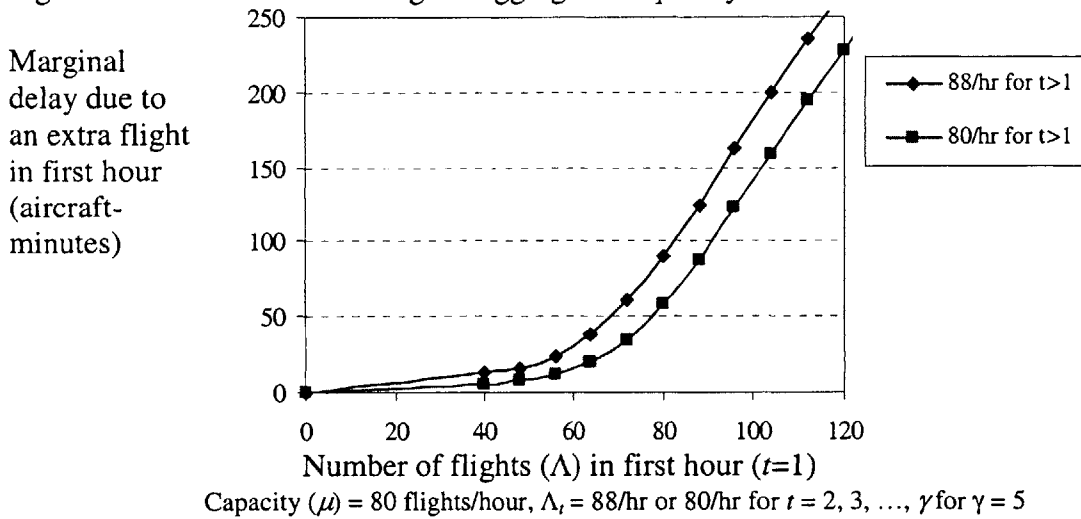
Given the same number of consecutive congested hours ($\gamma = 4$), allowing the number of per-hour flight operations to vary later during this peak period has a slightly more complicated effect than merely changing the aggregate flight frequency. While this effect will not be examined further in the ensuing analyses, it is worthwhile to note as a consequence of the fact that congestion at airports can easily spans multiple hours (as distinct from a relatively short peak period for urban roads). Figure 4-3 compares this effect using two cases: in the $t = 1$ case, the number of flights in the first of four hours is allowed to vary, given that there will be $\Lambda = 80$ flights for $t = 2, 3$ and 4; in the $t = 3$ case, the number of flights in the third of four hours is allowed to vary, given that $\Lambda = 80$ flights for $t = 1, 2$ and 4. The number for flights at all other periods ($t = 0, 5, 6, 7$ ...etc.) in both cases is assumed to be zero. At low aggregate frequencies, each incremental flight in latter case ($t = 3$) incurs a higher marginal delay than in the former case $t = 1$ because there is more delay "spilled over" from the previous periods. However, beyond an

aggregate of 80 flights an hour, each incremental flight in the former case incurs more "spill-over" delay in later periods (3 more subsequent hours with 80 flights each) than the latter case (only 1 more hour with 80 flights).

Figure 4-3. Effect of changing the period of frequency change (*t*)

Number of flights (Λ_t) in first (*t*=1) or third (*t*=3) hour
Capacity (μ) = 80 flights/hour, Λ_t = 80/hr for all other $1 \le t \le \gamma$ for $\gamma = 4$
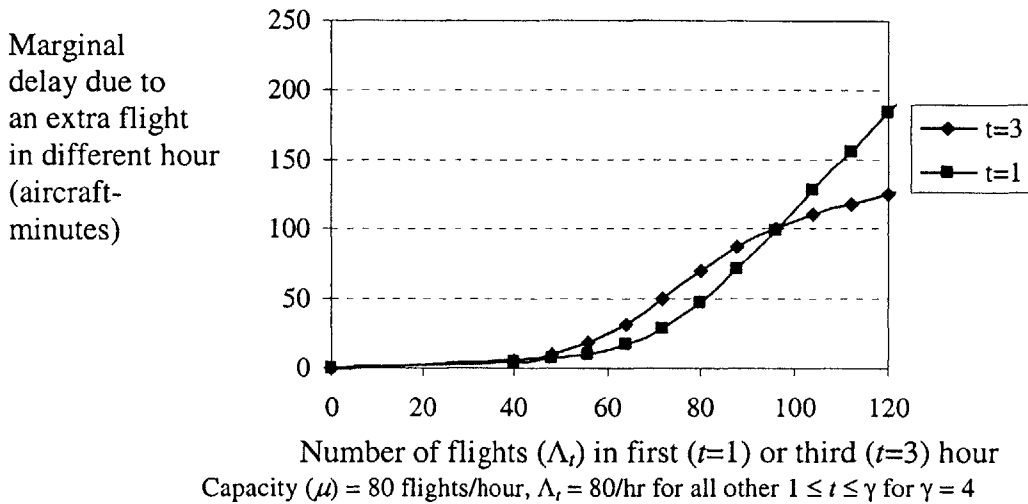
For simplicity, the demand for airport airside capacity is assumed to be insignificant (in terms of estimating congestion) in off-peak periods. As the number of consecutive hours of congestion (γ) increases, so does the *marginal delay* $(\lambda \cdot dW/d\lambda)$ due to an additional flight operation in the first hour.

In terms of carriers' net marginal revenue curves (*NMR*), Cao and Kanafani (2000) described a way to compute the value that a carrier derives from a specific runway time-slot from airlines' record of consumer enquiries and reservation data. They implicitly assume that, like this model, passengers' demand for transportation services from a particular carrier is independent of congestion costs and of other carriers' services. Their rationale is as follows: given equipment assignment constraints are satisfied, aircraft should be used to fly the route segment that brings the most incremental revenue to the airline. The same logic can be used to formulate a carrier's *NMR* curve.

As an illustration, suppose that after satisfying aircraft rotation constraints, a particular carrier can operate up to three flights within the same hour at a congested airport. The net marginal revenues (after deducting expected operating expenses) that can be obtained by operating different combinations of these specific flights are shown in the bottom of Figure 4-4.

Figure 4-4. Constructing an *NMR* curve with three sample flights
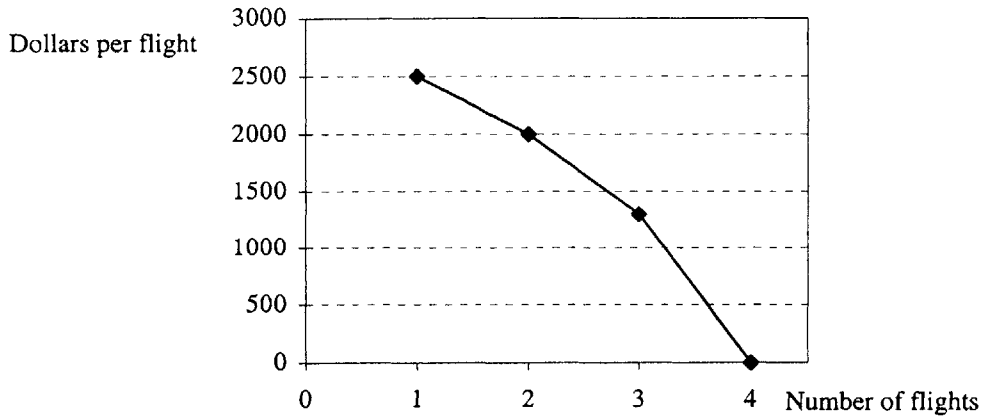
Feasible Combinations	Net Revenue	Comments
Flight A only	$2,300	
Flight B only	$2,500	←*Best choice if only 1 flight allowed*
Flight C only	$2,200	
Flights A and B only	$3,800	
Flights B and C only	$4,000	
Flights C and A only	$4,500	←*Best choice if only 2 flights allowed*
Flights A, B and C	$5,800	←*Best choice if 3 or more flights allowed*

If the carrier in Figure 4-4 is allowed to operate only one flight, it naturally chooses to operate Flight B. If it is allowed to operate only two flights, it should then choose to operate Flights A and C, since this combination gives the highest net revenue. If this carrier is allowed to operate three or more flights, then it rationally chooses to operate all three. To derive the *NMR* curve for this carrier, note that the first available flight that it chooses to operate, i.e., Flight B, yields $2,500 in net revenue. If the carrier is allowed to operate two flights, it rationally chooses the combination of Flights A and C, resulting in an *incremental* net revenue of $4,500 − $2,500 = $2,000 for this second "flight slot". If the carrier is allowed to operate three or more flights, the *incremental* net

revenue for this third "flight slot" is $5,800 − $4,500 = $1,300 for this third "flight slot". The *NMR* curve for this carrier is thus the line that joins these incremental net revenues together, as shown in Figure 4-4. In other words, this *NMR* curve represents the best flight choices for any particular carrier contingent upon the available capacity at the congested airport. To methodically apply this principle, however, privileged access to a large amount of past passenger demand and revenue data is required. Instead of relying on extensive sensitive data, crude generalizations of linear *NMR* curves will be used to draw qualitative insights.

Two features of any *NMR* curve are of particular importance: the location where it intersects with the *MSCC* curve (in relation to the airport capacity), and its approximate slope with respect to the market frequency. The location where the market *NMR* intersects with *MSCC* determines the surplus-maximizing optimum, which, together with the frequency shares of the carriers determines the surplus-maximizing congestion tolls. The slope of the *NMR* curve relates to the sensitivity with which carriers change their frequency of service as the congestion toll changes. A relatively "flat" market *NMR* curve means that carriers will change the frequency of flights substantially in response to a small change in the per-flight toll levied compared with a relatively "steep" market *NMR* curve.

To compare the relative impact of congestion pricing at airports with different carrier frequency mixes, the individual *NMR* curves for each carrier *i* for period *t* can be expressed as:

$$NMR_{i,t}(\lambda_{i,t}) = \max[k_t - \psi(\lambda_{i,t}),0] \qquad \text{for } t = 1, 2, 3, ..., \gamma \quad ...[4\text{-}33]$$

where ψ represents some function (linear for simple illustration) of $\lambda_{i,t}$.

The market marginal contribution can then be found by aggregating the individual *NMR* curve for each carrier. For simplicity, a linear functional form is chosen, and the

aggregate *NMR* is represented using parameters k_t and m to represent its intercept and slope respectively:

$$NMR_t (\Lambda_t) = \max(k_t - m \cdot \Lambda_t, 0) \qquad \text{for } t = 1, 2, 3, ..., \gamma \quad ...[4\text{-}34]$$

where k_t can be adjusted to specify an appropriate surplus-maximizing frequency (Λ_t^S);

m can be adjusted to examine the effect of different market structures (m is related to ψ in [4-33]); and

t is the index of hours falling within a period of γ consecutive hours of congestion.

In the following subsections, numerical experiments using various forms of [4-33] and [4-34] will be examined. The discussion will be divided into three parts: i) symmetric carriers, ii) asymmetric carriers with identical vertical *NMR* intercept [NMR_i (1) = NMR_j (1) for i ≠ j], and iii) asymmetric carriers with identical slope ($dNMR_i/d\lambda_i$).

4.5.1 The Case of Symmetric Carriers

The individual *NMR* curves for each of the n symmetric carriers can be expressed using the same parameters k and m (as in [4-34]) in the following fashion:

$$NMR_{i,t} (\lambda_{i,t}) = \max(k_t - m \cdot n \cdot \lambda_{i,t}, 0) \qquad \text{for } t = 1, 2, 3, ..., \gamma \quad ...[4\text{-}35]$$

In other words, the parameter ψ becomes $m \cdot n$ for the case of symmetric carriers. Summing the NMR_i's across the frequencies offered by individual carriers yields an aggregate $NMR_{\Sigma i}$ of:

$$NMR_{\Sigma i,t} (\Lambda_t) = \max(k_t - m \cdot \Lambda_t, 0) \qquad \text{for } t = 1, 2, 3, ..., \gamma \quad ...[4\text{-}36]$$

79

For comparison, three different market *NMR* curves are selected. The key features of these market *NMR* curves are listed in Table 4-1. Cases E and F have their respective market *NMR* curves set such that the surplus-maximizing frequency is equal to the airport capacity (μ = 80 flights/hour, as explained in Chapter 3) for all hours t (parameters k_e, k_f are chosen accordingly). Between these two cases, the slope parameter m is the only difference in the market *NMR* curves.

Table 4-1. Market Demand Curves Selected for Symmetric Operators

Case	Market Contribution (D)	Slope of D	Surplus-maximizing Frequency
E	$D(\lambda) = \max (b_e - 5\,\lambda, 0)$	"Flat"	Same as capacity μ for all j
F	$D(\lambda) = \max (b_f - 20\,\lambda, 0)$	"Steep"	Same as capacity μ for all j
G	$D(\lambda) = \max (b_g - 5\,\lambda, 0)$	"Flat"	10% above capacity for $j = 1$; at capacity for $j = 2, 3, ..., \gamma$

Case G is a third case that is based on the parameters used in Case E. However, as distinct from Case E, the surplus-maximizing frequency for Case G in the first hour ($t = 1$) of the congestion period is 10% above the airport capacity (k_g is chosen to achieve this), while that for subsequent hours ($k = 2, 3, ...\gamma$) it remains the same as the airport capacity. As a result, the congestion delay in subsequent hours is worse than Case E. Thus, we have for Case G (as distinct from Cases E and F):

$$NMR_{i,1} \neq NMR_{i,t} \qquad \text{for } t = 2, 3, ..., \gamma, \text{ for Case G.} \qquad ...[4-37]$$

In the absence of a congestion toll, the carriers would operate as many flights as allowed according to Equation [4-12]. Table 4-2 documents the no-toll frequency as a function of the number of carriers and the length of the congested period for Case E. The monotonically increasing nature of the no-toll frequency (Λ^P) is a direct result of the assumption about the use of Λ^S as a standardizing feature: the higher the γ, the higher the *MSCC* curve, and hence the higher the *NMR* curves are required for a given Λ^S value.

Table 4-2. No-toll equilibrium frequencies, Case E

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
Number of perfectly symmetric carriers (*n*)	2	88	92	92	92	92	92
	3	93	101	102	102	102	102
	4	96	108	110	110	110	110
	5	98	113	116	117	118	118
	6	99	117	121	123	124	125
	8	101	123	130	133	135	138
	10	102	127	136	141	145	148
	12	103	130	142	148	152	156
	16	104	134	149	157	164	169
	20	105	137	154	164	172	179

Tables 4-3 and 4-4 respectively show the corresponding *SRBR* and the surplus-maximizing toll per flight for the first hour of congestion ($t = 1$) vary for Case E. The surplus-maximizing toll varies from one period to another, such that the tolled aggregate frequency reaches the surplus-maximizing frequencies in each hour. In general, the no-toll equilibrium frequency, *SRBR* and the surplus-maximizing toll increase as γ or n increases. This is primarily because of the fact that the more symmetric carriers there are at an airport, the less the *MSCC* is internalized by each carrier, and the more delay can be reduced per dollar of toll imposed. As n becomes very large, the market structure approaches that comprising of atomistic carriers, i.e., similar to urban commuters.

Table 4-3. *SRBR* for Symmetric Operators, Case E

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
Number of perfectly symmetric carriers (*n*)	2	0.07	0.13	0.13	0.14	0.14	0.14
	3	0.14	0.29	0.33	0.35	0.36	0.36
	4	0.19	0.46	0.56	0.59	0.60	0.61
	5	0.22	0.60	0.76	0.84	0.92	0.93
	6	0.24	0.73	0.94	1.10	1.18	1.25
	8	0.28	0.94	1.34	1.59	1.76	1.98
	10	0.30	1.09	1.63	2.05	2.40	2.64
	12	0.32	1.21	1.96	2.50	2.89	3.24
	16	0.34	1.38	2.38	3.14	3.85	4.34
	20	0.37	1.52	2.70	3.68	4.55	5.30

Shaded areas represent *SRBR* at or above 1.

From Table 4-3, it can be observed that for a congested airport with as few as six symmetric carriers (each with frequency share of 17%), the *SRBR* exceeds 1.0 for as few as four consecutive hours of congestion ($\gamma = 4$). The *SRBR* is more sensitive to changes in the number of symmetric carriers (n) than to the length of the congestion period (γ). In other words, the degree to which carriers internalize their marginal congestion delay cost appears to be more of a determinant on *SRBR* than the length of the congestion period.

Table 4-4. Surplus-maximizing Tolls ($/flight), Symmetric Operators, Cases E & F

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
Number of perfectly symmetric carriers (n)	2	122	344	502	630	745	840
	3	162	459	669	840	992	1120
	4	182	516	752	945	1116	1260
	5	195	550	803	1008	1191	1344
	6	203	573	836	1050	1240	1400
	8	213	602	878	1103	1302	1470
	10	219	619	903	1134	1339	1512
	12	223	630	919	1155	1364	1540
	16	228	645	940	1181	1395	1575
	20	231	653	953	1197	1414	1595

Tolls expressed in dollars per flight.

Cross-referencing Tables 4-4 and 4-3 also shows that the absolute amount of congestion toll by itself is not an adequate indicator of *SRBR*. For example, the surplus-maximizing per-flight toll for $n = 5$, $\gamma = 2$ is about $550, which is almost half of the toll of $1,050 for $n = 6$, $\gamma = 4$, but the *SRBR* for the former is 60% compared with 110% for the latter.

Table 4-5 shows how the *SRBR* changes with a steeper market *NMR* curve (Case F). Contrasting this with Table 4-3 shows that with a steeper *NMR* curve, the *SRBR* is reduced. Note that since the surplus-maximizing toll depends solely on the location where the *MSCC* curve intersects the market *NMR* curve, the tolls that correspond to Case F are the same as the ones that correspond to Case E (in Table 4-4). In other words, if the airlines' frequency of service is highly sensitive to congestion delays (flatter *NMR* curve, as in Case E), imposing the same amount of toll eliminates a larger number of

82

flights that would otherwise be operated, and hence the amount of surplus gain per dollar of toll collected would be larger. Note that given a surplus-maximizing toll, one does not know whether better or worse off airlines become without knowing their *NMR* curves.

Table 4-5. SRBRs for Symmetric Operators, Case F

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
	2	0.03	0.08	0.09	0.09	0.11	0.11
Number of	3	0.05	0.13	0.19	0.21	0.21	0.24
perfectly	4	0.06	0.17	0.27	0.30	0.36	0.36
symmetric	5	0.06	0.21	0.34	0.40	0.48	0.52
carriers (n)	6	0.08	0.25	0.39	0.49	0.57	0.66
	8	0.07	0.29	0.47	0.62	0.75	0.85
	10	0.09	0.31	0.52	0.73	0.92	1.07
	12	0.09	0.34	0.59	0.80	1.00	1.21
	16	0.08	0.36	0.65	0.92	1.18	1.41
	20	0.08	0.38	0.73	1.01	1.33	1.63

Between Cases E and F, the slope of the *NMR* curve is changed. Between Cases E and G, however, the slope of the *NMR* curve remains the same, but the surplus-maximizing frequency is increased slightly by 10% (above the capacity, at $\lambda^s = 1.10 \times \mu$ = 88 flights/hour for all $t \leq \gamma$). Note that because the aircraft take-off and landing queues are now longer than in Case E, the *MSCC* curves need to be re-generated (sloping upwards at a steeper angle).

Tables 4-6 and 4-7 respectively show how the *SRBR* and the surplus-maximizing toll vary as a function of γ and the number of symmetric operators (n) for Case G. Comparing with Tables 4-3 and 4-4, a higher per-flight toll is required to arrive at the surplus maximum from a higher overall demand (higher surplus-maximizing frequency relative to the airport's sustainable capacity), yet the corresponding *SRBR* may be higher still. Note, however from Table 4-6 that with the slightly higher level of surplus-maximizing frequency in Case G (compared with E and F), the *SRBR* exceeds 1.0 for as few as five symmetric carriers.

Table 4-6. SRBRs for Symmetric Operators, Case G

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
Number of perfectly symmetric carriers (n)	2	0.10	0.15	0.18	0.19	0.17	0.19
	3	0.15	0.35	0.44	0.46	0.46	0.49
	4	0.21	0.54	0.71	0.77	0.82	0.89
	5	0.26	0.71	0.97	1.13	1.23	1.32
	6	0.27	0.87	1.24	1.47	1.64	1.78
	8	0.33	1.12	1.70	2.13	2.45	2.67
	10	0.35	1.32	2.14	2.75	3.18	3.56
	12	0.36	1.46	2.48	3.26	3.94	4.43
	16	0.41	1.69	3.06	4.21	5.14	5.94
	20	0.43	1.85	3.46	4.91	6.24	7.22

Table 4-7. Surplus-maximizing Tolls ($/flight), Symmetric Operators, Case G

Number of consecutive congested hours (γ)

		1	2	3	4	5	6
Number of perfectly symmetric carriers (n)	2	170	491	740	951	1139	1296
	3	227	654	987	1268	1518	1728
	4	255	736	1110	1426	1707	1944
	5	272	785	1184	1521	1821	2073
	6	283	818	1233	1584	1897	2160
	8	297	858	1295	1664	1992	2268
	10	305	883	1332	1711	2049	2332
	12	311	899	1357	1743	2087	2376
	16	318	920	1387	1782	2134	2430
	20	322	932	1406	1806	2162	2462

Summarizing the findings so far with respect to symmetric operators, given everything else the same:

i) The *SRBR* tends to increase with an increase in the number of consecutive hours of congestion (γ).

ii) *SRBR* is higher when the aggregate *NMR* curve is flatter, or more "elastic" (i.e., carriers respond to a small increase in toll with a large reduction in the number of flights operated).

iii) Airports with more symmetric carriers (larger n) tend to have higher *SRBR* than those with fewer symmetric carriers.

iv) With the same steepness in the net marginal revenue curve, a higher overall demand (as reflected in a higher surplus-maximizing frequency) tends to

result in higher *SRBR* in spite of the possibly higher tolls, given the same airport capacity, aircraft operating cost, etc.

4.5.2 The Case of Asymmetric Carriers with Identical NMR_i Intercept

If the same linear form of market marginal *NMR* curve is applied to asymmetric carriers with identical intercepts, the net marginal revenue curves for individual carriers can be expressed as:

$$NMR_i = \max[k - (\frac{m}{\alpha_i^S} \cdot \lambda_i), 0] \qquad \qquad ...[4\text{-}38]$$

where α_i^S denotes the frequency share of carrier *i* under surplus-maximizing fine tolls.

Summing these individual NMR_i's yields the aggregate *NMR* in the same form as [4-37]. Note also that for carriers with identical intercepts, each maintains a constant frequency share (α_i) in the absence of congestion effects as a profit-maximizing decision in view of the decisions from other carriers. In other words, if the number of total flight frequency is arbitrarily limited via a uniform tax per flight, each carrier still maintains the same frequency share. As such, these α_i's are the same as the surplus maximizing frequency shares α_i^S.

For the purpose of illustration, two scenarios with different congestion-free frequency shares among asymmetric carriers are used, and these are shown in Tables 4-8 and 4-9. Table 4-8 documents Case M, where carriers have similar congestion-free frequency shares, and where the surplus-maximizing frequency (Λ_t^S) is set to 110% of the airport capacity ($\mu = 80$ flights/hour) for the first of three consecutive hours ($t = 1$, $\gamma = 3$). In particular, three carriers share an identical *NMR* curve, each with 25% congestion-free frequency share (i.e., each carrier has $\lambda_i^S = 0.25$ Λ^S). The remaining two carriers also share an identical *NMR* curve, each with 12.5% congestion-free frequency share. In other words, these two frequency-weak carriers have steeper *NMR* curves than the three other carriers, and hence are expected to face higher surplus-maximizing tolls than their

85

counterparts. The "slope" parameter of the congestion-free market (aggregate) *NMR* curve is -$5 per incremental flight (relatively flat).

Table 4-8. Fine/Coarse Tolls for Asymmetric Carriers with Identical Intercept (Case M)

Carrier	NMR slope ($/flight)	No-toll frequency	Fine toll per flight ($)	Fine-toll frequency	Coarse toll per flight ($)	Coarse-toll frequency
1	-20	33	1390	22	1372	21
2	-20	33	1390	22	1372	21
3	-20	33	1390	22	1372	21
4	-40	25	1622	11	1372	14
5	-40	25	1622	11	1372	14
Total	-5	149	127424	88	124852	91
Gains($)*				152461		151849
SRBR				1.20		1.22

For $t = 1$; $\gamma = 3$; NMR_i's with identical intercepts; $\mu = 80$ flights/hour; $\Lambda^S_t = 88$ flights/hour
*Gains refer to surplus gains from the no-toll equilibrium.
A coarse toll of $1,440 is required to limit the aggregate frequency to 88 flights/hour (with a corresponding surplus gain from the no-toll equilibrium of $151,658 and *SRBR* of 1.18).

Because of the relatively flat congestion-free market *NMR* curve, an imposition of a surplus-maximizing fine toll reduces the no-toll market frequency by 40% (from 149 to 88), with much of this reduction borne by carriers with low congestion-free frequencies. If coarse tolling is adopted, the required toll is $1,372 per flight, resulting in a tolled aggregate frequency of 91 flights in the first hour. In terms of surplus gains per dollar of congestion toll, the *SRBR* for coarse tolling is quite similar to that under fine tolling.

Case N is very similar to Case M, but with different congestion-free frequencies. As shown in Table 4-9, Carrier 1 in Case N has a congestion-free or surplus-maximizing frequency share of 72% (64 flights out of 88), while each of the four identical carriers has about a 7% (6 flights out of 88) frequency share. Implementing a fine toll at the no-toll equilibrium reduces the total number of frequencies by roughly one-third (from 126 to 88), compared with almost one-half for Case M. Much of this reduction takes place among the "weaker" or more "toll-sensitive" carriers, who in the absence of tolls have significantly expanded their service from a congestion-free (same as fine-toll) scenario to a congestion-present (no-toll) scenario. Most notably, the dominant carrier (Carrier 1) in

Case N (Table 4-9) even increases its frequency of service from the no-toll equilibrium upon the imposition of a fine toll.

Table 4-9. Fine/Coarse Tolls for Asymmetric Carriers with Identical Intercept (Case N)

Carrier	NMR slope ($/flight)	No-toll frequency	Fine toll per flight ($)	Fine-toll frequency	Coarse toll per flight ($)	Coarse-toll frequency
1	-6.875	50	506	64	1290	40
2	-73.33	19	1727	6	1290	11
3	-73.33	19	1727	6	1290	11
4	-73.33	19	1727	6	1290	11
5	-73.33	19	1727	6	1290	11
Total	-5	126	73832	88	108360	84
Gains($)*				95419		88938
SRBR				1.29		0.82

For $t = 1$; $\gamma = 3$; NMR_i's with identical intercepts; $\mu = 80$ flights/hour; $\Lambda^S_t = 88$ flights/hour
*Gains refer to surplus gains from the no-toll equilibrium.
A coarse toll of $1,125 is required to limit the aggregate frequency to 88 flights/hour (with a corresponding surplus gain from the no-toll equilibrium of $88,159 and *SRBR* of 0.89).

Upon the imposition of a coarse toll instead of a fine toll, the weaker carriers in Case N are expected to almost double their service from 6 to 11 flights per hour while the dominant carrier will reduce its frequency from 64 to 40 flights per hour. The aggregate frequency operated under optimal coarse tolling reduces from 88 to 84 flights for the first hour. Meanwhile, the total amount of coarse toll that maximizes producer surplus is 47% higher than under fine tolling. In other words, imposing a coarse toll may i) require an excessive amount of tolls to be levied on some carriers, compared with the fine-toll scenario, and ii) lead to an undesirable distortion on the frequency shares from the surplus-maximizing level.

4.5.3 The Case of Asymmetric Carriers with Identical NMR_i Slope

The individual *NMR* curves for n carriers with identical slope can be expressed using the same parameters k and m in the following fashion:

$$NMR_{i,t}(\lambda_{i,t}) = \max(k_{i,t} - m \cdot n \cdot \lambda_{i,t}, 0), \text{ for } t = 1, 2, \ldots, \gamma \text{ and } i = 1, 2, \ldots, n \quad [4\text{-}39]$$

The relative positioning of the intercepts k_i can be arbitrarily chosen as an illustration. In practice, the relative positioning of this intercept reflects the attractiveness of the first flight operated by a particular carrier. A low positioning of this intercept is indicative of the fact that passengers in general do not find the services provided by this carrier attractive (relative to other carriers).

Two cases, Cases P and R, can illustrate how carriers' operating behaviour can be quite sensitive to the relative positioning of the intercepts k_i. The carrier-specific NMR_i's for these two cases are displayed in Figures 4-5 and 4-6 respectively. As shown in the Figures, the relative placements of the intercepts are very close to each other. In Case P, $k_{5,t}$ is placed at 85% of the other $k_{i,t}$'s, while in Case R, it is placed at 75% of the others.

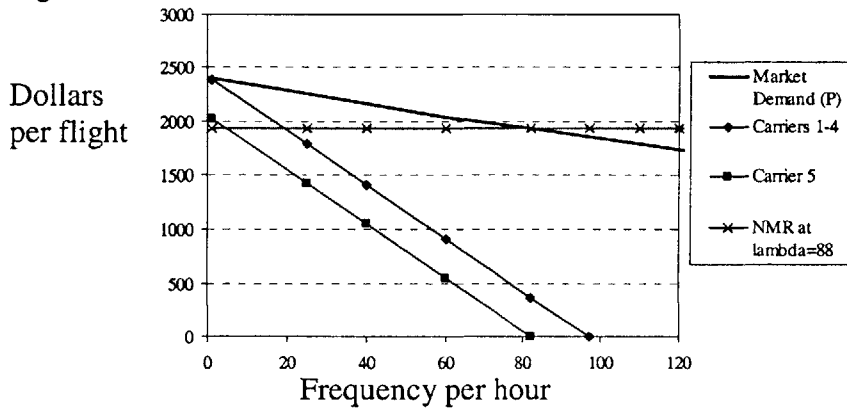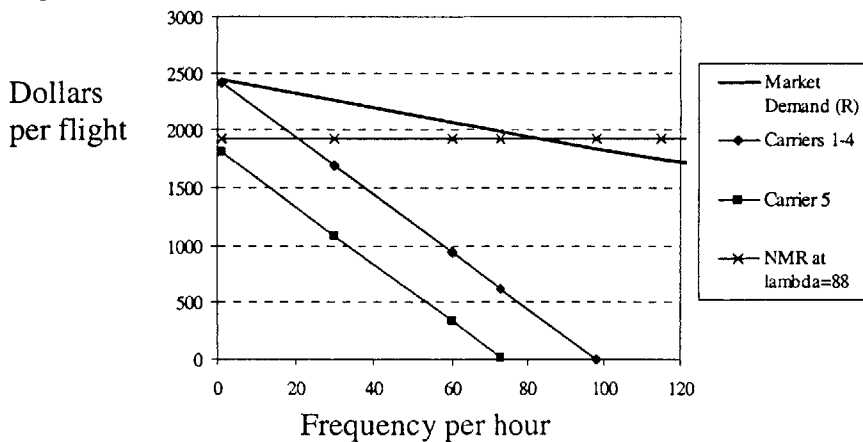Figure 4-5. *NMR* curves for Case P

Figure 4-6. *NMR* curves for Case R

As shown in Tables 4-10 and 4-11 respectively, the no-toll equilibrium frequencies for Cases P and R are not that different. Yet under congestion tolling (fine or coarse), carrier 5 completely withdraws from the market in Case R, while it only halves its frequency in Case P. Such sensitivity is attributable directly to the fact that in Case R, k_5 is just *below* the level of aggregate *NMR* that corresponds to the surplus maximum.

Table 4-10. Fine versus Coarse Tolls for Asymmetric Carriers with Same Slope (Case P)

Carrier	Relative NMR Intercept	No-toll frequency	Fine toll per flight ($)	Fine-toll frequency	Coarse toll per flight ($)	Coarse-toll frequency
1	100%	32	1367	21	1520	18.5
2	100%	32	1367	21	1520	18.5
3	100%	32	1367	21	1520	18.5
4	100%	32	1367	21	1520	18.5
5	85%	25	1790	4	1520	11
Total		153	121988	88	129200	85
Gains*				173397		172630
SRBR				1.35		1.34

For $t = 1$; $\gamma = 3$; $dNMR_i/d\lambda_i = -\$25/flt$; $dNMR_\Sigma/d\Lambda_i = -\$5/flt$; $\mu = 80$ flt/hour; $\Lambda^S_t = 88$ flt/hour
*Gains refer to surplus gains measured in dollars ($) from the no-toll equilibrium.
A coarse toll of $1,450 is required to limit the aggregate frequency to 88 flights/hour (with a corresponding surplus gain from the no-toll equilibrium of $172,917 and *SRBR* of 1.36).

Table 4-11. Fine versus Coarse Tolls for Asymmetric Carriers with Same Slope (Case R)

Carrier	Relative NMR Intercept	No-toll frequency	Fine toll per flight ($)	Fine-toll frequency	Coarse toll per flight ($)	Coarse-toll frequency
1	100%	31.5	1357	22	1357	22
2	100%	31.5	1357	22	1357	22
3	100%	31.5	1357	22	1357	22
4	100%	31.5	1357	22	1357	22
5	75%	22	1357	0	1357	0
Total		148	119416	88	119416	88
Gains*				149893		149893
SRBR				1.26		1.26

For $t = 1$; $\gamma = 3$; $dNMR_i/d\lambda_i = -\$25/flt$; $dNMR_\Sigma/d\Lambda_i = -\$5/flt$; $\mu = 80$ flt/hour; $\Lambda^S_t = 88$ flt/hour
*Gains refer to surplus gains measured in dollars ($) from the no-toll equilibrium.

Summarizing the discussion with respect to asymmetric operators:

i) The analysis of congestion tolling for asymmetric operators is more complicated than for symmetric operators.

ii)   The market and carriers' individual equilibrium frequencies under coarse tolling can be different from the frequencies corresponding to the surplus maximum associated with fine tolling.

iii)   For carriers with identical *NMR* intercepts and similar congestion-free or surplus-maximizing frequency shares, the difference of impact between fine and coarse tolling (versus no toll) can be small.

iv)   For carriers with identical *NMR* intercepts and dissimilar congestion-free or surplus-maximizing frequency shares, coarse tolling can substantially distort carriers' behaviour (i.e., forces a carrier to operate even further from the surplus-maximizing frequency).

v)   For carriers with identical *NMR* slopes, those with intercepts below the level that corresponds to the surplus maximum will choose to completely withdraw from the congested airport upon the imposition of the surplus-maximizing congestion toll. As such, the relative positioning of the *NMR* intercepts can translate into different sensitivities to congestion tolling.

4.6   Discussion

In this Chapter, an analytical framework was introduced to explore the impact of demand management in a variety of competitive market environments. With the help of a series of simplifying assumptions, several numerical examples were also presented. In particular, starting with the seemingly contrived case of symmetric carriers, some of the assumptions were gradually relaxed, resulting ultimately in a set of tools to analyze a wide range of competitive conditions at any congested airport.

One of the simplifying assumptions, for instance, requires that all carriers have the same cost structure. In reality, this need not be the case, and different cost structures among carriers, as well as different cost structures for in-flight versus ground delays, can in fact be incorporated in the model. The *MSCC* and *MPCC* functions can be recomputed to reflect these differences.

To design a realistic congestion pricing policy for multiple hours of congestion, multiple iterations may be required to determine the surplus-maximizing frequency in one hour relative to that in another. For cases where the surplus-maximizing frequency does not appear to vastly exceed the sustainable capacity at an airport, one can first attempt to work backward in time to first analyze the approximate *NMR*'s for $t = \gamma$, and then for $t = \gamma - 1$, etc., until to $t = 1$. This will yield a series of congestion tolls that should be charged at different congested periods within the day.

Alongside the development of analytical tools, a policy performance metric was introduced. In essence, the *SRBR* metric presents an order-of-magnitude estimate of the short-run welfare gains of a carrier or society as a whole with and without congestion pricing. Depending on the ultimate use of the collected tolls, the *SRBR* metric also serves as a rough guide to determine how much of the collected amount should be re-distributed back to the operators to make them economically indifferent to the imposition of the congestion fee.

Certainly, the collected tolls can be used to bring further social benefits, e.g., by expanding runway or terminal capacity at the congested airports, by expanding capacity at alternative airports or by improving other modes of transport. These benefits have not been considered in the development of the *SRBR* metric, but could conceivably be incorporated. As such, it would be *wrong* to simply assume that any demand management policy is inappropriate should the *SRBR* fall below 1.0.

One of the most important sources of uncertainty in the preceding analyses is due to the assumptions about the shape and position of the *NMR* curves relative to the MSCC curves. In practice, it is often difficult to attribute revenues gained in a complex airline network to individual flight segments, and there is likely substantial non-linearity as to how the passenger or freight demand for a particular flight may change if another flight is cancelled. Cao and Kanafani (2000) provide a framework for airlines to compute their value of runway time slots within a complex route network, by searching for the next

most profitable fleet assignment plan according to fixed travel demand. More importantly, the computation of the *SRBR* does not depend so much on the flights with high net marginal revenue, but mostly on those with low net marginal revenue. This suggests that computing the net marginal revenue for a small set of flights may be sufficient in enumerating the *SRBR*'s for most congested airports.

For carriers with different levels of *NMR* intercepts, carriers' sensitivity toward congestion tolling depends critically on the placement of their intercepts relative to the level that corresponds to the surplus optimum. Given the same *NMR* intercepts, as the frequency share of each carrier decreases, the *SRBR* tends to increase. This makes sense once it is realized that much of the *MSCC* for carriers with low frequency shares is simply not internalized among them. A low degree of internalization of *MSCC* is a valid reason for governmental intervention, or at least a clearer definition of property rights (in this case the right to use the take-off and landing capacity of congested airports).

Chapter 5

Implications for Public Policy

In Chapter 4, analytical tools and numerical experiments were presented to develop the case for and examine the potential impact of congestion pricing in a number of competitive scenarios. In this Chapter, several policy insights and implications from these experiments will be examined in greater detail. Note again that the discussion here pertains primarily to the efficient utilization of scarce airport capacity, and only sketches some arguments on the imperfect competition that may exist on select city-pair markets. A comprehensive solution that incorporates both efficient capacity utilization and remedies to imperfect competition requires substantially more information on consumers' preferences than have been assumed.

In Section 5.1, general policy-oriented insights from the analysis of congestion pricing in Chapter 4 will be presented. In Section 5.2, the discussion will move toward a more general, market-based demand management policy for airport airside traffic, including general aviation operations. In Section 5.3, some proposals on demand management at New York's LaGuardia Airport from the FAA will be reviewed in the light of the findings of this thesis, while in Section 5.4 potential uses of funds collected by a market-based demand management system will be discussed.

5.1   Insights from the Analysis of Congestion Pricing

From the development of analytical tools in Chapter 4, a number of policy-related insights emerged. Some of these will be recast and discussed at greater length in this

section. For now, the discussion will focus on congestion pricing as the only means of market-based demand management and on non-cooperative, profit-maximizing aircraft operators as its target group. A discussion on other means of market-based demand management will begin in Section 5.2.

First, the specific market structure at an airport affects how receptive aircraft operators would be to a congestion pricing policy. With a short-term horizon (without accounting for any potential benefits generated by appropriately using the congestion tolls collected), aircraft operators as a whole tend to benefit more per dollar of toll paid when i) congestion pricing is implemented at a relatively atomistic airport market (like New York LaGuardia, LGA, where the top carrier accounts for only 35% of the frequencies), and ii) the demand for airport access is high (i.e., either the number of consecutive congested hours is high, or the economically optimal flight frequency is high compared with the airport's sustainable capacity, or both). In other words, a congestion pricing policy would be most attractive, even to the target toll-payers, under these conditions.

At more concentrated airport markets (e.g., with one dominant carrier), the situation is less clear. Frequency-dominant carriers tend to internalize more congestion externalities than their weaker counterparts, and therefore tend to restrict their output (in terms of frequency) to a larger extent. To arrive at a surplus-maximizing operating condition on the utilization of scarce airport capacity, the dominant carriers appear to rightly qualify for lower congestion tolls than their weaker counterparts.

In practice, however, imperfect competition may arise: network economics can help explain why a frequency-dominant carrier may be more than proportionately attractive to passengers than weaker carriers. The market power that results in the dominant carrier may lead it to restrict output. In return, increased price competition brought about by the weaker carriers may encourage more output by the dominant carrier, and to encourage this a lower congestion toll for them may be applicable. This picture can complicate even further, as frequency does not necessarily equate with the level of

competition by carriers in their city-pair markets. In other words, frequency-weaker carriers may serve select markets with a high degree of market power, while frequency-dominant carriers may find themselves competing with many other carriers in large markets. As such, how remedies for imperfect competition should be incorporated into congestion tolls depends on the specific competitive circumstances. From the facility utilization standpoint alone, the tentative recommendation that frequency-dominant carriers should qualify for reduced congestion tolls still stand.

One caveat in this discussion is that the airplane size has thus far been kept constant. Congestion tolling may encourage carriers to use larger airplanes. As a result, congestion tolling may not necessarily change the size of final outputs in terms of seats provided for specific markets.

In view of these considerations, congestion at carrier-concentrated airports like Atlanta-Hartsfield (ATL), where the top carrier accounts for 74% of the frequency share, warrants a graduated congestion tolling scheme, with the frequency-dominant carrier charged at a much reduced (or, if necessary, even zero or negative) toll. Appropriate measures to counter-act monopolistic pricing by the frequency-dominant carrier at these airports should also be in place. Demand management with more uniform tolling is much more warranted at airports where carriers have more similar frequency shares. Figure 5-1 summarizes this discussion by graphically illustrating the market conditions under which demand management would be most needed. Recall also that Figure 3-8 provides more detail on the frequency shares of dominant carriers at several major airports in the US.

Second, as demonstrated in Chapter 4, coarse congestion tolling is likely to induce some distortions to the carriers' equilibrium frequencies. This occurs even if the coarse-toll surplus-maximizing frequency coincides exactly with that under fine tolling. This is largely because those flights with low marginal contribution would still be viable for carriers with a small degree of internalization of the congestion cost, while those with relatively high marginal contribution may not be viable for the frequency-dominant

carriers. This point is often ignored in the discussion of congestion pricing in the airport
context, and presents a notable difference from the urban transport context.

Figure 5-1. Appropriateness for Congestion Pricing in Different Market Conditions

Most desirable for congestion pricing

e.g. LGA

Increasingly
fragmented
carriers

Graduated tolls
warranted

e.g. ATL

Increasing demand relative to capacity ⟶

■ Most desirable for congestion pricing
▢ Least desirable for congestion pricing
LGA New York LaGuardia Airport, NY
DFW Dallas/Fort Worth Airport, TX (a hub for American/Delta)
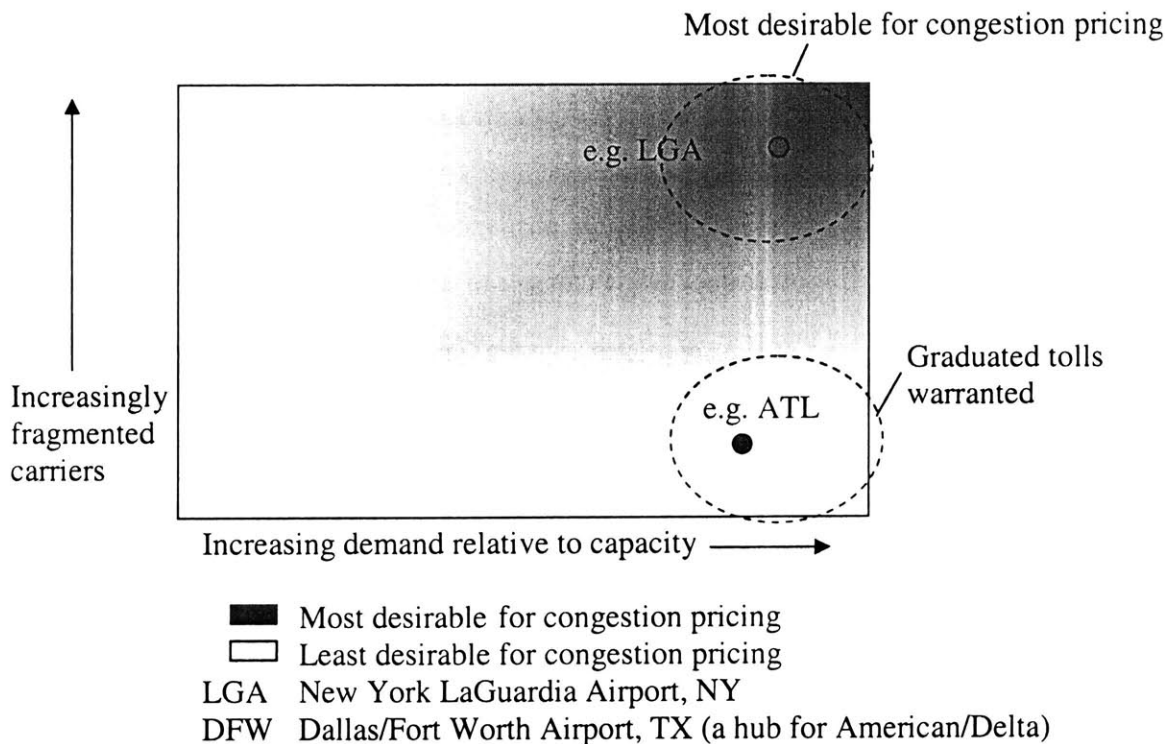
Third, while coarse tolling induces undesirable market distortions compared with
fine tolling, for cases where the frequency-dominant carrier has less than, say, a 50%
frequency share, the amount of distortion created by coarse tolling is still small. For
example, in Case M (see Table 4-8), the surplus gain achieved by surplus-maximizing
coarse tolling is within 90% of that achievable under fine tolling.

The approach taken in Chapter 4 is that the coarse toll is set such that producer
surplus is maximized. Note that this requires no less information than in fine tolling. In
particular, the aggregate frequency that results from this is somewhat different from both
the sustainable airport capacity and the surplus-maximizing frequency with fine tolling.
Alternatively, for administrative expediency, a more convenient aggregate frequency
(e.g. one that is close to the sustainable capacity at the congested airport) can be used as a

target frequency, and the congestion toll continually adjusted until this target is reached. As shown in the example of Case M (bottom of Table 4-8), such a target-oriented coarse toll is still expected to achieve surplus gains tantalizingly close to those achievable under fine tolling.

This kind of target-oriented coarse tolling does not require the policy administrator to have nearly as much information on carriers' individual *NMR* curves as does fine tolling. In fact, given the no-toll equilibrium, it can be argued that the policy administrator does not need to know anything about carriers' individual *NMR* curves in order to implement coarse tolling, as long as they are monotonically decreasing and have similar slopes, i.e., sensitivity to tolls ($dNMR/d\lambda$). A frequent adjustment of the amount of toll levied can be used to arrive at the coarse toll that corresponds with an surplus-maximizing frequency of service. As long as the NMR_i's and the *MSCC* functions are monotonic, Λ is also monotonic as a function of the toll T and a simple Newtonian adjustment can be used to iteratively estimate a new coarse toll:

$$T_{i+2} = T_i + (T_{i+1} - T_i) \cdot \left(\frac{\Lambda^S - \Lambda_i}{\Lambda_{i+1} - \Lambda_i} \right) \qquad \qquad ...[5\text{-}1]$$

where the superscript S refers to the surplus-maximizing frequency, or a proxy target.

Note that equation [5-1] does not depend on *any* specific values of the *NMR* curves. As the no-toll equilibrium provides one set of initial parameters (i.e., $T = 0$, and firms' profit-maximizing frequencies totalling Λ^P), only one other set of toll-operating parameters is required to enable administrators to gradually adjust the amount of congestion toll to reach the economically desirable level. In other words, given an initial trial toll, further adjustments of the congestion toll toward the economic optimum can be conducted on an informed basis.

Tables 5-1 and 5-2 document how [5-1] can be applied starting from the no-toll equilibrium for Case M (asymmetric carriers with equal intercepts) from Chapter 4. This

97

illustration differs slightly from those in Jansson (1998) in that the latter assumes constant demand elasticity in terms of flight frequency from the part of carriers, while no such assumption is made here. In particular, Table 5-1 describes how the coarse toll and the equilibrium frequencies evolve from an initial coarse toll of \$500/flight (for $t = 1$, $\gamma = 3$), well below the level corresponding to the surplus-maximizing frequency; while Table 5-2 starts with an initial toll of \$2,000/flight, well above the surplus optimum. In both cases, the equilibrium frequency converges relatively quickly to the surplus optimum of 88 flights/hour (with \$1,440 in toll per flight). However, a higher initial toll appears to result in a considerably more drastic response from carriers than a lower initial toll.

Table 5-1. Evolution of Coarse Tolls with a Low Initial Toll (Case M)

| Trial | Coarse Toll | Aggregate frequency | Carriers' individual frequencies | | | | |
			1	2	3	4	5
0	\$ 0/flight	149	33	33	33	25	25
1	\$ 500	126	28	28	28	21	21
2	\$ 1,326	93	21	21	21	15	15
3	\$ 1,372	91	21	21	21	14	14
4	\$ 1,441	88	20	20	20	14	14

For $t = 1$; $\gamma = 3$; NMR_i's with identical intercepts; $\mu = 80$ flights/hour; $\Lambda^S_t = 88$ flights/hour

Table 5-2. Evolution of Coarse Tolls with a High Initial Toll (Case M)

| Trial | Coarse Toll | Aggregate frequency | Carriers' individual frequencies | | | | |
			1	2	3	4	5
0	\$ 0/flight	149	33	33	33	25	25
1	\$ 2,000	55	13	13	13	8	8
2	\$ 1,298	96	22	22	22	15	15
3	\$ 1,435	88	20	20	20	14	14

For $t = 1$; $\gamma = 3$; NMR_i's with identical intercepts; $\mu = 80$ flights/hour; $\Lambda^S_t = 88$ flights/hour

Fourth, even if the congestion toll is by *fiat* limited to a level below that which corresponds to the surplus-maximizing aggregate frequency, there is still a net benefit to society (and often to the carriers as well). Table 5-3 documents how the series of coarse

tolls imposed in Case M according to Table 5-1 can result in net benefits to society before the surplus maximum is reached. For example, even the imposition of an initial congestion toll of $500/flight in the first hour ($t = 1$) can result in net surplus gains in excess of 40% of the maximum potential (and an *SRBR* of greater than 1). In particular, the reduction in delay costs more than compensates for the loss of revenue from curtailed flight operations.

Table 5-3. Benefits to Society at Different Levels of Coarse Tolls (Case M)

Trial	Coarse Toll	Aggregate frequency	Change Delay cost*	from Revenue*	no-toll Net benefit	equilibrium SRBR*
0	$ 0/flight	149				
1	$ 500	126	-$131,479	-$ 60,376	+$ 71,103	1.13
2	$ 1,326	93	-$248,234	-$ 93,302	+$154,932	1.26
3	$ 1,372	91	-$252,604	-$ 96,878	+$155,726	1.25
4	$ 1,441	88	-$258,610	-$100,534	+$158,076	1.25
Fine	Fine Tolls	88	-$258,610	-$ 79,572	+$179,038	1.41
$\Lambda=\mu$	$ 1,625	80	-$271,587	-$102,193	+$169,394	1.30

For $t = 1$; $y = 3$; NMR_i's with identical intercepts; $\mu = 80$ flights/hour; $\Lambda^S_t = 88$ flights/hour

* The change in the delay cost for all flights in the three ($y = 3$) hours of congestion due to changes in the first hour ($t = 1$) only. The change in the revenue is due to changes in the number of flights operated in the first hour ($t = 1$) only. The *SRBR* is calculated by dividing the net benefits (due to changes in the frequency in $t = 1$) by the amount of tolls levied for the same period (to implement this change). *SRBR*'s for subsequent hours must be computed separately in the same fashion.

The imposition of either coarse or fine tolling will result in the same reduction in delay costs at the surplus-maximizing frequency, but fine tolling minimizes the amount of revenue loss in a way that maximizes the producer surplus. In addition, if the toll is larger than the surplus-maximizing toll, there can still be significant net benefits to the carriers. The last entry in Table 5-3 shows that if the toll is charged such that the aggregate frequency in the first period is equal to 80 flights/hour (below the surplus maximum), the *SRBR* is still a respectable 1.30.

In other words, even if *neither* the exact optimal frequency *nor* the optimal toll are known, imposing a congestion toll at an airport with the right competitive structure will still likely result in net benefits to society and, very probably, *SRBR*'s greater than 1.

This means that the loss in net revenues can usually be overcome by increases in aggregate surplus, i.e., at least one mechanism exists such that reduction of net revenue (following the imposition of congestion tolling) can be fully compensated for through a combination of delay reduction and/or redistribution of collected tolls. This is indeed an observation that can serve as a strong argument in countering criticisms of congestion pricing.

Summarizing our discussion so far, congestion pricing as an effective demand management policy instrument:

i)   is most desirable at airports where no carrier has a dominant, overriding frequency share;

ii)   can be beneficial to carriers even if the target frequency does not equal the sustainable capacity or the surplus-maximizing aggregate frequency; and

iii)   can still be quite effective under many conditions in practice, even if the coarse tolling method is chosen.

5.2   Slot Lease Auctions

So far, the discussion of market-based demand management has been limited to congestion pricing. In reality, this is not the only policy option. In this section, slot lease auctions, as an alternative policy instrument to congestion pricing, will be discussed.

In congestion pricing, the policy administrator selects a congestion toll that induces carriers to operate limited flight frequencies to/from a congested airport. In a slot auction, the administrator chooses the desirable number of aggregate flight frequencies (i.e., slots) and lets in the highest bidders to these. Given perfect information and certainty, there is theoretically no difference between using either price or quantity as the planning instrument to arrive at the desirable market equilibrium. As Weitzman (1974) pointed out, however, airlines do not have perfect information or certainty on their future

revenues and costs. Therefore, slot lease auctions and congestion pricing may cater to different planning preferences among aircraft operators. For example, an emerging carrier with an uncertain expansion plan may want to have the right to operate more flights, and hence may favour a policy that lets carriers "pay as they go" through a congestion pricing scheme. On the contrary, an established carrier serving mature markets may have a far better idea on how many flights it would like to offer in the future than a new entrant, and hence be more confident in appraising the worth of slots and placing bids for these.

Recent advances in the auctioning of electromagnetic spectra in the U.S. provide important insights into how airport takeoff and landing slot auctions may be held.[9] Incidentally, this view is also shared by the US Department of Justice (2002). Electromagnetic spectra for primarily telecommunication purposes share similar properties as airport slots in that they are valued differently in different aggregations. For example, electromagnetic spectra covering contiguous regions may be valued more than geographically disjoint ones, and a single collection of spectra covering a number of major metropolitan areas may be worth more than the sum of individual spectra covering individual cities. Likewise, a single slot in the morning peak hour and another single slot in the evening peak hour may be worth more if they can be used by the same carrier than if each has to be used by different carriers.

McAfee and McMillan (1996) and Milgrom (2000) describe in detail the principles and implementation of simultaneous ascending auctions, while Cramton (1997) documents the relatively successful experience with these auctions of the Federal Communications Commissions (FCC) between 1994 and 1996. In short, multiple licences are open for bidding at the same time in simultaneous ascending auctions, and bid prices can only increase during the course of the auction. To prevent bidders from holding out till the last moment, an auction is divided into several rounds, and the maximum number of licences that a bidder can qualify to "win" at the end of each round depends on the number of licences on which the bidder has already placed bids by that round. The final

---

[9] Interesting insights on the set-up of auction schemes for sulfur dioxide emissions [see for example, Hauser (1992) and Joskow *et al* (1998)] are also relevant in this regard.

award of each licence is based on the highest bid placed at the end of a predetermined number of rounds, as well as the number of licences that a bidder is qualified to "win". For example, if a company placed the highest bids in the last round of bidding but had not been qualified to win any bids in previous rounds, it would still not be awarded any licence.

The primary advantage of this auction is that each bidder can see how much each licence and different combinations of licenses are worth before finalizing its bid. Moreover, spectra with close substitutability have a high chance of arriving at comparable prices, thereby reducing the "winner's curse", i.e., the amount of premium that a winner has to pay (above what others are willing to pay) in order to acquire the winning licence.

In the context of airport slot auctions, a predetermined number of "slots" can be put up for auction in a similar fashion. Because of the amount of effort involved in physically holding each auction, with all qualifying parties involved (instead of having just one administrative organization that sets the congestion toll in the case of congestion pricing), the tenure of the slot can conceivably be longer than, say, that of a particular congestion toll. Based on past evidence of a relatively inactive buy-sell market for slots under High-Density Rule (see for example, Gleimer, 1996), it is advisable that the slot tenure has a finite length of, say, about five years. This finite tenure will force the temporary holders of the slots to re-appraise the economic value of these slots, and will give an opportunity to other interested parties to acquire these slots.

As such, the slot auction approach, if properly implemented as suggested here, should be more appropriately referred to as a slot *lease* auction. To better contain the element of unpredictability in any auction, the slot lease tenures can be staggered, such that only a portion of all slot leases is up for auction at any time.

To visualize the equivalence between congestion pricing and slot lease auctions, consider a carrier that is deciding whether to acquire access to a congested airport by

paying the congestion toll every period versus leasing a slot for a considerable number of periods. The congestion toll (T) payable by the congestion-priced traffic would be a function of the aggregate frequency during time period t, i.e., Λ^t (the superscript here denotes a longer time frame, e.g. a period which is, may be, a month long instead of an hour long):

$$T^t = T(\Lambda^t) \qquad \text{for } t = 0, 1, 2,\dots \text{ (with } t=0 \text{ as the current period)} \qquad \dots[5\text{-}2]$$

where T^t denotes the congestion toll in period t.

With perfect and certain information about how many flights each carrier would operate, the future congestion tolls are known. Then, with a term structure of discount rates such that the one-period, zero-coupon bond prices (starting from $t = 1$, 2, etc.) are Z^1, Z^2, etc., carriers should be indifferent between using the congestion-priced capacity versus placing a bid for a slot lease for n periods starting at $t = 0$ in the amount of:

$$Highest\ Equivalent\ Bid = \sum_{t=1}^{n} Z^t \cdot T(\Lambda^t) \qquad \dots[5\text{-}3]$$

Assuming the demand curves faced by one carrier are not affected by the number of flights it and its competitors operate, a rational carrier should bid neither more nor less than what it would otherwise pay in present value under the congestion pricing policy during the tenure of the slot lease. In fact, a profit can be made for certain with no net initial investment, i.e., arbitrage opportunities would occur, if the price for any substitutable slot lease for n periods, L^n, (given perfect and certain information) is not equal to:

$$L^n = \sum_{t=1}^{n} Z^t \cdot T(\Lambda^t) \qquad \dots[5\text{-}4]$$

For example, if a carrier is willing to bid a higher price $L^h > L^n$, then an intermediary selling this slot lease at L^h can take advantage of the arbitrage opportunity by paying for the congestion fee every period on the carrier's behalf. (Any excess balance from $t = 0$ should be put into riskless bonds as described above.) The present value of this arbitrage profit will be equal to $L^h - L^n$. Similar arguments apply for $L^h < L^n$. This implicitly assumes that auction slot leases are fully transferable, and that transaction costs are negligible.

Given full information and certainty on future operational trends, there would be a single price for each pair of substitutable slot leases, i.e., carriers will be "taxed" the same amounts per slot lease. In other words, the overall effect to carriers would be that of a coarse congestion toll. As such, the applicability of coarse tolling also applies for the case of slot lease auctions. In other words, while it is applicable for airports with atomistic carriers (each with a small frequency share), it calls for additional modification to be used at airports with highly asymmetric carriers (in terms of frequency shares).

The more interesting case is the one where carriers only have partial or uncertain information on future operational trends. For example, consider the situation in which a particular carrier has secretly (i.e., unbeknownst to the other carriers) decided to increase its number of flights in the middle of its slot lease tenure (therefore increasing the expected present value of congestion tolls it would have to pay), while all other carriers remain committed to the currently known number of flights under all circumstances. Acquiring a slot lease at the price given by equation [5-3] would then be a less expensive option for that carrier than using the congestion-priced scheme, since the congestion toll would increase if there were more flights than originally expected.

Conversely, if this carrier secretly plans to reduce its number of flights during the slot lease term (therefore reducing the expected present value of congestion tolls), with all else being equal, it would be worthwhile using the congestion-priced capacity instead. (Similar arguments apply if the carrier in question expects that *other* carriers will increase or decrease the number of their flights in the future.) In other words, changes in the prices

of slot leases represent changes in expectations about carriers' future operating decisions. Therefore, adding slot lease auctions as an alternative option to congestion pricing has the potential of channelling additional information to the administrator as to how future congestion tolls may have to be adjusted. It also gives carriers additional options for fine-tuning their strategies.

To summarize the discussion in this section:

i)   Slot lease auctions are a viable alternative demand management approach, in addition to congestion pricing; and

ii)   The principles of simultaneous ascending auctions currently in use in telecommunications spectrum auctions can be applied reasonably well to the allocation of scarce airport capacity among competing carriers.

What "currency" should be used in a surplus-maximizing auction is in fact open for debate. The most common denominator for this is the amount in dollars a carrier is willing to pay. Borenstein (1988) points out that because different carriers may face different demand functions, slot auctions do not automatically guarantee social, as oppose to economic, efficiency. To see how this may happen, consider two carriers that face different demand elasticities (i.e., own-price elasticities, ignore cross-elasticities for the moment). The more elastic the demand is, the more sensitive the total revenue is with respect to changes in prices. As a result, the carrier that faces the less elastic demand will generally be willing to pay a higher bid than the other carrier with the more elastic demand, everything else being equal. Since a higher willingness to pay does not automatically equal social preference, the auction result does not guarantee that social objectives will be met.

One convenient remedy to this pitfall is to use the monetary-based auction as a common denominator, but allow part of the capacity to be allocated according to some other socially desirable criteria. This adjustment for social objectives will be discussed in

greater detail in Section 5.4, while the essence and practical implications of the auctioning approach will be explored in the remainder of this section.

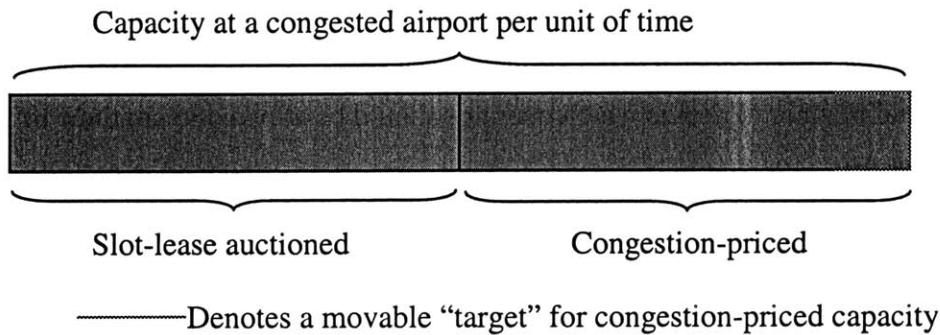## 5.3 Toward a Market-Based Demand Management Policy

At this point, the basic ingredients of a market-based demand management policy have been introduced. In this section, these ingredients will be combined to suggest a policy framework. Some further adjustments to fulfill specific social objectives will be discussed in the next section.

Congestion pricing and slot auctions can be seen as orthogonal to each other in that one fixes a price for airport access, while the other fixes the quantity to be accessed. Broadly speaking, a market-based demand management policy may include some combination of these two policy instruments. The phrase "some combination" includes the cases where the policy may involve the exclusive use of only one of these instruments, especially fine congestion tolling only at airports with grossly asymmetric carriers (in terms of frequency shares).

Figure 5-2 shows graphically how the capacity of an airport with symmetric (possibly with each carrier having significant frequency shares) or atomistic carriers may be segmented and allocated using these two instruments. Note that there is now no specific limit for the total number of flights allowed, since congestion pricing does not explicitly prohibit flight movements beyond certain limits. Depending on the length of tenure of the slot leases, some minimum level of congestion-priced capacity may be specified to ensure the existence of a liquid, spot market. Meanwhile, the slot lease auctions can provide valuable information to the airport authority as to carriers' future expectations of flight frequencies. As noted also by comments from the Department of Justice (2002), this will reduce the informational burden on the airport authority.

In other words, the congestion-tolled capacity acts as a "spot market" for airport capacity. By having a spot market for airport capacity, the need to specify a minimum utilization criterion, also known as a "use-it-or-lose-it" criterion, is not necessary as long as the congestion toll T is determined by the actual and anticipated delay (which in turn is driven by the number of flights operated, Λ).

Figure 5-2. A dual-channel market-based demand management policy

Capacity at a congested airport per unit of time

Slot-lease auctioned   Congestion-priced

————Denotes a movable "target" for congestion-priced capacity

If they are faced with massive uncertainties about the future, carriers may shy away from slot leases and prefer congestion tolling. As the bid prices for slot leases drop below a certain minimum amount, the administrator of the demand management policy may wish to increase the proportion of capacity administered under congestion tolling and decrease the proportion of slot lease auctions. If the uncertainties persist, the policy may eventually degenerate to one where only congestion pricing is practiced.

5.4   Achieving Social Objectives

As demand management is instituted at a congested airport with fairly symmetrical carriers and hitherto unregulated access, certain flights would be consolidated or dropped. In the presence of a benign social planner, the marginal congestion delay caused by a particular flight can be weighed against the schedule convenience it provides, i.e., how much less passengers would expect to wait to travel to that destination given that this incremental flight is provided. If the delay cost to society

(i.e., subsequent flights and passengers) outweighs the scheduled convenience of a particular flight, then dropping this flight would result in net benefits to society, and should therefore be warranted. This is the approach suggested in Hansen (2002) for the case of Los Angeles International Airport (LAX).

For cases where each carrier operates numerous flights to the same destinations, as in LAX, this approach can be implemented without much technical difficulty. Consider, however, the case where one of two closely spaced flights to the same destination has to be dropped, but each is operated by a different carrier. It would then be difficult for the social planner to decide which operator should be allowed to continue to operate its flights, or that both should be required to cooperate, i.e., condoned to collude, on this service.

The use of a purely market-based approach eliminates this ambiguity by directly or indirectly allocating the scarce airport capacity to those flights with the highest economic value. However, such surplus or efficiency maximization in the airport capacity allocation process does not automatically guarantee the optimization of social values such as reduced schedule convenience to underserved destinations. Taken one step further, while a once-a-day, 20-seat flight to a small, remote township may generate smaller profits than a 100-seat flight which is part of a 12-flight-a-day shuttle service to a large city, the contribution to the local economy of the remote township by the 20-seat flight may be of critical importance, compared with perhaps the only marginal importance of the 100-seat flight. In other words, society may value the less profitable, 20-seat flight more than the 100-seat flight.

To allow for this kind of social objectives to be properly reflected in the market-based demand management policy, financial subsidies can be provided. In FAA's "Notice of Alternative Policy Options for Managing Capacity at LaGuardia Airport and Proposed Extension of the Lottery Application" (2000), Policy Option B for congestion pricing explicitly incorporates this thinking. In this Option, two different congestion tolls would be levied: a lower fee for "all flights operating between LGA and any small hub or

non-hub airport qualifying for AIR-21 service, as well as general aviation flights", and a higher fee for all other flights. Effectively, an operational subsidy, equal to the difference between these two fees, is "built into" the fee structure to expressly encourage those flights deemed to be "socially attractive". This is Borenstein's argument (1988).

To avoid gaming efforts by carriers, the subsidy would be best designed as a "lump-sum subsidy". In this scheme, the total amount of subsidy to be granted to a particular type of flight operation (e.g. to underserved destinations) is fixed or determined by all parties at regular intervals, and the actual per-flight subsidy to which each flight or carrier is entitled depends on the number of flights operated by each carrier or carrier group. For example, the actual subsidy disbursed per eligible carrier i, s_i, may be related to the congestion toll (T), the total amount of designated subsidy (S), and the total number of flights flown by all carriers in this category (Λ_f) in this manner:

$$s = \min(T \cdot \lambda_{f,i}, \frac{S}{\Lambda_f} \cdot \lambda_{f,i}, \varphi \cdot S) \qquad \qquad ...[5\text{-}5]$$

where $\lambda_{f,i}$ is the number of subsidy-eligible flights operated by carrier i;

φ is a maximum portion of S that can be allocated to a single carrier.

T, the congestion toll per flight is included simply to avoid a net subsidy to certain carriers. It is certainly conceivable, however, that a net subsidy might be allowed, in which case the first term in the bracket would not be included in the formula. In any case, if the total number of flights in this subsidy category turns out to be many times larger than what was originally intended, the case *for* continuing such subsidy would be weaker than originally thought, and the actual subsidy disbursed per flight would in fact be so small that each flight in this category would essentially bear the full market rate. Mathematically, if the subscript f denotes the number of flights operated in the subsidy category,

$$\text{as } \Lambda_f \to \infty, \ \frac{S}{\Lambda_f} \to 0, \text{ and hence } s \to 0 \qquad \qquad ...[5\text{-}6]$$

Alternatively, if the congestion toll rises as a result of an increase in overall demand, the subsidy-eligible flights will automatically have to bear some of this increase, unless society demonstrates that the total subsidy should also increase. Table 5-4 presents three examples on how this fixed-sum subsidy scheme works if there are a) more flights in the subsidy category but the same total number of flights, and b) more flights outside of the subsidy category but the same number of flights in the subsidy category. As shown in the Table, at the base case where 50 of the 150 flights qualify for the subsidy, each flight qualifies for $1,000 of subsidy. This amount will not change as long as there are 50 eligible flights. However, if there are more non-subsidized flights ("more total demand"), the congestion toll increases and hence carriers will need to incur a higher out-of-pocket cost (i.e., subsidy-eligible flights are not entirely immune to changes in the market for airport access). If carriers substitute flights that are ineligible for subsidy with ones that are eligible (e.g., in anticipation of worse economic conditions in the short term), then the per-flight subsidy will decrease. This is illustrated under "more subsidy flights" in Table 5.4. The net out-of-pocket cost per flight for subsidy-eligible flights will also be higher in this case (despite the fact that the congestion toll remains the same).

Table 5-4. A Lump-sum Subsidy Scheme

	Base case	More subsidy flights	More total demand
Sum of subsidized flights (λ_f)	50	100	50
Total number of flights (λ)*	150	150	200
Total subsidy (S)	$ 50,000	$ 50,000	$ 50,000
Assumed congestion toll (T)*	$ 1,200	$ 1,200	$ 1,700
Subsidy per flight (s)	$ 1,000	$ 500	$ 1,000
Carriers pay this per flight	$ 200	$ 700	$ 700

*Figures in these categories are intended for the purpose of illustration only.
Assumes that the total subsidy allocated to each carrier does not exceed an allowable limit.

So far, the discussion of the lump-sum subsidy scheme has been limited to the congestion-priced capacity. In fact, if carriers operating the subsidized routes prefer slot leases, there is no economic reason why choosing to participate in the slot auctions should affect their eligibility toward their share of subsidy. To illustrate this point, Figure

5-2 showed how the capacity of an airport may be divided into a congestion-priced regime and a lease-auctioned regime. Consider the case where the operators of the first Λ-1 flights have decided whether to use the lease-auctioned or congestion-priced capacity. The operator of the Λth flight is now considering whether to use the lease-auctioned or congestion-priced capacity (assuming that such flexibility is allowed). Whether this operator chooses to acquire a slot lease or be congestion-priced does not change the fact that the congestion toll, T, will be based on a total of Λ flights. As a consequence, as long as all the airport slots are considered to be substitutable for one another, and there is a certainty on future trends, each slot lease should be priced exactly as in equation [5-4].

To summarize the discussion so far,

i)   Lump-sum subsidy schemes can be used to address specific policy goals in a flexible demand management strategy; and

ii)   The lump-sum subsidy scheme should be applicable regardless of a carrier's preference for congestion pricing or slot leases.

5.5   Uses of Funds

One of the more politically contentious issues in a market-based demand management policy concerns the use of any funds thus collected by the airport operator or other facility provider. As mentioned in Chapter 4, a portion of the congestion toll or slot lease bids may need to be distributed back to the carriers to ensure they would be appropriately compensated for their potential net loss of business. Small (1992) suggests a host of different uses of revenues from congestion pricing on highways, while FAA (2001a) outlines potential uses for funds collected from airport congestion tolls. Some of these are highlighted and discussed here.

Under a users-pay principle where airport systems are treated as isolated entities, excess funds collected from congestion-mitigating measures can be used for longer-term

capacity expansion. Along this line, Morrison (1983) modelled the optimal long-run investment in airport capacity by minimizing the sum of capacity expansion and delay costs, while Oum and Zhang (1990) modelled the effect of lumpiness of airport capacity (e.g., one runway versus no runway). While using these funds for long-term capacity expansion is a desirable goal, the very reason why certain airports become congested has to do with the difficulty of implementing any expansion plans at all.

In cases where longer-term capacity expansion is not practical, the collected funds can still be used to make sure that the congested airport does not act as an unreasonable deterrent to competitive entry. In this regard, several entrant carriers have raised a number of possibilities in their correspondences with the FAA. Most of these relate to the difficulty these carriers encounter at congested airports even after a runway slot is granted. For example, entrant carrier Vanguard Airlines (2001) stated that "operations at [LaGuardia] by new entrant carriers are severely restricted due to the continued unavailability of airport gates". Likewise, Spirit Airlines (2001) stated that "reliance on a single gate means that [it] suffers especially severe service disruptions when it encounters ground delays at [LaGuardia]." To further the goal of easing market entry and levelling the playing field for competitive carriers, a portion of the funds collected from any demand management policy may thus be used for, say, building or expanding common-use facilities on the landside at congested airports.

In cases where alternative airports are viable alternatives to a congested one, revenues collected under a demand management policy can also serve to fund such capital-intensive projects as improving ground access to these alternative airports. As an example, American Airlines (2000) and United Airlines (2001) assign their flight numbers to high-speed train service from Paris Charles de Gaulle Airport to a number of cities in France, and even award frequent flyer miles for such trips. Air France even abandoned its Paris-Brussels service in 2001 in favour of partnering with the French National Railroad Company to offer a high-speed rail link between these two cities (Air France, 2001). Lufthansa (2002) has done the same for a number of German city-pairs and even publishes the complementary train service schedule alongside flight

alternatives. Depending on the eventual popularity of the German high-speed train alternative, up to 5% of airport slots at Frankfurt-am-Main Airport dedicated to short-haul destinations may be freed in this manner (Baker and Field, 2001).

In short, a comprehensive demand management policy that aims at an efficient allocation of scarce airport capacity should be integrated within the framework of a broader policy that aims at increasing the congested airport's airside capacity, at comparable improvements of potentially congested landside facilities, as well as at coordinated regional transportation planning.

Chapter 6

Conclusions

Congestion at airports presents a more difficult analytical problem than congestion on urban highways. The mere fact that operators decide whether or not to operate a block of flights, and therefore potentially internalize different portions of their external congestion costs, vastly complicates the economics of congestion pricing.

These complexities, however, are no excuse for steadfastly adhering to non-market-based demand management policies, which for certain, will push airport users further away from competitive equilibria or Pareto optimality. Nevertheless, several elements of the more sophisticated variations of purely administrative demand management policies, as described in Chapter 2, can be useful in a market-based demand management policy.

The main thrust of this thesis is the modeling of the complexities of congestion pricing with operators of varying characteristics, from symmetric to asymmetric, and from frequency-dominant to atomistic. The analytical model developed focuses on the efficient utilization of scarce airport capacity, and abstracts away from the issue of imperfect competition. A probabilistic queuing model was used to generate realistic congestion delay estimates. Some of the main findings of the numerical experiments are as follows:

i)   The gains in carrier surplus relative to the amount of congestion toll collected tends to increase with an increase in the length of the congestion period (γ),

with an increase in the number of symmetric carriers (n), and with a flatter net marginal revenue curve (more "elastic" with respect to tolls).

ii)   Owing to the different degrees to which congestion cost is internalized, the analysis of congestion tolling for asymmetric operators is far more complicated than for symmetric operators.

iii)   For carriers with identical intercepts of their net marginal curves (*NMR*) and similar congestion-free or surplus-maximizing frequency shares, the difference in the impact of fine versus coarse tolls can be small.

iv)   For carriers with identical intercepts of their net marginal curves but dissimilar congestion-free or surplus-maximizing frequency shares, coarse tolling can substantially distort carriers' behaviour (i.e., forces a carrier to operate even further away from the surplus-maximizing frequency). Whether a frequency-dominant carrier would expand or reduce its frequency upon imposition of a surplus-maximizing toll depends on the relative frequency shares of other carriers.

v)   The relative positioning of the intercepts for carriers' net marginal curves can translate into different sensitivities to congestion tolling.

From a policy perspective, the analysis presented in this thesis argues against instituting uniform congestion pricing (i.e., coarse tolling) at airports where a single carrier strongly dominates in flight frequency. Instead, a graduated congestion tolling structure that mimics fine tolling is warranted at these airports.

At other airports where carriers' frequency shares are less disparate (i.e., less asymmetric), a combination of congestion pricing and slot lease auctions is advised. Given perfect information, there is no economic difference between these two approaches. In practice, these two approaches can cater toward different planning preferences by the airport authority and/or carriers. In either approach, any adverse effects on specific classes of flights deemed socially desirable (e.g. those to small and non-hub cities) can be mitigated through lump-sum subsidies. An important and helpful finding on the practical side is that the successful implementation of a market-based

demand management policy at these airports does not require precise knowledge of the net-marginal-revenue curves of the carriers at congested airports, as long as it is known that these curves satisfy certain reasonable assumptions.

Whether the application of demand management measures is warranted for most of the day or for only a few peak hours depends on the characteristics of the demand-to-capacity relationship at the specific airport studied. For the case of New York LaGuardia airport (LGA), after the relaxation of the historical slot-based demand management system called HDR, carriers consistently scheduled more flights than the optimal capacity from morning till late evening. After a temporary slot lottery, the number of flights operating at LGA on weekdays was reduced by roughly 10%. Yet this relatively small reduction resulted in an 80% reduction in the estimated average delays suffered by flights during the evening peak period and an 85% reduction in the estimated total delays for a typical weekday. This indeed presents a convincing case *for* demand management.

The analytical model developed in this thesis has not adopted the more microscopic viewpoint of starting with preferences at the passenger level. This is a complex task. Further research in this direction may be useful in understanding more precisely the trade-off in social benefits between instituting a market-based demand management policy versus none at all (e.g. for congested airports with a frequency-dominant carrier).

Bibliography

Air France, 2001. *Horaires/Timetable/Flugplan/Horarios*, 25 March – 27 October, 16.

American Airlines, 2000. *Worldwide Timetable*, July 2, 62-63.

Arnott, R., Palma, A.D. and Linsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *American Economic Review*, 83, 161-179.

Arnott, R. and Kraus, M., 1995. When are anonymous congestion charges consistent with marginal cost pricing? *Journal of Public Economics*, 67, 45-64.

Baker, C., 2001. Slot machines. *Airline Business*. 17 (6), 91-93.

Baker, C. and Field, D., 2001. Jumping the tracks. *Airline Business*, 17(6), 106-108.

Barnhart, C., P. Belobaba and A. R. Odoni, 2003. Applications of Operations Research in the Air Transport Industry, to appear in *Transportation Science*.

Baumol, W.J. and Bradford, D.F., Optimal departures from marginal cost pricing, *American Economic Review*, 60 (3), June, 265-283.

Borenstein, S., 1988. On the efficiency of competitive markets for operating licenses. *Quarterly Journal of Economics*, 103(2), May, 357-385.

Borenstein, S. 1989. Hubs and high fares: dominance and market power in the US airline industry. *Rand Journal of Economics*, 20, 344-365.

Borenstein, S., and Rose, N., 1994. Competition and price dispersion in the US airline industry. *Journal of Political Economy*, 103, 653-683.

Borenstein, S., and Netz, J., 1999. Why do all flights leave at 8 am? Competition and departure-time differentiation in airline markets. *International Journal of Industrial Organization*. 17, 611-640.

Bratu, S., 2003 (forthcoming). *Airline passenger on-time service reliability – analysis, algorithms and optimization decision models*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Brueckner, J.K., 2002. Internalization of Airport Congestion. *Journal of Air Transport Management*, 8, May, 141-147.

Brueckner, J.K., 2003. Airport congestion when carriers have market power. *American Economic Review*, 93 (1), 1357-1375.

Brussels International Airport Company, 2003. *Tariffs for the use of Aeronautical Services at the Brussels National Airport as from April 1, 2003.*

Cao, J-M. and Kanafani, A., 2000. The value of runway time slots for airlines. *European Journal of Operational Research*, 126, 291-500.

Carlin, A. and Park, R.E., 1970. Marginal cost pricing of airport runway capacity. *American Economic Review*, 60, 310-319.

Coase, R.H., 1960. The problem of social cost. *Journal of Law and Economics*, 3, 1-44.

Cramton, P., 1997. The FCC spectrum auctions: an early assessment. *Journal of Economics and Management Strategy*, 6(3), Fall, 431-495.

Daniel, J.I., 1995. Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues, *Econometrica*, 63, 327-370.

Daniel, J.I. and Pahwa, M., 2000. Comparison of three empirical modes of airport congestion pricing. *Journal of Urban Economics*, 47, 1-38.

Department of Justice (US), 2002. *Comments of the United States Department of Justice On the FAA Notice of Alternative Policy Options for Managing Capacity at LaGuardia Airport and Proposed Extension of Lottery Allocation.* US Department of Transportation Docket No. FAA-2001-9854-76.

Department of Transportation (US), 2000/1. *Air Travel Consumer Report.* Office of Aviation Enforcement and Proceedings. February.

Diamond, P.A., 1973. Consumption externalities and imperfect corrective pricing, *Bell Journal of Economics*, 4 (2), 526-538.

FAA, 2000. High density airports – Notice of lottery of slot exemptions at LaGuardia. *Federal Register*, 65 (233), December, 75765-75771; Docket FAA-2000-8278.

FAA, 2001a. Notice of alternative policy options for managing capacity at LaGuardia airport and proposed extension of the lottery allocation. *Federal Register*, 66 (113), June, 31731-31748; Docket FAA-2001-9852.

FAA, 2001b. *Airport Capacity Benchmark Report*, 2001.

Fan, T.P.C., 2002. Smaller aircraft for more profits? A preliminary examination of airlines' fleet size decision with fare and demand distributions. *Journal of the Transportation Research Forum*, 56 (3), summer, 77-93.

Fan, T.P.C. and Odoni, A.R., 2002. Airport Demand Management – A Practical Perspective, *Air Traffic Control Quarterly*, 10 (3), Special Issue: Capacity Improvements and Demand Management in the National Airspace System, 285-306.

General Accounting Office, 1990. *Airline Competition: Industry Operating and Market Practices Limit Market Entry*, U.S. General Accounting Office Report to Congressional Requesters, RCED 90-147.

Gleimer, E., 1996. Slot regulation at high density airports: how did we get here and where are we going? *Journal of Air Law and Commerce*, 61 (4), May-June, 877-931.

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.

Hauser, K., 1992. The politics and economics of auction design in the market for sulfur dioxide pollution. *Journal of Policy Analysis and Management*, 11(4), 553-572.

Honeywill, T., 2002. Stop-start for Road Pricing. *Professional Engineering*, 15 (5), 13 March, 26-27.

IATA, 2000. *Worldwide Scheduling Guidelines*, Effective 26 March, First edition.

Jansson, M., 1998. *Marginal cost congestion pricing under approximate equilibrium conditions*. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Joskow, P.L., Schmalensee, R., and Bailey, E.M., 1998. The market for sulfur dioxide emissions. *American Economic Review*, 88 (4), 669-685.

Kahn, A., 1993. The competitive consequences of hub dominance: a case study. *Review of Industrial Organization*, 8, 381-405.

Kivestu, P.A., 1976. *Alternative Methods of Investigating the Time-Dependent M/G/k Queue*, Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Knight, F.H., 1924. Some fallacies in the interpretation of social cost. *Quarterly Journal of Economics*, 38, August, 582-606.

Koopman, B.O., 1972. Air Terminal Queues under Time-Dependent Conditions, *Operations Research*, 20, 1089-1114.

Levine, M., 1969. Landing fees and the airport congestion problem. *Journal of Law and Economics*, 12, 79-108.

Levine, M., 1994. Airline competition in deregulated markets: theory, firm strategy and public policy. *The Economics of Transport*, 2, 283-284.

Lufthansa, 2002, *Flugplan/Timetable*, 31 March – 30 June, 24-25.

Malone, K., 1995. *Dynamic Queuing Systems: Behaviour and Approximations for Individual Queues and for Networks*, Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.

Mayer, C., and Sinai, T., 2002. Network effects, congestion externalities and air traffic delays: or why all delays are not evil. NBER (National Bureau of Economic Research) working paper W8701.

McAfee, R.P. and McMillan, J., 1996. Analyzing the airwaves auction. *Journal of Economic Perspectives*, 10(1), Winter, 159-175.

McDonald, J.F., 1995. Urban high congestion – an analysis of second-best tolls. *Transportation*, 22(4), 353-369.

McShan, S. and Windle, R., 1989. The implications of hub-and-spoke routing for airline costs and competitiveness. *Logistics and Transportation Review*, 25, 209-230.

Milgrom, P., 2000. Putting auction theory to work: the simultaneous ascending auction. *Journal of Political Economy*, 108(2) April, 245-272.

Mohring, H. and M. Harwitz. 1962. *Highway Benefits: An Analytical Framework*. Northwestern University Press, 81-86.

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

Morrison, S.A., 1987. The equity and efficiency of runway pricing. *Journal of Public Economics*, 34, 45-60.

de Neufville, R., and Odoni, A.R., 2003. *Airport Systems – Planning, Design and Management*. McGraw-Hill, New York, 461-497.

Official Airline Guide, 2000. MAXX desktop product.

Oum, T.H. and Zhang, Y., 1990. Airport pricing: congestion tolls, lumpy investment and cost recovery. Journal of Public Economics, 43 (3), December, 353-374.

PANYNJ (Port Authority of New York and New Jersey), 2001. *Schedule of Charges for Air Terminals*, Issued by the Aviation Department, January, 7.

Pagliari, R., 2001. Selling grandfather: an analysis of the latest EU proposals on slot trading. *Air and Space Europe*, 3 (1/2), 33-35.

Pigou, A.C., 1920. *The Economics of Welfare*. Macmillan. London.

Ramsey, E.P.A., 1927. A contribution to the theory of taxation. *Econ~mic Journal*, 37, March, 47-61.

Small, K.A., 1992. *Urban Transportation Economics*, Harwood Academic Press, Chur, Switzerland.

Shumsky, R.A., 1993. *The Response of the U.S. Air Carriers to the DoT's On-Time Disclosure Rule*, Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Spirit Airlines, 2001. Comments of Spirit Airlines, Inc., U.S. Department of Transportation Docket FAA-2001-9852-19, 3.

Stamatopoulos, M.A., 2000. *A Decision Support System for Airport Planning*, Doctoral Dissertation, Athens University of Economics and Business, Athens, Greece.

United Airlines, 2001. *Worldwide Timetable*, 5, July 9 – September 4, 82-83.

Vanguard Airlines, 2001. Comments of Vanguard Airlines, Inc., U.S. Department of Transportation Docket FAA-2001-9852-7, 3.

Vickrey, W., 1963. Pricing in urban and suburban transport. *American Economic Review*, 53, May, 452-465.

Vickrey, W., 1969. Congestion theory and transport investment. *American Economic Review*, 59, 251-260.

Walters, A.A., 1978. Airports – an economic survey. *Journal of Transport Economics and Policy*, 12 (2), May, 125-160.

Weitzman, M., 1974. Prices versus quantities. *Review of Economic Studies*, 41(4), October, 477-491.

Zhang, A., 1996. An analysis of fortress hubs in airline networks. *Journal of Transport Economics and Policy*, 30, 293-307.