

INTERNALIZATION OF AIRPORT CONGESTION: A NETWORK ANALYSIS

JAN K. BRUECKNER

CESIFO WORKING PAPER NO. 756

CATEGORY 1: PUBLIC FINANCE

JULY 2002

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the CESifo website:* www.CESifo.de

INTERNALIZATION OF AIRPORT CONGESTIONS: A NETWORK ANALYSIS

Abstract

The likely resurgence of air traffic in the U.S. means that airport congestion is a problem that must soon be confronted by policy makers. As part of their policy response, it is probable that some form of congestion pricing will be imposed at selected U.S. airports in the relatively near future. The theory developed in this paper, which extends the results of Brueckner (2002), provides an important guide for the formulation of congestion pricing rules. In particular, the theory says that the congestion tolls levied on the various airlines at a particular airport should generally be different, with the tolls being inversely related to a carrier's airport flight share. Internalization of airport congestion is the reason for this inverse relationship. In operating another peak flight, a carrier takes account of the congestion damage imposed on the other flights it operates. If these flights account for a large share of the airport's traffic, then most of the congestion created by the additional flight is internalized, justifying a low toll. By contrast, if the carrier operates only a few of the airport's flights, then little internalization occurs, and a high toll is needed to force the carrier to take into account the congestion damage it causes. The resulting flight-share rule is easy to implement, and it could help policymakers design proper toll systems at U.S. airports.

JEL Classification: H0, L9.

Jan K. Brueckner
Department of Economics and Institute
of Government and Public Affairs
University of Illinois at Urbana
1206 South Sixth St.
Champaign, IL 61820
U.S.A.
jbrueckn@uiuc.edu

Internalization of Airport Congestion: A Network Analysis

by

Jan K. Brueckner*

1. Introduction

Airport congestion and flight delays became serious problems in the U.S. at the end of the 1990s. Although the fall-off in air traffic after the terrorist attacks of September 2001 reduced congestion levels at U.S. airports, flight delays are likely to become a problem again once air traffic returns to its long-term growth trend. While the expansion of airport capacity offers a potential remedy for the congestion problem, the high cost and long gestation periods of such investment projects limit their attractiveness. More-immediate relief is offered by measures that allow better use of existing airport infrastructure. Such measures include improvements in air traffic control technology, which allows busy airports to handle more traffic while generally increasing the capacity of the airspace. In addition, airport congestion pricing, which diverts flights to off-peak hours by raising the level of landing fees during peak periods, is now frequently discussed as a remedy for the delay problem (for example, see Transportation Research Board, 2000). Indeed, the Federal Aviation Administration is considering the imposition of congestion pricing at New York's La Guardia airport, one of the most congested in the nation.

The theory of congestion pricing has been developed mainly in the context of road pricing (see Small (1992) for a survey). The theory shows that peak usage of a road or other congested facility is excessive because each user does not take into account the delays he imposes on fellow users. Peak usage can be optimally restricted by imposing a congestion toll equal to the cost of the external delays that each user generates.

The earliest contributions applying this principle to the case of airports include Levine (1969), Carlin and Park (1970), Morrison (1983), Morrison and Winston (1989), and Oum and Zhang (1990). However, more recent work by Daniel (1995, 2002), Daniel and Pahlwa (2000), and Brueckner (2002) recognizes a crucial difference between the airport and road contexts not appreciated by the early papers. In particular, while road users are properly viewed as

atomistic, with each user accounting for a tiny share of total traffic, airlines must be viewed as nonatomistic given that one or two carriers operate most of the flights at the highly congested U.S. airports. For example, United and American each operate around 40 percent of the flights at Chicago-O'Hare, while Delta operates over 70 percent of Atlanta's flights. As a result, an atomistic model of congestion will be inappropriate when applied in an airport context.

Brueckner (2002) shows that, when the atomistic model is abandoned, the verdict on congestion is softened. Although atomistic users of a congested facility ignore their external effects, his analysis shows that a nonatomistic airline takes into account a portion of the congestion caused by each of its flights. In particular, the airline *internalizes the congestion each flight imposes on the other flights it operates*. In the monopoly case, all congestion is internalized, while in the symmetric oligopoly case, each carrier internalizes a fraction α of the congestion caused by an extra flight, where α equals each carrier's airport flight share. In the model, airlines internalize the operating cost of congestion as well as passenger time costs, which are captured because congestion is capitalized into lower fares.

These conclusions suggest that, in the airport case, the overallocation of flights to the peak period may not be as severe as the atomistic model would predict. Correspondingly, the optimal congestion toll is lower than the one that would be generated by the standard road-pricing formula. In the monopoly case, no toll is needed since all congestion is internalized. In the symmetric oligopoly case, the toll should equal the congestion damage caused by an extra flight times one minus each carrier's airport flight share (i.e., $3/4$ of this damage if the airport has four symmetric carriers). The atomistic model would imply that the toll equals the full congestion damage from an extra flight.

Brueckner (2002) provides rudimentary empirical support for these results by showing that, in a sample of the 25 most-congested airports, the number of delays falls as airport concentration rises, holding the other determinants of delays constant. As predicted, this outcome reflects greater internalization of congestion as carrier market power grows.¹

Brueckner's (2002) analysis is limited, however, because it portrays an unrealistic route structure, where the congested airport is connected to a single other airport, which is uncongested. Given this limitation, it is natural to wonder whether the conclusions of the analysis

generalize to a route structure that more closely resembles a system of actual airline networks. The purpose of the present paper is to carry out such a generalization.

The assumed network structure is shown in Figure 1. Two airlines, denoted 1 and 2, serve four cities, A, B, H, and K. City H is a hub for airline 1, while city K is a hub for airline 2. Airline 1 serves the city-pair markets AH, BH, AB, HK, with passengers in market AB required to connect at the hub H. Similarly, airline 2 serves the city-pair markets AK, BK, AB, and HK, with its AB passengers connecting at K. For simplicity, cities are symmetric in the sense that demand for traffic in all city-pair markets is the same. It is assumed that airports A and B are uncongested in equilibrium, while the hub airports H and K experience congestion.

While including the type of routes analyzed by Brueckner (2002), this network structure contains a number of features not present in his model. The common elements are city-pair markets such as AH, where a congested airport is connected to an uncongested airport, with the route served by a monopoly carrier (airline 1). An analogous observation applies to markets BH, AK, BK. Unlike in Brueckner's model, however, the network structure contains a city-pair market, HK, that connects two congested airports, with this market subject to competition between carriers 1 and 2. Finally, market AB connects two uncongested airports, but passengers in this market generate congestion because of their need to connect at either hub H or K. In addition, this market is subject to competition between the two airlines.

The paper analyzes whether the internalization principle, and hence the congestion-pricing rule, of Brueckner (2002) apply in this more complex setting. The analysis shows that the answer is affirmative, and the results can be most easily grasped when stated in terms of the congestion-pricing rule. As in Brueckner's model, this rule states that the congestion toll levied on each one of an airline's flights at a given airport equals *the congestion damage from an extra flight times one minus the carrier's airport flight share*. Thus, zero congestion tolls are charged at the uncongested airports, A and B. Positive tolls are charged at the congested airports H and K, but these tolls are asymmetric across carriers because of their different usage of the airports. Since airline 1 operates more flights at hub H, thus internalizing more congestion, each of its flights (those operating on routes AH, BH and HK) is charged a lower toll at H than each of airline 2's flights, which operate on the HK route. The analogous pattern holds

at airport K, with airline 2 paying a low toll per flight and airline 1 facing a high toll for its HK flights. This toll structure is potentially controversial because, despite concerns about the hub dominance and its anticompetitive effects (see Borenstein (1999)), a low toll is charged to an airport’s dominant carrier, with a high toll paid by the airline with the lower flight share.

The paper is organized as follows. Section 2 reviews Brueckner’s (2002) model and develops its main implications. Section 3 analyzes the network model, characterizing the socially optimal allocation of traffic. Section 4 characterizes the equilibrium traffic allocation in the network model and derives the congestion pricing rule. Section 5 offers conclusions.

2. The Model and Benchmark Results

2.1. The setup

To develop the model, the discussion focuses for simplicity on the case where a single congested airport is connected to one uncongested airport. In this case, there is only one city-pair market and one set of passengers to consider. The analysis of the network case in Figure 1 requires several modifications to this setup, which are introduced in Section 3 below.

The model distinguishes between two travel periods at a given airport, denoted peak and off-peak. The peak period consists of a set of relatively short time intervals containing the day’s most desirable travel times, such as early morning or late afternoon. The off-peak period represents travel times not included in the peak.

To avoid inessential complications, the off-peak period is assumed to be uncongested at both airports over the range of passenger allocations examined in the model. In effect, the demand for off-peak travel is assumed to be small enough relative to airport capacity that off-peak congestion never occurs. By contrast, the congested airport always experiences peak-period congestion over the range of relevant allocations, while the other airport is again uncongested during the peak.

Unfortunately, these assumptions do not explicitly capture the traffic patterns at actual hub airports, where flights are concentrated in evenly-spaced traffic “banks” that allow the interchange of connecting passengers between arriving and departing flights. Even though these banks lead to a cyclical intraday traffic pattern, the morning and late afternoon banks

tend to be larger than the midday banks.² As a result, a peak/off-peak distinction like that in the model exists at hubs despite the overall cyclical traffic pattern.

At the congested airport, peak congestion depends on the number of flights operating during the peak period, denoted n_p . Congestion raises an airline's operating costs, so that cost per flight is given by c in the off-peak period and by $c + g(n_p)$ in the congested peak period, where g is nondecreasing and convex. The function $g(\cdot)$ must equal zero when n_p is sufficiently small, but its positive range (where g is increasing) is assumed to be relevant. All flights are assumed to use identical aircraft with fixed seat capacity s , and a 100 percent load factor is assumed, so that all seats are filled.

To see the effect of congestion on passenger time costs, consider the demand side of the model. Passengers are represented by a continuum with index θ . For simplicity, θ is uniformly distributed between zero and one with unit density, so that the total mass of passengers is unity. Passenger utility is given by the sum of consumption x and travel benefits B , with $u = x + B$. Since consumption is equal to income minus the airfare, it follows that travel decisions can be based on the difference between benefits B and the fare.

Travel benefits, which depend on θ , differ for peak and off-peak travel. The benefits from off-peak travel are given by the function $b_o(\theta)$. The benefits of peak travel, which are affected by congestion and thus by n_p , are represented by $B_p(\theta, n_p) \equiv b_p(\theta) - t(n_p)$. The function b_p thus represents the “gross” benefit of peak travel, which would apply in the absence of congestion. The function t represents the additional passenger time costs resulting from travel during the congested peak period. The additively separable form of B_p is a key assumption that simplifies the analysis. The implication of separability, which may be unrealistic, is that time costs $t(n_p)$ do not depend on θ , which makes them identical for all consumers. See Brueckner (2002) for a discussion of the effects of relaxing the separability assumption. Like g , $t(\cdot)$ is assumed to be nondecreasing and convex, and its positive range (where the function is increasing) is assumed to be relevant.

Both peak and off-peak benefits are assumed to be increasing across the passenger continuum, with $b'_o(\theta), b'_p(\theta) > 0$ holding for all θ . In addition, to ensure a simple division of passengers between the peak and off-peak periods, the benefit functions are assumed to satisfy

a “single-crossing” assumption. In particular, one of the following relationships holds for all $\theta \in [0, 1]$: $b'_p(\theta) > b'_o(\theta)$, $b'_p(\theta) = b'_o(\theta)$, or $b'_p(\theta) < b'_o(\theta)$. For simplicity, the analysis focuses on the first case, where the peak benefit function is always steeper than the off-peak function. While this case is the most intuitively plausible one, as argued below, Brueckner (2002) discusses the effect of adopting an alternate assumption.

To understand the intuition underlying the assumptions on the benefit functions, θ can be viewed as an index of the passenger’s tendency to travel on business. Since business travel, associated with a high θ , is a crucial job requirement, both peak and off-peak travel benefits should be high relative to benefits for a low- θ leisure traveler. As a result, $b'_p, b'_o > 0$ should hold. Moreover, since business travel must occur during the early and late peak hours to avoid disruption of the work day, peak travel benefits should increase relative to off-peak benefits as θ increases, yielding $b'_p > b'_o$.³

2.2. The social optimum

The social optimum is an allocation of passengers to the peak and off-peak periods that maximizes welfare, which equals the difference between travel benefits for passengers and airline costs. Given the single-crossing assumption, Brueckner (2002) shows that the optimal allocation has the natural property that high- θ passengers use the peak period, with low- θ passengers traveling off-peak. The optimization problem then involves choosing the critical point θ^* that separates the two groups of passengers. In addition, a lower bound $\underline{\theta}$ is chosen, below which consumers do not travel.

The welfare measure (travel benefits minus airline costs) can be written

$$W = \int_{\underline{\theta}}^{\theta^*} b_o(\theta) d\theta + \int_{\theta^*}^1 [b_p(\theta) - t(n_p)] d\theta - n_o c - n_p [c + g(n_p)], \quad (1)$$

where n_o is the number of off-peak flights (recall that the density of θ is unitary). The discreteness of peak and off-peak flights is ignored, with both n_p and n_o chosen in a continuous fashion to satisfy the relations $sn_p = 1 - \theta^*$ and $sn_o = \theta^* - \underline{\theta}$ (recall that s gives seats per

flight). Substituting in (1), W can then be rewritten as

$$\int_{\underline{\theta}}^{\theta^*} b_o(\theta)d\theta + \int_{\theta^*}^1 \{b_p(\theta) - t[(1 - \theta^*)/s]\}d\theta - c(1 - \underline{\theta})/s - [(1 - \theta^*)/s]g[(1 - \theta^*)/s]. \quad (2)$$

The first-order condition for choice of θ^* is given by

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)] - g(n_p)/s - n_p[t'(n_p) + g'(n_p)/s] = 0, \quad (3)$$

where $n_p = (1 - \theta^*)/s$.⁴ The first expression in (3) gives the change in travel benefits for a passenger who is switched from the off-peak to the peak period, who gains $b_p(\theta^*) - t(n_p)$ in peak benefits while losing $b_o(\theta^*)$ in off-peak benefits. Because the presence of the extra peak passenger requires the airline to operate $1/s$ additional flights, costs rise by $(1/s)g(n_p)$, the second term in (3). In addition, the extra passenger generates a congestion effect. Because of the required increase in peak flights, time cost rises by $(1/s)t'(n_p)$ for each of $1 - \theta^*$ existing peak passengers, for a total increase of $n_p t'(n_p)$. The added congestion also raises operating costs for each existing peak flight by $(1/s)g'(n_p)$, for a total increase of $(n_p/s)g'(n_p)$. The congestion effect caused by the extra passenger is thus $n_p[t'(n_p) + g'(n_p)/s]$. With all these considerations taken into account, the optimal θ^* thus balances the individual gain from additional peak travel against the incremental congestion and operating costs it generates.

At an interior solution, it can be shown that $\underline{\theta}$ satisfies $b_o(\underline{\theta}) = c/s$. Thus, for the lowest- θ passenger to fly, travel benefits are just equal to the cost of providing a seat.

2.3. Oligopoly equilibrium

To develop the oligopoly equilibrium conditions for this simple model, consider the case where two airlines compete. The fares for peak and off-peak travel are allowed to differ, and they are denoted f_p and f_o . To determine the allocation of passengers between the periods in any market equilibrium, observe that, at the point where the continuum divides between the peak and off-peak, the relevant passenger (whose type is again denoted θ^*) is indifferent between travel in the two periods. Thus, θ^* must satisfy $b_p(\theta^*) - t(n_p) - f_p = b_o(\theta^*) - f_o$, indicating that travel benefits net of the fare are equal across periods. Note that n_p in this

relationship represents total peak flights for the two carriers, $n_p^1 + n_p^2$. Note also that since $b'_p > b'_o$ holds, all passengers with $\theta > \theta^*$ strictly prefer the peak while all passengers with $\theta < \theta^*$ strictly prefer the off-peak.

When carriers have market power, the off-peak fare, which is paid by the lowest- θ passenger to fly, will be set to exhaust travel benefit for that passenger, with $f_o = b_o(\underline{\theta})$. Substituting into the above indifference relationship, the peak fare can then be written $f_p = b_p(\theta^*) - t(n_p) - b_o(\theta^*) + b_o(\underline{\theta})$. Letting n_o^1 and n_o^2 denote the carriers' off-peak flights, θ^* and $\underline{\theta}$ then satisfy $1 - \theta^* = s(n_p^1 + n_p^2)$ and $\theta^* - \underline{\theta} = s(n_o^1 + n_o^2)$. Solving for θ^* and $\underline{\theta}$, and substituting into the above fare solutions, f_p and f_o can be expressed as functions of the four flight variables.

Carrier i 's profit is written

$$f_o s n_o^i + f_p s n_p^i - c(n_o^i + n_p^i) - n_p^i g(n_p^1 + n_p^2). \quad (4)$$

Its goal is to maximize (4) by choice of n_o^i and n_p^i subject to the above relationships between the fares and the n variables. In doing so, carrier i treats the other carrier's flight choices as parametric. Assuming interior solutions for all variables, and imposing symmetry, the first-order condition for n_p^i reduces to⁵

$$\begin{aligned} [b_p(\theta^*) - t(n_p) - b_o(\theta^*)] - g(n_p)/s - (n_p/2)[t'(n_p) + g'(n_p)/s] \\ - [(1 - \theta^*)/2][b'_p(\theta^*) - b'_o(\theta^*)] = 0. \end{aligned} \quad (5)$$

Comparing (5) to the optimality condition (3), similarities and differences are apparent. First, note that the first line of (5) is almost identical to the expression in (3), with the key difference being that the congestion terms are multiplied by $1/2$. Thus, each carrier internalizes *only half* of the congestion imposed by an additional flight. But this fraction represents the congestion that the carrier *imposes on itself and its own passengers*, which is therefore taken into account in its decisions.

While the carrier naturally internalizes the impact of congestion on its own operating costs, exploitation of market power accounts for the internalization of passenger time costs. To see

this conclusion, note from the indifference condition that the first term in (5) equals the fare differential $f_p - f_o$ between the peak and off-peak periods. Now suppose that the carrier moves a passenger from the off-peak to the peak period, with appropriate changes in its peak and off-peak flights. Referring to (5), the revenue earned from this passenger increases by an amount equal to the fare differential $f_p - f_o$, represented by the first term. But to induce the given passenger to move between the periods, the carrier must accept a peak fare reduction of $t'(n_p)/s$, which offsets the increase in passenger time costs from higher peak congestion. This fare reduction in turn generates a revenue loss of $(n_p/2)t'(n_p)$ on the carrier's inframarginal peak passengers, which must be subtracted from the revenue gain from the given passenger. In this way, the carrier internalizes the effect of congestion on the time costs of its own passengers. However, the effects of congestion on the costs incurred by the other airline and its passengers are not considered.

A “residual” market-power effect, which is not beneficial, is captured by the last term in (5). This residual effect arises because to increase peak flights, the airline must convince lower- θ passengers, who value the peak relatively less, to use that period, an inducement that requires a further decline in f_p . In other words, because $b'_p > b'_o$ holds, peak benefits fall by more than off-peak benefits as lower- θ passengers are added to the peak, necessitating a further reduction in f_p to maintain the indifference condition.

It is easily seen that the effect of uninternalized congestion is to make θ^* too small, so that too many passengers tend to use the peak period (recall that their number is $1 - \theta^*$). However, because the fare reduction underlying the residual market-power effect is unappealing to the carrier, it tends also to allocate too few passengers to the peak period. As a result, the net effect on peak usage is indeterminate, making the comparison between the equilibrium and optimum ambiguous.

A congestion toll can be used to force the carriers to take account of the congestion that they do not internalize. The toll per flight should equal

$$R(n_p) = \left(1 - \frac{1}{2}\right) [sn_p t'(n_p) + n_p g'(n_p)], \quad (6)$$

where the term in brackets is the congestion damage caused by an extra flight and the mul-

tiplicative factor is one minus each carrier’s flight share. It is easily seen that, when this toll expression is subtracted from carrier profit, the $1/2$ factors in the first-order condition (5) disappear, so that congestion is fully taken into account.

As explained in Brueckner (2002), imposition of a congestion toll may not be welfare-improving. To see this conclusion, observe that if the residual market-power effect dominates, leading to underuse of the peak period, then by further restricting peak usage, the toll pushes the equilibrium farther away from the optimum. Only if the residual effect is very small in magnitude compared to the congestion effect can we be assured that a toll system is desirable. This outcome, of course, reflects second-best considerations, with the toll correcting only one of the distortions reflected in (5).

3. Network Analysis: The Social Optimum

With the above background, the discussion now turns to the analysis of traffic allocations in the realistic network depicted in Figure 1. Several general observations regarding this network setting are useful at the outset. First, in contrast to the simple model, where there was no distinction between an airline route and a city-pair market, this distinction is important in the network context. In particular, while a route such as AH serves passengers in the AH city-pair market, the route is used by passengers in the AB market as well (these passengers also fly on route BH). A related point is that, in the network context, each city-pair market will have its own θ^* and $\underline{\theta}$ critical values. These values are therefore *market-specific*, not route-specific, a distinction that did not arise in the simple model.

An additional observation concerns the routing for passengers in city-pair markets AH and BH. While these passengers could also make connecting trips through hub K on airline 2 (traveling from A or B to K and then to H), they are assumed to favor nonstop travel on carrier 1. Similarly, AK and BK passengers could travel through hub H, but they shun such connecting trips. These routing assumptions can be justified by assuming that passengers will never take connecting flights when a nonstop flight is available, and relaxing them would add considerable complexity to the analysis.

A final observation concerns the rationale for the network’s hub-and-spoke structure. As

is well recognized, such a network structure emerges in practice because of “economies of traffic density,” under which cost per passenger on a route declines as traffic density rises (reflecting increasing returns at the route level).⁶ Unfortunately, explicit incorporation of this density effect would make analysis of the network case exceedingly cumbersome. As a result, the simple assumption that cost per passenger in the absence of congestion is a constant, equal to c/s , is used instead. Under this assumption, however, a formal rationale for the hub-and-spoke structure is missing. Indeed, a preferred arrangement would be to serve AB passengers with nonstop service between the uncongested airports A and B, avoiding the congested hubs. To finesse this issue, allowing a manageable analysis of congestion within a hub-and-spoke structure, the analysis takes a shortcut. It arbitrarily imposes the assumption that AB traffic must pass through the hub, even though the microfoundations for this pattern are not explicitly present. The lessons of the analysis would clearly generalize, however, to a model with economies of density, but the cost in terms of analytical complexity would be high.

To begin the analysis of the network case, the first maintained assumption is that demand for travel in each city-pair market is the same. Thus, all of the markets AH, BH, AK, BK, AB, and HK share the common travel benefit functions b_p and b_o , and each market has a uniform distribution of θ and a unitary mass of passengers.

However, because of the network structure, the time cost of congestion is computed differently for some passengers than in the simple model of section 2. Note first that, because passengers traveling in markets AH and BH use only one congested airport, their time costs are analogous to those in the simple model, being given by $t(m_p^H)$, where m_p^H gives total peak traffic at hub H. Similarly, time costs for passengers in markets AK and BK are given by $t(m_p^K)$, where m_p^K is total peak traffic at hub K. However, since passengers traveling in market AB must connect at one of the hubs, they endure twice as much congestion as an AH passenger or other similar passenger. In other words, the AB passenger experiences congestion on landing at the hub and endures further congestion when his connecting flight departs. Therefore, the time costs for an AB passenger connecting at hub H are given by $2t(m_p^H)$, while costs for a trip through hub K are $2t(m_p^K)$. Passengers traveling in market HK, who take a nonstop flight, also experience congestion twice, but these experiences occur at different airports, once at the

origin and again at the destination. Therefore, time costs for an HK passenger are equal to $t(m_p^H) + t(m_p^K)$.

It should be noted that this last time cost expression involves the implicit assumption that a flight departing H in the peak period arrives at K during that airport's peak period. This outcome is guaranteed, however, only if the flight is instantaneous. At the cost of somewhat greater complexity, the setup could be made more realistic by explicitly viewing the model as portraying round-trip travel. Then, for a round trip originating during H's peak period, time cost on the outbound leg would be $t(m_p^H)$, with a zero cost incurred at K since the flight arrives after that airport's peak has passed. A peak-period return departure from K would generate a time cost of $t(m_p^K)$ and a zero cost at H, implying a total time cost for the round trip of $t(m_p^H) + t(m_p^K)$. While this expression is the same as the one above, the round trip orientation would require doubling the time costs for the other types of trips considered above. Because the implications of this alternate model differ only slightly from those developed below, the simpler approach based on instantaneous flight times is retained.

The above reasoning can also be applied to derive the congestion costs incurred by the airlines. Since flights on the AH and BH routes use only one congested airport (hub H), the peak operating cost per flight is given by $c + g(m_p^H)$. Similarly, peak operating cost on the routes AK and BK is given by $c + g(m_p^K)$. Because HK flights use both congested hub airports, peak operating cost is given by $c + g(m_p^H) + g(m_p^K)$.

Additional notation is required for the analysis, as follows. Let θ_{AH}^* denote the critical θ value for market AH. By symmetry, this value can be used to represent critical values for the other analogous markets, BH, AK, and BK. Similarly, let θ_{AB}^* and θ_{HK}^* represent the critical θ values for markets AB and HK, and let the $\underline{\theta}$ values for the various markets be labeled analogously. Let n_p be the total number of peak flights operated on the routes between A and B and hub H (i.e., the number of AH flights plus the number of BH flights), and let n_o be the number of off-peak flights on these routes. By symmetry, these same variables can be used to denote total peak and off-peak flights on the routes AK and BK. Finally, let k_p and k_o denote the number of peak and off-peak flights on the HK route.

Using all of the above information, and exploiting symmetry, an expression for social

welfare (travel benefits minus airline costs) is easily derived. Welfare equals

$$\begin{aligned}
W &= 4 \left(\int_{\underline{\theta}_{AH}}^{\theta_{AH}^*} b_o(\theta) d\theta + \int_{\theta_{AH}^*}^1 [b_p(\theta) - t(m_p)] d\theta \right) \\
&+ \int_{\underline{\theta}_{AB}}^{\theta_{AB}^*} b_o(\theta) d\theta + \int_{\theta_{AB}^*}^1 [b_p(\theta) - 2t(m_p)] d\theta \\
&+ \int_{\underline{\theta}_{HK}}^{\theta_{HK}^*} b_o(\theta) d\theta + \int_{\theta_{HK}^*}^1 [b_p(\theta) - 2t(m_p)] d\theta \\
&- 2(n_o c + n_p [c + g(m_p)]) - (k_o c + k_p [c + 2g(m_p)])
\end{aligned} \tag{7}$$

To see the use of symmetry in (7), note first that travel benefits in markets AH, BH, AK and BK are just four times the AH value. In addition, observe that $m_p^H = m_p^K$ will hold at the optimum, with the common value denoted m_p . This fact implies that HK time costs $t(m_p^H) + t(m_p^K)$ can be written $2t(m_p)$, with HK operating cost $c + g(m_p^H) + g(m_p^K)$ written $c + 2g(m_p)$. Finally, note that the second to last expression in (7) gives operating costs on routes AH, BH, AK, and BK, which are just double the costs through one hub.

While the constraints relating n_p to θ^* and $\underline{\theta}$ in the model of section 2 were simple, the appropriate constraints in the present context are more complex. The relevant constraints are as follows:

$$m_p = n_p + k_p \tag{8}$$

$$n_p = [1 - \theta_{AB}^* + 2(1 - \theta_{AH}^*)]/s \tag{9}$$

$$k_p = (1 - \theta_{HK}^*)/s \tag{10}$$

$$n_o = [\theta_{AB}^* - \underline{\theta}_{AB} + 2(\theta_{AH}^* - \underline{\theta}_{AH})]/s \tag{11}$$

$$k_o = (\theta_{HK}^* - \underline{\theta}_{HK})/s \tag{12}$$

The constraint in (8) says that the number of flights at a given hub airport is equal to total flights on the AH and BH (or AK and BK) routes, n_p , plus total flights on the HK route, k_p .

To understand (9), observe first that AB connecting traffic is split between the hubs, a division that is efficient because it limits congestion. As a result, $(1 - \theta_{AB}^*)/2$ peak passengers use each of the two hub routings in this market. But since each connecting passenger uses two flights, the number of peak passengers to be handled on each hub routing is double this amount, implying that the number of peak flights required to accommodate them is $(1 - \theta_{AB}^*)/s$. In addition, $(1 - \theta_{AH}^*)$ passengers must be served in each of the city-pair markets AH and BH, which means that $2(1 - \theta_{AH}^*)/s$ additional flights must be operated on the AH and BH routes (an analogous observation applies to routes AK and BK). Of course, nonstop and connecting passengers will be commingled on the n_p flights that are operated through each hub. An analogous discussion applies to routes AK and BK. Total peak traffic in market HK is $(1 - \theta_{HK}^*)$, so that the required number of HK flights is given by (10). Recalling that off-peak traffic is given by the market's θ^* value minus its $\underline{\theta}$ value, (11) and (12) give off-peak flights in analogous fashion to (9) and (10).

The social optimum is found by choosing θ_{AH}^* , θ_{AB}^* , θ_{HK}^* , $\underline{\theta}_{AH}$, $\underline{\theta}_{AB}$, $\underline{\theta}_{HK}$, n_p , n_o , k_p , k_o , and m_p to maximize (7) subject to (8)–(12). It can be shown that, after some manipulation, the first-order conditions for the first three variables reduce to

$$[b_p(\theta_{AH}^*) - t(m_p) - b_o(\theta_{AH}^*)] - g(m_p)/s - m_p[t'(m_p) + g'(m_p)/s] = 0 \quad (13)$$

$$[b_p(\theta_{AB}^*) - 2t(m_p) - b_o(\theta_{AB}^*)] - 2g(m_p)/s - 2m_p[t'(m_p) + g'(m_p)/s] = 0 \quad (14)$$

$$[b_p(\theta_{HK}^*) - 2t(m_p) - b_o(\theta_{HK}^*)] - 2g(m_p)/s - 2m_p[t'(m_p) + g'(m_p)/s] = 0. \quad (15)$$

The interpretation of (13)–(15) follows that of the optimality condition (3) from the simple model. The first term in each equation gives the change in travel benefits when a passenger in the relevant city pair market is switched from the off-peak to the peak period. Note that, because they encounter congestion twice, the travel-benefit expressions for passengers in the AB and HK markets involve subtraction of $2t(m_p)$ rather than $t(m_p)$. Accommodation of an extra peak passenger in type-AH and AB markets requires $1/s$ and $2/s$ additional flights, respectively, generating extra operating costs of $g(m_p)/s$ and $2g(m_p)/s$, which appear in (13)

and (14). Since each of the $1/s$ additional flights required by an extra HK passenger encounters congestion twice, operating costs rise by $2g(m_p)/s$ (see (15)). Lastly, the congestion caused by an extra type-AH passenger equals $m_p[t'(m_p) + g'(m_p)/s]$ in (13), an expression analogous to the congestion term in (3), but where m_p replaces n_p . However, each extra AB or HK passenger generates twice as much congestion, so that the corresponding terms in (14)–(15) are multiplied by two.

Because the conditions in (14) and (15) have exactly the same form, it follows that $\theta_{AB}^* = \theta_{HK}^*$. Thus, the peak traffic levels in markets AB and HK are the same at the optimum, a consequence of the parallel features of these markets. However, because the type-AH benefit differential $b_p(\theta_{AH}^*) - b_o(\theta_{AH}^*)$ in (13) is equated to an expression half as large as the corresponding expressions for the other markets, it follows that this differential is smaller than those in (14) and (15). Recalling that $b'_p - b'_o > 0$, it then follows that $\theta_{AH}^* < \theta_{AB}^* = \theta_{HK}^*$. Thus, because a type-AH peak passenger generates lower operating and congestion costs and incurs a lower time cost than passengers in the other two markets, the optimal volume of such passengers $(1 - \theta_{AH}^*)$ is greater.

Assuming interior solutions, it can be shown that the first-order conditions for the $\underline{\theta}$ variables reduce to $b_o(\underline{\theta}_{AH}) = c/s$, $b_o(\underline{\theta}_{AB}) = 2c/s$, and $b_o(\underline{\theta}_{HK}) = c/s$. Again, these conditions say that, for the lowest- θ passenger to fly in each market, travel benefit equals the cost of providing the required seat(s). Note that because two seats are needed for an AB passenger, it follows that $\underline{\theta}_{AB}$ is larger than the $\underline{\theta}$ values for the other markets. This fact in turn implies that the total number of AB passengers (peak plus off-peak) is smaller than in the type-AH and HK markets. Because $\underline{\theta}_{AH} = \underline{\theta}_{HK}$, these latter markets have identical passenger totals but a different peak/off-peak split (see above).

4. Network Analysis: Equilibrium

To begin the characterization of equilibrium, consider first the type-AH markets, where the airlines operate as monopolists. In these markets, the fares and critical θ values need not be the same across airlines, although symmetry will hold in equilibrium. As before, the off-peak fares in such markets must exhaust the travel benefits of the lowest- θ passengers to fly. As a result,

$f_{AH,o}^i = b_o(\underline{\theta}_{AH}^i)$ must hold, where $f_{AH,o}^i$ denotes off-peak fare in type-AH markets charged by airline i , $i = 1, 2$, and $\underline{\theta}_{AH}^i$ denotes the corresponding $\underline{\theta}$ value. As before, the θ^* values in the type-AH markets, denoted θ_{AH}^{*1} and θ_{AH}^{*2} for the two airlines, must satisfy indifference conditions. For carrier 1, θ_{AH}^{*1} must satisfy $b_p(\theta_{AH}^{*1}) - t(m_p^H) - f_{AH,p}^1 = b_o(\theta_{AH}^{*1}) - f_{AH,o}^1$, where $f_{AH,p}^1$ gives the carrier's peak type-AH fare. Substituting the above expression for $f_{AH,o}^1$ and rearranging, this condition allows $f_{AH,p}^1$ to be written as

$$f_{AH,p}^1 = b_p(\theta_{AH}^{*1}) - t(m_p^H) - b_o(\theta_{AH}^{*1}) + b_o(\underline{\theta}_{AH}^1). \quad (16)$$

An analogous manipulation for carrier 2 yields

$$f_{AH,p}^2 = b_p(\theta_{AH}^{*2}) - t(m_p^K) - b_o(\theta_{AH}^{*2}) + b_o(\underline{\theta}_{AH}^2), \quad (17)$$

where $f_{AH,p}^2$ denotes its peak fare in type-AH markets.

In market HK, where the airlines compete, they must charge the same fares. The off-peak fare satisfies $f_{HK,o} = b_o(\underline{\theta}_{HK})$. As above, substitution of this equality into the indifference condition governing θ_{HK}^* allows the peak fare HK to be written

$$f_{HK,p} = b_p(\theta_{HK}^*) - t(m_p^H) - t(m_p^K) - b_o(\theta_{HK}^*) + b_o(\underline{\theta}_{HK}). \quad (18)$$

Note that the time cost of an HK trip in (18) is given by the earlier asymmetric expression $t(m_p^H) + t(m_p^K)$.

Since off-peak connecting trips in market AB are identical regardless of whether the passenger flies on airline 1 or 2, the airlines must charge identical off-peak fares for such trips. The common value, denoted $f_{AB,o}$, satisfies $f_{AB,o} = b_o(\underline{\theta}_{AB})$. However, since the two hub airports could experience different degrees of congestion, peak AB trips may not be equivalent, and as a result, AB fares may differ across airlines. Using the $f_{AB,o}$ solution, the indifference condition for airline 1's passengers allows its peak AB fare to be written

$$f_{AB,p}^1 = b_p(\theta_{AB}^*) - 2t(m_p^H) - b_o(\theta_{AB}^*) + b_o(\underline{\theta}_{AB}), \quad (19)$$

while airline 2's peak fare can be written

$$f_{AB,p}^2 = b_p(\theta_{AB}^*) - 2t(m_p^K) - b_o(\theta_{AB}^*) + b_o(\underline{\theta}_{AB}). \quad (20)$$

Note from (19) and (20) that any AB fare difference between the carriers reflects a difference in the time cost of the trip, a consequence of a potential difference between m_p^H and m_p^K .⁷

The next step is to derive the constraints relating the numbers of flights to the θ^* and $\underline{\theta}$ variables. These constraints have a structure analogous to that of (8)–(12), adjusted so that symmetry is not imposed at the outset, but they require additional notation. Let n_p^1 and n_o^1 denote total peak and off-peak flights operated by airline 1 on the routes AH and BH, with n_p^2 and n_o^2 denoting airline 2's total peak and off-peak flights on routes AK and BK. In equilibrium, these variables will be symmetric across airlines, with the common values denoted n_p and n_o (consistent with the notation used in the analysis of the social optimum). In addition, let k_p^1 and k_o^1 denote airline 1's peak and off-peak flights on route HK, with k_p^2 and k_o^2 denoting airline 2's flights. Note that $k_p^1 + k_p^2$ equals the total flight variable k_p from above, while $k_o^1 + k_o^2 = k_o$.

Using this notation, the additional constraints are as follows:

$$m_p^H = n_p^1 + k_p^1 + k_p^2 \quad (21)$$

$$m_p^K = n_p^2 + k_p^1 + k_p^2 \quad (22)$$

$$n_p^1 + n_p^2 = [2(1 - \theta_{AB}^*) + 2(1 - \theta_{AH}^{*1}) + 2(1 - \theta_{AH}^{*2})]/s \quad (23)$$

$$k_p^1 + k_p^2 = (1 - \theta_{HK}^*)/s \quad (24)$$

$$n_o^1 + n_o^2 = [2(\theta_{AB}^* - \underline{\theta}_{AB}) + 2(\theta_{AH}^{*1} - \underline{\theta}_{AH}^1) + 2(\theta_{AH}^{*2} - \underline{\theta}_{AH}^2)]/s \quad (25)$$

$$k_o^1 + k_o^2 = (\theta_{HK}^* - \underline{\theta}_{HK})/s \quad (26)$$

Eqs. (21), (22), (24) and (26) are analogous to (8), (10) and (12), with total HK flights written as the sum of the individual carrier flights on the route. To understand (23), observe that peak AB passengers generate a need for $2(1 - \theta_{AB}^*)$ total seats on the routes AH, BH, AK and BK

(two seats for each passenger). In addition, AH and BH passengers create a need for $2(1 - \theta_{AH}^{*1})$ seats on these two routes, with AK and BK passengers generating a need for $2(1 - \theta_{AH}^{*2})$ seats on these two routes. Eq. (23) sums these quantities and divides by s to generate the total peak flights needed on the four routes connecting A and B to the two hubs. A parallel interpretation applies to (25).

Profit for airline 1 is written

$$\begin{aligned}
& 2f_{AH,p}^1(1 - \theta_{AH}^{*1}) + 2f_{AH,o}^1(\theta_{AH}^{*1} - \underline{\theta}_{AH}^1) + f_{HK,p}sk_p^1 + f_{HK,o}sk_o^1 \\
& + f_{AB,p}^1[sn_p^1 - 2(1 - \theta_{AH}^{*1})]/2 + f_{AB,o}^1[sn_o^1 - 2(\theta_{AH}^{*1} - \underline{\theta}_{AH}^1)]/2. \quad (27)
\end{aligned}$$

While the first line of (27) is self-explanatory (recall that the airline serves two type-AH markets), the terms multiplying the fares in the second line, which give peak and off-peak traffic levels in market AB, require explanation. To interpret the first of these terms, note that total peak seats offered by airline 1 on routes AH and BH equals sn_p^1 . Of these seats, $2(1 - \theta_{AH}^{*1})$ are taken by passengers in markets AH and BH, leaving the rest for AB connecting passengers. Since each such passenger uses two seats, division by two yields their number. The second multiplicative term has a similar interpretation.

Airline 1's goal is to maximize profit in (27), taking as given the choices of airline 2. Its decision variables include the flight totals n_o^1 , n_p^1 , k_o^1 , k_p^1 . In addition, since the airline is a monopolist in markets AH and BH, it can choose traffic levels in these markets directly, which adds θ_{AH}^{*1} and $\underline{\theta}_{AH}^1$ to its list of choice variables.

To generate first-order conditions for this problem in the presence of the many constraints considered above, the following procedure is used. First, eqs. (23)–(26) are solved to give the four market-level quantities θ_{AB}^* , $\underline{\theta}_{AB}$, θ_{HK}^* and $\underline{\theta}_{HK}$ as functions of the decision variables (i.e., of θ_{AH}^{*1} and $\underline{\theta}_{AH}^1$ and the flight variables). Then, these solutions are substituted into the peak fare expressions in (19)–(21) and into the previous $f_{AB,o}$ and $f_{HK,o}$ solutions, which eliminates the market-level variables θ_{AB}^* , $\underline{\theta}_{AB}$, θ_{HK}^* and $\underline{\theta}_{HK}$ from the airline's objective function. Airline 1 then chooses n_o^1 , n_p^1 , k_o^1 , k_p^1 , θ_{AH}^{*1} and $\underline{\theta}_{AH}^1$ optimally, treating the analogous list of variables for airline 2 as parametric.

Making use of the first-order conditions for n_o^1 , k_o^1 , and θ_{AH}^1 , and imposing symmetry, the first-order conditions for θ_{AH}^{*1} , n_p^1 , and k_p^1 reduce to

$$\begin{aligned} [b_p(\theta_{AH}^*) - t(m_p) - b_o(\theta_{AH}^*)] - g(m_p)/s - (m_p - k_p/2)[t'(m_p) + g'(m_p)/s] \\ - (1 - \theta_{AH}^*)[b'_p(\theta_{AH}^*) - b'_o(\theta_{AH}^*)] = 0 \end{aligned} \quad (28)$$

$$\begin{aligned} [b_p(\theta_{AB}^*) - 2t(m_p) - b_o(\theta_{AB}^*)] - 2g(m_p)/s - 2(m_p - k_p/2)[t'(m_p) + g'(m_p)/s] \\ - [(1 - \theta_{AB}^*)/2][b'_p(\theta_{AB}^*) - b'_o(\theta_{AB}^*)] = 0 \end{aligned} \quad (29)$$

$$\begin{aligned} [b_p(\theta_{HK}^*) - 2t(m_p) - b_o(\theta_{HK}^*)] - 2g(m_p)/s - m_p[t'(m_p) + g'(m_p)/s] \\ - [(1 - \theta_{HK}^*)/2][b'_p(\theta_{HK}^*) - b'_o(\theta_{HK}^*)] = 0. \end{aligned} \quad (30)$$

To interpret these conditions, observe that the first and second terms in each equation, which give the change in travel benefits as well as the cost of operating additional flights when an extra passenger is shifted into the peak period, are the same as the corresponding terms in the optimality conditions (13)–(15). As in the equilibrium analysis of section 2, the last term in each of the conditions represents a residual market-power effect, caused by the need to persuade passengers with lower relative valuations of peak travel to enter the peak period.⁸

The key comparison, however, is between the congestion terms in (28)–(30) and those in the optimality conditions (13)–(15), a comparison that tells the extent to which congestion is internalized. First, compare (13) and (28) to gauge internalization of the congestion caused by type-AH passengers. The congestion term in (13) includes the multiplicative factor m_p , indicating that a type-AH passenger imposes congestion on the m_p total flights at the hub he uses. However, the corresponding term in (28) is $m_p - k_p/2$, which indicates that some of the type-AH passenger's congestion is not internalized. The uninternalized portion represents the congestion borne by $k_p/2$ flights, and these flights are those operated by the *other airline* on the route HK, which represent half the k_p total. The given airline has no incentive to internalize the congestion imposed on these flights since they are operated by the other carrier.

Similarly, a comparison of (14) and (29) shows that part of the congestion caused by AB

passengers is not internalized. Because such a passenger connects at the hub, flying both in and out of the airport, he effectively congests the hub's m_p flights twice, accounting for the $2m_p$ factor in (14). The smaller factor $2(m_p - k_p/2)$ in (29) shows that the airline ignores the impact of this double congestion on the $k_p/2$ flights operated on route HK by the other airline, as in the case of AH passengers.

It is interesting to note that, even though competition occurs in the AB market, the congestion caused by AB passengers is internalized to the same extent as that caused by type-AH passengers, over the whom the airline has monopoly power. In other words, in both cases, the airline fails to internalize the congestion imposed on the HK flights operated by the other carrier. This fact suggests a conclusion that is not fully apparent in the simple model. In particular, even though the market power of carriers (which leads them to recognize the congestion-sensitivity of fares) is the key to internalization, the degree of market power appears not to be crucial. Instead, regardless of how much market power passengers experience, the extent to which the congestion they cause is internalized depends only on the airline's flight share at the congested airport.

Finally, a comparison of (15) and (30) shows the same lack of internalization of part of the congestion created by HK passengers. Because an HK passenger flies between two hubs, he imposes congestion on $2m_p$ flights, accounting for the appearance of this factor in (15). However, the airline ignores the congestion imposed by this passenger on the other carrier's type-AH flights, which is created by serving that airline's hub. This congestion affects n_p such flights. In addition, the airline ignores the fact that an extra HK passenger twice congests the other carrier's $k_p/2$ flights on this route, once at each endpoint. Summing these effects, the extra HK passenger effectively congests $n_p + k_p = m_p$ of the other carrier's flights. This congestion is not internalized, accounting for the m_p factor in (30), which is too small by half.

Observe that, because the inequality $m_p < 2(m_p - k_p/2)$ holds after simple rearrangement, a smaller portion of congestion is internalized for HK than for AB passengers (compare (29) and (30)). With the perceived cost of serving AB passengers thus higher, it follows that $\theta_{AB}^* > \theta_{HK}^*$, implying a lower peak traffic volume in market AB than in HK.⁹

As in the case of the simple model of section 2, peak-period tolls can be levied to ensure

full internalization of congestion. Given the above discussion, internalization requires raising cost per peak passenger by the following amounts: by $(k_p/2)[t' + g'/s]$ for AH passengers, by $k_p[t' + g'/s]$ for AB passengers, and by $m_p[t' + g'/s]$ for HK passengers. Note, however, that since the AB passenger takes two flights, the correct adjustment is given by raising his cost by $(k_p/2)[t' + g'/s]$ per flight.

These adjustments can be implemented by the following toll system. First, each airline should be charged a congestion toll of $(k_p/2)[st' + g']$ for each peak flight operated at its own hub airport. Dividing by s , this toll raises cost per passenger by the appropriate amount for type-AH and AB passengers (see above). Then, each airline should be charged a toll of $(n_p + k_p/2)[st' + g']$ for each peak flight it operates at the *other carrier's* hub. Combined with the own-hub toll, the total toll charge incurred by an HK flight then equals $(n_p + k_p)[st' + g'] = m_p[st' + g']$, which yields the correct adjustment in cost per HK passenger.

This toll system can be represented by a direct generalization of the rule already derived for the simple model of section 2. In particular, the toll paid by an airline at a given airport is equal to *the congestion damage caused by an extra flight times one minus the carrier's airport flight share*. To see that the above system embodies this rule, recall that the damage from an extra flight at either hub equals $m_p[st'(m_p) + g'(m_p)]$. Then consider the following relationships:

$$\begin{aligned}
\text{Own-hub toll} &= (1 - \text{flight share}) * (\text{damage from an extra flight}) \\
&= [1 - (n_p + k_p/2)/m_p] * m_p[st'(m_p) + g'(m_p)] \\
&= (k_p/2)[st'(m_p) + g'(m_p)]
\end{aligned} \tag{31}$$

Similarly,

$$\begin{aligned}
\text{Other-hub toll} &= (1 - \text{flight share}) * (\text{damage from an extra flight}) \\
&= [1 - (k_p/2)/m_p] * m_p[st'(m_p) + g'(m_p)] \\
&= (n_p + k_p/2)[st'(m_p) + g'(m_p)]
\end{aligned} \tag{32}$$

Of course, since congestion damage is zero at an uncongested airport, no tolls should be levied at airports A and B.

As in the simple model, this toll system can be viewed as providing a peak toll *schedule*, which relates the proper toll to the airport traffic level m_p and to the (endogenous) flight shares of the carriers. Facing such a schedule, carriers would adjust traffic so as to internalize all the congestion they create. The same equilibrium could be achieved, however, by evaluating the above toll expressions at the optimum, and charging a flat peak toll that would be unresponsive to airport activity. However, computing the flat toll requires knowing the optimum, information which is not needed to generate the toll schedule. Computing that schedule only requires knowledge of the congestion damage from an extra flight, which could be computed using airline operations and cost data along with estimates of the value of passenger time.

Several additional observations regarding this toll system are useful. First, as in the simple model, imposition of such a system may not be welfare improving because it corrects only one distortion, leaving the residual market-power effect in place. Use of tolls is guaranteed to be desirable only if the latter effect is small. Second, the toll system may be controversial. The reason is that it charges a low toll to the hub airport's dominant carrier while charging a high toll to the carrier with the low flight share. While this pattern reflects respectively high and low internalization of congestion at the airport by the two types of carriers, it might appear to be the wrong response to concerns about market power by the dominant hub airline (see Borenstein (1989)). However, since such market power is fully represented in the model, the conclusions of the analysis take it into account. As a result, concerns about the anticompetitive effect of the toll system would appear to be misplaced.

4. Conclusion

The likely resurgence of air traffic in the U.S. means that airport congestion is a problem that must soon be confronted by policy makers. As part of their policy response, it is probable that some form of congestion pricing will be imposed at selected U.S. airports in the relatively near future. The theory developed in this paper, which extends the results of Brueckner (2002), provides an important guide for the formulation of congestion pricing rules. In particular, the

theory says that the congestion tolls levied on the various airlines at a particular airport should generally be different, with the tolls being inversely related to a carrier's airport flight share. Internalization of airport congestion is the reason for this inverse relationship. In operating another peak flight, a carrier takes account of the congestion damage imposed on the other flights it operates. If these flights account for a large share of the airport's traffic, then most of the congestion created by the additional flight is internalized, justifying a low toll. By contrast, if the carrier operates only a few of the airport's flights, then little internalization occurs, and a high toll is needed to force the carrier to take into account the congestion damage it causes. The resulting flight-share rule is easy to implement, and it could help policymakers design proper toll systems at U.S. airports.

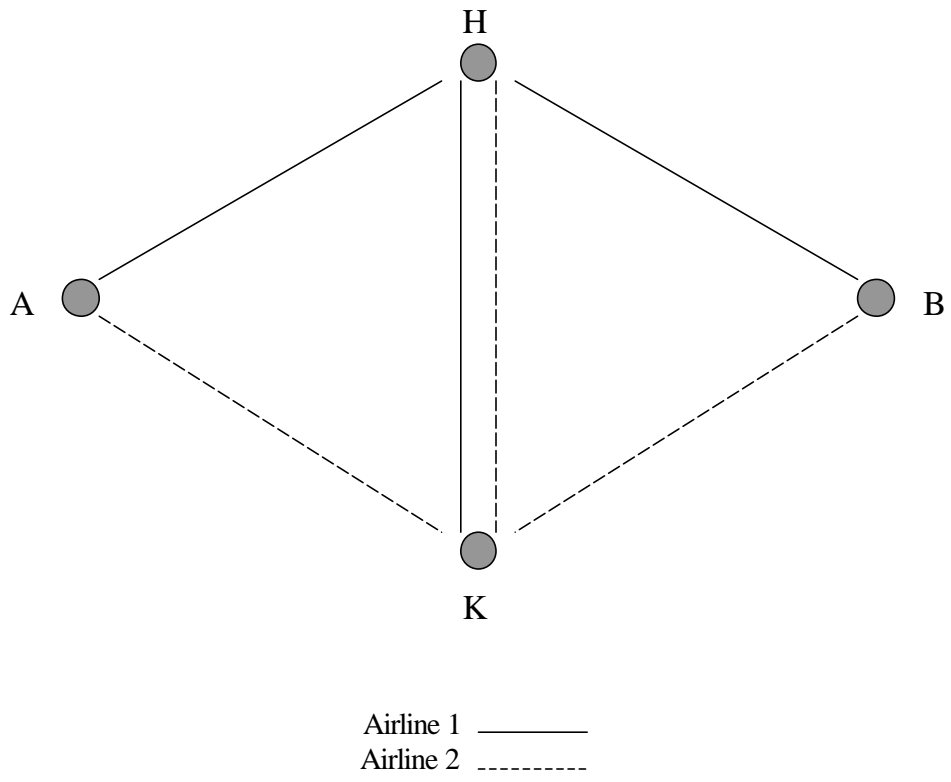


Figure 1: Network Structure

References

- CARLIN, A., PARK, R.E., 1970. Marginal cost pricing of airport runway capacity, *American Economic Review* 60, 310-319.
- BORENSTEIN, S., 1989. Hubs and high fares: Market power and dominance in the U.S. airline industry, *Rand Journal of Economics* 20, 344-365.
- BRUECKNER, J.K., SPILLER, P.T., 1994. Economies of traffic density in the deregulated airline industry, *Journal of Law and Economics* 37, 379-415.
- BRUECKNER, J.K., 2001. The economics of international codesharing: An analysis of airline alliances, *International Journal of Industrial Organization* 19, 1475-1498.
- BRUECKNER, J.K., 2002. Airport congestion when carriers have market power, *American Economic Review*, forthcoming.
- 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., PAHWA, M., 2000. Comparison of three empirical models of airport congestion pricing, *Journal of Urban Economics* 47, 1-38.
- DANIEL, J.I., 2002. Distributional consequences of airport congestion pricing, *Journal of Urban Economics*, forthcoming.
- LEVINE, M.E., 1969. Landing fees and the airport congestion problem, *Journal of Law and Economics* 12, 79-108.
- MAYER, C., SINAI, T., 2002. Network effects, congestion externalities, and air traffic delays: Or why all delays are not created evil, NBER Working paper #8701.
- MORRISON, S.A., 1983. Estimation of long-run prices and investment levels for airport runways, *Research in Transport Economics* 1, 103-130.
- MORRISON, S.A., WINSTON, C., 1989. Enhancing the performance of the deregulated air transportation system, *Brookings Papers on Economic Affairs, Microeconomics*, 61-123.
- OUM, T.H., ZHANG, Y., 1990. Airport pricing, *Journal of Public Economics* 43, 353-374.
- SMALL, K.A., 1992. *Urban Transportation Economics*, Harwood Academic Publishers, Chur, Switzerland.

TRANSPORTATION RESEARCH BOARD, 1999. *Entry and Competition in the U.S. Airline Industry: Issues and Opportunities*, National Academy Press, Washington, D.C.

Footnotes

*This paper was written while I was a visitor at the Center for Economic Studies at the University of Munich. I thank CES for its support and hospitality. In addition, I thank Kangoh Lee for a number of very helpful comments.

¹Using a different approach that focuses on a single airport, Daniel (1995) finds contradictory evidence. In particular, the intraday traffic patterns generated by Daniel's simulation model fit his data for the Minneapolis-St. Paul airport best when the hub airline (Northwest) is assumed to behave atomistically, ignoring the congestion it imposes on its own flights. However, Brueckner's (2002) findings are supported by the results of Mayer and Sinai (2002), who use a much larger sample to show that airport delays decline with the level of airport concentration.

²This pattern is revealed by the graphs in Daniel (1995) and Mayer and Sinai (2002).

³Note that this single-crossing inequality actually says nothing about the levels of the benefit functions. However, to avoid a degenerate equilibrium, the levels of the functions must be such that they intersect at an intermediate value of θ . Thus, $b_p(\theta) > (<) b_o(\theta)$ must hold for high (low) θ , indicating that peak benefits are higher (lower) than off-peak benefits for business (leisure) passengers. See Brueckner (2002) for details.

⁴Under the maintained assumptions, it is easily seen that the second-order condition for the optimization problem is satisfied.

⁵It is assumed that the second-order conditions for the the profit-maximization problem are satisfied. These conditions involve b_p'' and b_o'' , which have no natural sign.

⁶See Brueckner and Spiller (1994) for discussion and evidence.

⁷Models of hub-and-spoke networks typically impose a fare-arbitrage condition for connecting passengers. In the present context, this condition says that an AB passenger cannot travel more cheaply by purchasing separate tickets for the two legs of the AB trip. For passengers traveling on airline 1, the arbitrage condition for peak fares is $f_{AB,p}^1 < 2f_{AH,p}^1$, with the off-peak condition given by $f_{AB,o} < 2f_{AH,o}^1$. Although such arbitrage conditions are commonly satisfied automatically in other hub-and-spoke models (see Brueckner (2001), for example), satisfaction cannot be verified in the present setting and must be assumed.

⁸Observe that, as in section 2, these terms include the number of peak passengers carried by the airline in the given market. This number is half the total in markets AB and HK $((1 - \theta_{AB}^*)/2$ and $(1 - \theta_{HK}^*)/2$), but the entire traffic in the type-AH markets $(1 - \theta_{AH}^*)$.

⁹To see this conclusion, let the LHS of (29) be evaluated at θ_{HK}^* . With $2(m_p - k_p/2) > m_p$, it follows that the resulting expression is less than the expression in (30), which in turn equals zero. The LHS of (29) is thus negative at θ_{HK}^* , indicating that an extra peak AB passenger lowers profit. As a result, θ_{HK}^* is too large to be optimal for market AB, implying $\theta_{AB}^* < \theta_{HK}^*$. Note that because (28) has a different form than the other equations given the absence of the 1/2 factor in the residual-market-power term, a similar comparison of θ_{AH}^* to the other values is not possible.

CESifo Working Paper Series

- 689 Amihai Glazer and Vesa Kannianen, The Effects of Employment Protection on the Choice of Risky Projects, March 2002
- 690 Michael Funke and Annekatrin Niebuhr, Threshold Effects and Regional Economic Growth – Evidence from West Germany, March 2002
- 691 George Economides, Apostolis Philippopoulos, and Simon Price, Elections, Fiscal Policy and Growth: Revisiting the Mechanism, March 2002
- 692 Amihai Glazer, Vesa Kannianen, and Mikko Mustonen, Innovation of Network Goods: A Non-Innovating Firm Will Gain, March 2002
- 693 Helmuth Cremer, Jean-Marie Lozachmeur, and Pierre Pestieau, Social Security, Retirement Age and Optimal Income Taxation, April 2002
- 694 Rafael Lalive and Josef Zweimüller, Benefit Entitlement and the Labor Market: Evidence from a Large-Scale Policy Change, April 2002
- 695 Hans Gersbach, Financial Intermediation and the Creation of Macroeconomic Risks, April 2002
- 696 James M. Malcomson, James W. Maw, and Barry McCormick, General Training by Firms, Apprentice Contracts, and Public Policy, April 2002
- 697 Simon Gächter and Arno Riedl, Moral Property Rights in Bargaining, April 2002
- 698 Kai A. Konrad, Investment in the Absence of Property Rights: The Role of Incumbency Advantages, April 2002
- 699 Campbell Leith and Jim Malley, Estimated General Equilibrium Models for the Evaluation of Monetary Policy in the US and Europe, April 2002
- 700 Yin-Wong Cheung and Jude Yuen, Effects of U.S. Inflation on Hong Kong and Singapore, April 2002
- 701 Henry Tulkens, On Cooperation in Musgravian Models of Externalities within a Federation, April 2002
- 702 Ralph Chami and Gregory D. Hess, For Better or For Worse? State-Level Marital Formation and Risk Sharing, April 2002
- 703 Fredrik Andersson and Kai A. Konrad, Human Capital Investment and Globalization in Extortionary States, April 2002
- 704 Antonis Adam and Thomas Moutos, The Political Economy of EU Enlargement: Or, Why Japan is not a Candidate Country?, April 2002

- 705 Daniel Gros and Carsten Hefeker, Common Monetary Policy with Asymmetric Shocks, April 2002
- 706 Dirk Kieseewetter and Rainer Niemann, Neutral and Equitable Taxation of Pensions as Capital Income, April 2002
- 707 Robert S. Chirinko, Corporate Taxation, Capital Formation, and the Substitution Elasticity between Labor and Capital, April 2002
- 708 Frode Meland and Gaute Torsvik, Structural Adjustment and Endogenous Worker Recall Probabilities, April 2002
- 709 Rainer Niemann and Caren Sureth, Taxation under Uncertainty – Problems of Dynamic Programming and Contingent Claims Analysis in Real Option Theory, April 2002
- 710 Thomas Moutos and William Scarth, Technical Change and Unemployment: Policy Responses and Distributional Considerations, April 2002
- 711 Günther Rehme, (Re-)Distribution of Personal Incomes, Education and Economic Performance Across Countries, April 2002
- 712 Thorvaldur Gylfason and Gylfi Zoega, Inequality and Economic Growth: Do Natural Resources Matter?, April 2002
- 713 Wolfgang Leininger, Contests over Public Goods: Evolutionary Stability and the Free-Rider Problem, April 2002
- 714 Ernst Fehr and Armin Falk, Psychological Foundations of Incentives, April 2002
- 715 Giorgio Brunello, Maria Laura Parisi, and Daniela Sonedda, Labor Taxes and Wages: Evidence from Italy, May 2002
- 716 Marta Aloi and Huw Dixon, Entry Dynamics, Capacity Utilisation and Productivity in a Dynamic Open Economy, May 2002
- 717 Paolo M. Panteghini, Asymmetric Taxation under Incremental and Sequential Investment, May 2002
- 718 Ben J. Heijdra, Christian Keuschnigg, and Wilhelm Kohler, Eastern Enlargement of the EU: Jobs, Investment and Welfare in Present Member Countries, May 2002
- 719 Tapio Palokangas, The Political Economy of Collective Bargaining, May 2002
- 720 Gilles Saint-Paul, Some Evolutionary Foundations for Price Level Rigidity, May 2002
- 721 Giorgio Brunello and Daniela Sonedda, Labor Tax Progressivity, Wage Determination, and the Relative Wage Effect, May 2002
- 722 Eric van Damme, The Dutch UMTS-Auction, May 2002

- 723 Paolo M. Panteghini, Endogenous Timing and the Taxation of Discrete Investment Choices, May 2002
- 724 Achim Wambach, Collusion in Beauty Contests, May 2002
- 725 Dominique Demougin and Claude Fluet, Preponderance of Evidence, May 2002
- 726 Gilles Saint-Paul, Growth Effects of Non Proprietary Innovation, May 2002
- 727 Subir Bose, Gerhard O. Orosel, and Lise Vesterlund, Optimal Pricing and Endogenous Herding, May 2002
- 728 Erik Leertouwer and Jakob de Haan, How to Use Indicators for ‘Corporatism’ in Empirical Applications, May 2002
- 729 Matthias Wrede, Small States, Large Unitary States and Federations, May 2002
- 730 Christian Schultz, Transparency and Tacit Collusion in a Differentiated Market, May 2002
- 731 Volker Grossmann, Income Inequality, Voting Over the Size of Public Consumption, and Growth, May 2002
- 732 Yu-Fu Chen and Michael Funke, Working Time and Employment under Uncertainty, May 2002
- 733 Kjell Erik Lommerud, Odd Rune Straume, and Lars Sørgard, Downstream Merger with Oligopolistic Input Suppliers, May 2002
- 734 Saku Aura, Does the Balance of Power Within a Family Matter? The Case of the Retirement Equity Act, May 2002
- 735 Sandro Brusco and Fausto Panunzi, Reallocation of Corporate Resources and Managerial Incentives in Internal Capital Markets, May 2002
- 736 Stefan Napel and Mika Widgrén, Strategic Power Revisited, May 2002
- 737 Martin W. Cripps, Godfrey Keller, and Sven Rady, Strategic Experimentation: The Case of Poisson Bandits, May 2002
- 738 Pierre André Chiappori and Bernard Salanié, Testing Contract Theory: A Survey of Some Recent Work, June 2002
- 739 Robert J. Gary-Bobo and Sophie Larribeau, A Structural Econometric Model of Price Discrimination in the Mortgage Lending Industry, June 2002
- 740 Laurent Linnemer, When Backward Integration by a Dominant Firm Improves Welfare, June 2002
- 741 Gebhard Kirchgässner and Friedrich Schneider, On the Political Economy of Environmental Policy, June 2002

- 742 Christian Keuschnigg and Soren Bo Nielsen, Start-ups, Venture Capitalists, and the Capital Gains Tax, June 2002
- 743 Robert Fenge, Silke Uebelmesser, and Martin Werding, Second-best Properties of Implicit Social Security Taxes: Theory and Evidence, June 2002
- 744 Wendell Fleming and Jerome Stein, Stochastic Optimal Control, International Finance and Debt, June 2002
- 745 Gene M. Grossman, The Distribution of Talent and the Pattern and Consequences of International Trade, June 2002
- 746 Oleksiy Ivaschenko, Growth and Inequality: Evidence from Transitional Economies, June 2002
- 747 Burkhard Heer, Should Unemployment Benefits be Related to Previous Earnings?, July 2002
- 748 Bas van Aarle, Giovanni Di Bartolomeo, Jacob Engwerda, and Joseph Plasmans, Staying Together or Breaking Apart: Policy-makers' Endogenous Coalitions Formation in the European Economic and Monetary Union, July 2002
- 749 Hans Gersbach, Democratic Mechanisms: Double Majority Rules and Flexible Agenda Costs, July 2002
- 750 Bruno S. Frey and Stephan Meier, Pro-Social Behavior, Reciprocity or Both?, July 2002
- 751 Jonas Agell and Helge Bennmærker, Wage Policy and Endogenous Wage Rigidity: A Representative View From the Inside, July 2002
- 752 Edward Castronova, On Virtual Economies, July 2002
- 753 Rebecca M. Blank, U.S. Welfare Reform: What's Relevant for Europe?, July 2002
- 754 Ruslan Lukach and Joseph Plasmans, Measuring Knowledge Spillovers Using Patent Citations: Evidence from the Belgian Firm's Data, July 2002
- 755 Aaron Tornell and Frank Westermann, Boom-Bust Cycles in Middle Income Countries: Facts and Explanation, July 2002
- 756 Jan K. Brueckner, Internalization of Airport Congestion: A Network Analysis, July 2002