TI 2002-078/3
Tinbergen Institute Discussion Paper

Airport Pricing

Eric Pels
*Erik T. Verhoef**

Department of Spatial Economics, Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam,

** Tinbergen Institute*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Airport Pricing:

Network Congestion Pricing with Market Power and Endogenous Network Structures

Eric Pels Erik T. Verhoef
Free University Amsterdam, Department of Spatial Economics
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Email: apels@feweb.vu.nl everhoef@feweb.vu.nl

Abstract.
Conventional economic wisdom suggests that congestion pricing would be an appropriate response to cope with the growing congestion levels currently experienced at many airports. Several characteristics of aviation markets, however, may make naive congestion prices equal to the value of marginal travel delays a non-optimal response. This paper develops a model of airport pricing that captures a number of these features. The model in particular reflects (1) that airlines typically have market power and are engaged in oligopolistic competition at different sub-markets; (2) that part of external travel delays that aircraft impose are internal to an operator and hence should not be accounted for in congestion tolls; (3) that the airlines' consumers may impose external benefits of increased frequencies upon one-another; (4) that different airports in an international network will typically not be regulated by the same authority; and (5) that an individual airline's network will not be exogenous but may instead be affected by congestion levels and tolls, which may create discontinuities. We present an analytical treatment for an undetermined number of nodes, links and operators in a network of undetermined size and shape, and some numerical exercises for a small triangular network. Some main conclusions are that second-best optimal tolls are typically lower than what would be suggested by congestion costs alone and may even be negative, that pricing may induce changes in network configurations, and that cooperation between regulators need not be stable but that non-cooperation may lead to welfare losses, also when compared to a no-tolling situation.

1 INTRODUCTION

Most economic studies in network congestion pricing are inspired by the case of a congested road network, and consider the relatively simple case of atomistic price-taking users of a network, and congestion that arises on the links of the network, as opposed to at its nodes. In many network markets, however, the primary users are operators that may possess substantial market power, in particular if economies of scale, scope, density or networks allow large operators to produce services at relatively low average costs. And in such cases – aviation being the one to be considered in this paper – congestion often occurs at nodes of the network, rather than along its links, due to the relatively large concentration of flows near the operators' 'home nodes'. This paper studies the economic implications for congestion pricing in such situations. Although the discussion will be cast in terms of a congested aviation network, the insights developed may of course often carry over to congestion pricing at nodes for different modes than aviation, and possibly even to congested networks other than for transport, provided market conditions are similar to those considered here.

Congestion pricing in aviation is a relevant issue because many airports are facing capacity problems. European airports are usually slot-constrained; slots are allocated by a slot coordinator, while most U.S. airports (except the high-density rule airports Washington Ronald Reagan, New York LaGuardia, New York Kennedy and Chicago O'Hare) are unconstrained, i.e. flights are served on a first-come, first-serve basis. When capacity is limited, arriving aircraft cause delays and thus costs for other arriving aircraft. The slot-allocation mechanism at most airports is economically inefficient. The users of capacity may pay less than the marginal social cost – congestion costs are not paid at unconstrained airports, entry is deterred at constrained airports – and are not necessarily the (potential) users that attach the highest economic value to the capacity.

In the literature on the allocation of scarce capacity for congested airports and other types of infrastructure, various methods are proposed to allocate capacity, or slots. All methods intend to allocate the capacity to the parties that attach the highest economic values to them. Slot auctions appear an attractive since the "bidders" have an incentive to reveal their valuation for the capacity (if the auction is properly designed), and have received attention from both policy makers (the European Commission) and academics (see e.g. Rassenti et al, 1982, for a simulation study). However, complications with slot auctions include the fact that operators need "networks" of slots: multiple slots at a node, and complementary slots at other nodes, so that auction designs for such configurations are highly complex.

Airport congestion pricing also aims to allocate scarce capacity to those parties who attach the highest economic value to it. It has already received some attention in the literature. Carlin and Park (1970) estimated the external cost of a peak-period landing at LaGuardia was \$2000 (in 1969 \$); about twenty times the actual landing fee, although this number should not be interpreted as an equilibrium congestion toll. Oum and Zhang (1990) examine the relation between congestion tolls and capacity costs, and find that when capacity investment is lumpy, the cost recovery theorem (which states that congestion toll revenues just cover amortized capacity (expansion) costs under constant returns to scale) no longer holds. Daniel (1995,

2001) combines stochastic queuing theory with a Vickrey-type bottleneck model, and simulation results show that congestion pricing causes a redistribution of flights over the day, where smaller aircraft may divert to other airports because they value their use less than the social cost of using the congested airport. Brueckner (2002) analyzes airport congestion pricing when airlines are nonatomistic, and concludes that there may be only a limited or even no role for congestion pricing when the number of airlines using the node decreases, as the share of internalized congestion costs increases.

This paper also analyzes airport congestion pricing, taking a network perspective. Most studies of (second-best) congestion pricing in transport networks concern road traffic and consider link-based tolls. A question that naturally arises is whether insights from studies on link-based pricing are directly transferable to node-based pricing, especially under second-best circumstances where multiple market distortions exist simultaneously. The nature of these other market distortions, additional to congestion externalities, will often be different at nodes than along links. The economic nature of transport nodes and their primary user(s) may thus imply important deviations from the economic conditions governing a congested road network. Specifically, individual road users would typically not have any market power, and can thus be assumed to take travel times and tolls (if any) as given (in practice, this would normally also hold for transport firms that may have multiple trucks using the same congested network simultaneously). Transport nodes, in contrast, and especially the more congested hubs, will typically have spatial monopolistic power, while the primary user(s) will often be a limited set of operators that compete under oligopolistic conditions. Moreover, when positive network externalities (or economies of density) induce these operators to use a hub-and-spoke type of networks, with different operators using different hubs, these oligopolies may be asymmetric. A substantial share of congestion costs may then in fact not be external effects, but internal instead, in the sense that the travel delays imposed by one service upon other services would often concern services of that same operator, who can be assumed to take these firm internal congestion effects already into account when designing a profit-maximizing price and frequency schedule for the hub (Brueckner, 2002).

A further implication of oligopolistic competition would be that another distortion, besides congestion, is likely to be present, namely that of strategic interaction between competitors with the result of non-competitive pricing. Absent congestion, consumer prices may then exceed marginal costs, implying that an economic argument for subsidization rather than taxation would exist. As pointed out by Buchanan (1969) and Baumol and Oates (1988) in the context of a polluting monopolistic firm, the implication for Pigouvian externality pricing is that the second-best optimal tax would be below the marginal external costs and may even become negative. This would provide a second argument, in addition to the point raised by Brueckner (2002), of why optimal congestion charges at a hub would be below marginal external congestion costs if straightforwardly defined as the value of a single service's marginal delay costs for all other services.

A third market failure that can be identified when switching from a benchmark situation with atomistic users to consumers of oligopolistic competitors derives from the fact that the consumers would typically value higher frequencies on offer, as these on average allow departures and arrivals closer to the desired times. Analogous to what Mohring (1972) identified in the context of bus transit, a positive externality among consumers may then exist that arises from the fact that increased passengers' demand would have a positive effect on aggregate frequency, which then in turn positively affects the *ceteris paribus* valuation of other consumers for the service. Under second-best circumstances where no attempt is made of an internalization of the implied external benefits, it is conceivable that another downward adjustment of second-best congestion tolls would be in order.

This paper aims to investigate such issues in a network environment, by developing a model that is cast in terms of aviation and considers second-best congestion pricing for incoming and outgoing flights at airports. The model extends an earlier model of second-best pricing in congested road networks (Verhoef, 2002). The second-best circumstances under which congestion tolls have to be set are those just mentioned. We consider a network with multiple nodes, where airlines may offer direct services as well as indirect services between different origin-destination pairs – the latter representing hub-and-spoke type of networks. Both operators and consumers suffer from congestion at nodes. Network operators may change their network as a result of congestion pricing (*e.g.* abandon indirect services). Three types of interacting players are present in our model: nodal regulators, operators, and consumers; each having their own objective. Congestion tolls can be determined by a single regulator for all nodes in the network, but also by "local" regulators of specific airports. "Competition" between local regulators then becomes an issue.

The general analytical model is presented as a static, one-period model. However, because in the analytical part, no restrictions are imposed on the size and configuration of the network, a dynamic (discrete time) problem could readily be analyzed in the same framework, by considering a so-called 'hyper-network' that encompasses a copy of the physical network for every time period to be considered, and that adds 'virtual' links to represent departure time decisions and scheduling costs (or, possibly, waiting costs, if passengers are free to skip the earliest connection(s) during a stop-over). In other words, the model to be presented is certainly not confined to static network problems alone.

The structure of the paper is as follows. First, the notation and assumptions will be presented in Section 2. Section 3 contains the (profit) maximization model for the network operators (airlines). Section 4 contains the regulator's optimization problem. Section 5 presents a simple numerical solution, and Section 6 concludes.

2 NOTATION AND ASSUMPTIONS

In the model, we distinguish three different parties. *Passengers* wish to travel between an origin and destination (a formulation with freight transport with atomistic demanders would be comparable to the one given here). In order to do so, services of a *transport network operator* are necessary. Transport network operators, in turn, need the services of two (origin and destination) or more (in case of indirect services) transport nodes. Prices for the use of the node may be set by a profit maximizing *nodal operator* or a welfare maximizing *regulatory authority*. Because we are concerned with (second-best) optimal nodal prices, we will be considering a regulatory authority alone. An extension of the model to four types of players (regulators, nodal operators, network operators and users) is considered as an interesting option for future work.

We analyze a static transport network with $H \geq 3$ nodes. Only return markets are considered (consistent with the stationary state equilibrium approach), so that the number of markets or origin-destination pairs in the network is $J = H \cdot (H-1)/2$. The notation to be used is summarized in Table 1.

The set of network operators may be (considered to be) divided in subsets of different modes, although this is not done explicitly in our analysis. These different modes could be identified for instance by the elements of $\boldsymbol{a}_{i,j}$ (on longer segments of a trip, airlines will typically be faster than trains) or the cost function (airlines will have a higher cost per transport movement than for instance a bus).

h	index for nodes (airports), with h=1,…,H
i,l	possible indices for network operators (airlines), with i=1,…,I or l=1,…,I
j,m	possible indices for O-D pairs (markets), with j=1,…,J or m=1,…,J
k	index for links, denoted k = 1,…,K
$p_{i,j}$	fare of operator i in market j
$f_{i,k}$	frequency of operator i on link k
$f_{i,j}$	frequency of operator i in market j
$q_{i,j}$	the number of passengers transported by operator i in market j
$D_j\left(\sum_{i=1}^{I} q_{i,j}\right)$	inverse demand function for market j
ϕ_h	average (per passenger) node specific time losses due to congestion
$\phi_{i,j}$	carrier-market specific time losses due to congestion
$\phi_{i,k}$	carrier-link specific time losses due to congestion
η_h	parameter denoting the slope of the congestion delay function at node h
vot_p	passenger value of time
vot_l	airline value of time
$\mathbf{a}_{i,j}$	vector of operator-market specific attributes (*e.g.* trip-time)
$c^f_{i,k}$	carrier i's cost per transport movement on link k
$c^q_{i,k}$	carrier i's cost per passenger on link k
$g_{i,j}(p_{i,j}, f_{i,k}, a_{i,j}, \phi_{i,j})$	generalized user cost (for a passenger) for the use of operator i in market j
λ_i	operator i's passenger load (load factor multiplied by seat capacity)
t_h	congestion toll at node h
$\theta_{i,j,k}$	dummy equal to 1 if $f_{i,k}<f_{i,o}$, $\forall o \neq k$
$\delta_{i,j}$	dummy equal to 1 if operator i is active in market j
$\delta_{i,j,k}$	dummy equal to 1 if link k is used in market j by operator i
$\delta_{k,h}$	dummy equal to 1 if link k has node h as origin or destination
$\delta_{j,h}$	dummy equal to 1 if node h is origin or destination in market j

Table 1. Notation

For the general specification of the model, a number of assumptions are made that will now be presented.

Assumption 1. A given passenger's trip in an origin- destination j pair will involve one network operator only. The inverse demand function in each market is linear in form:

$$D_j\left(\sum_{i=1}^{I} q_{i,j}\right) = \alpha_j - \beta_j \sum_{i=1}^{I} q_{i,j} \tag{1}$$

where α_j and β_j >0; α_j represents the maximum gross valuation by consumers in market j; $q_{i,j}$ is the number of passengers transported by operator i in market j, and β_j is the demand sensitivity parameter. A linear form is convenient in the numerical version of the model; for the analytical exposition it saves somewhat on the notation as the slope β_j is constant. If this model were used with multiple modes, as mentioned above, trips or trip segments between two nodes with different modes are then seen as pure substitutes. Complementarity between modes could occur *en route*, if a mode-switch occurs at a stop-over during the trip.

Assumption 2. Operator i of course services at most J markets. Operator i's seat capacity and load factor are constant. Operators keep the load factor constant over all markets. The frequency on link k can then be expressed as

$$f_{i,k} = \frac{1}{\lambda_i}\sum_{j=1}^{J}\delta_{i,j,k}q_{i,j} \tag{2}$$

where $\delta_{i,j,k}$ is a dummy equal to 1 if link k is used in market j by operator i and λ_i is the product of the load factor and the seat capacity. In what follows, λ_i will be referred to as the "passenger load" for simplicity. λ_i are exogenous; operators maximize profits *given* λ_i. In a more ambitious exercise, one could make the passenger load (or even the load factor and/or the seat capacity) endogenous, and use vehicles (aircraft, trains, trucks, cars etc.) with different capacities on different links. The frequency offered in a market is the minimum of the frequencies on the links used in that specific market:

$$f_{i,j} = \min(f_{i,k} | \delta_{i,j,k} = 1) \tag{3}$$

which is equivalent with

$$f_{i,j} = \sum_{k=1}^{K}\theta_{i,j,k}\delta_{i,j,k}f_{i,k} = \frac{1}{\lambda_i}\sum_{k=1}^{K}\theta_{i,j,k}\delta_{i,j,k}\sum_{m=1}^{J}\delta_{i,m,k}q_{i,m} \tag{3'}$$

where $\theta_{i,j,k}=1$ if $f_{i,k}<f_{i,o},\ k\neq o\ \ \forall\ k,o,\delta_{i,j,k}=\delta_{i,j,o}=1$. This means that the frequency in a given market is directly dependent on the frequency in other markets using the same link(s). Note that we assume that when $\theta_{i,j,k}=1$ then $\theta_{i,j,k}\neq\theta_{i,j,o}$, $\forall\ k,o$. For each market, there can be only one link with the smallest frequency.

Assumption 4. Congestion occurs at nodes only (*i.e.*, not on links). This is of course a drastic simplification, because the "capacity in the air" or the capacity of the air traffic control system may also be scarce. This could be a (future) extension of the model. The average congestion costs per passenger or per flight (measured in additional travel time) at node h are assumed to increase linearly with the total frequency at the node:

$$\phi_h = \eta_h \sum_{k=1}^{K}\sum_{i=1}^{I}\delta_{k,h} f_{i,k} = \eta_h \sum_{k=1}^{K}\sum_{i=1}^{I}\delta_{k,h}\frac{1}{\lambda_i}\sum_{j=1}^{J}\delta_{i,j,k} q_{i,j} \tag{4}$$

where η_h is the constant slope of the congestion function and $\delta_{k,h}$ indicates the nodes used on links h. Note that arriving and departing movements need not contribute equally to, and suffer equally from congestion. However, as we only consider return markets, we do not have to make this distinction. The congestion term (in time units) to be included in the passengers' generalized cost function for alternative i in market j then is

$$\phi_{i,j} = \sum_{k=1}^{K}\sum_{h=1}^{H}\delta_{k,h}\delta_{i,j,k}\phi_h \tag{5}$$

where $\delta_{i,j,k}$ denotes the links used in market j and ϕ_h the congestion (measured in time) suffered at these nodes. Multiplying this term by the passengers' value of time yields the monetized congestion delay cost to passengers. Likewise, the term to be included in the airline cost function (over all flights, using link k, to passengers) is:

$$\phi_{i,k} = \sum_{j=1}^{J}\sum_{h=1}^{H}\delta_{k,h}\delta_{i,j,k}\phi_h \tag{6}$$

Assumption 5. The different alternatives i in market j are characterized by a generalized user cost function $g_{i,j}(p_{i,j},\ \sum_{k=1}^{K}\theta_{i,j,k}\delta_{i,j,k} f_{i,k}\ ,\ \boldsymbol{a}_{i,j},\ vot_p\times\phi_{i,j})$ where $vot_p\times\phi_{i,j}$ represents the monetized average congestion costs per passenger (vot_p is the passenger's value of time), $p_{i,j}$ is the fare and $\boldsymbol{a}_{i,j}$ is the vector of carrier-market specific attributes. The generalized user cost function is linearly additive in form:

$$g_{i,j} = p_{i,j} + \frac{\gamma\lambda_i}{\rho + \sum_{k=1}^{K}\theta_{i,j,k}\delta_{i,j,k}\sum_{m=1}^{J}\delta_{i,m,k} q_{i,m}} + \mathbf{a}_{\mathrm{i,j}} + vot_p \times \varphi_{i,j} \tag{7}$$

where γ represents the passenger sensitivity to frequency. An increase in frequency lowers the generalized cost. Note that this formulation introduces a second externality into the model, besides congestion as discussed under assumption 4. This second externality concerns the external benefit that passengers pose upon one another through the lower generalized costs that results from the higher frequency induced by an individual passenger's travel decision; this effect appears in the second term on the right hand side of (7). This external effect is analogous to the "Mohring-effect" as often considered in the context of transit (Mohring, 1972). This effect captures that an increased frequency implies that the expected deviation from the preferred arrival time will be smaller.

Note, however, that a higher frequency of one airline on a given origin-destination pair is assumed not to affect the passengers' valuation of another airline's frequency for that same origin-destination pair. This is a simplifying assumption, which is made to prevent further complications in the strategic game between airlines that would arise when an airline's increase in frequency would directly affect competing firms' customers' willingness to pay due to an increased overall frequency in the market. As a result, the Mohring effect, although present in our model, is purely firm internal. It will therefore not be used as a strategic instrument in the sense of directly affecting the attractiveness of the competitors' supply, and it will be optimized by the airline itself (see Appendix 2) and will hence not affect (first- and second-best) optimal tax rules.

Finally, ρ is a constant that increases the generality of equation (7); in addition, a (very small) value unequal to zero turned out to be convenient in the numerical optimization as it prevents division by zero.

Assumption 6. The operator's cost per passenger $c_{i,k}^{q}$ and per transport movement $c_{i,k}^{f}$ are constant on each link. t_h is the congestion toll at node h. Total operating costs for operator i are then:

$$\sum_{k=1}^{K}\left(f_{i,k}\left(c_{i,k}^{f} + \sum_{h=1}^{H}\delta_{k,h}t_h + vot_l \times \phi_{i,k} \right) + c_{i,k}^{q}\sum_{j=1}^{J}\delta_{i,j,k}q_{i,j} \right) \tag{8}$$

which may be rewritten as:

$$\sum_{k=1}^{K}\left(\left[\sum_{m=1}^{J}\delta_{i,m,k}q_{i,m} \right] \times \left[\frac{1}{\lambda_i}\left(c_{i,k}^{f} + \sum_{h=1}^{H}\delta_{i,k,h}t_h + vot_l \times \phi_{i,k} \right) + c_{i,k}^{q} \right] \right) \tag{8'}$$

where $vot_l \times \phi_{i,k}$ represents the monetized average congestion costs per flight (vot_l is the airline's value of time).

Assumption 7. Tolls on a given airport are not differentiated over airlines. Since the congestion at the node is caused by an excessive demand for transport movements (frequency), the network operators are assumed to pay the toll. Of course, network operators may consequently pass (part of) the tolls onto the passengers by increasing the ticket price(s).

Assumption 8. Competing operators on a specific market act as Cournot opligopolists (*i.e.* they choose an optimal output (and frequency) taking the others' outputs as given). Operators do not believe that by their actions, they can affect the regulator's tolls (*i.e.* regulators and operators are playing a Stackelberg type game, the regulators being the leader). Passengers are pure price takers.

Although these assumptions may seem restrictive, many of these assumptions are quite common in the aviation economics literature. The functional form of the cost function used in this paper is similar to the one used by Brueckner and Spiller (1991)[1]. Combined with a linear demand curve, the "Brueckner-Spiller" model has been used regularly in the literature to analyze aviation networks. Despite the conceptual simplicity (but, in case of large networks, computational complexity), recent trends in the aviation markets can easily be explained using this model; see e.g. Brueckner (2001) for an analysis of airline alliances, and Pels *et al.* (2001) for an analysis of optimal airline networks[2]. It is not the objective of this paper to calculate exact tolls for existing airports, for which these assumptions would clearly be too restrictive. This paper aims to develop theoretical insights into the consequences of airport congestion pricing, for which these assumptions suffice.

3 EQUILIBRIUM PRICES, FREQUENCIES AND TOLLS

With these assumptions, we can now turn to the derivation of optimal tolls. There are three types of players in the model (passengers, network operators and regulatory authorities), each with their own maximization problem. The model is solved in three steps. First, a passenger demand function for network operator *i* in market *j* is determined. Then, using this demand function, the operator's problem is specified, and the associated profit maximizing optimality conditions are derived. Finally, the regulator's problem is solved, again using the passenger demand function, and also using the operator optimality conditions as restrictions.

[1] Brueckner and Spiller (1991) do not include congestion in their cost function.
[2] See Brueckner (2001) for additional references.

The passenger optimization problem

The maximum willingness to pay for the marginal passenger in market *j* for alternative *i*, including monetized time costs, is given by $D_j\left(\sum_{i=1}^{I} q_{i,j}\right)$, while each pasenger's generalized user cost for the use of operator *i* are given by $g_{i,j}(\cdot)$ as defined in equation (7). Intra-marginal passengers' net benefits are determined according to the familiar Marshallian surplus. According to Wardrop's equilibrium conditions, marginal benefits are equal to the average generalized costs in equilibrium (or marginal net benefits are zero) for all used alternatives (operators in this case), so that $D_j(..)=g_{i,j}(..)\ \forall\ i$ in equilibrium, while the average generalized costs of unused alternatives cannot be lower than $D_j(..)$ and will typically be higher. Because operators incur costs for a service also when $q_{i,j}=0$ (see (8)), unused alternatives in our model will not actually be offered. By assumption, demand and generalized cost functions are linear, so that the equilibrium condition for used alternatives implies:

$$p_{i,j} = \alpha_j - \beta \sum_{l=1}^{I} q_{l,j} - \frac{\gamma}{\rho + \frac{1}{\lambda_i} \sum_{k=1}^{K} \theta_{i,j,k} \delta_{i,j,k} \sum_{m=1}^{J} \delta_{i,m,k} q_{i,m}} - \mathbf{a_{i,j}} - vot_p \times \varphi_{i,j} \qquad (9)$$

where α_j is the market specific constant in the inverse demand function and β is the slope of the inverse demand function. This operator specific inverse demand curve incorporates passengers' optimizing behavior, and is used in the next step to maximize operator profits. Note that the arguments of this inverse demand function include the quantities sold by competing transport network operators in market *j*.

The transport network operator maximization problem

As stated in assumption 8, we assume Cournot behavior in modeling the operators' competition. This is motivated by earlier (empirical) research.[3] In a Cournot oligopoly, excess

[3] For instance, in an empirical analysis of Chicago-based airline routes involving American Airlines and United Airlines, Oum et al. (1993) conclude that "the overall results indicate that the duopolists' conduct may be described as somewhere between Bertrand and Cournot behavior, but much closer to Cournot, in the majority of the sample observations". Brander and Zhang (1990), using similar data, find " strong evidence … against the highly competitive Bertrand hypothesis". Brander and Zhang (1990) find Cournot behavior plausible for the markets under consideration (Chicago-based routes where American Airlines and United Airlines together have a market share exceeding 75%). Based on these observations, we assume Cournot competition.

profits can be made when the number of suppliers is finite. For the alternative of Bertrand-competition, equilibrium prices would equal marginal costs without collusion, when marginal costs are constant (as they are in this model). The current financial problems of many airlines does not mean that Cournot oligopoly modeling would not be appropriate for this sector. High fixed costs – absent from our model – may contribute to financial problems, also under Cournot-competition.

Thus, the operators in this model maximize profits with respect to $q_{i,j}$, taking the competitors quantities as given (note that the assumption of a fixed passenger load implies that maximization with respect to frequencies independent of passenger numbers is neither possible nor necessary). In general, the maximization problem for operator i is:

$$\max_{q_{i,j}} \pi_i = \sum_{j=1}^{J} \delta_{i,j} p_{i,j}(..) q_{i,j} - \sum_{k=1}^{K} \left(\sum_{j=1}^{J} \delta_{i,j,k} q_{i,j} \left[\frac{1}{\lambda_i} \left(c_{i,k}^{f} + \sum_{h=1}^{H} \delta_{i,k,h} t_h + vot_l \times \varphi_{i,k} \right) + c_{i,k}^{q} \right] \right) \tag{10}$$

The first-order necessary conditions are:

$$\begin{aligned} &\delta_{i,j} p_{i,j}(..) + \sum_{m=1}^{J} \delta_{im} q_{i,m} \frac{\partial p_{i,m}(..)}{\partial q_{i,j}} - \\ &\sum_{k=1}^{K} \delta_{i,j,k} \left[\frac{1}{\lambda_i} \left(c_{i,k}^{f} + \sum_{h=1}^{H} \delta_{k,h} t_h + vot_l \times \varphi_{i,k} \right) + c_{i,k}^{q} \right] - \sum_{k=1}^{K} \sum_{m=1}^{J} \delta_{i,m,k} q_{i,m} \frac{vot_l}{\lambda_i} \frac{\partial \varphi_{i,k}}{\partial q_{i,j}} = 0, \quad \forall j \end{aligned} \tag{11}$$

Similar conditions can be derived for the other operators, *-i*. The full set of equalities for all relevant paths with $q_{i,j}>0$, define a Nash-equilibrium. This equilibrium is derived using the most simple strategy available (*i.e.* taking the competitor's price and frequency as given). Other strategies, for example a Stackelberg-model, are of course possible, but these can make the analysis much more complicated because we are dealing with transportation networks. An operator can be a market leader on certain markets, and a follower on other markets. This would complicate the analysis considerably, and is beyond the scope of the current paper.

It can be shown (by implicit differentiation) that $\partial q_{i,j} / \partial t_h < 0$. By writing $\mathrm{d}q_{i,j} = (\partial q_{i,j} / \partial t_h)\mathrm{d}t_h + (\partial q_{i,j} / \partial q_{l,j})\mathrm{d}q_{l,j}$ (and ignoring all other effects), where $l \neq i$, it becomes clear that the imposition of a toll may lead to an increase in the equilibrium level of $q_{i,j}$, since the second term on the right hand side may become positive.

It follows from (11) that the equilibrium value for $q_{i,j}$ is a function of the toll t_h if $\delta_{k,h}=1$. The optimal toll is determined by the regulator.

The regulator's maximization problem

We can make general assumptions on the behavior and objectives of nodal regulator(s). Two important dimensions are the type of objective (surplus versus profit maximization) and the number of regulators (a single "global" regulator for the entire network, versus node-specific regulators). In terms of objectives, we will consider only welfare-maximizing regulation, postponing the regulation of a private airport in a similar model set-up with four layers as a future research challenge. We will consider first a global regulator with the objective to maximize global surplus (taking into account benefits and costs over the entire network). The network can then be a national network with multiple operators and nodes, or an international regulatory office can attempt to maximize total surplus for the international network. Oum et al. (1996) show that "network pricing" of nodes (i.e. joint pricing of hub and spoke nodes) leads to a higher welfare than independent pricing of the nodes in the network. We will also consider the result of non-cooperating local regulators, each concerned with a local welfare measure alone.

The global regulator maximizes surplus for the entire network: the regulator considers consumer surplus in *all* markets and profits of *all* operators. It sets differentiated tolls t_h for all nodes h in the system. In a restricted version of this problem, the regulator could maximize global welfare by setting a common toll t, assumed equal for all h. The authority thus maximizes the following objective function:

$$\max_{t_h, q_{i,j} \forall i,j,h} \varpi_G = \sum_{j=1}^{J} \int_0^{\sum_{i=1}^{I} q_{i,j}} D_j(x)dx - \sum_{i=1}^{I}\sum_{j=1}^{J} q_{i,j}\left[\frac{\gamma\lambda_i}{\rho + \sum_{k=1}^{K}\theta_{i,j,k}\delta_{i,j,k}\sum_{m=1}^{J}\delta_{i,m,k}q_{i,m}} + \mathbf{a_{i,j}} + vot_p \times \varphi_{i,j}\right] - \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\delta_{i,j,k}q_{i,j}\left[\frac{1}{\lambda_i}\left(c_{i,k}^{f} + vot_l \times \varphi_{i,k}\right) + c_{i,k}^{q}\right] \tag{14}$$

subject to the operators' first order conditions for profit maximization (for all i and j). The first right hand side (rhs) term represents total benefits (as integral of the Marshallian inverse demand function). The second rhs term represents total generalized costs (excluding the airline fares, which cancel out against the airline revenues). The third rhs term represents airline operating costs (excluding the expenditures on tolls, which cancel out against toll revenues). The three terms together thus give social surplus. The regulator sets the toll t_h, given the airline (profit maximizing) optimality conditions. A change in t_h affects the optimal output, and thus total welfare. Adding the operators first-order conditions for profit maximization as restrictions to the regulator's objective, and maximizing over t_h and $q_{i,j}$ then yields the equilibrium in quantities and tolls.

Solving the first-order conditions for welfare maximization only yields an implicit solution (optimal tax rule). To analyze the effects of tolling, a numerical analysis will therefore be presented in the next section.

4 NUMERICAL ANALYSIS

In this section, the numerical solutions for an illustrative model are presented. We consider a symmetric network with three nodes, three links and three operators, as depicted in Figure 1. Operator *a* uses node *A* as hub, and offers direct services on markets *AB* and *AC*, using links 1 and 2 respectively. Operator *a* also offers an indirect service in market *BC*, using both links 1 and 2. Comparable services are offered by operators *b* and *c*, both offering three services, of which two direct and one indirect. The parameterization of the model is arbitrary; no attempt was made to mimic the conditions in any existing network[4]. This reflects the purpose of the current exercise, namely to illustrate the analytical results provided earlier, and to "check" their applicability and correctness in a network of a manageable configuration.

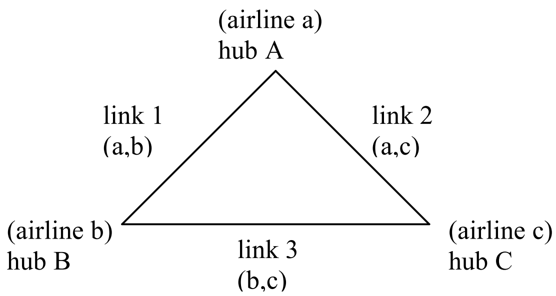

Figure 1. Network configuration

The demand characteristics are given in Table 2; airline characteristics in Table 3, and airport characteristics in Table 4.

[4] The empirical plausibility of the equilibrium values of some key indicators, however, will be investigated.

α_{AB}	70000
α_{AC}	100000
α_{BC}	90000
$\beta_{AB}=\beta_{AC}=\beta_{BC}$	4
γ	100000

Table 2. Demand characteristics

λ_a	200	c_a^f	95000	c_a^p	95
λ_b	225	c_b^f	105000	c_b^p	105
λ_c	200	c_c^f	100000	c_c^p	100

Table 3. Transport network operator characteristics

vot_l	55	η_A	1.54
vot_p	18	η_B	1.7
		η_C	1.65

Table 4. Node characteristics

We thus assume that for each individual operator, the cost per flights and the cost per passenger are equal on all links used. Moreover, $\mathbf{a}_{i,j}$, reflecting constant operator-market specific attributes, is set at 0 for simplicity. Using these inputs, a number of different node pricing schemes are compared. These involve the maximization of global welfare using a common toll (not differentiated over airports) and differentiated (airport-specific) tolls, and a no-cooperation regime in which three local regulators aim to maximize local welfare in a Nash-type of game (*i.e.*, taking the tolls at other airports as given).

Maximizing global surplus

Base scenario: no tolling

First, a "base scenario" is presented, where tolls are fixed at 0; see Table 5 (see Appendix 1 for the definition of "local" welfare[5]).

	q_{AB}	q_{AC}	q_{BC}	Profits ($\times10^8$)	Consumer benefits ($\times10^8$)	Toll rev's	Welfare ($\times10^8$)
a, A	4275	5997	2761	2.680	6.780	0	6.094
b, B	4306	3400	5297	2.508	7.855	0	5.372
c, C	1451	6114	5568	2.934	9.625	0	7.198
Total	10032	15511	13626	8.122	24.261	0	18.663

Table 5. Operator equilibrium and welfare, no tolling

[5] In brief, local welfare comprises the local airline's profits, all toll revenues raised at the local airport, and half of the consumers' surplus for flights to and from the node. The latter implicitly assumes that the equilibrium number of return flights to and from the airport are equally divided between local and non-local residents.

		No toll			Common toll			Specific toll			Local optimization		
	market	AB	AC	BC	AB	AC	BC	AB	AC	BC	AB	AC	BC
	generalized cost	29866.7	37949.7	35486.9	30191.8	38263.6	35813.4	30176.7	38289.7	35794	45949.2	64427.3	55149.7
Share in generalized cost:													
operator a	frequency	9%	6%	8%	10%	6%	8%	10%	6%	8%	17%	7%	-
	congestion	29%	23%	49%	28%	23%	48%	28%	23%	48%	9%	7%	-
	fare	62%	71%	43%	62%	71%	44%	62%	71%	4%	74%	85%	-
	fare elasticity (abs.)	1.079	1.120	1.383	1.103	1.137	1.470	1.100	1.140	1.475	3.24	3.14	-
operator b	frequency	10%	8%	7%	10%	8%	7%	10%	8%	7%	14%	10%	8%
	congestion	29%	46%	25%	28%	45%	25%	28%	45%	25%	9%	14%	9%
	fare	61%	46%	67%	62%	47%	68%	62%	47%	68%	77%	76%	83%
	fare elasticity (abs.)	1.067	1.284	1.128	1.088	1.346	1.145	1.159	1.339	1.143	2.62	89.07*	2.40
operator c	frequency	10%	7%	8%	10%	7%	8%	10%	7%	8%	-	7%	9%
	congestion	60%	23%	25%	58%	23%	25%	58%	23%	25%	-	7%	9%
	fare	30%	70%	67%	32%	70%	67%	32%	70%	67%	-	86%	82%
	fare elasticity (abs.)	1.587	1.084	1.061	1.766	1.100	1.079	1.773	1.103	1.076	-	3.18	2.89

Table 6.a *Specification of generalized costs and fare elasticities*

	No toll	Common toll	Node specific toll	Local optimization
Consumer benefits	2.42606×10^{9}	-0.342%	-0.344%	-39%
Airline cost	4.22165×10^{7}	+49.048%	+49.116%	+1805%
Frequency costs	1.06062×10^{8}	+0.380%	+0.391%	+17%
Congestion costs	4.11464×10^{8}	-2.012%	-2.023%	-74%
Total welfare	1.86632×10^{9}	+0.005%	+0.005%	-8%

Table 6.b *Specification of total welfare*

* This high elasticity is explained by the very low output in this market

Operator *c* makes the highest profits, which is consistent with the fact that this operator can serve the two markets with the highest levels of demand with direct flights. Operator *b* has the largest number of indirect passengers. This is consistent with the fact that the indirect market for operator *b*, AC, is the one with the highest level of demand. Also the local welfare in node *c* is highest, which probably is the combined result of node *c* relatively high levels of demand (Table 2), and a relatively large capacity at the airport (Table 4).

Table 6 provides a breakdown of the consumers' equilibrium generalized costs into the three segments (frequency discount, congestion costs and fare), for all market-airline combinations, and reports equilibrium fare elasticities.[6] These figures are given to help in assessing the plausibility of the parameter constellation used.

The share of the frequency in total generalized costs may seem low, but note that we consider a significant congestion effect. Increasing the frequency effect (by increasing γ) is only possible within certain boundaries. If γ is too high, the model cannot converge to an equilibrium. On the one hand, passengers have a strong preference for a large number of flights, as indicated by γ. On the other hand, a large number of flights increases generalized costs as long as congestion is significant.[7]

The fare elasticities reported in Table 6.a all seem to have reasonable values. Indirect services have more elastic demand. On indirect connections, the fare makes up a relatively small part of the passengers' generalized costs. This would suggest a relatively low fare elasticity for indirect flights: a larger percentage fare decrease is needed to obtain the same absolute increase in demand. However, as indirect demands are relatively small, the same absolute increase implies a relatively large relative increase, suggesting a relatively high fare elasticity for indirect flights. Apparently, the latter effect dominates in our numerical model.

On a final note, optimality of the solution was verified by plotting the objective functions for an interval around the optimum values. For example, π_A was plotted for values of $q_{AB,A}$ (i.e. passengers in market *AB* transported by airline *A*) in the interval <-1000,10000>, keeping all other $q_{i,j}$ fixed, and was strictly concave.

---

[6] Given that $D=\alpha-\beta Q$, we can determine the fare (sub) elasticity as $\Delta Q/\Delta P \times P/Q$, where $\Delta Q/\Delta P=1/\beta$.

[7] The treatment of density economies in this model is in essence similar to the model used by Brueckner and Spiller (1991) (and many authors following Brueckner and Spiller). In the Brueckner and Spiller model, it is quite easy to determine the feasible parameter space using the second order conditions for profit maximization and non-negativity constraints. It is not possible to analytically determine the feasible parameter space in this model. An upper boundary for γ can therefore also not be determined. Optimality of the solution is verified by substituting "non-optimal" values. Details are available upon request.

A common toll

When a common toll, not differentiated over airports, is implemented, the number of indirect passengers (who visit more airports during a trip) is affected more strongly in a relative sense than the number of direct passengers. Table 7 shows that the numbers of transfer passengers transported by operators *a*, *b* and *c* decrease by 3, 2 and 5.5%, respectively.

The level of the toll corresponds on average with about half of the airlines' constant marginal cost per flight. The toll increases the network operators' cost, and the operators compensate partly by increasing the fares (Table 6.a; the fare share in generalized costs increases somewhat). As a result, consumer surplus decreases (Table 6.b). Reduced demand and the possibility that not all "extra" costs (due to the toll) are transferred to the passenger have a negative impact on profits; total airline profits decrease by 1%. The increase in the fare and the costs due to reduced frequencies outweigh the congestion effects in Table 6.a, so that in all markets, the generalized cost increases. The higher overall generalized cost reduces demand somewhat in most markets, as a result of which the generalized costs net of the fare decreases (because congestion decreases) and airline costs (net of the tolls) decrease. These total cost savings outweigh the loss of consumer surplus, so there is an overall increase in total welfare. At node A, however, the cost savings do not outweigh the loss in consumer surplus, so that local welfare decreases.

	q_{AB}	q_{AC}	q_{BC}	profits	consumer benefits	toll rev's (abs.)	welfare
a, A	-0.14	-0.08	-3.04	-1.36	-0.39	7.138×10^6	-0.09
b, B	0.09	-2.03	0.13	-0.95	-0.35	6.771×10^6	0.10
c, C	-5.51	-0.07	-0.07	-0.88	-0.31	7.347×10^6	0.02
Total	-0.82	-0.50	-0.59	-1.06	-0.34	21.256×10^6	0.01

Table 7. Percentage change in equilibrium outputs and welfare (relative to base case in Table 5), second-best optimal common toll = 47760.8

The welfare effects of tolling seem to be relatively small. This is due to the parameter values used. By changing the congestion parameters, larger welfare effects may be obtained (see also below under 'parameter variations'), and tolls may actually be larger than the constant marginal cost per flight. The overall picture, however, does not change.

Brueckner (2002) argued that airlines internalize congestion they impose on themselves. When the number of airlines is limited, a relatively "small" congestion toll would be optimal. A similar argument is valid here. Carriers already internalize congestion they impose on themselves (because congestion is included in the profit functions), and that its passengers impose upon one-another (because this is consistent with profit maximization; see Appendix 2). Furthermore, the two other distortions present in the current setting may also be expected to affect second-best optimal toll levels (differentiated or not). The first is the oligopolistic market structure, implying that quantities offered will be below those that would be offered in absence of market power. This would call for a downward adjustment of tolls, and – when considered in isolation – would call for a subsidy, rather than a tax. The second is the external benefits that passengers impose upon one-another through the depressing impact of higher frequencies on generalized costs. Appendix 2 shows, however, that airlines optimally internalize this externality insofar as occurring among its own passengers. As only firm-internal Mohring effects are considered in our model (see equation 7), this externality will therefore not affect tax levels.

However, in this (network) setting, higher tolls (compared to the situation in the Brueckner paper) may be in order, since airlines impose congestion costs on other airlines and passengers in markets in which they serve only indirectly; operator *a*'s activities on links AB and AC cause delays for all passengers in market BC, and for operators B or C on link BC. Such effects are not included in operator A's profit function. In practical terms, a Continental flight between Newark and Boston causes congestion costs for passengers (on all flights on all airlines) in the (direct and indirect) Boston – Chicago market.

As already noted, local welfare at node A decreases. The redistribution of welfare seems more important than an overall increase in welfare in this case. A regulator responsible for all nodes in a network may thus increase *total* welfare by setting a optimal common toll, but then for some nodes welfare may actually decrease if only locally raised toll revenues can be kept.

Again, optimality of the solution was verified by plotting the objective functions. For different values of the toll, the airline maximization problem was solved, and the regulator's objective function was plotted for tolls falling in the interval <0,80000> (roughly the optimal toll minus and plus 100%), and was strictly concave.

Node-specific tolls

Setting node-specific tolls will generally be the preferred option from an efficiency viewpoint (note that these may always be set at an equal level, so that the welfare level of a common toll is always achievable). The results of the simulations of the node specific tolls are presented in Table 8.

	q_{AB}	q_{AC}	q_{BC}	profits	consumer benefits	toll rev's (abs)	tolls (abs)	welfare
a, A	0.05	-0.32	-3.48	-1.69	-0.37	8.872×10^6	59417.7	0.03
b, B	0.16	-1.38	0.25	-0.43	-0.35	3.847×10^6	27093.9	-0.16
c, C	-6.00	-0.29	0.13	-1.02	-0.31	8.506×10^6	55331.9	0.11
total	-0.78	-0.54	-0.56	-1.06	-0.34	21.225×10^6		0.01

Table 8. Percentage change in equilibrium outputs and welfare (relative to base case in Table 5), optimal node specific tolls

When the regulator sets node-specific tolls, there is a very modest increase in welfare compared to the common toll regime, despite the rather pronounced differentiation in tolls (the toll at node *B* is roughly half of the tolls at nodes *A* and *C*). The welfare gain with differentiated tolls is only 0.0001% higher than with a common toll[8]. The differentiation reflects that with a common toll, the toll will be too high at some nodes and too low at others (unless demand and cost conditions are the same for all nodes and operators). The small relative efficiency gain from introducing differentiation, however, shows that the induced efficiency losses need not always be large.

The local toll leads to a decrease in local welfare at node *B* relative to the base case. The profits of network operator *b* do not decrease as much as profits of either operator *a* or *c*, and are higher than in the case of a common toll. This is primarily caused by the relatively low toll level at airport *B*. For the same reason, however, toll revenues in *B* are relatively low with node-specific tolling, which leads to a relatively large local welfare decrease of (compared to the two other nodes) under the assumption that a node will only keep the locally raised toll revenues. A reverse pattern is observed in *A* and *C*. This example thus shows that even if the overall net welfare effect of a switch from common to differentiated tolling is relatively small, the distributional impacts may be substantial, both over nodes as over actors at these nodes.

Optimality of the solution was verified as above.

Maximizing global surplus, parameter variations

It is of course useful to consider numerical results for different parameter constellations. As the fare elasticities and specifications of generalized costs reported in Table 5 appear reasonable, the demand parameters will not be changed. Instead, we will look at a variation in

[8] In the next section simulations with larger effects will be presented.

congestion parameters; we multiply η_A (the slope of the congestion function) by 0.3, so that the difference in airport capacities becomes relatively large. Results are reported in Table 9.

	q_{AB}	q_{AC}	q_{BC}	profits ($\times 10^8$)	consumer benefits ($\times 10^8$)	toll rev's	welfare ($\times 10^8$)
a, A	4596	6307	4164	3.247	6.940	0	6.992
b, B	4506	3510	4780	2.343	8.062	0	5.449
c, C	1561	6330	5084	2.784	9.811	0	7.360
Total	10663	16147	14028	8.374	24.813	0	19.801

Table 9.a Equilibrium outputs and welfare, no tolling, $\eta_A = 0.462$

	q_{AB}	q_{AC}	q_{BC}	profits	consumer benefits	toll rev's (abs.)	welfare
a, A	1.13	0.95	14.60	10.38	2.24	-0.524×10^8	1.41
b, B	-0.38	11.88	-1.61	6.00	1.96	-0.455×10^8	-0.91
c, C	31.77	0.60	-0.18	5.67	1.78	-0.494×10^8	-0.24
Total	4.98	3.19	3.72	7.59	1.97	-1.473×10^8	0.16

Table 9.b. Percentage change in equilibrium outputs and welfare (relative to base case in Table 9.a), second-best optimal common toll = -287238

	q_{AB}	q_{AC}	q_{BC}	profits	consumer benefits	toll rev's (abs)	tolls (abs)	welfare
a, A	64.51	70.43	-100	112.62	11.59	-6.009×10^8	-3.4529×10^6	6.64
b, B	49.62	-100	74.60	94.95	12.50	-2.633×10^8	-1.7638×10^6	27.54
c, C	-100	65.86	74.69	183.41	11.03	-7.818×10^8	-4.1643×10^6	2.65
total	34.14	31.59	22.80	131.21	11.67	-16.46×10^8		10.91

Table 9.c. Percentage change in equilibrium outputs and welfare (relative to base case in Table 9.a), optimal node specific tolls

When comparing Table 9.a with Table 5, we see that all services involving node *A* as one of the nodes increase in patronage following the expansion, while only the direct services between *B* and *C* suffer, due to the increased competition of the indirect service of airline *a* on this market. Consumer benefits and welfare increase in all nodes, and so do the profits of the most advantaged airline *a*, while profits of *b* and *c* reduce. These changes are rather intuitive.

The reduction in congestion makes the market power component in the second-best common toll dominant, so that it becomes negative (Table 9.b). The effects mirror those in Table 7, in the sense that indirect connections now show the greatest increase due to the subsidy. All consumers benefit, as expected, and so do all airlines, albeit that airline *a* enjoys the greatest relative (further) increase in profits. Apparently, its competitive advantage, caused by its home airport's large capacity, also allows it to profit from the subsidization more strongly than do its competitors. In line with this, node *A* sees an increase in local welfare, while the other nodes see small decreases.

The greatest effects are, however, found for node specific 'tolling' – subsidization, in fact. This induces a change in networks. Surprisingly, and even more so because indirect connections increase the most strongly under the common toll, all indirect connections are terminated with differentiated tolling. It is instructive to follow the sequence in which services are withdrawn from the market, and sometimes re-introduced, in the sequence of local network equilibria that will arise if the regulators, for every given set of services, set tolls optimally.

When differentiated tolls are first implemented following the no-toll equilibrium of Table 9.a, a first naïve calculation produces the expected results of a negative toll at *A* and positive ones at *B* and *C*, reflecting the relative congestion levels. However, in this naïve calculation, *b*'s optimal outputs in markets *BC* and *c*'s optimal outputs in markets *AB* and *BC* are negative, indicating that these airlines will in fact pull out of these markets. When the airlines have done so and tolls are adjusted accordingly, airline *b*'s optimal output in market *AC* turns out to be negative. Fixing also this output at zero leads to a local optimum with positive q's only, where the tolls at nodes *B* and *C* are "highly negative", due to the reduced competition. Tolls at *A*, *B* and *C* are 2.944×10^6, -7.581×10^6 and -9.969×10^6 respectively: the highest toll thus applies at the airport with the highest capacity. This counterintuitive result stems from the fact that competition is more severe at this attractive airport. The low tolls at the other two airports reflect the reduced number of competitors on flights originating or terminating on these airports, so that the distortion due to market power, in itself calling for a subsidy, becomes more severe (only operator *a* serves *BC*, while *AB* is served by *a* and *b*; and *AC* by *a*, *b* and *c*).

This equilibrium, however, is a local optimum only, conditional on this network type. At the margin, the airlines *b* and *c* indeed will not have an incentive to offer services in market *BC* (the market they both served directly, and in which airport *A* is not necessary), so the equilibrium appears stable. Although the airlines would receive a subsidy, entering the market with an arbitrarily small frequency means that the share of "frequency costs" in the passengers' generalized costs are relatively high compared to those of airline *a*, so that a negative price is necessary. Offering a more substantial frequency, however, reduces the share of the frequency costs in the total generalized costs, so that in turn the fare can be high enough to allow a profit in this market[9]. In other words, the networks described above are not globally

[9] Remember that in equilibrium, $D_j(..)=g_{i,j}(..)$. $D_j(..)$, and thus $g_{i,j}(..)$, is the same for all airlines, but the composition of $g_{i,j}(..)$ may differ.

optimal from the airlines' perspective. Airlines *B* and *C* therefore re-enter market *BC*. Because congestion costs are relatively high for airline *a* in market *BC* (because congestion is incurred twice due to the transfer), *a*'s fare then must be negative in the resulting new equilibrium, yielding a loss even after counting the subsidies, and *a* withdraws from its indirect market. All airlines now only serve their two direct markets. This is the final equilibrium, that is presented in Table 9[10]. A check with three other local equilibria, resulting from allowing one airline to serve its indirect market again, revealed that none of the airlines could increase profits by actually doing so, taking into account the equilibrium responses of the other airlines and the regulator.

In the new equilibrium, the withdrawal of all indirect services means that competition has reduced in all markets. As a result, all tolls are negative, so as to give the airlines an incentive to increase the outputs in each market. This more than offsets the loss in output due to reduced competition: outputs increase by at least 23% in each market. Although congestion increases, the negative tolls indicate that the competitive effect outweighs the congestion effect at the margin.

In this case where airport capacities are more diverse, there is a substantial welfare change of 10% compared to the base case following the implementation of differentiated tolls; which by far exceeds the increase in welfare from a common toll (< 1%). The welfare gains result both from an increase in profits and in consumer surplus, which outweigh the necessary tax payments.

Whether or not negative tolls are feasible in practice is not the relevant issue here. The results show that positive tolls may reduce welfare when market power is important, so that for example the most efficient non-negative common toll would be zero, and welfare would be falling in the toll level in that equilibrium. These results further show that tolls may induce network changes. In the new networks, tolls may be negative due to reduced competition. Although one cannot say in general, based on these results, that competition between hub-and-spoke networks is economically inefficient when there are relatively large differences in airport capacity (and prices for the use of the airports do not reflect social marginal costs), it is interesting to see that other studies also find that competition between hub-and-spoke networks is not always an equilibrium outcome. Hendricks et al. (1999) for instance find that

[10] The same optimum is reached when after the first optimization round, only airline *c*'s output in market *BC* is fixed at zero, all optimum outputs are positive. Then all outputs are positive, but airline *b*'s fare in market AC is negative. Airline *b* can obtain a higher profit if it cancels services in market *AC*.

"aggressive" (Bertrand) competition between two carriers may lead to an outcome where a monopoly hub-spoke network dominates or where there is competition between a hub-spoke network and nun-hub networks.

Maximizing local surplus

Although in the above simulations a toll always increases *total* welfare, *local* welfare, as defined in Appendix 1, may actually decrease. As pointed out, this partly depends on the definition of the local welfare (the distribution of consumer surplus over the different nodes is not straightforward, and the assumption is made that only locally raised toll revenues will be kept locally), but there are also other economic reasons for this phenomenon. Local authorities may therefore prefer to set a toll that maximizes local rather than global welfare. This section therefore considers the results of non-cooperative regulation, where all three local regulators maximize local welfare (as defined in Appendix 1) for the network in Figure 1, taking the tolls set by other regulators as given.

The resulting equilibrium (for the base-case parameters) shows a change in networks: operators *a* and *c* withdraw from their indirect markets (Table 10).[11] Operator *b* still has a positive optimal output in the indirect market, although it is relatively small, and the corresponding elasticity is extraordinary large (Table 6.a).

	q_{AB}	q_{AC}	q_{BC}	profits	consumer benefits	toll rev's (abs)	Tolls (abs)	welfare
a, A	-38.36	-26.51	-100	-71.84	-27.67	2.6812×10^{8}	3.69432×1^{6}	-24.70
b, B	-21.57	-95.94	-9.65	-57.45	-31.27	1.51047×10^{8}	2.14806×1^{6}	-31.16
c, C	-100	-28.88	-29.49	-61.93	-29.17	3.66565×10^{8}	4.2977×10^{6}	-12.02
total	-40.07	-42.67	-36.06	-63.82	-29.43	7.85732×10^{8}		-21.67

Table 10. Percentage change in equilibrium outputs and welfare (relative to base case in Table 5), local welfare optimization

The first thing that becomes apparent in Table 10 is that the tolls are much higher than in the case of global welfare optimization (Table 8). The explanation is simple. Whereas toll revenues cancel out in the global maximization problem, they do matter in the local optimization problem. Only toll revenues from the local airline are considered as transfers, but those from the other airlines contribute to the local objective. This leads to relatively high toll

[11] The interior solution for the full model entailed negative outputs on the indirect markets for operators *a* and *c*. The equilibrium in Table 10 is the interior solution for the model with these ouputs set exogenously at zero. It was checked whether either operator *a* (*c*) could make higher profits by waiting for *c* (*a*) to leave the market; but when airline *c*'s (*a*'s) output in market *AB* (*BC*) is fixed at 0, airline *a*'s (*c*'s) optimal output in market *BC* (*AB*) is still negative.

levels, and as a result, airline operational costs increase. The airlines respond by decreasing the number of passengers (flights) in their networks. The "remaining" passengers pay a much higher ticket price, to cover the higher airline operational costs (these costs increase by a staggering 1800%, see Table 6.b).

Looking at Table 6.a, we see that the share of congestion costs in the generalized costs is significantly lower in the case of local welfare optimization. This also becomes clear in Table 6.b; congestion costs decrease by as much as 87%. The share of the fare increases in generalized costs, indicating that the network operators are able to pass (part) of the congestion tolls onto the passengers. The high tolls cause a decrease in network operator profits, but this decrease alone does not explain the decrease in welfare. Consumer surplus also decreases, and the decrease in both aggregate profits and consumer surplus outweigh the increase in toll revenues, so that overall welfare reduces.

Given that local (and global) welfare is much lower under this maximization scheme, the question then is whether local regulators have an incentive to "cooperate". Table 11 presents the local welfare measures under 5 different maximization scenarios: one scenario where all regulator cooperate, one scenario where regulators do not cooperate, and three scenarios where all but one regulators cooperate. The highest possible welfare for each regulator is marked in bold. In each of the scenarios, regulators that jointly maximize welfare do so be setting differentiated tolls.

welfare$\times 10^8$	A, B and C joint	A separate, B and C joint	B separate, A and C joint	C separate, A and B joint	A, B and C separate
A	6.09	**6.83**	5.24	4.41	4.59
B	5.36	4.46	**6.46**	3.90	3.70
C	7.20	5.93	5.92	**8.11**	6.33

Table 11. Local welfare under different welfare maximization scenarios

When the base-scenario is the Nash-equilibrium between local regulators (column 6), all regulators could improve local welfare by entering an agreement with *all* other regulators to jointly maximize welfare. However, *all* regulators could obtain an even higher welfare level when they wait for the *other* (two) regulators to cooperate; in that case, they can reach an even higher welfare level. All regulators have a first-mover disadvantage. For example, when *A* and *B* move first, they can reach a welfare level of 4.41 and 3.90 respectively rather than 6.83 and 6.46 respectively. The regulators thus seem to be caught in sort of a prisoner's dilemma. By entering a global agreement, they can all improve their situation, but they can improve their situation even further by waiting for the other regulators to enter an agreement. By waiting for the others to enter the agreement, in the end no agreement will be negotiated, so that all nodes are worse off.

When base-scenario is an historical agreement between regulators, all regulators could improve local welfare by leaving the agreement, as long as the other regulator(s) will not do so. For example, when *A* leaves the agreement (without any cost), *B* and *C* can only pay *A* to stick to the agreement, or *B* can threaten to also leave the agreement, in which case *A* will be worse off. This, however, could hardly be seen as a credible threat, since *B* would be worse off than in an agreement with only *C*. *C*, however, will have an incentive to leave the agreement if *A* moves first. In any case, there is a first-mover advantage. When *both A* and *C* leave the agreement, thinking the other will not do so, all nodes are worse off. Again, the players are caught in sort of a prisoner's dilemma.

CONCLUSION

Conventional economic wisdom suggests that congestion pricing would be an appropriate response to cope with the growing congestion levels currently experienced at many airports. Several characteristics of aviation markets, however, may make naïve congestion prices equal to the value of marginal travel delays a non-optimal response. This paper has developed a model of airport pricing that captures a number of these features. The model in particular reflects that airlines typically have market power and are engaged in oligopolistic competition at different sub-markets; that part of external travel delays that aircraft impose are internal to an operator and hence should not be accounted for in congestion tolls; that the airlines' consumers may impose external benefits of increased frequencies as well as external congestion costs upon one-another, which however were shown to be internalized by a profit-maximizing firm insofar as these effects occur within its own population of consumers; that different airports in an international network will typically not be regulated by the same authority; and that an individual airline's network will not be exogenous but may instead be affected by congestion levels and tolls, which may create discontinuities. We presented an analytical treatment for an undetermined number of nodes, links and operators in a network of undetermined size and shape, which through the use of 'hyper-networks' would be readily applicable to dynamic problems (in discrete time) such as peak – off-peak differences, and some numerical exercises for an admittedly hypothetical small triangular network, which was only designed to illustrate the possible comparative static impacts of tolling, in addition to marginal equilibrium conditions as could be derived for the general model specification. Although the paper considered a congested aviation network, the insights developed may of course often carry over to congestion pricing at nodes for different modes than aviation, and possibly even to congested networks other than for transport, provided market conditions are similar to those considered here. Certainly, economies of scale, scope, density and networks will make the occurrence of market power in networks not be a feature specific to aviation, and the fixedness of capacity in relation to variability of demand makes congestion a phenomenon relevant for most contemporary network markets.

Some main conclusions are that second-best optimal tolls are typically lower than what would be suggested by congestion costs alone and may even be negative, that pricing may induce changes in network configurations, and that cooperation between regulators need not be stable, but that non-cooperation may lead to welfare losses, also when compared to a no-tolling situation.

The joint occurrence of multiple market distortions, the network environment, and the discreteness of an airline's network choice, may lead to considerable complications in the determination of optimal tolls, and sometimes unexpected outcomes. One example encountered was where the increase in an airport's capacity may make the tolls on other airports reduce more than on that airport itself. The reason is that network changes may increases market power and reduce congestion on these other airports, both calling for downward adjustments of tolls.

While Brueckner (2002) has made clear that congestion tolls on airports may be smaller than expected when congestion costs among aircraft are internal for a firm, our analysis adds to this that a further downward adjustment may be in order due to market power, and that unexpected effects from tolling may occur due to the endogeneity of airlines' network choices. As an aside, we have demonstrated that an effect similar to that identified by Brueckner (2002) occurs insofar as an airline's customers impose externalities upon one-another, the cases considered being congestion and the Mohring effect. As profit maximizing firms internalize these externalities, government intervention would be warranted only insofar as similar externalities would exist between competing firms' customers (Appendix 2).

The various downward adjustments in tolls may well cause the optimal values of these to be negative. Insofar as subsidization is considered unacceptable for whichever reason, our results warn that the most efficient among the non-negative tolls may actually be a zero toll.

Other immediate policy implications include the demonstrated potential need for toll coordination over an entire network, and for the serious consideration of impacts of tolling on airlines' network choices and on the related competitive conditions on the various sub-markets over a network.

The model in this paper contains a few simplifying assumptions that may be relaxed in future work. Load factors and aircraft capacity are fixed in this model for simplicity. In a more advanced version of this model, load factors and aircraft capacity can be endogenized. This makes the derivation of the optimality conditions far more complicated, but it should be feasible in a numerical experiment. One can also add a fourth layer to the model, describing the airport's optimization problem. For example, the airport can maximize profits under a cost recovery constraint. The model then deals with interactions between four types of agents. Finally, no distinction is made between peak and off-peak traffic in this paper. This distinction is quite common in the literature (see e.g. Brueckner (2002), Daniel (1995)) and could, as discussed, make a straightforward but important extension of the model in this paper.

REFERENCES

Adler, N. and J. Berechman (2001), Evaluating optimal multi-hub networks in a deregulated aviation market with an application to Western Europe, *Transportation Research*, **35A**, 373-390.

Baumol, W.J. and W.E. Oates (1988), *The theory of environmental policy*, second edition, Cambridge University Press

Brander, J.A. and A. Zhang (1990), Market conduct in the airline industry: an empirical investigation, *Rand Journal of Economics*, **21**(4), 567-583

Brueckner, J.K. (2001), The economics of international codesharing: an analysis of airline alliances, *International Journal of Industrial Organization*, **19**, 1475-1498

Brueckner, J.K. (2001), Airport congestion pricing when carriers have market power, *American Economic Review*, forthcoming.

Brueckner, and P.T. Spiller (1991), Economics of traffic density in a deregulated airline industry, *International Journal of Industrial Organization*, **37**, 323-342

Buchanan, J.M. (1969), External diseconomies, external taxes, and market structure, *American Economic Review*, **59**, 174-177

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

Daniel, J.I. (2001), Distributional consequences of airport congestion pricing, *Journal of Urban Economics*, **50**, 230-258.

Daniel, J.I. (1995), Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues, *Econometrica*, **63**, 103-130

Hendricks, K., M. Piccione and G. Tan (1999), Equilibria in networks, *Econometrica*, **67**, 1407-1434

Mohring, H. (1972), Optimization and scale economies in urban bus transportation, *American Economic Review*, **62**, 591-604

Oum, T.H., A. Zhang and Y. Zhang (1996), A note on optimal airport pricing in a hub-and-spoke system, *Transportation Research*, **30B**(1), 11-18

Oum, T.H., A. Zhang and Y. Zhang (1993), Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets, *Journal of Transport Economics and Policy*, **27**, 171-192

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

Pels, E., P. Nijkamp and P. Rietveld (2000), A note on the optimality of airline networks, *Economics Letters*, **69**(3), 429-434

Rassenti, S.J., V.L. Smith and R.L. Buffin (1982), A combinatorial auction mechanism for airport time slot allocation, *Bell Journal of Economics*, **13**(2), 402-417

Verhoef, E.T., P. Nijkamp and P. Rietveld (1996), Second-best congestion pricing: the case of an untolled alternative, *Journal of Urban Economics*, **40**(3), 279-302

Verhoef, E.T. (2002), Second-best congestion pricing in general static transportation networks with elastic demands, *Regional Science and Urban Economics*, **32**, 281-310

APPENDIX 1 LOCAL WELFARE

In the analysis, a measure of "local welfare" is used. This is half of the consumer surplus in markets having airport h as the origin or destination (we conveniently assume that the aggregate demand for return trips on each origin-destination pair is always made up by equal amounts of passengers from both cities, so that the other half of consumer surplus in a market is maximized by the authority regulating the complementary airport), plus the profits of the airline i using airport h as its hub, plus the locally raised tax revenues:

$$\varpi_{\mathrm{h}}^{\mathrm{L}} = \frac{1}{2}\left[\sum_{\mathrm{j=1,j}\in J(\mathrm{h})}^{\mathrm{J}} \delta_{\mathrm{j,h}} \int_{0}^{\sum_{\mathrm{i=1}}^{\mathrm{I}} \mathrm{q_{i,j}}} D_{\mathrm{j}}(x)\mathrm{d}x - \sum_{\mathrm{i=1}}^{\mathrm{I}} \sum_{\mathrm{j=1,j}\in J(\mathrm{h})}^{\mathrm{J}} \mathrm{q_{i,j}} D_{\mathrm{j}}(..)\right] + \sum_{\mathrm{i=1,}\in I(\mathrm{h})}^{I} \pi_{\mathrm{i}} + \sum_{\mathrm{i=1}}^{\mathrm{I}} \sum_{\mathrm{k=1}}^{\mathrm{K}} \left(\sum_{\mathrm{j=1}}^{\mathrm{J}} \delta_{\mathrm{i,j,k}} \mathrm{q_{i,j}} \frac{\delta_{\mathrm{k,h}} \mathrm{t_h}}{\lambda_{\mathrm{i}}} \right) \quad \text{(A1.1)}$$

where J(h) is the set of markets with node h as origin or destination and I(h) is the set of carriers with node h as hub. The first term on the RHS is passenger surplus. The second term between brackets is total generalized costs, which, for alternative (i,j) equals $q_{i,j}$ multiplied by the inverse demand function for market j; in equilibrium $g_{i,j}(..)=D_j(..)$. The second term is carrier surplus. The final term on the RHS are the toll revenues.

This function could be the objective function of a local regulator. An implicit solution here is that cities are of equal size; hence consumer surplus can be divided equally over two nodes. Other specifications are of course possible, but then may require assumptions on the redistribution of toll revenues. In all cases $\sum_{\mathrm{h=1}}^{\mathrm{H}} \varpi_{\mathrm{h}}^{\mathrm{L}} = \varpi$; this was verified numerically.

APPENDIX 2 PROFIT MAXIMIZATION AND THE INTERNALIZATION OF THE MOHRING EFFECT, CONGESTION EXTERNALITIES BETWEEN A FIRM'S PASSENGERS, AND CONGESTION EFFECTS AMONG A FIRM'S AIRCRAFT

This appendix will consider, in a simple setting, the question of whether a profit maximizing firm will optimally internalize the Mohring externality, congestion externalities between its passengers, and congestion effects between its aircraft (the latter has been shown already by Brueckner, 2002). To focus on the essentials, we consider operations on a single market with a direct connection only, and ignore competition between airlines by only considering a profit maximizing airline that faces a downward sloping inverse demand function $D(q)$, for services offered at a frequency $f=q/\lambda$ (as in the main text, we consider a fixed passenger load λ).

Customers' generalized costs are the sum of the ticket price p, per-passenger 'frequency costs' $c_f^p(f,\cdot)$ with $\partial c_f^p(f,\cdot)/\partial f<0$ that represent the Mohring effect, and per-passenger congestion costs $c_c^p(f,\cdot)$ with $\partial c_c^p(f,\cdot)/\partial f>0$.

The airline, too, suffers from congestion, which is reflected by the per-flight term $c_c^a(f,\cdot)$ with $\partial c_c^a(f,\cdot)/\partial f>0$, and in addition incurs constant variable costs per flight equal to c_f^a. The airline's profit function can therefore be written as:

$$\begin{aligned}\Pi &= p\cdot q - f\cdot c_c^a(f,\cdot) - f\cdot c_f^a \\ &= \left(D(q) - c_c^p(\frac{q}{\lambda},\cdot) - c_f^p(\frac{q}{\lambda},\cdot)\right)\cdot q - \frac{q}{\lambda}\cdot c_c^a(\frac{q}{\lambda},\cdot) - \frac{q}{\lambda}\cdot c_f^a\end{aligned} \qquad \text{(A2.1)}$$

The profit-maximizing output can be found by setting the derivative with respect to q equal to zero:

$$\frac{\partial \Pi}{\partial q} = \left(D - c_c^p - c_f^p\right) + q\cdot\left(\frac{\mathrm{d}D}{\mathrm{d}q} - \frac{1}{\lambda}\cdot\frac{\partial c_c^p}{\partial f} - \frac{1}{\lambda}\cdot\frac{\partial c_f^p}{\partial f}\right) - \frac{c_c^a}{\lambda} - \frac{f}{\lambda}\cdot\frac{\partial c_c^a}{\partial f} - \frac{c_f^a}{\lambda} = 0 \qquad \text{(A2.2a)}$$

Observing that the first term on the rhs, in brackets, is equal to the market price, this implies the following profit maximizing price p^{Π}:

$$p^{\Pi} = q\cdot -\frac{\mathrm{d}D}{\mathrm{d}q} + \frac{1}{\lambda}\cdot\left(q\cdot\frac{\partial c_c^p}{\partial f} + q\cdot\frac{\partial c_f^p}{\partial f} + c_c^a + f\cdot\frac{\partial c_c^a}{\partial f} + c_f^a\right) \qquad \text{(A2.2b)}$$

Social surplus can be written as:

$$W = \int_0^q D(x)dx - q\cdot\left(c_c^p(\frac{q}{\lambda},\cdot) + c_f^p(\frac{q}{\lambda},\cdot)\right) - \frac{q}{\lambda}\cdot c_c^a(\frac{q}{\lambda},\cdot) - \frac{q}{\lambda}\cdot c_f^a \qquad \text{(A2.3)}$$

Surplus maximization requires:

$$\frac{\partial W}{\partial q} = D - c_c^p - c_f^p - q\cdot\left(\frac{1}{\lambda}\cdot\frac{\partial c_c^p}{\partial f} + \frac{1}{\lambda}\cdot\frac{\partial c_f^p}{\partial f}\right) - \frac{c_c^a}{\lambda} - \frac{f}{\lambda}\cdot\frac{\partial c_c^a}{\partial f} - \frac{c_f^a}{\lambda} = 0 \qquad \text{(A2.4a)}$$

Observing that the first three terms on the rhs again equal the market price, this implies the following surplus maximizing price p^W:

$$p^W = \frac{1}{\lambda}\cdot\left(q\cdot\frac{\partial c_c^p}{\partial f} + q\cdot\frac{\partial c_f^p}{\partial f} + c_c^a + f\cdot\frac{\partial c_c^a}{\partial f} + c_f^a\right) \qquad \text{(A2.4b)}$$

A comparison between (A2.2a) with (A2.4a), and (A2.2b) with (A2.4b), reveals that the profit maximizer not only optimally internalizes the congestion effects that its aircraft impose on each other (as in Brueckner, 2002), but also the congestion externalities and the Mohring effect that its passengers impose upon each other. Surely the marginal values of these externalities will generally differ between the optimum and profit maximizing equilibria as q and f will take on different equilibrium values – but given these values, perfect internalization will prevail. The difference between the two equilibria is caused by the first term in (A2.2b), which represents the monopolistic mark-up in the profit-maximizing price.

The intuition behind this result is similar to that behind the internalization of congestion externalities by a private road operator (*e.g.* Verhoef *et al.*, 1996). Internalization by pricing enables the extraction of additional revenues from intra-marginal users in exchange of reduced congestion costs without affecting their generalized price. And revenues add to profits, while congestion costs do not. As the Mohring effect is analytically identical to a congestion externality – albeit that it has the opposite sign – it is not surprising that the same principle applies.