Airport Congestion When Carriers Have Market Power

by

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

May 2001

Abstract

This paper analyzes airport congestion when carriers are nonatomistic, showing how the results of the road-pricing literature are modified when the economic agents causing congestion have market power. The analysis shows that when an airport is dominated by a monopolist, either discriminating or nondiscriminating, congestion is fully internalized, provided that a separability assumption on travel benefits is satisfied. The analysis thus yields no role for congestion pricing under monopoly conditions. Under a Cournot oligopoly, however, carriers are shown to internalize only the congestion they impose on themselves. A toll that captures the uninternalized portion of congestion can then improve the allocation of traffic. The toll is equal to the congestion cost from an extra flight times one minus a carrier's flight share. At an airport like Chicago-O'Hare, this rule would imply that United and American would be charged for only about half of the congestion created by an additional flight.

Airport Congestion When Carriers Have Market Power

by

Jan K. Brueckner*

1. Introduction

Air traffic delays in the U.S. have grown dramatically in the last few years, becoming a major public policy issue. In Europe, delays have plagued airline traffic for an even longer period. A delay occurs when an airline flight arrives more than 15 minutes late. Under this definition, U.S. flights experienced 374,116 delays in 1999, and delays grew to 450,289 in the year 2000, for a stunning 20.4 percent single-year increase.

Table 1 shows delay data for the 15 U.S. airports with the most delays in 1999.[1] While the delay totals represent delays attributable to local operations at the given airport, the on-time figures also capture the effect of delays elsewhere in the system (which affect arrivals, as well as subsequent departures, at the given airport). The poor on-time records of Newark and New York-La Guardia, which are widely recognized, are clear in the table, but the poor performance of other airports (Boston, Philadelphia, Washington-Dulles) is also evident.

Table 1 shows that weather is the major reason for delays, accounting for well over half of the total in most cases. Typically, the second largest source of delays is "volume" (i.e. traffic exceeding airport capacity), whose share is also shown in the table. However, because weather-related delays often arise from restricted airport operations during bad weather, both sources actually reflect the same problem: too many flights attempting to take off and land relative to airport capacity.[2]

Solutions to the delay problem are now widely discussed in policy circles and in the press. Increasing the size of congested airports by investing in new runways is one remedy, although the long gestation period of such projects means that the benefits lie far in the future. Improvement in air traffic control technology, which can increase the capacity of the nation's airspace while also allowing busy airports to handle more flights, is another remedy that is slowly being implemented. A third remedy for the delay problem is the imposition of congestion pricing at U.S. airports. With such a pricing mechanism, the landing fees paid by airlines would vary with the level of congestion at the airport. This approach stands in sharp contrast to current practice, where the landing fee incurred by a flight depends only on the weight of the aircraft, being unrelated to time of day or airport conditions. Under congestion pricing, operating costs at peak hours would rise substantially relative to off-peak costs, encouraging airlines to shift their flights away from the peak. With a more-even distribution of traffic across time, airport congestion would fall, reducing the number of delays. Although no U.S. airport has implemented congestion pricing, endorsements of such a system are now frequently heard. For example, the influential 1999 monograph by the Transportation Research Board included a call for congestion pricing among its host of airline policy recommendations.

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

Transportation researchers have long recognized that this principle applies to airports as well as urban highways. Levine (1969) and Carlin and Park (1970) offer the earliest discussions of airport congestion pricing, with later treatments given by Morrison (1983) and Morrison and Winston (1989). The most sophisticated analysis, however, is provided by Daniel (1995, 2000, 2001), who develops a detailed simulation model that shows the effects of congestion pricing in a realistic setting.

Given the intense current focus on the airport congestion problem and possible remedies, it is important to ensure that our understanding of the congestion problem is sound on a theoretical level. However, most of the previous literature can be criticized for simply extrapolating the lessons of road-pricing models to the case of airports without recognizing a critical difference between these contexts. In particular, although road users are appropriately viewed as atomistic, with each user small relative to total traffic, this view is incorrect in the case of airlines. As seen in Table 1, one or two airlines 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.

The purpose of the present paper is to show that, when the atomistic model is abandoned, the verdict on congestion is softened. While atomistic users of a congested facility ignore their external effects, the 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.* This internalization suggests that the over-allocation of flights to peak hours may not be as severe in practice as an atomistic model would predict. Indeed, in the case where an airport is served by a single monopolistic carrier, the analysis shows that the peak/off-peak allocation of flights may be efficient. These findings are important because they suggest a more-limited role for airport congestion pricing than that envisioned by many analysts.

Since Daniel (1995) recognizes the possible internalization of airport congestion, this basic insight has appeared before in the literature. However, an important contribution of the present paper is to develop the insight in the context of a transparent model (Daniel's simulation framework is highly complex). Moreover, the current model has an important feature that is missing from previous work. In particular, the internalization of congestion in the model occurs partly through *congestion's effect on airfares.*

To see this contribution more clearly, note that two types of costs from airport congestion are recognized in the model. These are additional passenger time costs, which capture the value of time lost to delays, and increased airline operating costs, reflecting the greater crew expenses and reduced aircraft utilization resulting from congestion. In Daniel's model, the airline is assumed to allocate its flights across peak and off-peak periods so as to *minimize the sum of these costs.* While this is a convenient behavioral assumption, its drawback is that passenger time costs are not actually a direct cost to the airline. Carriers may end up taking such costs into account, but the proper avenue for such an effect must be through airfares. In other words, the need to offer a fare discount for travel during a congested period forces the carrier to take passenger time costs into account. The present model follows this approach, portraying the airline as a profit maximizer charging congestion-sensitive fares. When the carrier has market power, it takes this fare sensitivity into account in allocating flights between the peak and off-peak periods. Relative to a cost-minimizing model, the result is a more-realistic framework for analyzing internalization of airport congestion.

The model assumes that the day consists of two travel periods, a narrow, congestible peak period and an broader off-peak period where congestion never arises. Passengers are represented by a continuum, with the benefits of peak and off-peak travel differing for each individual, but with both benefits increasing across the continuum. When the two benefit functions satisfy a single-crossing condition, the continuum divides at a critical point between the peak and off-peak periods. Fares are set taking travel benefits and the time cost of peak travel into account, and the airline then maximizes profit.

The analysis begins by characterizing the socially optimal traffic allocation between the peak and off-peak periods. To compare equilibrium outcomes to the social optimum, the discussion first considers the benchmark competitive case, where carriers are atomistic. Then, the analysis turns to the case of a perfectly discriminating monopolist, who can charge a different fare to each passenger. The more-realistic case where the monopolist cannot discriminate is then considered, with fares differing only between the peak and off-peak periods. Using the same assumption on fares, the analysis concludes with a discussion of oligopoly models, considering both the Cournot and Stackelberg cases. In each of the nonatomistic cases, congestion is partially or fully internalized, suggesting a limited role for congestion pricing.

2. Basic Analysis

2.1. The model

The analysis is based on the simplest possible model capable of illustrating the relevant issues. As noted above, the model distinguishes between two travel periods at a given airport, denoted peak and off-peak.[3] The peak period consists of a relatively short time window containing the day's most desirable travel times, such as early morning or late afternoon. To be realistic, the peak could consist of a collection of disjoint time intervals representing these desirable times. 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 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 peak period is always congested over the range of relevant allocations.

Peak congestion depends on the number of flights operating during the peak period, denoted n_p. Congestion raises an airline's operating costs, with cost per flight 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 be zero for n_p sufficiently small, but its positive range (where the function is increasing) is assumed to be relevant. All flights are assumed to use identical aircraft with fixed seat capacity s, and for simplicity, a 100 percent load factor is assumed, so that all seats are filled.

To characterize the effect of congestion on passenger time costs, the demand side of the model must be developed. First, it is assumed that 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 depend on θ, and they 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 $B_p(\theta, n_p)$. For most of the analysis, an important restriction is placed on this function. In particular, the benefit function is assumed to be additively separable in its arguments, with $B_p(\theta, n_p) \equiv b_p(\theta) - t(n_p)$. 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 key implication of separability is that these time costs do not depend on θ. 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.

As the analysis will demonstrate, the additive separability of B_p yields clearcut conclusions regarding internalization of congestion. Once these conclusions are derived, the discussion considers the nonseparable case, showing how the results become less clearcut. It is shown, however, that the overall lesson of the analysis (substantial internalization of congestion) continues to apply in this more general case.

In order to make the analysis tractable, the properties of the benefit functions b_p and b_o are also restricted by imposing a number of simplifying assumptions. The first assumption is that no two passengers have the same off-peak benefits, with the same statement applying to peak benefits. Exploiting the first of these properties, let consumers be sorted in increasing order of b_o, so that $b_o'(\theta) > 0$ holds for all $\theta \in [0,1]$. The second assumption is that peak benefits are also increasing in θ, with $b_p'(\theta) > 0$ also holding over the unit interval. This assumption is natural because it says that peak and off-peak travel benefits increase in step with one another across the passenger continuum. Thus, a high value for peak travel is associated with a high value for off-peak travel, indicating a natural linkage in a passenger's valuation of the two travel periods.

A third assumption is imposed in order to easily characterize the allocation of passengers between the peak and off-peak periods. This assumption says that the benefit functions $b_p(\theta)$ and $b_o(\theta)$ exhibit a *single-crossing property*. 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 $b_p'(\theta) < b_o'(\theta)$. Thus, the peak benefit function is either steeper everywhere than the off-peak function, has the same slope, or is flatter everywhere.

It should be noted that, in other continuum-based models of consumer sorting, a single-crossing property can be generated from primitive assumptions on preferences rather than being imposed arbitrarily, as is necessary here. Epple and Romer (1991), for example, rely on a normality assumption to generate single crossing in their local public finance model, which guarantees a simple pattern of consumer sorting by income across communities.

To make the above single-crossing assumptions more concrete, θ could 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 $b_p' > b_o'$.

Note that this single-crossing inequality actually says nothing about the levels of the benefit functions. However, to avoid a degenerate equilibrium, it is shown below that 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.

The other two single-crossing cases are less easily rationalized in the business-leisure context. If $b_p' \equiv b_o'$, then the differential between peak and off-peak benefits is constant across passenger types, which would be true if the leisure passenger shared the business passenger's preference for peak hours. If $b_p' < b_o'$ holds, on the other hand, then peak benefits decline relative to off-peak benefits as the tendency for business travel rises, a counterintuitive pattern. Because of the lower plausibility of these latter cases, the bulk of the analysis is carried out under the assumption that the first single-crossing case, where $b_p' > b_o'$ holds, is relevant. However, recognizing that the alternate cases may be appropriate under some other scenario, the discussion shows how they effect the results of the analysis.

It should be noted that, under the business-leisure interpretation of the passenger continuum, the separability assumption for peak benefits may be unrealistic. In particular, the higher time valuation of business passengers would suggest that, rather than depending on n_p alone, time cost t should be an increasing function of θ. In this case, the benefit function $B_p(\theta, n_p)$ cannot be expressed in a separable fashion, an outcome whose consequences are explored below.

2.2. The social optimum

The social optimum consists of an allocation of passengers to the peak and off-peak periods that maximizes welfare, measured as the difference between travel benefits for passengers and airline costs. Given the single-crossing assumption, the optimal allocation will have the property that high-θ passengers use the peak period, with low-θ passengers traveling off-peak. The 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.

To show that passengers are allocated as claimed between the peak and off-peak, suppose to the contrary that a high-θ passenger with $\theta = \theta_h$ were allocated to the off-peak period, while a low-θ passenger with $\theta = \theta_l < \theta_h$ were allocated to the peak. Combined travel benefits for the two passengers is then $b_o(\theta_h) + b_p(\theta_l) - t(n_p)$. Now consider switching the assignments of the two passengers, which would yield the new benefit expression $b_o(\theta_l) + b_p(\theta_h) - t(n_p)$. Note that since the number of peak and off-peak passengers is unaffected, $t(n_p)$ as well as airline costs are unchanged. Subtracting the first expression above from the second, the change in travel benefits is equal to $[b_p(\theta_h) - b_p(\theta_l)] \; - \; [b_o(\theta_h) - b_o(\theta_l)]$. Since b_p' exceeds b_o' over the interval $[\theta_l, \theta_h]$, this expression is positive, implying that the initial assignment of passengers was not optimal.

With high-θ passengers assigned to the peak, the welfare measure (travel benefits minus airline costs) can be written

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

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

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

Differentiating (2), the first-order condition for choice of $\underline{\theta}$ is

$$b_o(\underline{\theta}) \;\geq\; c/s, \tag{3}$$

with equality holding if $\underline{\theta} > 0$. This condition says that, for the lowest-θ consumer who travels, benefits are at least as large as the cost an airline seat (cost per flight divided by seats per flight). If $b_o(0) > c/s$, so that benefits for the consumer at the bottom of the continuum exceed the seat cost, then (3) is satisfied as an inequality, and all consumers travel. If the reverse inequality holds, then (3) is satisfied as an equality at some positive $\underline{\theta}$, and consumers at the bottom of the continuum do not travel.

Assuming an interior solution, the first-order condition for choice of θ^* is given by

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)] \;-\; n_p t'(n_p) \;-\; (1/s)[g(n_p) + n_p g'(n_p)] \;=\; 0, \tag{4}$$

where $n_p = (1-\theta^*)/s$. It can be shown that, for an interior θ^* to emerge, the benefit functions b_p and b_o must intersect between $\underline{\theta}$ and $\theta = 1$, as noted above.[4] To interpret (4), note that the first expression gives the change in travel benefits for a passenger who is switched from the off-peak to the peak period. As a result of this switch, which corresponds to a reduction in θ^*, the passenger gains $b_p(\theta^*) - t(n_p)$ in peak benefits while losing $b_o(\theta^*)$ in off-peak benefits. The extra peak passenger also generates a congestion effect. His presence requires a $1/s$ increase in peak flights, yielding $(1/s)t'(n_p)$ in additional time cost for each peak passenger. Since there are $1-\theta^*$ such passengers, the total effect is given by the product of the last two expressions, or $n_p t'(n_p)$ (the second term in (4)). The congestion caused by higher peak traffic also raises operating costs for existing flights. This effect is captured by $(1/s)n_p g'(n_p)$ in (4), which equals the increase in cost per flight ($(1/s)g'$) times the number of flights affected. Finally, the need to offer more peak flights also raises airline costs. While the offsetting reduction in off-peak flights means that the part of the cost expression involving c is unchanged, the extra peak-period flights generate additional costs of $(1/s)g(n_p)$. Note that this term is not part of the congestion effect, which equals $n_p t'(n_p) + (1/s)n_p g'(n_p)$. With all these effects taken into account, the optimal θ^* thus balances the individual gain in travel benefits against incremental congestion and operating costs as passengers switch to the peak period.[5] This rule is similar to those emerging from the road-pricing literature.

Under the maintained assumptions on the functions b_p, b_o, t and g, the second partial derivatives $W_{\theta^*\theta^*}$ and $W_{\underline{\theta}\underline{\theta}}$ from (3) are both negative, and the cross partial is zero. As a result, the second-order condition for the optimization problem is satisfied. For future reference, note that since the LHS of (4) equals $-W_{\theta^*}$, this expression is increasing rather than decreasing in θ^*.

3. Analysis of Equilibria

3.1. The competitive case

To analyze equilibria under different market structures, it is useful to begin by considering the competitive case, where carriers are atomistic. In this case, each carrier operates a single flight, and fares just cover operating costs. Thus, peak and off-peak fares are given by

$$f_p = [c + g(n_p)]/s \tag{5}$$

$$f_o = c/s. \tag{6}$$

To determine the allocation of passengers in the competitive case, the first step is to note 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. In other words, 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. Then, note 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. The minimum θ value among off-peak passengers satisfies $b_o(\underline{\theta}) - f_o \geq 0$, indicating that the gain from travel is nonnegative.

Rearranging the last condition and substituting f_o from (6), the resulting condition is the same as (3). Therefore, the margin separating travelers from non-travelers is determined efficiently in the competitive case. However, the former group is divided inefficiently between the peak and off-peak periods. To see this, the above indifference condition can be rewritten to read $b_p(\theta^*) - t(n_p) - b_o(\theta^*) - f_p + f_o = 0$. After substituting the fare expressions from (5) and (6), the condition determining θ^* then reduces to

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)] \; - \; (1/s)g(n_p) \;\; = \;\; 0, \tag{7}$$

where n_p again equals $(1 - \theta^*)/s$.

A comparison of (7) and (4) shows that the terms $n_p t'(n_p) + (1/s)n_p g'(n_p)$, which capture the congestion effect, are absent from the equilibrium condition. As a result, in determining θ^*, competitive carriers do not take into account the congestion caused by adding an extra passenger to the peak period. Since the congestion terms are subtracted in (4), it follows that the LHS is negative at the equilibrium θ^* value, where (7) equals zero. This fact implies that θ^* must be raised from the equilibrium to reach the optimum (recall that the LHS of (4) is increasing in θ^*). Thus, moving from the competitive equilibrium to the optimum means reallocating passengers from the peak to the off-peak period, indicating that the peak period is over-used in equilibrium.

To limit this over-use, the solution is to impose congestion pricing. The peak-period congestion toll, which depends on n_p, equals the congestion cost imposed by an additional flight, and it is given by

$$R(n_p) \;\; = \;\; sn_p t'(n_p) \; + \; n_p g'(n_p). \tag{8}$$

When the carriers are charged $R(n_p)$ per flight, f_p in (5) is augmented to include the additional expression $R(n_p)/s$. Eq. (7) is then modified to include the missing congestion terms $n_p t'(n_p) + (1/s)n_p g'(n_p)$, and the new equilibrium coincides with the social optimum.[6] The congestion toll achieves this outcome by putting upward pressure on the peak fare, which diverts passengers to the off-peak period. It can be shown that, once θ^* has adjusted to the new equilibrium, the peak fare remains higher than prior to imposition of the toll.

3.2. The case of a perfectly discriminating monopolist

Having developed the competitive benchmark, the discussion now turns to alternate cases where carriers have market power. The natural starting point is the case of a perfectly discriminating monopolist, who can charge a different fare to each passenger. Although such behavior on the part of firms is usually thought to be implausible, the extensive price discrimination practiced by the airline industry suggests that a perfectly discriminating model is far from outlandish (see, for example, Borenstein and Rose (1994)).

Letting $f_p(\theta)$ and $f_o(\theta)$ denote the peak and off-peak fares charged to a type-θ passenger, the monopolist sets these fares so as to make the passenger indifferent between traveling and not traveling. Thus, the fares are set so as to exhaust travel benefits, satisfying

$$f_p(\theta) \;=\; b_p(\theta) \;-\; t(n_p) \tag{9}$$

$$f_o(\theta) \;=\; b_o(\theta). \tag{10}$$

Using the previous argument along with (9) and (10), it is easily seen that the monopolist maximizes revenue by allocating high-θ passengers to the peak and low-θ passengers to the off-peak.[7] Profit is then given by $\int_{\underline{\theta}}^{\theta^*} f_o(\theta)d\theta \;+\; \int_{\theta^*}^{1} f_p(\theta)d\theta$ minus airline costs. However, after substituting for the fares using (9) and (10), the resulting objective function is identical to the social welfare measure W in (1). Thus, in maximizing profit, the perfectly discriminating monopolist replicates the social optimum, generating optimal values of $\underline{\theta}$ and θ^*.

While this conclusion is not especially surprising, it is important from a policy perspective. In particular, the conclusion suggests that airport congestion is fully internalized under conditions that are not particularly unrealistic. This outcome occurs when a carrier has substantial ability to price discriminate, and when it controls virtually all of an airport's traffic, as is the case at a number of dominated hubs (see Table 1). The analysis suggests that congestion levels at such airports may not be far from optimal, conditional on airport capacity. While capacity growth itself may be desirable, the analysis implies that there may be little or no role for government intervention in reallocating traffic away from peak periods. Indeed, imposition of a congestion toll under these conditions would lead to a *suboptimal* level of peak congestion, with inefficient under-use of the peak period.

Interestingly, Daniel's (1995) findings undermine this conclusion. Daniel found that his simulation results replicated existing Minneapolis traffic patterns better when the dominant hub carrier (Northwest) was assumed to behave atomistically, ignoring the congestion it imposed on itself, than when it took such congestion into account. While this finding sounds a note of caution in using the present model to draw policy conclusions, the finding is troublesome in that it seems inconsistent with optimizing behavior on the part of the airlines.

3.3. The case of a nondiscriminating monopolist

Recognizing that perfect price discrimination may not occur, it is useful to analyze the monopoly equilibrium when the firm cannot discriminate, charging uniform peak and off-peak fares (again denoted f_p and f_o). In this case, the distribution of passengers is governed by conditions analogous to those in the competitive case. The previous indifference condition for the type-θ^* passenger again applies, and this condition can be rearranged to read $f_p = b_p(\theta^*) - t(n_p) - b_o(\theta^*) + f_o$. In contrast to (6), the monopolist raises f_o to exhaust travel benefits for the lowest-θ passenger, so that $\underline{\theta}$ satisfies $f_o = b_o(\underline{\theta})$.

The monopolist's revenue is given by $f_p(1-\theta^*) + f_o(\theta^* - \underline{\theta})$. Substituting for the fares using the above relationships, and subtracting costs, profit equals

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)](1-\theta^*) \ + \ b_o(\underline{\theta})(1-\underline{\theta}) \ - \ c(1-\underline{\theta})/s \ - \ [(1-\theta^*)/s]g[(1-\theta^*)/s]. \quad (11)$$

Differentiating (11), the first-order condition for choice of $\underline{\theta}$ is

$$b_o(\underline{\theta}) \ - \ (1-\underline{\theta})b_o'(\underline{\theta}) \ \ \geq \ \ c/s, \quad (12)$$

with equality holding if $\underline{\theta} > 0$. Since $b_o(\underline{\theta}) > c/s$ holds at an interior monopoly equilibrium, it follows that $\underline{\theta}$ exceeds the optimal value, which either satisfies $b_o(\underline{\theta}) = c/s$ or is zero (recall that $b_o' > 0$). The higher $\underline{\theta}$ means the monopolist sets f_o high enough so that some consumers who would have traveled under the social optimum choose not to do so. The monopolist is exploiting market power, recognizing that a higher fare excludes some consumers while generating more revenue from inframarginal passengers.

Differentiating (11), the first-order condition for θ^* is given by

$$\begin{aligned} [b_p(\theta^*) - t(n_p) - b_o(\theta^*)] \ - \ n_p t'(n_p) \ - \ (1/s)[g(n_p) + n_p g'(n_p)] \\ - \ (1-\theta^*)[b_p'(\theta^*) - b_o'(\theta^*)] \ \ = \ \ 0, \end{aligned} \quad (13)$$

where n_p again equals $(1-\theta^*)/s$. This condition differs from the optimality condition (4) only in the appearance of the last term involving the benefit derivatives. Because the congestion terms continue to appear in (13), it follows that the nondiscriminating monopolist, like his discriminating counterpart, fully internalizes the congestion effect.[8]

Even though congestion is again internalized, the extra term in (13) means that the allocation of passengers between the peak and off-peak periods is not optimal. Because the single-crossing assumption implies $b_p' - b_o' > 0$, the extra term is negative, so that the LHS of (4) is positive at the equilibrium. It follows that θ^* must be decreased from the equilibrium to reach the optimum, implying that the monopolist allocates *too few* passengers to the peak period, relative to the social optimum.

Exploitation of market power accounts for the monopolist's internalization of congestion, while also explaining the appearance of the last term in (13). To see this, note that in adjusting θ^*, the monopolist must alter the peak fare to maintain the indifference condition, ensuring that passengers divide between the periods as intended. Because higher congestion reduces peak benefits, the monopolist must cut f_p as additional passengers are allocated to the peak period, an adjustment that maintains the indifference condition. This market-power effect is beneficial, causing the monopolist to restrict peak traffic.

A "residual" market-power effect, which is not beneficial, is captured by the last term in (13). This residual effect arises because lower-benefit passengers are added as peak traffic is raised, altering the difference between b_p and b_o for the type-θ^* passenger. Since $b_p' > b_o'$ holds, peak benefits fall by more than off-peak benefits as θ^* declines, which requires a further reduction in f_p to maintain the indifference condition. This fare reduction leads to a further restriction of peak traffic, generating under-use of the peak period.

The sign of the residual market-power effect, which underlies this result, obviously depends on the relationship between the benefit derivatives, a relationship that would be altered under a different single-crossing assumption. To see the effect of changing this assumption, suppose that the benefit slopes were identically equal, with $b_p' \equiv b_o'$. Then, the pattern of passenger assignments to the peak and off-peak periods would be a matter of indifference to both the social planner and the monopolist, so that the current assignment (high-θ passengers in the peak) could be retained. All the previous analysis is then unaffected, but the extra term in (13) is now identically zero. As a result, the nondiscriminating monopoly equilibrium coincides with the social optimum. In this case, residual market-power effect does not distort the allocation of passengers between the two periods.

The remaining single-crossing case, where $b_p' < b_o'$ holds, is considered in the appendix. In this case, passenger assignments are reversed, with the peak period used by low-θ, rather than high-θ, passengers. The analysis shows that, for the competitive case and the perfectly discriminating monopolist, the previous efficiency verdicts are unaffected. While the nondiscriminating monopolist continues to internalize congestion, the residual market-power effect may now lead to the assignment of *too many*, rather than too few, passengers to the peak period. Thus, the direction of the distortion in the nondiscriminating case depends on the form of the single-crossing condition.

Summarizing the analysis up to this point yields

Proposition 1. *While competitive carriers fail to internalize the congestion effect, leading to over-use of the peak period, congestion is fully internalized by a perfectly discriminating monopolist, who replicates the social optimum. A nondiscriminating monopolist again internalizes the congestion effect, but a residual market-power effect may distort the allocation of passengers, yielding over- or under-use of the peak period.*

The important implication of this result is that, even in the more-realistic nondiscriminating case, there appears to be no role for congestion pricing at an airport dominated by a monopolistic carrier. The airport's traffic allocation may not be optimal, but the culprit is not a failure to internalize the effects of congestion.

3.4. The case of Cournot oligopoly

Consider now the oligopoly case, where each of k identical firms behaves in Cournot fashion. While a monopolist can be portrayed as choosing θ^* and $\underline{\theta}$, each oligopoly firm can only choose the number of its own flights, with the aggregation of all flights determining θ^* and $\underline{\theta}$. Letting n_p^i and n_o^i give peak and off-peak flights for carrier i, θ^* and $\underline{\theta}$ satisfy $1 - \theta^* = s\sum_{i=1}^k n_p^i$ and $\theta^* - \underline{\theta} = s\sum_{i=1}^k n_o^i$. Rearranging, these relationships yield $\theta^* = 1 - s\sum_{i=1}^k n_p^i$ and $\underline{\theta} = 1 - s\sum_{i=1}^k (n_o^i + n_p^i)$. After substitution, the relations $f_o = b_o(\underline{\theta})$ and $f_p = b_p(\theta^*) - t(n_p) -$

$b_o(\theta^*) + b_o(\underline{\theta})$ can then be written as

$$f_o = b_o[1 - s\Sigma(n_o^i + n_p^i)] \tag{14}$$

$$f_p = b_p(1 - s\Sigma\, n_p^i) \ - \ b_o(1 - s\Sigma\, n_p^i) - t(\Sigma\, n_p^i) \ + \ b_o[1 - s\Sigma(n_o^i + n_p^i)]. \tag{15}$$

Carrier j's profit is then written

$$f_o s n_o^j \ + \ f_p s n_p^j \ - \ c(n_o^j + n_p^j) \ - \ n_p^j g(\Sigma\, n_p^i). \tag{16}$$

Carrier j chooses n_o^j and n_p^j to maximize (16) subject to (14) and (15), viewing other carriers' flight choices as parametric. Letting n_p denote $\sum_{i=1}^{k} n_p^i$, total peak flights, the resulting first-order condition for n_p^j reduces to

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)] \ - \ n_p^j t'(n_p) \ - \ (1/s)[g(n_p) + n_p^j g'(n_p)]$$
$$- \ s n_p^j [b_p'(\theta^*) - b_o'(\theta^*)] \ = \ 0, \tag{17}$$

This condition differs from monopoly condition (13) in two ways. First, the congestion effect is only partly internalized. Rather than being multiplied by total flights n_p, t' and $(1/s)g'$ are multiplied by n_p^j, carrier j's flights. Thus in choosing the number of flights, carrier j considers *only the congestion it imposes on itself*, which consists of additional time costs for its own passengers and extra own-operating costs. The second difference relative to (13) is that the residual market-power effect represented by the last term applies only to a subset of passengers, those using carrier j.

Symmetry of the equilibrium can be used to rewrite (17) as

$$[b_p(\theta^*) - t(n_p) - b_o(\theta^*)] \ - \ (n_p/k)t'(n_p) \ - \ (1/s)[g(n_p) + (n_p/k)g'(n_p)]$$
$$- \ [(1 - \theta^*)/k][b_p'(\theta^*) - b_o'(\theta^*)] \ = \ 0, \tag{18}$$

Written in this way, the first-order condition shows that each carrier internalizes a fraction $1/k$ of the congestion effect, accounting only for the congestion it imposes on itself. Note that as k becomes large, (17) converges to the competitive condition (7), verifying for the present model the usual conclusion that perfect competition is the limiting case of oligopoly.[9]

Although the residual market-power effect continues to distort the allocation of traffic, the distortion due to uninternalized congestion, which creates a tendency for over-use of the peak period, can be remedied by a congestion toll. To eliminate this distortion, the toll per flight should be set at

$$\widetilde{R}(n_p) \quad = \quad \left(1 - \frac{1}{k}\right)[sn_p t'(n_p) \; + \; n_p g'(n_p)], \tag{19}$$

As in the competitive case, imposition of the toll moves θ^* and n_p toward the social optimum, while raising the peak-period fare. Note that, because of the residual market-power effect, the equilibrium generated by the toll in (19) will not be exactly optimal unless $b'_p \equiv b'_o$.[10]

The analysis thus demonstrates that each airline is charged for congestion that it fails to internalize, with the toll equal to the congestion cost from an extra flight times one minus the carrier's flight share. In the case of a duopoly airport, roughly approximated by Chicago-O'Hare, each of the duopolists would be charged for *half* of the congestion associated with an additional flight.[11]

If this principle were extrapolated to an asymmetric model, it would imply that small airlines should be charged a larger toll per flight than large airlines, reflecting their limited internalization of congestion. Unfortunately, such a rule might have undesirable competitive effects, given that ability-to-pay may be lower for small than for large carriers. However, since efficiency analysis is problematic in an asymmetric setting, the basis for this rule may be questionable. To see this, note that one way of generating an asymmetric equilibrium is to assume cost differences across firms. But in such a setting, a planner would allocate all passengers to the efficient (large) firms, so that small, inefficient firms would disappear in the social optimum. Evidently, to generate a social optimum that resembles observed asymmetric equilibria, a richer model is needed. Only in the context of such a model would statements about asymmetric congestion tolls be reliable.

Summarizing the preceding analysis yields

Proposition 2. *In a Cournot oligopoly, each carrier internalizes only the congestion it imposes on itself, allowing a role for congestion pricing. In the symmetric model, the toll per flight equals the congestion cost from an extra flight times one minus a carrier's airport flight share. Imposition of such a toll, however, cannot eliminate any remaining traffic misallocation due to the residual market-power effect.*

3.5. The case of a Stackelberg oligopoly

It is interesting to explore how the oligopoly results change if Cournot behavior is replaced by the assumption that one carrier is a Stackelberg leader. Because a general analysis of this case is infeasible, additional simplifying assumptions are imposed. First, attention is restricted to a model with just two carriers, with carrier 1 being the leader and 2 the follower. Because of the problematic nature of models with asymmetric costs (see above), the two carriers are assumed to have identical costs. As a result, the realistic case where a large leader interacts with a group of small, fringe followers is not considered. In addition, all the functions in the model are assumed to be linear, with $t' \equiv \tau$, $g' \equiv \gamma$, and $b_p' = b_o' \equiv \beta$ (the benefit functions thus have a common slope).

Under these assumptions, the first-order conditions for carrier 2 are derived, using (14)-(16), and these conditions are differentiated to yield reaction functions, which show how the follower's decision variables, n_o^2 and n_p^2, change in response to the leader's variables, n_o^1 and n_p^1. Taking the follower's reactions into account, the equations (14)-(16) are again used to derive the first-order conditions for the leader. Then, combining the follower's and the leader's conditions, which constitute a linear equation system, the Stackelberg equilibrium is computed algebraically. The Cournot equilibrium is also computed, and the results are compared.

Before discussing the solutions, it is important to note how the leader's first-order conditions change relative to the Cournot case. Because $\partial n_p^2/\partial n_p^1 = -1/2$, the follower reduces his peak flights by half in response to a unit increase in the leader's flights. As a result, the additional congestion generated by an increase in the leader's flights is partly mitigated. This effect is manifested in the leader's first-order condition by a 50 percent reduction in the amount of his own congestion that is internalized. In other words, the congestion terms

$n_p^1 t'(n_p) + (1/s)n_p^1 g'(n_p)$ from (17), which reduce to $n_p^1\tau + (1/s)n_p^1\gamma$ under linearity, are replaced by $[n_p^1\tau + (1/s)n_p^1\gamma]/2$.

Although the residual market-power term in (17) is zero given equal benefit slopes, new market-power effects emerge in both of the leader's first-order conditions. As a result, it is difficult to predict the net effect of the reduced internalization of congestion. However, the solutions show that n_p^1 is larger in the Stackelberg than in the Cournot equilibrium, indicating that carrier 1's peak flights increase when it becomes a leader. Conversely, n_p^2 is smaller in the Stackelberg equilibrium, so that carrier 2's peak flights fall when it becomes a follower. Moreover, the increase in n_p^1 dominates the decline in n_p^2, so that the total number of peak flights, $n_p = n_p^1 + n_p^2$, is larger in the Stackelberg equilibrium. By contrast, the change in total off-peak flights is ambiguous. Because the details of this analysis are tedious and complex, they are not presented. Summarizing yields

Proposition 3. *In a linear duopoly model with symmetric firm costs, the tendency toward over-use of the peak period due to uninternalized congestion is exacerbated under Stackelberg behavior (relative to the Cournot case).*

3.6. The effects of nonseparability

The previous results have been derived under separability of the peak travel benefit function, and it is important to appraise the effects of eliminating this restriction. With nonseparability of $B_p(\theta, n_p)$, the second integral in the welfare function (2) becomes $\int_{\theta^*}^1 B_p[\theta, (1-\theta^*)/s]d\theta$. The benefit-differential term in the optimality condition (4) is then written $B_p(\theta^*, n_p) - b_o(\theta^*)$, while the $-n_p t'$ term is replaced by $n_p \int_{\theta^*}^1 B_p^n(\theta, n_p)d\theta/(1-\theta^*) < 0$, where the n superscript denotes B_p's negative partial derivative with respect to n_p.[12]

Although nonseparability has no effect on the conclusions of the competitive and discriminating-monopoly analysis, the results on internalization of congestion for the other market structures are altered. To see this, observe that the form of the equilibrium condition (13) for the nondiscriminating monopolist is unchanged aside from notation. The benefit-differential term is modified as above, the residual market-power effect involves B_p^θ instead of b_p', and $-n_p t'$ is replaced by $n_p B_p^n(\theta^*, n_p)$. Because this latter term differs from the expression replacing $-n_p t'$

in (4) (the above integral), congestion may not be exactly internalized by the nondiscriminating monopolist. The outcome depends on the direction of the following inequality:

$$B_p^n(\theta^*, n_p) \quad \leq \ (>) \quad \frac{1}{1-\theta^*} \int_{\theta^*}^{1} B_p^n(\theta, n_p) d\theta. \tag{20}$$

To evaluate (20) under the business-leisure interpretation of the passenger continuum, note that B_p^n in this case should become more negative as θ increases, reflecting the greater value of the time lost to congestion for high-θ passengers. The integral in (20), which represents the average value of B_p^n between θ^* and 1, is then more negative than $B_p^n(\theta^*, n_p)$ (for a given n_p). Because the integral appears in the social optimality condition while the latter expression appears in the equilibrium condition, it follows that the monopolist does not fully internalize congestion. Under other scenarios, of course, the inequality in (20) could be reversed, implying over-internalization of congestion.

These discrepancies arise because, in focusing on the type-θ^* passenger via the indifference condition, the monopolist does not consider congestion effects felt by inframarginal passengers. In the case where B_p is separable, the marginal, type-θ^* individual turns out to be a perfect representative for inframarginal passengers, and congestion is properly internalized. Otherwise, internalization is not exact.

Similar conclusions apply in the oligopoly case. Rather than exactly internalizing the congestion it imposes on itself, an oligopoly carrier may internalize a smaller or larger share when benefits are nonseparable. It is important to recognize that, as in the monopoly case, this conclusion tempers the results of the analysis without overturning its main lesson. This lesson is that the exercise of market power leads carriers to internalize a portion of the congestion they generate, with this portion rising as market power grows.

4. Conclusion

This paper has analyzed airport congestion when carriers are nonatomistic, showing how the results of the road-pricing literature are modified when the economic agents causing congestion have market power. The analysis shows that when an airport is dominated by a monopolist, either discriminating or nondiscriminating, congestion is fully internalized, provided that a separability assumption on travel benefits is satisfied. In allocating traffic between the peak and off-peak periods, the monopolist fully accounts for the passenger time cost associated with peak congestion (which alters fares), while also taking account of its impact on his own operating costs. The analysis thus suggests no role for congestion pricing under monopoly conditions. This conclusion is qualified in the absence of separability. Although the nondiscriminating monopolist may not exactly internalize congestion in this case, substantial internalization still occurs.

Under a Cournot oligopoly, carriers are shown to internalize only the congestion they impose on themselves (assuming separability). A toll that captures the uninternalized portion of congestion can then improve the allocation of traffic. The toll is equal to the congestion cost from an extra flight times one minus a carrier's flight share.

This result shows the flaw in a simple extrapolation of results from the road-pricing literature to the airport setting. Instead of being charged for *all* the congestion an added flight causes, as would occur if airlines were treated like road users, the toll should reflect only the costs imposed on other carriers. At an airport like Chicago-O'Hare, this rule would imply that United and American would be charged for only about half of the congestion created by an additional flight. At a monopoly airport, the rule implies a zero toll since all congestion is internalized. Given the likelihood that some form of congestion pricing will be implemented at U.S. airports, awareness of such results may be valuable.

Appendix

This appendix demonstrates that over-use of the peak period may occur when the alternate single-crossing assumption $b_p' < b_o'$ holds. Under this assumption, low-θ passengers are assigned to the peak period, necessitating changes in the welfare measure (2). The integrands must be switched in the two integrals, and $1 - \theta^*$ must be replaced by $\underline{\theta} - \theta^*$ in each instance. The social optimality condition (4) still governs choice of θ^*, although n_p is given by $(\underline{\theta} - \theta^*)/s$. However, since the choice of $\underline{\theta}$ now affects traffic in the congested peak period, the relevant optimality condition has a more-complex form than (3).[13]

As before, the perfectly discriminating monopolist satisfies both social optimality conditions, while the peak period is again over-used in the competitive case.[14] For the nondiscriminating monopolist, a market-power effect again distorts the choice of $\underline{\theta}$ in an upward direction.[15] As for the choice of θ^*, the monopolist's optimality condition is again given by (13), with $n_p = (\theta^* - \underline{\theta})/s$, so that congestion is once again internalized. However, since $b_p' < b_o'$ now holds, the modified expression in (4) is negative when evaluated at the equilibrium. Since this expression is now decreasing rather than increasing in θ^*, it follows that θ^* must once again be reduced from its equilibrium value to reach the optimum. Recalling that the peak period now contains low-θ passengers, this fact means that the residual market-power effect tends to allocate *too many* passengers to the peak. However, since the reverse effect occurs at the bottom of the continuum, with $\underline{\theta}$ too large in equilibrium, the net effect on peak usage relative to the optimum is ambiguous. The conclusion is that peak usage may be larger than optimal under the given single-crossing assumption, although this outcome is not assured.

Table 1.
Delays in 1999 at Major U.S. Airports[¤]

Airport	Delays	% Weather	% Volume	Percent on Time: Arrival	Percent on Time: Departure	Carrier and % Flight Share: 1st Carrier	2nd Carrier	3rd Carrier
Chicago-O'Hare	49,202	73.8	12.7	66.4	70.1	United (44.5)	American (38.9)	Northwest (2.2)
Newark	36,553	76.4	9.1	61.6	69.0	Continental (57.2)	United (7.9)	Delta (5.7)
Atlanta	32,737	79.8	9.4	69.1	73.2	Delta (73.5)	Air Tran (10.9)	US Airways (2.3)
NY-La Guardia	28,474	56.1	13.0	59.9	71.1	US Airways (37.6)	Delta (18.8)	American (16.8)
San Francisco	21,187	82.5	8.3	67.9	78.5	United (58.2)	American (7.4)	Delta (4.8)
Dallas-Ft. Worth	16,731	75.5	15.8	78.3	76.3	American (68.5)	Delta (17.3)	United (1.9)
Boston	14,989	76.0	1.2	62.3	70.7	US Airways (25.9)	American (25.7)	Delta (15.3)
Philadelphia	14,516	72.6	6.4	59.6	62.1	US Airways (65.8)	Delta (6.1)	American (6.0)
NY-Kennedy	13,547	74.6	8.5	72.0	81.0	American (27.8)	Delta (20.4)	TWA (15.8)
Phoenix	11,919	35.2	46.2	70.4	69.2	America West (48.4)	Southwest (26.2)	United (5.4)
Detroit	11,522	46.1	21.4	75.4	73.7	Northwest (79.8)	Delta (3.3)	Southwest (2.8)
Los Angeles	10,646	84.5	1.2	73.9	79.2	United (36.3)	American (17.2)	Southwest (11.5)
St. Louis	9,631	85.5	1.3	78.0	74.1	TWA (73.2)	Southwest (12.9)	Delta (2.5)
Houston	9,524	84.8	4.8	71.2	75.4	Continental (81.6)	Delta (3.3)	American (2.9)
Washington-Dulles	9,248	63.9	17.1	63.6	70.1	United (62.4)	US Airways (19.2)	Delta (4.4)

[¤]1999 is the most recent year for which the data in the first three columns are available at the airport level. The figures in these columns are taken from the FAA web site (http://www.faa.gov/newsroom.htm). The on-time percentages, which pertain to July 1999, are taken from DOT's *Air Travel Consumer Report* (http://www.dot.gov/airconsumer/). The flight-share data are from Baker (2000).

References

Baker, C., 2000. Airports: Top 100 ranking, *Airline Business*, July, 55-66.

Borenstein, S. and N.L. Rose, 1994. Competition and price dispersion in the U.S. airline industry, *Journal of Political Economy* 102, 653-683.

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

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 M. Pahwa, 2000. Comparison of three empirical models of airport congestion pricing, *Journal of Urban Economics* 47, 1-38.

Daniel, J.I., 2001. Distributional consequences of airport congestion pricing, *Journal of Urban Economics*, forthcoming.

Epple, D.N. and T. Romer, 1991. Mobility and redistribution, *Journal of Political Economy* 99, 828-858.

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

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. and C. Winston, 1989. Enhancing the performance of the deregulated air transportation system, *Brookings Papers on Economic Affairs, Microeconomics*, 61-123.

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

*I thank Hadi Esfahani, Kangoh Lee and Eric Verhoef for helpful comments. Any errors or shortcomings in the paper, however, are my responsibility.

[1]1999 is the most recent year for which the data in the first three columns of Table 1 are available at the airport level. The figures in these columns (as well as the numbers in the text) are taken from the FAA web site (http://www.faa.gov/newsroom.htm). The on-time percentages, which pertain to July 1999, are taken from DOT's *Air Travel Consumer Report* (http://www.dot.gov/airconsumer/). The flight-share data are from Baker (2000).

[2]Weather conditions far from an airport can also create delays by blocking air-traffic corridors, reducing the capacity of the air space.

[3]By focusing on a single airport, the analysis does not consider the link between congestion at the given airport and congestion at other airports that are connected to it. If each such airport is uncongested, this omission would appear to have little consequence. However, if the other airports are congested, then altering the peak/off-peak traffic allocation at the given airport would affect their congestion levels by changing the time pattern of traffic to and from that airport. By ignoring such interrelationships, the present analysis follows the prior literature. However, analysis of airport congestion in a network context would be a useful avenue for future research.

[4]Recognizing that the LHS of (4) equals $-W_{\theta^*}$, necessary conditions for an interior solution are as follows. For θ^* to be less than unity (so that the peak period is indeed used), $b_p(1) > b_o(1)$ must hold, indicating that passengers at the top of the continuum prefer peak travel. Otherwise, W is increasing in θ^* at $\theta^* = 1$ (i.e., the LHS of (4) is negative). Similarly, for θ^* to be greater than $\underline{\theta}$ (so that the off-peak period is used), $b_p(\underline{\theta}) < b_o(\underline{\theta})$ must hold, indicating that a type-$\underline{\theta}$ passenger (where $\underline{\theta}$ satisfies (3)) prefers off-peak travel. Otherwise, W is decreasing in θ^* at $\theta^* = \underline{\theta}$ (with the LHS of (4) positive). These two inequalities imply that the benefit functions must intersect at an intermediate value of θ.

[5]As stated earlier, it is assumed that the solution to (4) involves positive values for $t(\cdot)$ and $g(\cdot)$.

[6]Alternatively, a fixed congestion toll (independent of n_p), which is equal to (8) evaluated at the socially optimal n_p, can be charged. While this toll also generates the optimum, the expression in (8), which constitutes a *schedule* giving the toll as a function of the current number of peak flights, does not require the planner to compute the social optimum.

[7] Passengers can be directed to such an allocation if they are charged more than the amount in (9) or (10) in the period where the monopolist does not want them to travel.

[8] It is assumed that the second-order conditions for the monopolist's optimization problem are satisfied. These conditions involve b_p'' and b_o'', which have no natural sign.

[9] Once symmetry is imposed, the first-order condition for choice of n_o^j reduces to $b_o(\underline{\theta}) - [(1-\underline{\theta})/k]b_o'(\underline{\theta}) \geq c/s$, indicating that the market-power effect in the choice of $\underline{\theta}$ applies only to a carrier's own passengers.

[10] With the maintained single-crossing assumption, the peak-period will be under-used, as in the nondiscriminating monopoly case. As before, this conclusion could be reversed under an alternate assumption.

[11] It is interesting to note that this conclusion overturns the well-known self-financing rule for congested facilities, which says that toll revenue exactly covers the construction cost of a congested facility built with constant returns to scale (see Small (1992)). To see this, let the t and g functions depend on airport capacity H, and assume that the congestion-cost expression is appropriately homogeneous of degree zero in H and n_p. Then, the optimality condition for H can be written $n_p[sn_pt^n(n_p, H) + n_pg^n(n_p, H)] = HF'(H)$, where F is the cost function for capacity and the n superscripts denote derivatives with respect to n_p. Using (19), congestion-toll revenue is equal to the LHS of the previous expression times $(1 - 1/k)$, while the RHS equals $F(H)$ under constant returns. Thus, toll revenue fails to cover the cost of the optimal-size airport.

[12] To avoid complications, the modified single-crossing inequalty $B_p^\theta(\theta, n_p) > b_o'(\theta)$ must hold for all values of n_p as well as for all θ (the θ superscript denotes partial derivative).

[13] The optimality condition is $[b_p(\underline{\theta}) - t(n_p)] - c/s - n_pt'(n_p) - (1/s)[g(n_p) + n_pg'(n_p)] \geq 0$, with equality holding if $n_p > 0$.

[14] Since $\underline{\theta}$ now satisfies $b_p(\underline{\theta}) - t(n_p) \geq c/s$ in the competitive case, it follows that the expression in footnote 13 is negative at the competitive equilibrium. Since that expression is increasing in $\underline{\theta}$, it follows that $\underline{\theta}$ must be raised above the equilibrium value to reach the optimum. Similarly, since (7) again holds at the competitive equilibrium, the LHS of the optimality condition (4) is negative at the equilibrium. Because that expression is now decreasing in θ^*, it follows that θ^* must be reduced from the equilibrium to reach the optimum. With the optimum having a higher $\underline{\theta}$ and lower θ^* than the equilibrium, it follows that equilibrium involves over-use of the peak period.

[15] The equilibrium condition consists of the optimality condition from footnote 13 with the term $-(1-\underline{\theta})b_p'(\underline{\theta})$ added. At an interior equilibrium, where the modified expression equals zero, the expression in footnote 13 is positive. Since that expression is increasing in $\underline{\theta}$, it follows that the optimal value (which makes the expression zero or is itself zero) is less than the equilibrium value.