Price vs. Quantity-Based Approaches to Airport Congestion Management

by

Jan K. Brueckner
Department of Economics
University of California, Irvine
3151 Social Science Plaza
Irvine, CA 92697
e-mail: jkbrueck@uci.edu

May 2008, revised September 2008, February 2009

Abstract

This paper analyzes price and quantity-based approaches to management of airport congestion, using a model where airlines are asymmetric and internalize congestion. Under these circumstances, optimal congestion tolls are differentiated across carriers, and a uniformity requirement on airport charges (as occurs when slots are sold or tolls are uniform) distorts carrier flight choices. Flight volumes tend to be too low for large carriers and too high for small carriers. But quantity-based regimes, where the airport authority allocates a fixed number of slots via free distribution or an auction, lead carriers to treat total flight volume (and thus congestion) as fixed, and this difference generates an efficient outcome as long as the number of slots is optimally chosen.

Price vs. Quantity-Based Approaches to Airport Congestion Management

by

Jan K. Brueckner*

1. Introduction

Flight delays caused by airport congestion are a growing problem in both the US and Europe, and policymakers have struggled to formulate a response. To address rising congestion at Chicago's O'Hare Airport, the Federal Aviation Administration (FAA) took a micromanagement approach, prevailing on the airport's two major carriers (United and American) to cut their peak flight volumes while prohibiting smaller carriers from adding flights to fill the gap. More-systematic FAA interventions have occurred at New York airports, initially at LaGuardia and most recently at John F. Kennedy and Newark airports, where the FAA capped peak hour flight operations. All of these interventions followed surges in flights spurred by relaxation of long-standing slot constraints at O'Hare, LaGuardia and JFK, three airports where FAA-allocated slots give airlines the right to operate at particular times.[1]

In announcing the New York flight caps, the FAA also proposed a system where carriers would each year relinquish a portion of their slots for redistribution via an auction. This proposal is expected to stimulate the existing secondary slot market, where slots are traded among carriers, but it is strongly opposed by the airlines, who expect it to raise their costs. Contemporaneously, a position paper issued by the U.S. Department of Justice (Whalen et al., 2007) endorsed slot auctions as a mechanism for addressing airport congestion. Following these New York policy decisions, the FAA took another significant step by changing its rules on landing fees, which are charged to carriers for each flight operation at an airport. While landing fees traditionally depended only on aircraft weight, the new rules effectively allow the fees to vary by time of day. This change permits airports to implement congestion pricing, with high landing fees charged during peak hours and lower fees charged in off-peak periods.

With these developments, several new and distinct policy approaches for attacking airport congestion have gained credibility, mirroring recent progress in implementing congestion pricing for roads (London and Stockholm are prominent examples). The purpose of the present paper is to provide a comparative analysis of these different possible approaches to airport congestion management, with the goal of helping to inform the policy debate.

The congestion-management approaches are divided into two main categories: price-based and quantity-based. Under the price-based approaches, the airport authority announces a charge per flight that airlines must pay to use a congested airport, with quantity decisions (flight totals) chosen entirely by the carriers. The first price-based approach is a regime of differentiated, carrier-specific congestion tolls, while the second approach levies a uniform per-flight charge for airport access. The latter regime corresponds either to a uniform-toll system, where differentiation across carriers is ruled out, or to a slot-sale regime, where airport slots are sold to carriers at a common price. Under the quantity-based approaches, the airport authority announces a total desired flight volume, which is enforced through allocation of a corresponding number of airport slots. Under a slot-distribution regime, the slot allocation is achieved by distributing the fixed total number of slots free of charge, while allowing carriers to make adjustments in their slot holdings through trade. This regime closely corresponds to the slot system in existence at slot-controlled US airports prior to the latest FAA proposals, where trading occurs in secondary markets.[2] Under a slot-auction regime, the airport allocates the fixed slot total through an auction process. As modeled, this regime differs from the FAA's New York auction proposal by auctioning all, rather than a portion, of the slots in each period. It should be noted that, although the two-quantity based approaches actually involve prices (which govern slot trading and the slot auction), their key element is a fixed, announced volume of total flights, which carriers take into account in their decisions.

Using the simplest possible model, the analysis begins by characterizing the socially optimal allocation of flights and showing that the optimum can be generated by the differentiated-toll regime. The need for differentiated tolls arises from the assumed asymmetry of the carriers combined with internalization of airport congestion. Internalization occurs because a nonatomistic carrier, in scheduling an extra flight, takes into account the additional congestion costs imposed on the other flights it operates. The appropriate congestion toll then captures only the congestion imposed on other carriers, excluding the congestion the carrier imposes on itself. With carrier asymmetry, a carrier with a large flight share should then pay a low toll, given that it internalizes most of the congestion from its operation of an extra flight, while a small carrier, which internalizes little of the congestion it creates, should pay a high toll. Note that this pattern of tolls, though required for efficiency, may be controversial given the inverse relationship between carrier size and the toll per flight. While Daniel (1995) was the first to recognize the potential for internalization of airport congestion, this pricing rule was advanced by Brueckner (2002, 2005) and further explored by other authors.[3]

Differentiated charges are inefficiently ruled out under the slot-sale/uniform toll regime, and the analysis explores the consequences of this approach. By failing to account for differences in the internalization of congestion, the uniform charge under this regime excessively penalizes large carriers and insufficiently penalizes small carriers for the congestion they create. Given this pattern, large carriers tend to operate too few and small carriers too many flights under the slot-sale/uniform-toll regime, and the analysis (with the help of some numerical examples) makes this conclusion precise. Although the regime's inefficiency disappears when carriers are symmetric (in which case differentiated tolls end up being uniform), asymmetry characterizes most airports, making the inefficiency a practical concern.

With differentiated tolls controversial and the slot-sale/uniform-toll regime inefficient, the analysis then asks whether the quantity-based approaches might be superior. The initial focus is on the slot-distribution regime, where slots are distributed and then traded between carriers. Remarkably, the analysis shows that, even though the trading process involves a uniform slot price, like that in the slot-sale/uniform-toll regime, the slot-distribution regime is efficient provided that the airport authority distributes the right number of slots, equal to the socially optimal total flight volume. The key difference between the regimes is that, under the slot-distribution regime, carriers understand that the total flight volume (and hence the level of airport congestion) is fixed by the number of distributed slots, while carriers under the price-based regimes expect total flights (and thus congestion) to respond to their flight-volume choices. The analysis shows that, by generating this different view, the slot-distribution regime yields an efficient outcome.

The slot-auction regime is the final focus of the analysis. The discussion assumes that slots are allocated via a uniform-price, multi-unit auction, and that the total volume of slots to be auctioned is announced by the airport authority. Since the total flight volume is then viewed as fixed by the carriers, as under the slot-distribution regime, the same efficiency result emerges. Thus, as long as the proper number of slots is auctioned, the social optimum is achieved.

The main lesson of the paper is that efficiency can be achieved through either of these quantity-based regimes, obviating the need for a potentially controversial system of differentiated congestion tolls. This outcome emerges despite the fact that both regimes involve prices (the slot-trading price or the auction price) that are uniform across carriers, as in the inefficient slot-sale/uniform-toll regime. The efficiency conclusion is welcome given that a slot-distribution system is currently in place at congested US airports.

The analysis relies on a highly stylized model, but the main conclusions should be robust to generalizations that offer greater realism. Adapting the framework of Brueckner and Van Dender (2008), the model portrays a congested airport served by two asymmetric carriers, with peak and off-peak periods collapsed into a single period that is always congested. In the analysis, carriers treat congestion tolls and prices as parametric and uninfluenced by their chosen flight volumes. This view is consistent with the usual approach to Pigouvian taxation, where the government, faced with a market distortion, computes the social optimum and levies taxes at a fixed rate to reach it. Economic agents, even if they otherwise enjoy market power, treat such Pigouvian taxes as parametric. This non-manipulative assumption is used, perhaps in a more controversial fashion, in the auction analysis, where carriers are assumed to make bids based on their true valuations of slots.

In an actual implementation of a congestion-management regime, different behavior could emerge. For example, implementation of tolls might rely on an iterative approach, where peak-period tolls are initially computed based on current traffic volumes and then adjusted downward as traffic shifts toward off-peak periods. The carriers, perceiving a connection between flight volumes and tolls, would then have an incentive to manipulate the system, acting on the basis of false, understated demands for airport usage with the goal of depressing the toll. Similarly, under a slot-sale regime, the airport authority might take a trial-and-error approach in setting the slot price, encouraging the airlines to view the price as endogenous and thus subject to manipulation. The same incentive might arise in the trading process under the slot-distribution regime, and strategic behavior may also be present in the slot-auction regime. If such manipulation occurs, then the results of the present analysis are not strictly relevant, calling into question their usefulness as a guide for public policy. If, however, the extent of manipulation is "small," then the results may still have some practical value. Whatever view of carrier behavior is correct, the urgency of the airport congestion problem makes any analysis of congestion-management policies, including one based on standard Pigouvian assumptions, a high-priority undertaking.[4]

It should be noted that the discussion in the paper parallels the analysis of Verhoef (2008) while using a more tractable set of assumptions.[5] The main simplification, which follows Brueckner and Van Dender (2008), is the assumption that carriers face perfectly elastic demands for air travel. This approach eliminates the flight-reducing distortion arising from the exercise of market power, allowing a sole focus on the distortion arising from the congestion externality, which tends to make flight volumes excessive. With market power eliminated and a constant-returns assumption modified, the analysis is able to derive parallel results that are simpler and more clearcut than those of Verhoef (2008). Once the analysis is complete, the two sets of results are compared.

A final issue concerns the connection between the present analysis and the classic paper of Weitzman (1974), which investigates price vs. quantity regulation. Weitzman's main goal is to compare these approaches when costs and benefits are uncertain, which sets his paper apart from the present analysis. But Weitzman begins his discussion by noting that price and quantity regulation are equivalent under certainty in a fairly general model, a conclusion that appears to contradict the current results. The present nonequivalence between the two types of regulation appears to arise because congestion causes firms to impose an externality on one another, rather than on other agents. While this difference is immaterial when carriers in the present model are symmetric, in which case the four regimes considered are equivalent, it matters under asymmetry.[6]

The plan of the paper is as follows. Section 2 characterizes the social optimum and the laissez-faire equilibrium. Section 3 analyzes the price-based regimes for congestion management, while section 4 analyzes the quantity-based regimes. Section 5 compares information requirements and revenue adequacy across the regimes, while section 6 contrasts the results with those of Verhoef (2008). Section 7 offers conclusions.

2. Basic Analysis

2.1. The setup

The analysis focuses on a single congested airport served by two airlines, denoted 1 and 2, who interact in Cournot fashion. Following Pels and Verhoef (2004), the model combines the peak and offpeak periods from Brueckner's (2002) analysis into a single congested period, an assumption that rules out reallocation of traffic between periods as a response to price-based congestion remedies. While the airlines experience common congestion, they are assumed to serve separate markets out of the congested airport, thus charging different fares. This assumption serves to generate asymmetry between the carriers, a crucial component of the ensuing analysis.

In order to maintain the simplest possible focus on the congestion phenomenon, the analysis suppresses the market-power element found in many previous models, including Verhoef (2008). In these models, a reduction in a carrier's flight volume reduces the level of airport congestion while also raising fares through a standard market-power effect. As a result, airline choices involve both the exploitation of market power and the desire to limit congestion. To focus solely on the congestion issue, market power is eliminated from the model by assuming that carriers face perfectly elastic demands for air travel.

Accordingly, it is assumed that the passengers of airlines 1 and 2 are willing to pay fixed "full prices" of p_1 and p_2 for travel in and out of the congested airport, reflecting horizontal demand curves in the two markets. Airline 1 is assumed to serve the higher-price market, so that $p_1 \geq p_2$. Since passengers dislike airport congestion, which imposes additional time costs, the actual fares that the airlines charge must be discounted below these full prices.[7] To derive the discount, let f_1 and f_2 denote flight volumes for the two carriers, and let $t(f_1 + f_2)$ denote the extra time cost per passenger due to congestion and the resulting delays, a cost that depends on total flights at the congested airport. The function t satisfies $t(0) = 0$, $t' \geq 0$ (equality may hold over a range of low traffic levels), and $t'' \geq 0$ over the function's positive range. Taking account of passenger congestion cost, airline 1 is then able to charge a fare equal to $p_1 - t(f_1 + f_2)$, with airline 2 charging $p_2 - t(f_1 + f_2)$. When congestion cost is added to these fares, the resulting full prices are p_1 and p_2.

Letting s denoted the fixed seat capacity of an aircraft and assuming that all seats are filled, the total number of seats sold by carrier i is $sf_i,\ i = 1, 2$. For simplicity, s is normalized to unity, so that revenue for airline i is

$$[p_i - t(f_1 + f_2)]f_i, \quad i = 1, 2. \tag{1}$$

Note that, with the normalization of s, p_i becomes the full price per flight.

In addition to raising passenger time cost, airport congestion raises an airline's operating cost by $g(f_1 + f_2)$ for each flight. Like $t(\cdot)$, the function g satisfies $g(0) = 0$ and $g', g'' \geq 0$. An airline also incurs operating costs per flight that depend on its own flight volume but are unrelated to airport-level congestion. These costs, given by $\tau(f_1)$ and $\tau(f_2)$, are assumed to increase with a carrier's flight volume, reflecting decreasing returns to scale ($\tau' > 0,\ \tau'' \geq 0$ hold along with $\tau(0) > 0$). More generally, it could be assumed that returns to scale are initially constant (so that $\tau' = 0$ holds at low f values), with a range of decreasing returns, the one relevant for the analysis, encountered at higher flight volumes.

While the analogous τ function in Brueckner and Van Dender (2008), Verhoef and Pels (2004), and Verhoef (2008) is constant, reflecting constant returns to scale (a constant cost per flight), the assumption of decreasing returns is needed to generate sensible results in the presence of perfectly elastic demands when full prices differ across carriers. If the cost per flight were instead constant, the social optimum would involve a degenerate solution in which only the carrier serving the high-price market operates. Decreasing returns may, in any case, be a plausible assumption for a carrier operating at a congested airport. While intense use of runways and other airport infrastructure used jointly by both carriers is the source of airport congestion (an effect captured by the $t(\cdot)$ and $g(\cdot)$ functions), a busy airport will also have intense usage of carrier-specific facilities such as gates and baggage systems (captured by $\tau(\cdot)$). Such usage may well be subject to decreasing returns at high levels.

Using the above functions, total costs for the two airlines are given by $[\tau(f_1)+g(f_1+f_2)]f_1$ and $[\tau(f_2)+g(f_1+f_2)]f_2$. Airline 1's profits can then be written

$$\pi_1 \;=\; [p_1 - t(f_1+f_2)]f_1 \;-\; [\tau(f_1)+g(f_1+f_2)]f_1 \tag{2}$$

and rewritten as

$$\pi_1 \;=\; [p_1 - \tau(f_1)]f_1 \;-\; c(f_1+f_2)f_1 \tag{3}$$

where

$$c(f_1+f_2) \;\equiv\; t(f_1+f_2) \;+\; g(f_1+f_2) \tag{4}$$

gives passenger plus airline congestion cost per flight (note that t is multiplied by $s = 1$). Given the properties of the t and g functions, $c(0) = 0$ holds and $c' \geq 0,\ c'' \geq 0$. Analogously, carrier 2's profit is given by

$$\pi_2 \;=\; [p_2 - \tau(f_2)]f_2 \;-\; c(f_1+f_2)f_2. \tag{5}$$

In the analysis, it is assumed that the positive range of the $c(\cdot)$ function is relevant, with flight volumes being large enough to generate airport congestion.

2.2. The social optimum

Consider first the social optimum. With perfectly elastic demands, consumer surplus is zero, which means that the social optimum maximizes the combined profits of the carriers. After adding the profit expressions in (3) and (5), differentiation with respect to f_1 and f_2 yields the first-order conditions

$$p_1 \;-\; \tau(f_1) \;-\; f_1\tau'(f_1) \;-\; c(f_1+f_2) \;-\; (f_1+f_2)c'(f_1+f_2) \;=\; 0 \tag{6}$$

$$p_2 \;-\; \tau(f_2) \;-\; f_2\tau'(f_2) \;-\; c(f_1+f_2) \;-\; (f_1+f_2)c'(f_1+f_2) \;=\; 0. \tag{7}$$

Computation of the Hessian determinant of total profit shows that satisfaction of the second-order condition is not guaranteed and must be assumed.[8]

From (6) and (7), a carrier's flight volume is optimal when the full price p_i per flight equals the marginal social cost of a flight, which is given by $\tau + f_i\tau' + c$ plus the marginal congestion damage from an extra flight. This latter cost is computed taking into account the congestion cost imposed on both carriers when an extra flight is operated. In particular, when f_1 is increased, passenger plus airline congestion costs for airline 1 (given by f_1c) increases by $c + f_1c'$, while these costs for airline 2 (given by cf_2) increase by f_2c'. The sum of the terms involving c', equal to $(f_1 + f_2)c'(f_1 + f_2) \equiv$ MCD, gives the marginal congestion damage from an extra flight.

Inspection of (6) and (7) shows that airline 1, which serves the high-price market, operates more flights than airline 2 at the optimum. Denoting the social optimally values with an asterisk, $f_1^* > f_2^*$ then holds. This conclusion follows because $\tau(f) + f\tau'(f)$ is increasing in f under the maintained assumptions, implying that $f_1 > f_2$ must hold for both (6) and (7) to be satisfied given $p_1 > p_2$. For future reference, let $q = f_1 + f_2$ denote the total flight volume, and let $q^* = f_1^* + f_2^*$ denote its optimal value.

2.3. The laissez-faire equilibrium

Consider next the laissez-faire equilibrium. Each airline, behaving in Cournot fashion, maximizes profit viewing the other airline's flight volume as fixed, yielding the first-order conditions

$$p_1 - \tau(f_1) - f_1\tau'(f_1) - c(f_1 + f_2) - f_1c'(f_1 + f_2) = 0 \tag{8}$$

$$p_2 - \tau(f_2) - f_2\tau'(f_2) - c(f_1 + f_2) - f_2c'(f_1 + f_2) = 0. \tag{9}$$

The carriers' second-order conditions are satisfied, and it is easily seen that (8) and (9) generate downward-sloping reaction functions. Airline 1's reaction function has a slope between -1 and 0 (f_1 is on the vertical axis) and is thus flatter than 2's function, which has a slope less than -1. As a result, the laissez-faire equilibrium is unique and stable. As in the case of the social optimum, $f_1 > f_2$ holds in the equilibrium.[9]

Focusing on airline 1, the difference between conditions (8) and (6) is the absence of f_2c' in the last term. This absence shows that, in scheduling an extra flight, airline 1 takes into account the additional congestion costs imposed on its own flights (f_1c'), ignoring the congestion imposed on airline 2 (f_2c'). Thus, while the airline internalizes some of the congestion from an extra flight, it ignores the impact on the other carrier. Airline 2 behaves in analogous fashion.

With both carriers ignoring a portion of the congestion they create, the total flight volume in the laissez-faire equilibrium is excessive relative to the social optimum. This conclusion follows from noting that the locii generated by (8) and (9) in (f_2, f_1) space, whose intersection determines the optimum, are both lower than the reaction functions generated by (6) and (7), a consequence of the larger multiplicative factor in the last term. As result, the socially optimal point must lie below both reaction functions. This conclusion in turn implies that the optimum lies below the line where $f_1 + f_2$ is constant at the equilibrium level, a line that lies between the reaction functions on either side of the equilibrium (it passes through the equilibrium point and has a slope of -1). However, even though the socially optimal point must lie below this line (yielding a smaller flight total), both individual flight volumes need not be smaller than the equilibrium levels, as would occur in the symmetric case. For example, the socially optimal point could lie to the northwest of the equilibrium (making f_1 larger and f_2 smaller than the equilibrium levels). A location to the southwest of the equilibrium is ensured if the degree of asymmetry between the carriers (the difference between p_1 and p_2) is sufficiently small.[10] Summarizing yields

Proposition 1. *The laissez-faire equilibrium has a larger total flight volume than the social optimum. If the degree of carrier asymmetry is small, flight volumes for the two carriers are individually larger than the optimal levels.*

3. Price-Based Approaches to Congestion Management

3.1. The differentiated-toll regime

The divergence between the laissez-faire equilibrium and the optimum can be eliminated by imposition of differentiated congestion tolls, the first of two price-based approaches to be considered. The toll per flight is equal to that portion of the congestion damage from an extra flight not internalized by a carrier. The congestion tolls thus differ across carriers, being equal to $T_1 = f_2 c'$ for carrier 1 and $T_2 = f_1 c'$ for carrier 2, with both expressions evaluated at the optimum. After imposition of these tolls, $f_1 T_1$ and $f_2 T_2$ are subtracted from the profit expressions in (3) and (5). Assuming that the carriers view the tolls as parametric, as discussed in the introduction, T_1 and T_2 are then subtracted in the airline first-order conditions (8) and (9), and the solutions to these modified conditions coincide with the social optimum.

More precisely, the differentiated tolls are given by

$$T_1 \;=\; f_2^* c'(f_1^* + f_2^*) \;=\; (1-\phi)(f_1^* + f_2^*)c'(f_1^* + f_2^*) \;=\; (1-\phi)\mathrm{MCD}^* \tag{10}$$

$$T_2 \;=\; f_1^* c'(f_1^* + f_2^*) \;=\; \phi(f_1^* + f_2^*)c'(f_1^* + f_2^*) \;=\; \phi\,\mathrm{MCD}^* \tag{11}$$

where

$$\phi \;\equiv\; \frac{f_1^*}{f_1^* + f_2^*} \;>\; \frac{1}{2} \tag{12}$$

is airline 1's airport flight share and MCD^* is marginal congestion damage, both evaluated at the optimum. A carrier's toll thus equals one minus its flight share times marginal congestion damage. The key feature of the resulting toll structure is summarized in

Proposition 2. *With its flight share greater than* $1/2$, *the large carrier pays a lower congestion toll than the small carrier under the differentiated-toll regime.*

In other words, using (10) and (11), $\phi < 1/2$ implies $T_1 = (1-\phi)\mathrm{MCD}^* < \phi\mathrm{MCD}^* = T_2$. The intuitive explanation is that, since airline 1 has more flights, it internalizes more of the congestion damage from its operation of an extra flight than does airline 2. Since less congestion damage then goes uninternalized, airline 1 can be charged a lower toll. While this toll pattern is required for efficiency, the inverse association between a carrier's size and the toll it pays would be controversial in practice, generating concerns about equity across airlines and potential political opposition from smaller carriers.[11] This point, recognized by Brueckner (2002), is stressed by Morrison and Winston (2007).

3.2. The slot-sale/uniform-toll regime

The second-priced based approach to congestion management replaces the differentiated-toll regime's unequal charges with a levy per flight that is uniform across carriers. This approach has two possible interpretations. On the one hand, it could correspond to a congestion-toll regime subject to a second-best constraint of toll uniformity, a response to the political pressure noted above. Alternatively, the approach could correspond to a regime under which the airport authority sells airport access via a slot system, with airlines required to purchase a slot for each flight they operate. Under such a slot-sale regime, the slot price would necessarily be uniform across carriers, matching the structure of the uniform-toll regime. Carriers would be allowed to purchase as many slots as they like at the announced price.

Let z denote the charge per flight under this slot-sale/uniform-toll regime. Then, since the terms $f_1 z$ and $f_2 z$ are subtracted from the profit expressions in (3) and (5), the regime's equilibrium is characterized (assuming z is viewed as parametric) by the following first-order conditions:

$$p_1 - \tau(f_1) - f_1\tau'(f_1) - c(f_1+f_2) - f_1 c'(f_1+f_2) = z \tag{13}$$

$$p_2 - \tau(f_2) - f_2\tau'(f_2) - c(f_1+f_2) - f_2 c'(f_1+f_2) = z. \tag{14}$$

The conditions (13)–(14) generate solutions for f_1, f_2, and $q = f_1 + f_2$ conditional on z, denoted $f_1(z)$, $f_2(z)$, and $q(z)$. Taking account of these dependencies, the airport authority then chooses an optimal value for the charge per flight, selecting z to maximize welfare. Substituting $f_1(z)$, $f_2(z)$ and $q(z)$ into the total profit (welfare) expression, equal to $\pi_1 + \pi_2$ from (3) and (5), the first-order condition for choice of z is

$$[p_1 - \tau(f_1) - f_1\tau'(f_1)]f_1'(z) + [p_2 - \tau(f_2) - f_2\tau'(f_2)]f_2'(z) = [c(q) + qc'(q)]q'(z), \tag{15}$$

Note that the z-derivatives f_1', f_2', and q' in (15) are computed by total differentiation of (13) and (14), and that z arguments are selectively suppressed in the equation to save space. Let the optimal value for z, which solves (15), be denoted $\widehat{z}$ and let the corresponding optimal values for f_1, f_2 and q be denoted $\widehat{f}_1 = f_1(\widehat{z})$, $\widehat{f}_2 = f_2(\widehat{z})$, and $\widehat{q} = q(\widehat{z})$.

Given the complexity of the expressions that emerge for f_1', f_2', and q', little can be said about magnitude of $\widehat{z}$ from (17) without imposing further assumptions, as is done below. However, even without knowledge of z's optimal value, it is obvious that the slot-sale/uniform-toll regime is inefficient, as can be seen by contrasting (13) and (14) with the analogous conditions for the differentiated-toll regime. Since its tolls have different magnitudes, whereas carriers pay a common charge under the slot-sale/uniform-toll regime, the latter regime will not be able to generate the social optimum. By not taking into account airline 1's greater internalization of congestion, the regime will tend to penalize airline 1 too much and airline 2 not enough for the congestion they create. Thus, the regime will tend to make the flight volume too small for the large carrier and too large for the small carrier.

Whether these tendencies end up making $\widehat{f}_1$ smaller and $\widehat{f}_2$ larger than the first-best optimal values f_1^* and f_2^* depends on the relationship between $\widehat{q}$, the optimal total flight volume under the slot-sale/uniform-toll regime, and q^*, the socially optimal volume. While the relationship between $\widehat{q}$ and q^* is ambiguous in general, if $\widehat{q}$ happens to equal q^*, then the above tendencies will indeed make $\widehat{f}_1$ too small and $\widehat{f}_2$ too large under the slot-sale regime. However, if $\widehat{q} \geq q^*$, then airline 2's insufficient congestion penalty combined with an excessive total number of flights will again make $\widehat{f}_2$ too large. But with airline 1's penalty too severe, $\widehat{f}_1$ could be either larger or smaller than f_1^*. Conversely, if $\widehat{q} \leq q^*$, then airline 1's excessive penalty combined with an insufficient total flight volume will make $\widehat{f}_1$ too small, with no implication emerging for $\widehat{f}_2$. A formal statement, including some sharper results for a special case, is as follows:

Proposition 3. *The slot-sale/uniform-toll regime is inefficient, being unable to generate socially optimal flight volumes for the individual carriers. Specifically,*

> *(i) If the regime's optimal total flight volume $\widehat{q}$ is less than or equal to the socially optimal total flight volume q^*, then the large carrier operates too few flights ($\widehat{f}_1 < f_1^*$).*
>
> *(ii) If $\widehat{q} \geq q^*$ holds, then the small carrier operates too many flights ($\widehat{f}_2 > f_2^*$).*
>
> *(iii) Combining (i) and (ii), if $\widehat{q} = q^*$, then $f_1^* > \widehat{f}_1 > \widehat{f}_2 > f_2^*$.*

(iv) If the $\tau(\cdot)$ and $c(\cdot)$ functions are linear, then $\widehat{q} = q^*$ holds, so that the inequalities in case (iii) obtain.

(v) With linearity, the regime's optimal charge is given by $\widehat{z} = \frac{1}{2}\mathrm{MCD}^*$.

Parts (*i*)–(*iii*) of the proposition, which are proved in the appendix, formalize the above intuition by showing that at least one of the inequalities $\widehat{f}_1 < f_1^*$ and $\widehat{f}_2 > f_2^*$ must hold, with both holding when $\widehat{q} = q^*$. Part (*iv*) shows that the latter case applies when the model's cost functions are linear, yielding $f_1^* > \widehat{f}_1 > \widehat{f}_2 > f_2^*$ and showing the slot-sale/uniform toll regime inefficiently narrows the dispersion of the individual carrier flight volumes across the high and low-price markets.

Explicit solutions for the linear case are presented in the appendix. Although these solutions can be used to directly establish part (*iv*) of the proposition, a less-algebraic approach is more illuminating, as follows. Suppose that the cost functions are linear, with $\tau(f_i) \equiv \theta + \alpha f_i$ and $c(f_1 + f_2) = c(q) \equiv \beta q$. In this case, it can be shown that $f_1'(z) = f_2'(z) = -1/(2\alpha + 3\beta) \equiv -\delta$, with $q'(z) = -2\delta$. Then dividing through by $q'(z)$ in (15), the equation reduces to

$$[p_1 - \tau(f_1) - f_1\tau'(f_1)]/2 + [p_2 - \tau(f_2) - f_2\tau'(f_2)]/2 = c(q) + qc'(q). \tag{16}$$

Substituting the linear functions, (16) becomes

$$\frac{1}{2}(p_1 + p_2) - (\theta + \alpha q) - 2\beta q = 0, \tag{17}$$

an equation that directly determines $\widehat{q}$, the optimal total flight volume under the slot-sale/uniform-toll regime. However, after imposing linearity in the social optimality conditions (6) and (7), adding the equations, and dividing by 2, the resulting equation (which determines q^*) is the same as (17). Thus, $\widehat{q} = q^*$ holds, establishing part (*iv*) of the proposition.

Part (*v*) of the proposition shows that the optimal charge per flight $\widehat{z}$ under the slot-sale/uniform-toll regime is the average of the differentiated tolls in (10) and (11). This result is easily established using (17) along with the first-order conditions in (13) and (14).[12]

As a final point, it should be noted that inefficiency of the slot-sale/uniform-toll regime vanishes when the carriers are symmetric (when $p_1 = p_2$). In this case, the differentiated-toll regime itself generates uniform charges, which means that explicit imposition of this requirement has no welfare effect.[13]

3.3. Numerical examples

Narrowing of the dispersion of flight volumes across markets under the slot-sale/uniform-toll regime is an intuitive result, and this section presents numerical examples to determine whether the result is robust to cost nonlinearity. In the examples, $\tau(f_i) \equiv \theta + 4f_i^4$ and $c(f_1 + f_2) \equiv 4(f_1 + f_2)^4$. Variation in the multiplicative constants in these functions (which are arbitrarily set equal to the quartic exponents) has little qualitative effect on the solutions. However, use of quadratic rather than quartic functions does not generate enough decreasing returns to sustain an interior social optimum (with both flight volumes positive) unless the multiplicative constants are very large.

Table 1 shows numerical results in two cases with different values of $p_i - \theta$. In the first case, $p_1 - \theta = 9$ and $p_2 - \theta = 5$, while in the second case $p_1 - \theta$ equals 7 instead of 9. Starting with a comparison between q values, the upper panel shows that $\widehat{q} < q^*$ holds when $p_1 - \theta$ is high, with the reverse inequality holding in the second panel where $p_1 - \theta$ is lower (the values are close in each case, however). This reversal shows that the total flight volume under the slot-sale/uniform-toll regime may be either larger or smaller than the socially optimal flight volume, with the linear case representing a particular instance where the two values are equal.

In each panel, comparison of the flight volumes reveals the narrowing of the flight dispersion seen in the linear case. In both cases, the inequalities $f_1^* > \widehat{f}_1 > \widehat{f}_2 > f_2^*$ hold even though $\widehat{q}$ differs from q^*. This tendency may therefore be fairly general, holding under a variety of functional specifications.

Table 1 also shows the laissez-faire equilibrium, illustrating its excessive total flight volume (Proposition 1) and lower welfare level. Note that the slot-sale/uniform-toll regime achieves around 70 to 75 percent of the welfare gain generated by the differentiated-toll regime.[14] Observe also that, in both cases, the socially optimal f_1 is larger than the laissez-faire equilibrium level, with the opposite relationship holding for f_2. Thus, the social optimum lies to the northwest of the equilibrium point, a possibility recognized in the earlier discussion. Additional computations confirm that a more-natural southwesterly location for the optimum (with both flight volumes smaller than the equilibrium levels) arises when the difference between p_1 and p_2 is sufficiently small.[15]

4. Quantity-Based Approaches to Congestion Management

4.1. The slot-distribution regime

Given that differentiated tolls are likely to be controversial and that the slot-sale/uniform-toll regime is inefficient, the analysis now turns to the analysis of quantity-based approaches. Under such an approach, the airport authority allocates a fixed number of slots to the airlines, with the allocation achieved either by distributing the slots directly to the carriers, with trading then possible, or through a slot auction.

Consider first the slot-distribution regime, where the airport authority distributes slots free of charge to the carriers, who then trade at an established price in order to adjust flight volumes. The slot-trading price could be set by a central clearing house to balance slot purchases and sales. The ensuing analysis of this regime follows Verhoef (2008) while using the current, more-tractable set of assumptions.

Let n_1 and n_2 denote the numbers of slots allocated to the two carriers, with $n_1 + n_2 = n$. Letting w denote the price at which the carriers trade slots, profits for carriers 1 and 2 are then equal to

$$\pi_1 = [p_1 - \tau(f_1)]f_1 - c(n)f_1 - w(f_1 - n_1) \tag{18}$$

$$\pi_2 = [p_2 - \tau(f_2)]f_2 - c(n)f_2 - w(f_2 - n_2). \tag{19}$$

Several features of these expressions deserve note. First, if $f_i > n_i$, then carrier i is a buyer of slots and makes an outlay of $w(f_i - n_i) > 0$, while $f_i < n_i$ means that the carrier is a slot seller, earning revenues equal to the negative of the previous expression. Second, because a carrier understands that total flight volume remains constant when it trades a slot with the other carrier, the $f_1 + f_2$ argument of the congestion-cost function c remains constant at n, the fixed slot total, a crucial difference relative to the price-based approaches.

Assuming no manipulation of the trading process, the equilibrium conditions for the slot-distribution regime are

$$p_1 \; - \; \tau(f_1) \; - \; f_1\tau'(f_1) \; - \; c(n) \;\; = \;\; w \tag{20}$$

$$p_2 \; - \; \tau(f_2) \; - \; f_2\tau'(f_2) \; - \; c(n) \;\; = \;\; w \tag{21}$$

$$f_1 \; + \; f_2 \;\; = \;\; n. \tag{22}$$

The first two conditions determine a carrier's desired flight volume (and thus the magnitude of its slot purchase or sale) when faced with the price w, while (22) says that the slot trades balance. This requirement implies $f_1 - n_1 = -(f_2 - n_2)$, which reduces to (22).

As in the slot-sale conditions (13)–(14), a uniform price (here w) appears on the RHS of equations (20)–(21). But the latter conditions are missing the previous terms involving marginal congestion cost (the f_ic' terms), which differ across carriers. This absence reflects the carriers' understanding that congestion remains fixed when slots are traded. The upshot is that the expressions $p_i - \tau(f_i) - f_i\tau'(f_i),\ i = 1, 2$, are set equal to a common value $(w - c(n))$ under the slot-distribution regime. This same property holds at the social optimum, where the above expressions are set equal to the common value $c(q) + qc'(q)$ (see (6) and (7)). As a result, the slot-distribution regime is potentially efficient. By contrast, because the carrier-specific congestion terms f_ic' are present under the slot-sale/uniform-toll regime (see (13) and (14)), the $p_i - \tau(f_i) - f_i\tau'(f_i)$ expressions assume different values, accounting for the inefficiency of that regime.

To generate the social optimum under the slot-distribution regime, the airport authority sets n, the number of slots distributed, equal to q^*, the socially optimal flight volume, and distributes the slots in any arbitrary fashion to the carriers. Then, it is easily seen that the solution to (20)–(22) has $f_1 = f_1^*$ and $f_2 = f_2^*$, with $w = w^* = q^*c'(q^*) = \text{MCD}^*$. This conclusion follows because (20) and (21) evaluated at these values are identical to the optimality conditions (6) and (7) evaluated at the social optimum. Since the latter conditions hold by definition, it follows that the above values represent a solution to the equilibrium conditions (20)–(22). Note that the equilibrium slot price of MCD^* is double the previous optimal charge under the slot-sale/uniform-toll regime ($\frac{1}{2}\mathrm{MCD}^*$). Summarizing yields

Proposition 4. *The slot-distribution regime is efficient. To reach the social optimum, the airport authority distributes q^* slots to the carriers and allows them to trade, leading to an equilibrium slot selling price of* MCD^*.

Note that the efficiency of the slot-distribution regime is independent of the pattern of distribution of slots to the carriers. That pattern affects equilibrium profit levels but not the chosen flight volumes.[16]

The essential behavioral difference between this setup and the price-based regimes is the carriers' recognition that, *regardless of their choices, the total flight volume is fixed at a value equal to the total number of slots distributed.* As seen above, this recognition is the key to the efficiency of the slot-distribution regime. Note that it allows efficiency to be achieved even though the regime involves a price w that is common across carriers, as in the inefficient slot-sale/uniform-toll regime.[17]

One could ask whether this same recognition could be exploited under a slot-sale regime. The airport authority could announce its intention to sell n slots at a price z, which would lead carriers to treat total congestion as fixed and thus apparently generate an efficient outcome. However, such an approach gives inconsistent assurances to the carriers. On the one hand, carriers would expect to be able to purchase as many slots as they wish at the announced price. However, the existence of a cap on aggregate slot purchases may make the carriers doubt their ability to exercise this right. For a carrier to believe that it has an unfettered ability to make slot purchases, it would have to believe that the other carrier would reduce its own slot purchases in a one-for-one fashion. For these reasons, a slot-sale regime with an announced slot total seems logically untenable and thus not worthy of consideration.

4.2. The slot-auction regime

Consider now an auction regime as a means of allocating slots. Since a large number slots must be allocated, a multi-unit auction is the appropriate setup, and two variants exist. A "pay-as-bid" multi-unit auction is analogous to a standard first-price auction, but for simplicity, the airport authority is assumed to use the "uniform-price" variant, which is analogous to a second-price auction. Under a uniform-price auction, carriers offer unit-specific bids for slots, specifying different prices for first, second, and subsequent slots requested. The bidders with the highest (unit-specific) bids win slots, but they pay a common price equal to the highest bid not accepted.[18] Since, in the event of winning at least one slot, a carrier's bids on later units may set the price that is paid (potentially being the highest unaccepted bid), an incentive exists to understate the valuations of these later units (see Ausubel and Cramton (2008)).[19] Despite this conclusion, the maintained assumption of non-manipulative behavior is applied again in the ensuing analysis, so that carriers are assumed to bid on the basis of their true valuations for slots. Even though this assumption is at variance with the standard approach to auction analysis, it may yield a rough approximation to the actual outcome under a uniform-price auction while facilitating a simple discussion.

Without manipulative behavior, carrier i reports to the auctioneer a bid function giving its true willingness-to-pay for incremental slots, equal to $p_i - \tau(f_i) - f_i\tau'(f_i) - c(n)$, where n now represents the number of slots to be auctioned. This function gives the marginal profit increase from the carrier's f_i^{th} slot, thus representing its willingness-to-pay for that slot. Note that, since the auction size n is announced in advance by the auctioneer, a carrier's bid function assumes that the total flight volume is fixed at n (yielding a fixed level of congestion), as in the slot-distribution regime.

Based on these bid functions, the auctioneer then sets a price y such that carriers bidding at least y for incremental slots receive them, with a total of n slots being allocated.[20] It is easy to see that the resulting equilibrium conditions are identical to those from the slot-distribution regime, being given by (20)–(22) with y in place of w. In particular, given that the bid functions are downward sloping, the new versions of (20)–(21) ensure that a carrier receives all the incremental slots for which its bid exceeds or equals y, while (22) ensures that the resulting slot total equals n. Given this equivalence, the previous argument establishing efficiency of the slot-distribution regime applies as well to the slot-auction regime under the present assumptions.[21] Because of the equivalence of the two regimes, the price emerging from the auction of q^* slots equals the realized trading price under the distribution regime, so that $y^* = w^* = \text{MCD}^*$. Summarizing yields[22]

Proposition 5. *With non-strategic behavior, a uniform-price, multi-unit slot auction is efficient. To reach the optimum, the airport authority auctions q^* slots, leading to an equilibrium price of* MCD^*.

5. Other Comparisons Across Regimes

5.1. Information requirements

Having so far considered four different price and quantity-based approaches to congestion management, it is useful to ask whether their information requirements are also different. The answer is no: the airport authority needs the same information to implement the four different approaches, despite their different forms. In particular, the authority needs all the information required to compute the social optimum: the levels of p_1 and p_2 and the form of the $\tau(\cdot)$ and $c(\cdot)$ functions. This information is required to compute the optimal tolls under the differentiated-toll regime, the charge per flight under the slot-sale/uniform toll regime, and the number of slots to distribute or auction under the two quantity-based regimes. Therefore, despite the apparent simplicity of quantity-based regimes, they offer no information advantage over the price-based alternatives.[23]

5.2. Revenues relative to airport cost

It is also useful to compare the revenue raised by the different regimes, and to ask whether it covers the cost of airport, which has so far been ignored. As noted by Brueckner (2002), the standard self-financing theorem, which says that optimal tolls exactly pay for an optimal-size road when road capacity is produced with constant returns, fails in the airport context. To see this conclusion, let K denote the physical size of the airport, which is an argument (previously suppressed) of the congestion-cost function. Suppose that, as usual, the rewritten function $c(f_1+f_2,K)$ satisfies $c_K<0$ and exhibits zero-degree homogeneity, and that physical capacity is produced with constant returns, yielding an airport cost of μK. Then, the condition for optimal choice of K, given by $-(f_1+f_2)c_K=\mu$, must hold along with the optimality conditions (6) and (7). But c's zero-degree homogeneity implies $c_K=-(f_1+f_2)c'/K$, so that the above optimality condition becomes $(f_1+f_2)^2c'=\mu K$. Substituting MCD and evaluating at the social optimum, the condition can then be written $(f_1^*+f_2^*)\text{MCD}^*=\mu K^*$ But, recalling (10) and (11), revenue at the optimum under the differentiated-toll regime equals $[f_1^*(1-\phi)+f_2^*\phi]\text{MCD}^*<(f_1^*+f_2^*)\text{MCD}^*$, and this inequality implies that toll revenue falls short of airport cost.

Recalling that $\widehat{z}=\frac{1}{2}\text{MCD}^*$ holds in the linear case under the slot-sale/uniform-toll regime, the same conclusion holds under that regime. In other words, revenue equals $(q^*/2)\text{MCD}^*$, while airport cost at the optimum equals $\mu K^*=q^*\text{MCD}^*$ from above, so that revenue falls short of cost. Under the slot-distribution regime, the trading price equals $w^*=\text{MCD}^*$, but since all trading revenue is retained by the airlines, the airport suffers a 100 percent shortfall relative to cost under this regime. However, since the auction price y^* also equals MCD^* while all the revenue accrues to the airport, it follows that costs are exactly covered under the slot-auction regime. In other words, $q^*y^*=q^*\text{MCD}^*=\mu K^*$. Summarizing yields[24]

Proposition 6. *While slot-auction revenue exactly covers airport cost at the optimum under standard assumptions, all the other regimes generate revenue shortfalls.*

6. The Effect of Market Power

In a closely related paper, Verhoef (2008) analyzes slot-based regimes when demand is imperfectly elastic, allowing carriers to exercise market power. Uninternalized congestion again tends to make flight volumes excessive, but carriers have a new incentive to limit their flight volumes in order to raise fares. The net effect of these two distortions is ambiguous, so that flight volumes in the laissez-faire equilibrium could either be too large or too small. Since Verhoef (2008) assumes that, under the toll regime, the two distortions must be addressed using a single toll instrument, the possibility of insufficient flight volumes means that optimal tolls could be negative.[25]

In Verhoef's model, two carriers serve the same market (facing a downward-sloping, linear demand curve), and asymmetry between them is generated by a difference in costs. The $\tau(\cdot)$ function is constant in his model (yielding constant returns), the $c(\cdot)$ function is linear, and the functions differ across carriers in a way that gives one carrier lower costs at any flight volume. Given the assumed cost structure and service to a single market, the high-cost carrier does not operate at the social optimum, an outcome that is supported by levying a positive toll on that carrier. By contrast, the low-cost carrier faces a negative toll (receives a subsidy), which raises its flight volume and corrects the market-power distortion. Note that uninternalized congestion vanishes with the absence of the other carrier, making this the only distortion present.

As in the present analysis, a slot-sale/uniform-toll regime eliminates the differentiation of tolls required to support the social optimum. One consequence is that the high-cost carrier may operate under that regime rather than being forced out of business, an undesirable second-best outcome. This pattern matches one of the conclusions of Proposition 3 (too large a flight volume for the smaller carrier), although Verhoef is not able to state an equally clearcut result given the greater complexity of his model. Nevertheless, the two papers offer a similar verdict by highlighting the inefficiency of a structure that imposes a common charge on asymmetric carriers.

Verhoef also analyzes a slot-distribution regime and shows that the trading equilibrium efficiently removes the high-cost carrier from the market. However, unless the airport authority can compel the low-cost carrier to use all of the slots it ends up holding, the social optimum cannot be generated. In other words, the authority can issue total slots equal to the socially optimal flight volume, but the low-cost carrier, even though it ends up holding this many slots, can choose instead the smaller flight volume corresponding the best monopoly output, leaving some slots unused.

This discussion shows that Verhoef's analysis and results parallel those of the present paper. However, the greater tractability of the present model, a consequence of a different set of simplifying assumptions, yields results that are more transparent and clearcut.

7. Conclusion

This paper has analyzed price and quantity-based approaches to management of airport congestion, using a model where airlines are asymmetric and internalize congestion. Under these circumstances, optimal congestion tolls are differentiated across carriers, and a uniformity requirement on airport charges (as occurs when slots are sold or tolls are uniform) distorts carrier flight choices. Flight volumes tend to be too low for large carriers and too high for small carriers. But quantity-based regimes, where the airport authority allocates a fixed number of slots via free distribution or an auction, lead carriers to treat total flight volume (and thus congestion) as fixed, and this difference generates an efficient outcome as long as the number of slots is optimally chosen.

All of these conclusions presume the absence of manipulative or strategic behavior on the part of the carriers, a view that may be inaccurate. Nevertheless, the results provide a benchmark for policy evaluation that may give an approximation to actual outcomes under more-sophisticated behavior by the carriers. Other aspects of the model are also highly stylized, but the principles underlying the results would also emerge in more detailed and realistic frameworks.

A key lesson of the analysis is that a slot-distribution regime, where slots are distributed to the carriers and then traded through a clearing house, is equivalent to an efficient regime of differentiated congestion tolls. Since such a toll regime is bound to be controversial given its inverse relation between tolls and carrier size, the analysis generates a clear presumption in favor of the equivalent slot-distribution regime. This conclusion is welcome, given that the slot-distribution approach is already in place at slot-constrained US airports, although trading occurs at low volumes.[26] To foster more-active trading, the current bilateral system could be replaced by a central clearing house, and airlines could be given clearer property rights, replacing the current tenuous arrangement in which slots are ultimately the property of the FAA (Whalen et al. (2007) stress this point).

Although the analysis shows that a slot-auction regime can also achieve efficiency, free distribution of slots could be preferable given the beleaguered airline industry's strenuous opposition to new, cost-increasing charges. A downside, however, is that the slot-distribution regime generates no revenue, which then must come from other charges such as the weight-based landing fees that carriers currently pay. To ensure that free distribution works well, hoarding of unused slots, meant to deny airport access to a carrier's competitors, must be prevented by "use-or-lose" requirements, which already exist. It might appear that new entrants (who hold no slots) are disadvantaged under this system, but their status is no different from that of an incumbent carrier seeking to increase its flight volume, which must purchase new slots to do so.[27]

It should be noted that, because this endorsement of the slot-distribution and auction regimes is based on an exceedingly simple model, it should be viewed with some caution. Daniel and Harback (2008c), for example, argue that the 30-minute time intervals used in typical slot regimes are too crude for efficient management of airport congestion. These authors argue that airport access must be controlled instead on a minute-by-minute basis to achieve an optimal outcome. For example, a 10-minute widening of the arrival intervals for flight banks at congested hubs can greatly reduce congestion, but such fine-tuning is hard to achieve under a slot system. In Daniel and Harback's view, congestion tolls with a finely varying time structure are required to generate an optimal traffic pattern. To address this limitation of the analysis, peak and off-peak periods could be added to the model, although an appropriate treatment may require a realistic simulation framework like the one used in Daniel's work.

While the paper also abstracts from network issues, Brueckner (2005) analyzes congestion pricing in a simple model where two carriers operate partially ovelapping hub-and-spoke networks out of different, concentrated hubs. The current simplifications (perfectly elastic demand, no peak/off-peak distinction) could be imposed in that model, and all of the present results would emerge. Thus, the lessons of the paper appear fully robust to a network generalization.

Another limitation of the analysis is its assumption that the airport authority has the information needed to implement the various regimes, including information on demand, operating costs, and both passenger and airline congestion costs. While good estimates of the relevant parameters are available, as seen in the carefully calibrated numerical models of Daniel (1995) and Daniel and Harback (2008a), it seems likely that actual policy implementation would involve much less precision, with the price-based regimes possibly relying on iterative or trial-and-error procedures that invite manipulation.[28] Such implementation problems were illustrated by the recent controversy over flight caps at the New York airports, where the carriers argued (possibly with some justification) that FAA caps were much too tight.[29] While these obstacles may make achievement of the optimum difficult, conceptual clarity regarding effects of different approaches to airport congestion management is nevertheless important, and this paper has attempted to provide it.

Appendix

Proof of Proposition 3, parts (i)–(ii)

To establish part (i) of the proposition, first note that, after combining (8) and (9) and eliminating f_2, q^* and f_1^* must satisfy

$$p_1 \;-\; \tau(f_1^*) \;-\; f_1^*\tau'(f_1^*) \;\;=\;\; p_2 \;-\; \tau(q^*-f_1^*) \;-\; (q^*-f_1^*)\tau'(q^*-f_1^*). \qquad (a1)$$

Similarly, after combining (13) and (14), $\widehat{q}$ and $\widehat{f}_1$ must satisfy

$$\begin{aligned} p_1 \;-\; \tau(\widehat{f}_1) \;-\; \widehat{f}_1\tau'(\widehat{f}_1) \;\;&=\;\; p_2 \;-\; \tau(\widehat{q}-\widehat{f}_1) \;-\; (\widehat{q}-\widehat{f}_1)\tau'(\widehat{q}-\widehat{f}_1) \;+\; (\widehat{f}_1-\widehat{f}_2)c'(\widehat{q}) \\ &>\;\; p_2 \;-\; \tau(\widehat{q}-\widehat{f}_1) \;-\; (\widehat{q}-\widehat{f}_1)\tau'(\widehat{q}-\widehat{f}_1), \qquad (a2) \end{aligned}$$

where the inequality uses $\widehat{f}_1 > \widehat{f}_2$. Now suppose that $\widehat{q} \leq q^*$ holds while $\widehat{f}_1 \geq f_1^*$. Since $\tau(f) + f\tau'(f)$ is increasing in f, the LHS of $(a2)$ is then no larger than the LHS of $(a1)$. Similarly, the last expression in $(a2)$ is then at least as large as the RHS of $(a1)$. But the equality in $(a1)$ then implies that the last expression in $(a2)$ should be at least as large as the LHS expression, and the resulting contradiction establishes that $\widehat{f}_1 < f_1^*$ must hold when $\widehat{q} \leq q^*$. Part (ii) is established in parallel fashion.

Solutions for the linear case

In the linear case, the optimal total flight volume (which equals the optimal number of slots sold) is given by

$$q^* \;\;=\;\; \widehat{q} \;\;=\;\; \frac{p_1 + p_2 - 2\theta}{2\alpha + 4\beta}. \qquad (a3)$$

Note that positivity of this expression is ensured if the intercept θ of the τ function is smaller than both of the full prices, a condition that is assumed to hold. Carrier 1's flight volumes at the optimum and under the slot-sale/uniform-toll regime are given by

$$f_1^* \;\;=\;\; \frac{(p_1-\theta)(\alpha+\beta) \;-\; \beta(p_2-\theta)}{\alpha(2\alpha+4\beta)} \;\;>\;\; \frac{(8\alpha+10\beta)(p_1-\theta) \;-\; 6\beta(p_2-\theta)}{(4\alpha+8\beta)(4\alpha+2\beta)} \;\;=\;\; \widehat{f}_1, \; (a4)$$

where the inequality follows from some algebra. Since $f_1^* + f_2^* = \widehat{f}_1 + \widehat{f}_2$, $(a2)$ implies $f_2^* < \widehat{f}_2$. Note that, although f_1^* and $\widehat{f}_1$ are both positive, positivity of f_2^* and $\widehat{f}_2$ (given by the expressions in $(a4)$ with p_1 and p_2 reversed) requires a sufficiently large value of α. This fact shows the need for a sufficient degree of decreasing returns in order to generate interior solutions.

Table 1

Numerical Examples*

	Differentiated-Toll (social optimum)	*Slot-Sale/ Uniform-Toll*	*Laissez-Faire*
$p_1 - \theta = 9$ $p_2 - \theta = 5$			
q	0.707105	0.694426	0.842159
f_1	0.668742	0.550228	0.545672
f_2	0.038363	0.144198	0.296488
$\pi_1 + \pi_2$	4.96840	4.82512	4.49634
$p_1 - \theta = 7$ $p_2 - \theta = 5$			
q	0.706804	0.708284	0.816954
f_1	0.562943	0.458862	0.478234
f_2	0.143861	0.249401	0.338720
$\pi_1 + \pi_2$	3.72792	3.66090	3.46772

*The table shows total flights (q), flights by the individual carriers (f_1 and f_2), and total profit ($\pi_1 + \pi_2$, which gives welfare), for case where the cost functions take the nonlinear forms $\tau(f_i) = \theta + 4f_i^4$ and $c(f_1 + f_2) = 4(f_1 + f_2)^4$.

References

Ausubel, L.M., Cramton, P., 2002. Demand reduction and inefficiency in multi-unit auctions, Unpublished paper, University of Maryland.

Basso, L.J., Zhang, A., 2007. Congestible facility rivalry in vertical structures, *Journal of Urban Economics* 61, 218-237.

Braid, R.M., 1996. Peak-load pricing of a transportation route with an unpriced substitute. *Journal of Urban Economics* 40, 179-197.

Brueckner, J.K., 2002. Airport congestion when carriers have market power, *American Economic Review* 92, 1357-1375.

Brueckner, J.K., 2005. Internalization of airport congestion: A network analysis. *International Journal of Industrial Organization* 23, 599-614.

Brueckner, J.K., Van Dender, K., 2008. Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate. *Journal of Urban Economics*, forthcoming.

Brueckner, J.K., Verhoef, E.T., 2009. Manipulable congestion tolls. Unpublished paper, Tinbergen Institute, Amsterdam.

Burguet, R., 2000. Auction theory: A guided tour, *Investigaciones Economicas* 24, 3-50.

Czerny, A., 2008. Managing congested airports under uncertainty. In: Czerny, A., Forsyth, P., Gillen D., Niemeier, H.-M. (Eds.), *Airport Slots: International Experiences and Options for Reform.* Ashgate, Aldershot, UK, pp. 111-126.

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., Harback, K.T., 2008a. (When) Do hub airlines internalize their self-imposed congestion delays? *Journal of Urban Economics* , 63, 583-612.

Daniel, J.I., Harback, K.T., 2008b. Pricing the major hub airports. *Journal of Urban Economics*, forthcoming.

Daniel, J.I., Harback, K.T., 2008c. Pricing Canadian airports. Unpublished paper, University of Delaware.

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

Forsyth, P., Niemeier, H.-M., 2008. Setting the slot limits at congested airports. In: Czerny, A., Forsyth, P., Gillen D., Niemeier, H.-M. (Eds.), *Airport Slots: International Experiences and Options for Reform*. Ashgate, Aldershot, UK, pp. 63-83.

Gillen, D., 2008. Airport slots: A Primer. In: Czerny, A., Forsyth, P., Gillen D., Niemeier, H.-M. (Eds.), *Airport Slots: International Experiences and Options for Reform*. Ashgate, Aldershot, UK, pp. 41-59.

Hahn, R.W., 1984. Market power and transferable property rights. *Quarterly Journal of Economics* 99, 753-765.

Johnson, T., Savage, I., 2006. Departure delays, the pricing of congestion, and expansion proposals at Chicago's O'Hare airport. *Journal of Air Transport Management* 12, 182-190.

Kasper, D.M., 2008. Slots, property rights and secondary markets: Efficiency and slot allocation at US airports. In: Czerny, A., Forsyth, P., Gillen D., Niemeier, H.-M. (Eds.), *Airport Slots: International Experiences and Options for Reform*. Ashgate, Aldershot, UK, pp. 271-288.

Mayer, C., Sinai, T., 2003. Network effects, congestion externalities, and air traffic delays: Or why all delays are not evil, *American Economic Review* 93, 1194-1215.

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

Oum, T.H., Zhang, Y., 1990. Airport pricing: Congestion tolls, lumpy investment, and cost recovery. *Journal of Public Economics* 43, 353-374.

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

Rupp, N.G., 2008. Do carriers internalize congestion costs? Empirical evidence on the internalization question. *Journal of Urban Economics*, forthcoming.

Schofield, A., 2007. Rush hour revenue: Airlines could face fees for peak-time flights if the U.S. government has its way. *Aviation Week and Space Technology*, October 29, pp. 41-42.

Small, K.A., 1992. *Urban transportation economics*, Harwood Academic Publishers, Chur, Switzerland.

Starkie, D., 2008. Aviation markets, Ashgate, Aldershot, UK.

Verhoef, E.T., 2008. Congestion pricing, slot sales and slot trading in aviation. Unpublished paper, Free University of Amsterdam.

Weitzman, M., 1974. Prices vs. quantities. *Review of Economic Studies* 41, 477-491.

Whalen, W.T., Carlton, D.W., Heyer, K., Richard, O., 2007. Proposal for a market-based solution to airport delays. Economic Analysis Group Discussion Paper, Antitrust Division, U.S. Department of Justice.

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

Footnotes

*I am indebted to Ami Glazer for extremely helpful suggestions and to Volodymyr Bilotkach, Joe Daniel, Ricardo Flores-Fillol, Nick Rupp, and especially Ken Small for additional comments. Three referees also provided very helpful suggestions, and I also benefited from the comments of various seminar participants. Any shortcomings in the paper, however, are my responsibility.

[1]Washington Reagan (National) Airport is also slot-constrained. See Gillen (2008) and Starkie (2008, Ch. 15) for discussions of the history of the slot system at these airports, and see Forsyth and Niemeier (2008) for a good discussion of the economics of airport slots.

[2]Many European airports have slot systems, but while slot trading is well-established in the UK, it was illegal under recently at other EU airports (see Gillen (2008)).

[3]See Pels and Verhoef (2004), Zhang and Zhang (2006), Basso and Zhang (2007), and Morrison and Winston (2007). For an earlier contribution predating the analysis of internalization, see Oum and Zhang (1990). For representative studies in the road-pricing literature, see Small (1992) and Braid (1996).

[4]For an early analysis of market manipulation, focusing on the market for pollution permits, see Hahn (1984). See also Brueckner and Verhoef (2009) for an analysis of non-Pigouvian, "manipulable" congestion tolls, which involve a transparent toll rule that is designed to be manipulated. Carriers then choose flight volumes with full knowledge of the resulting impacts on toll liabilities.

[5]Erik Verhoef and I began discussing slot-based remedies for congestion several years ago, and he produced the main elements of a single-authored paper based on his preferred set of assumptions, a paper recently finalized as Verhoef (2008). Meanwhile, I recognized that the different approach of Brueckner and Van Dender (2008) could be applied fruitfully to the problem, leading to the present paper. The paper owes a debt to the original discussions with Verhoef, and it borrows directly from his treatment of slot trading, as explained further below.

[6]Czerny (2008) presents a variant of Weitzman's uncertainty analysis for the case of airports, contrasting quantity regulation (implemented by slots) to price regulation (implemented by some form of airport access charge).

[7]See Forbes (2008) for empirical evidence that airport congestion reduces fares.

[8]A sufficient condition for positivity of the Hessian determinant is $c'' \equiv 0$. However, when τ is a constant, the second-order condition fails, indicating that the optimum involves a corner solution with $f_2 = 0$.

[9]Since $p_1 > p_2$ holds and $\tau(f) + f\tau'(f) + fc'(q)$ is increasing in f holding q fixed, the f_1 value satisfying (8) must be larger than the f_2 value satisfying (9).

[10]This conclusion follows by continuity from the symmetric case, where the carriers' common optimal flight volumes are smaller than the common equilibrium level.

[11]This toll pattern also obtains in a model with explicit peak and off-peak periods, as seen in the analysis of Brueckner (2005). That model generates carrier asymmetry through an explicit network structure, with the carriers operating concentrated hubs at different airports. The model of Flores-Fillol (2008) adds generality in a different dimension by explicitly capturing carrier choices of flight frequency and aircraft size, showing how congestion tolls affect both. The present analysis could be modified to incorporate such generalizations without a fundamental effect on the results. However, the model, along with its predecessors, does not portray entry and exit of carriers, and the effect of generalizations in this dimension are less obvious.

[12]To establish the result, add the optimality conditions in (13) and (14) and divide by 2, yielding $(p_1 + p_2)/2 - (\theta + \alpha q) - 3\beta q/2 \;=\; z$. Substituting (17) yields $z = \beta q/2$, which equals $qc'(q)/2 = \text{MCD}/2$. Since $q = q^*$, $z = \text{MCD}^*/2$ follows, establishing part (v) of the proposition.

[13]The slot-sale/uniform-toll regime is also efficient when carriers do not internalize congestion, behavior that Daniel (1995) and Daniel and Harback (2008a) view as realistic. They argue that, when a Stackelberg leader interacts with competitive-fringe airlines, the offsetting behavior of these carriers (which increase their flights in one-for-one fashion as the leader reduces its flights) eliminates that carrier's incentive to take account of self-imposed congestion (see Brueckner and Van Dender (2008) for further discussion). Thus, even large carriers end up behaving "atomistically," viewing congestion as uninfluenced by their choices. Daniel (1995) and Daniel and Harback (2008a) provide evidence supporting this view of airline behavior, while Brueckner (2002) and Mayer and Sinai (2003) provide contrary evidence in favor of internalizing behavior. Rupp (2008) provides additional evidence against internalization using an extension of Mayer and Sinai's methodology. To incorporate non-internalizing behavior in the present model, suppose that both carriers, despite their nonatomistic sizes, treat airport congestion as parametric, ignoring the fact that the $c(\cdot)$ function depends on both flight volumes. With congestion viewed as parametric, the terms containing c' on the LHS of the laissez-faire conditions (8) and (9) vanish. The tolls required to generate the optimum are then the same across carriers and equal to $q^*c'(q^*) = \text{MCD}^*$ (the "atomistic toll"), so that each carrier is charged for the full congestion damage done by an extra flight

(including the damage done to its existing flights). Since differentiated tolls are uniform even though carriers are asymmetric, a slot-sale/uniform-toll regime can again generate the social optimum.

[14]This percentage gain is obviously higher the smaller is the degree of asymmetry between the carriers. For example, if $p_1 - \theta$ falls from 7 to 6, the percentage gain rises from 74 to 87 percent. Interestingly, the gain is independent of the magnitude of the multiplicative constants in the cost functions (provided that these constants remain equal). These welfare results bear some connection to the findings of Morrison and Winston (2007) and Daniel and Harback (2008b), who investigate the effect of imposing atomistic tolls, which are appropriate under non-internalizing behavior (see footnote 13), when carriers in fact internalize congestion. Their calibrated models differ substantially, as do the answers reached. Daniel and Harback find that inappropriate use of atomistic tolls is worse than no tolling at all, while Morrison and Winston find only a small welfare loss from use of the wrong tolls. By contrast, the results in Table 1 pertain to use of a uniform toll that is optimally smaller than the atomistic toll (only half as large in the linear case, as seen in part (v) of Proposition 3) and thus insured to generate a welfare gain.

[15]Note the large decrease in the flight volume of the smaller carrier means that moving toward the optimum generates substantial distributional effects.

[16]It could be argued that the initial distribution of slots should affect the final allocation across carriers. This view is incorrect under the present assumptions, but it might have some relevance under manipulative behavior.

[17]The perception of a fixed flight total has the same effect as atomistic behavior, under which carriers do not take into account their impact on congestion (see footnote 13). If carriers were assumed from the outset to behave atomistically, as argued by Daniel (1995), then the fixed flight total under the slot-distribution regime would have no additional bite, in constrast to its crucial role under the present assumptions.

[18]In a pay-as-bid auction, the auctioneer again allocates a fixed slot supply to the highest bidders but makes them pay their winning, unit-specific bids. Since bidders trade off the surplus gain from a lower price against a smaller chance of winning, the pay-as-bid system can elicit understatement of valuations.

[19]The useful survey by Burguet (2000) provides further discussion.

[20]As noted earlier, this auction setup, where all the available slots are auctioned, differs from the one proposed by the FAA to deal with congestion at the New York airports. Under that proposal, only a portion of the slots (expropriated from the carriers, who held them as a

result of a prior distribution) is auctioned each year.

[21]It is easy to see that efficiency also obtains under a pay-as-bid auction.

[22]A practical question under the quantity-based regimes concerns the terms of slot usage. Under one approach, a slot distribution or auction could be carried out relatively frequently (say, once a year), with all slots reverting to the airport at the end of the holding period. Under the previously existing system at US airports, the slot holding period was indefinite, although under the new FAA proposal, carriers each year would face a chance of losing a portion of their slots for reallocation via auction.

[23]Imperfect information, and the resulting likelihood of erroneous choices of the levels of the decision variables, could generate welfare losses of different magnitudes under price and quantity-based regimes, a subject that could be a subject for further investigation.

[24]The proposition's statement regarding the slot-sale/uniform-toll case assumes linearity. Also, note that the proposition applies more generally to the case where carriers are symmetric.

[25]Another approach, suggested by Brueckner (2005), is to correct the market-power distortion through subidies paid at the level of the city-pair market (which can vary according to the extent of market competition) combined with congestion tolls levied at the airport level.

[26]Kasper (2008) stresses the overall virtues, relative to congestion pricing, of a slot-disribution regime coupled with a secondary market.

[27]The analysis suggests that slot auctions are unnecessary if a slot-trading regime is already in place. The FAA's recent proposal for the New York airports does not conform to this view, and the rationale is that auctioning a portion of the available slots each year will stimulate the existing secondary market, which is viewed as underdeveloped.

[28]Johnson and Savage (2006) provide "back-of-the-envelope" toll calculations applied to O'Hare Airport, showing that even simple methods can be useful.

[29]See Schofield (2007).