AIRLINE ROUTE STRUCTURE COMPETITION AND NETWORK POLICY

Hugo E. Silva, VU University, h.silvamontalva@vu.nl
Erik T. Verhoef, VU University, e.t.verhoef@vu.nl
Vincent A.C. van den Berg, VU University, v.a.c.vanden.berg@vu.nl

ABSTRACT

This paper studies pricing policies and route structure choices by carriers with market power in a network setting and in the presence of congestion externalities. We account for passenger benefits from increased frequency, passenger connecting costs, airline endogenous hub location and route structure strategic competition. The analysis shows that an instrument directly aimed at regulating route structure choice may be needed to maximize welfare, in addition to per-passenger and per-flight tolls designed to correct output inefficiencies. This holds true even when the regulator is does not face any constraint on pricing.

Keywords: Route structure, hub-spoke networks, airport pricing

1 INTRODUCTION

Following the deregulation of the airline industry, several changes in aviation markets were observed. In addition to changes in fares, the most notorious change was in the way markets were served: the adoption of hub-and-spoke (H, henceforth) route structures by carriers became dominant. Such a decision by carriers has often been explained with three arguments: economies of density, frequency effects and strategic advantages. The first refers to the fact that average cost in a direct route decreases with the number of passengers, and the second to the fact that there are benefits for passengers of increased frequencies, e.g. reductions in schedule delay costs (the difference between desired and actual departure/arrival time). Both may be better exploited under H structures. The third argument, strategic advantages, refers to the fact that adopting H route structures may bring further advantages because of the effect it may have on competitors.

The outcomes of a deregulated environment where carriers can choose how to serve markets have been well studied in the literature. On the other hand, literature on pricing and regulation in aviation markets has mostly focused on either a single origin destination pair, hence ignoring network effects, or in networks where carriers have fixed route structures, hence ignoring its endogenous nature and its effect on optimal policy.

The objective of this paper is to extend the pricing and regulation analysis by elaborating on policy instruments that can induce the social efficient outcome in a network setting, where carriers with market power choose a route structure in the presence of congestion externalities. It is known from earlier literature, which abstracts away from endogenous route structure, that oligopolistic carriers partially internalize congestion and exert market power. This means that two inefficiencies need to be corrected: the dead-weight loss from market-power markups (e.g. with subsidies) and the excessive number of flights that are scheduled (e.g. with slot constraints or congestion pricing). In this paper, we study whether and how the inclusion of route structure choice by carriers changes these conclusions. Specifically, do regulators need an additional instrument, on top of the ones described above, to induce the socially desirable outcome? We carry out the analysis in the simplest possible setting that allows us to account for strategic interactions in route structure choice, endogenous hub locations, market power exertion by airlines, congestion externalities at airports, and passenger frequency benefits and transfer costs.

The main result of our analysis is that a regulatory instrument directly targeted on route structure choice may be needed to maximize welfare, in addition to tolls designed to induce the efficient outputs. We find that social welfare can be increased by using an additional policy instrument even when the regulator does not face any constraint on tolling. Specifically, the first-best optimal route structures and output levels cannot always be enforced by just using an airline- and market-specific per-passenger toll (to correct for market power), together with an airline- and link-specific per-flight toll (to correct for congestion), designed to induce the efficient output for the optimal route structure. Thus, the equilibrium with those tolls is not always efficient, even when the regulator can perfectly discriminate airlines and has no pricing constraints.

2 The model

In order to keep the simplest possible focus on the route structure choice by agents with market power in presence of externalities, we use a stylized model that follows Brueckner's (2004) in the basic assumptions, and extends it by considering congestion, airline competition and the analysis of how to enforce the social optimum.

We consider a symmetric duopoly of airlines that compete in each of the three symmetric markets that are shown in Figure 1. These markets $M = \{AB, BC, AC\}$ represent return-trips for simplicity (e.g. people travel from A to B and return). The links $L = \{ab, bc, ac\}$ are always available to any airline; that is, both airlines have permission to schedule flights between any city-pair. Each market m can be served by airlines either directly, flying non-stop from the origin airport to the destination airport, or via a hub airport that an airline chooses to use for the connection. As a result, the two possible route structures for an airline are: fully connected (henceforth F), which implies having flights in all three links; or hub-and-spoke, where they choose one airport as its hub, and fly only between the hub and the two remaining airports, serving two markets non-stop and one with connecting flights. We let each airline's hub to be endogenous, therefore asymmetric settings with hub-and-spoke structures may arise.

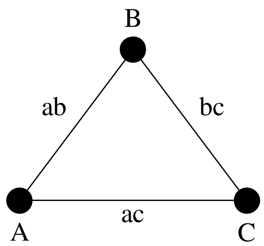


Figure 1: Network

We model the airlines' competition with a two-stage game where, first, carriers simultaneously choose route structure, and then they compete in output at a market level. Specifically, in the latter stage, airlines have the number of passengers in each market (q_m), number of flights in each link (f_l), and aircraft size (s_l) as strategic variables; this is an extension of the Cournot assumption that airlines take rival's quantities, instead of fares, as given.[1]

We assume that the full price faced by a passenger when traveling with airline i is:

$$\theta_i^m = p_i^m + D_i^m + g_i^m \, . \tag{1}$$

This is simply the sum of the fare, p_i^m, the congestion delay cost, D_i^m, and schedule delay cost, g_i^m.[2] We further assume that airlines are perceived as imperfect substitutes and that the passenger demand function for an airline i in market m, q_i^m, is linear in own and the rival's price. Therefore, the demand faced by the airline depends on its own full price, θ_i^m, as well as its rival's, θ_j^m (hereafter, when subscript j appears in the same expression with i, it refers to the rival airline). These assumptions are summarized in the following inverse demand structure:

$$\theta_i^m = A - B \cdot q_i^m - E \cdot q_j^m \, , \tag{2}$$

where A, B, and E are positive parameters satisfying $0 \leq E \leq B$. Note that we ignore demand dependencies between markets (city-pairs). This set of assumptions allows us to analyze the effect

[1]The assumption that airlines compete in a Cournot fashion is common in the airline literature and supported by empirical evidence by Brander & Zhang (1990) and Oum et al. (1993).

[2]The schedule delay is the time difference between a passengers desired departure time and the actual departure time.

of airline horizontal differentiation on route structure choice by means of varying the ratio of substitutability, E/B, that ranges from 0 when the airlines' outputs are independent in terms of demand interaction to 1 when the airlines' outputs are perfect substitutes; at the same time,we consider that not all passengers choose the airline with the most attractive fare-delay combination due to factors that may differ across carriers, such as service level (e.g. language), and make passengers perceive airlines as imperfect substitutes.

Following Brueckner (2004), we model the airlines' cost per flight as a function of aircraft size (s):

$$C(s) = c_f + c_q \cdot s \,, \tag{3}$$

where c_f is the fixed cost per flight, and c_s the marginal cost per seat. This formulation captures in a simple way that increasing the number of passengers per link may reduce average cost per passenger through economies of seats. We also assume a constant load factor of 100%.

As a natural benchmark, we consider a regulator that controls all airports and maximizes welfare, so that we analyze a three-stage game: in the first stage, the regulator set per-passenger tolls to each airline in each market (τ_i^m), and per-flight tolls to each airline in each link (τ_i^l). In the second and third stage, airlines choose route structure and output respectively. We look at sub-game perfect equilibria through backward induction, so we first analyze the airlines' Nash equilibria.

3 Airlines equilibrium

As airlines, in the second stage, have a discrete choice between alternative route structures (F or H at any of the airports) we need to look first at their profits taking route structures as given, and then analyze the equilibrium in route structure.

3.1 The fully connected route structure

In this setting, airline i uses F as its route structure, and has as strategic variables the frequency on each link, f_i^l; and the number of passengers in each market, q_i^m. The seats per flight are $s_i^l = q_i^l / f_i^l$, where q_i^l is the number of passengers in link l, because there is no gain of having spare capacity.

We assume that the average schedule delay depends only on the flight frequency of the airline in the link that connects that market, and that it decreases with frequency (e.g. $\partial g_i^{AB}/\partial f_i^{ab} < 0 \wedge \partial g_i^{AB}/\partial f_i^l = 0 \; \forall \; l \neq ab \wedge \partial g_i^{AB}/\partial f_j^l = 0 \; \forall \; l$). The assumption that schedule delay does not depend on the rival's frequency, as congestion does, reflects our view that, in the differentiated duopoly, frequency is perceived as an airline-specific attribute. We also assume that there is congestion at the origin and destination, that airport runway congestion depends on total number of flights at that airport, and that it increases in the total number of flights. For example, denoting $F^l = f_i^l + f_j^l$ the total number of flights on link l, the full price faced by a passenger of market AB flying with i is:

$$\theta_i^{AB} = p_i^{AB} + D(F^{ab} + F^{ac}, K_A) + D(F^{ab} + F^{bc}, K_B) + g_i^{AB}(f_i^{ab}) \,, \tag{4}$$

where D is the delay cost function, assumed common to all airports. The congestion at the origin A depends on the total number of flights operating at A ($F^{ab} + F^{ac}$) and the airport's capacity (K_A).[3]

[3]Without loss of generality, both delay functions, D and g, include the passengers valuation of time.

We look at the particular case where all airports have the same capacity, but this could easily be extended. The airline's profit, using (2), (3) and $s_i^l = q_i^l / f_i^l$, is:

$$\pi_i^F = \sum_{m \in M} q_i^m \cdot (A - B \cdot q_i^m - E \cdot q_j^m - g_i^m - D_i^m - c_q - \tau_i^m) - \sum_{l \in L} f_i^l \cdot (c_f + \tau_i^l) . \tag{5}$$

where the superscript F refers to the fully connected route structure. Airline profit indirectly depends also on the rival's route structure. That structure determines the rival's number of passengers and number of flights, which will affect demands and delays. What we do, in this section, is to look at the airlines' best response in output irrespective of the rival's quantities, and then, when deriving the equilibrium in route structure, compare the differences that arise from the different rival's route structure choices. The first-order conditions for q_i^m and f_i^l imply the following pricing and frequency setting rules:

$$\frac{\partial \pi_i^F}{\partial q_i^m} = 0 \Rightarrow \ p_i^m = c_q + \tau_i^m + B \cdot q_i^m , \tag{6}$$

$$\frac{\partial \pi_i^F}{\partial f_i^l} = 0 \Rightarrow \ - \sum_{m \in M} q_i^m \cdot \left(\frac{\partial D_i^m}{\partial f_i^l} + \frac{\partial g_i^m}{\partial f_i^l} \right) = c_f + \tau_i^l . \tag{7}$$

Equation (6) states that the fare charged by the airline in market m is the sum of the marginal cost per capacity unit (c_q), the airport charge per passenger in that market (τ_i^m), and a conventional markup reflecting carrier market power ($B \cdot q_i^m$). Equation (7) states that airline's marginal cost per flight (right-hand side of (7)) equals marginal revenue (left-hand side, marginal congestion costs plus marginal schedule delay benefits); therefore, airlines internalize own-passenger congestion. These rules are analogous to the rules obtained previously in Cournot competition (e.g. Pels & Verhoef 2004). The airline's profit using a fully connected route structure, in sub-game equilibrium, is given by:

$$\Pi_i^F = \sum_{m \in M} B \cdot (q_i^m)^2 - \sum_{l \in L} f_i^l \cdot (c_f + \tau_i^l) , \tag{8}$$

which is obtained by replacing Eq. (6) into Eq. (5).

3.2 The hub and spoke route structure

We now look at the case where an airline (i) chooses to serve the markets with a hub-and-spoke route structure. For illustration, we use airport B as the hub. Other cases are simply obtained by changing notation only. When we study the full game equilibrium, then it is necessary to explicitly model the choice of hub airport. The changes with respect to the fully connected case is that the market AC (the spoke market) is served with connecting flights at the hub. We assume, as in previous studies of hub-and-spoke route structures, that the fare for the spoke market is set independently; this implies that the fare for market AC is not restricted to be equal to the sum of the fares of the two hub markets (AB and BC). The fares must, however, satisfy the arbitrage condition: the sum of the fares of the hub markets (in this case, AB and BC) cannot be lower that the fare charged to the connecting passengers (market AC).

The number of seats per flight, on each link, changes in this case because the passengers from the spoke market are also traveling through links ab and bc. As a consequence, in this setting, aircraft sizes will satisfy:

$$s_i^{ab} = (q_i^{AB} + q_i^{AC}) / f_i^{ab} \ , \wedge s_i^{bc} = (q_i^{BC} + q_i^{AC}) / f_i^{bc} . \tag{9}$$

Full prices in the hub markets have the same structure as before (see (4)), but they change in the spoke market. We assume that the passengers' congestion delay is the sum of the delays at each leg, and that schedule delay cost for a "hubbing" passenger is the sum of the schedule delays of a passenger flying both legs. This is:

$$\widetilde{D}_i^{AC} = D_i^{AB} + D_i^{BC} \wedge \widetilde{g}_i^{AC} = g_i^{AB} + g_i^{BC} \; . \tag{10}$$

The definition of $\widetilde{g}_i^{AC}$ reflects our assumption that g captures the schedule delay cost (first term, g_i^{AB}), and that is also able to capture the transfer costs at the hub (second term, g_i^{BC}). This is a simple way to model the fact that a passenger incurs an additional cost from connecting and that the cost is lower when the frequency of connections is higher. Although there is no reason why the transfer cost should be exactly equal to the schedule delay cost incurred by a passenger traveling only the second leg, it is a convenient assumption as for unequal frequencies it is hard to express the frequency at the origin-destination level.[4] These assumptions shape profit in the following way:

$$\begin{aligned}\pi_i^H = & \sum_{m \in \{AB,BC\}} q_i^m \cdot (A - B \cdot q_i^m - E \cdot q_j^m - g_i^m - D_i^m - c_q - \tau_i^m) + \\ & q_i^{AC} \cdot (A - B \cdot q_i^{AC} - E \cdot q_j^{AC} - \widetilde{g}_i^{AC} - \widetilde{D}_i^{AC} - 2 \cdot c_q - T_i^{AC}) - \sum_{l \in \{ab,bc\}} f_i^l \cdot (c_f + \tau_i^l) \; ,\end{aligned} \tag{11}$$

where the superscript H refers to the profit when an airline is serving the markets with a hub-and-spoke route structure. The difference, besides the new definition of delays, is that, everything else constant, an additional passenger in the AC market requires an increase of aircraft size in both links, hence the cost per passenger is $2 \cdot c_q$.

The first-order conditions in this case lead to the following pricing and frequency setting rules:

$$\frac{\partial \pi_i^H}{\partial q_i^m} = 0 \Rightarrow p_i^m = c_q + \tau_i^m + B \cdot q_i^m \quad \forall m \in \{AB, BC\} \; , \tag{12}$$

$$\frac{\partial \pi_i^H}{\partial q_i^{AC}} = 0 \Rightarrow p_i^{AC} = 2 \cdot c_q + T_i^{AC} + B \cdot q_i^{AC} \; , \tag{13}$$

$$\frac{\partial \pi_i^H}{\partial f_i^l} = 0 \Rightarrow - \sum_{m \in \{AB,BC\}} (q_i^m + q_i^{AC}) \cdot \left(\frac{\partial D_i^m}{\partial f_i^l} + \frac{\partial g_i^m}{\partial f_i^l} \right) = c_f + \tau_i^l \; \forall \, l \in \{ab, bc\} \, . \tag{14}$$

These set of equations basically state that airlines apply a market power markup in each market and set frequency to equalize own marginal revenue with own marginal cost, hence partially internalizing congestion. In this setting, the sub-game equilibrium profit for an airline, using the first-order conditions (Eqs. (12) and (13) in Eq. (11)), can be written as:

$$\Pi_i^H = \sum_{m \in M} B \cdot (q_i^m)^2 - \sum_{l \in \{ab,bc\}} f_i^l \cdot (c_f + \tau_i^l) \, . \tag{15}$$

Just as in the previous case, it is the revenues from the markup minus frequency costs that are not charged to passengers.

[4] An alternative assumption used in the literature is that passengers incur a fixed cost when transferring. The results do not change significantly if we assume this instead.

3.3 Second-stage: the choice of route structure

In contrast to the third-stage, in this stage the airlines' decision variables are discrete. Airlines can either choose to serve the markets with a fully connected route structure or with a hub-and-spoke route structure. When an airline chooses to use a hub-and-spoke structure, it also chooses which airport to use as the hub. To characterize the equilibria, we need to compare the airlines' best responses, knowing the outcome of the third stage (quantity and frequency), for all the rival's possible route structures.

Denote the route structure choice of an airline i as a choice of $r_i \in RS = \{F, H_A, H_B, H_C\}$, where H_x refers to a hub-and-spoke structure with airport X as the hub. The relevant comparisons are the profits given the route structure of the rival. Let $\Pi_i(r,v)$ be the airline's i profit, evaluated at the outputs of the third-stage equilibrium, when it has chosen r as its route structure, and the rival uses v. Then, it is straightforward that a symmetric setting with both airlines using route structure r will be an equilibrium of the airlines' game if and only if:

$$\Pi_i(r,r) \geq \Pi_i(u,r) \;\; \forall\, u \in RS \,. \tag{16}$$

Because airlines are symmetric, whenever this holds true for one airline, it will hold true for the other as well, and both airlines having r will be a perfect sub-game equilibrium of the airlines' competition. Also note that, whenever both airlines choose H_A in equilibrium, both having H_B and both having H_C are also equilibria, because markets are symmetric as well. We will refer to this set of equilibria as (H,H): both airlines using hub-and-spoke route structures and both using the same airport as their hub. It follows that (F,F) is the equilibrium where both airlines choose the fully connected route structure. As hub location is endogenous, asymmetric equilibria where airlines use different hubs may arise. We denote this set of possible equilibria as (H_x,H_y), regardless of the location of the airlines' hubs. Asymmetric equilibria, with one airline choosing route structure u and the other v, will arise if and only if the following holds:

$$\Pi_i(u,v) \geq \Pi_i(w,v) \;\wedge\; \Pi_i(v,u) \geq \Pi_i(y,u) \;\; \forall\, w \in RS\,, \forall\, y \in RS \;\, u \neq v \,. \tag{17}$$

This implies that u is the best response when the rival chooses v and *vice versa*, which again, because of airline and market symmetry, implies that there are multiple equilibria for a particular u and v. We denote (F,H) to the asymmetric equilibria where one airline serves the markets with a fully connected route structure and the other with a hub-and-spoke route structure, regardless of the hub airport choice.

Note that the above conditions, (16) and (17), are not necessarily mutually exclusive, therefore multiple sets of equilibria may arise. For instance, it may be the case that for a certain parameter constellation (F,F) and (H_x,H_y) are both equilibria. Despite having a highly stylized model, these comparisons are hard to perform analytically. To surpass this, we look at some of the relevant equilibrium conditions that, together with numerical examples, allows us to solve the equilibrium, provide intuition and compare our results to those in previous literature.

First, we look at the expression that makes using a fully connected structure a best response to the rival using fully connected as well, $\Gamma_i \equiv \Pi_i(F,F) - \Pi_i(H_B,F)$. Using Eqs. (8) and (15), it is:

$$\Gamma_i = \sum_{m\in M} B \cdot \left[(q^m_{i|(F,F)})^2 - (q^m_{i|(H_B,F)})^2 \right] + c_f \cdot \left[\sum_{l\in\{ab,bc\}} f^l_{i|(H_B,F)} - \sum_{l\in L} f^l_{i|(F,F)} \right] , \tag{18}$$

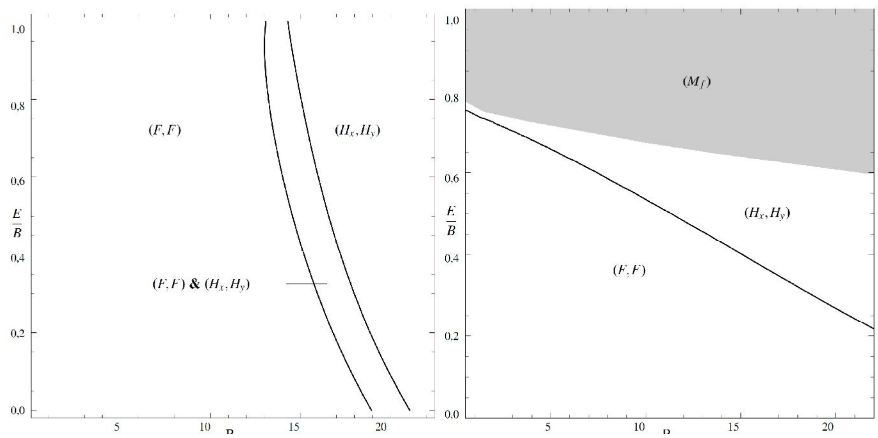
where the variables in Eq. (18) are evaluated at the untolled equilibrium with route structures indicated in parentheses. Eq. (18) shows that there are two effects driving the adoption of fully connected over hub-and-spoke: the change in revenues, as a result of the change in the number of passengers in all three markets (first bracketed term in the right-hand side), and the change in costs, due to variations in the total number of flights (second bracketed term in the right-hand side). Despite that is not possible to assess the sign of Γ_i analytically, the sign of each term is intuitive. Hub-and-spoke networks are meant to save airline costs through to a reduced number of links flown, thus, a natural expectation is that the total number of flights is reduced when moving from fully connected to hub-and-spoke (Brueckner 2004). Is is, therefore, expected that the second bracketed term is negative, so that it favors the adoption of a hub-and-spoke route structure over fully connected. The change in number of passengers in each market, however, may favor the point to point structure. To see this, note that in the connecting market (AC in this case), the passengers face a higher full price under H_B than under F, because they incur higher travel delays and the cost of connecting (see Eq. (10)). As a result, the equilibrium number of passengers in the connecting market (for a given route structure of the rival) should be higher under a fully connected route structure. On the other hand, the full price in the remaining two markets that are serve directly under both structures (markets AB and BC) may be higher or lower due to two counteracting effects: a higher (lower) frequency under hub-and-spoke with respect to fully connected in each link implies higher (lower) congestion, but decreased (increased) schedule delay costs. Therefore, its sign is, *a priori,* ambiguous. However, numerical analyses show that, for the considered parameters, the number of passengers in these markets that are higher under a hub-and-spoke route structure than under a fully connected structure, given that the rival uses a fully connected structure; therefore, this effect also favors hub-and-spoke route structures.

Brueckner (2004) already showed that this expression, for a monopoly and in absence of congestion, can be positive or negative depending on parameters, and that the own-demand sensitivity parameter plays a key role. Therefore, a meaningful exercise is to analyze the effect of competition on the indifference point between F and H. For this purpose, the case with independent products is very useful. When airlines are independent, the only interdependency is through congestion; thus, by looking at cases with $E = 0$, we can identify the indifference point in absence of strategic interaction. Then, by varying the ratio of substitutability E/B keeping all other parameters constant, it is possible to assess the effect of competition on the choice of route structures.

Figure 2a shows the equilibria for a wide range of the own-demand sensitivity B parameter (horizontal axis) and all possible values of the ratio of substitutability E/B (vertical axis), for a particular parameter constellation and functional forms (see A for details). In our model, the second-order conditions involving the cross derivatives do not hold globally, but we restrict the numerical analysis to the cases where they do hold. The lines divide the different parameter regions with a common set of equilibria, and we have set the tolls to zero in this case to describe the unregulated equilibrium.

Figure 2a reveals that, for the chosen parameters, two set of equilibria arise: both airlines using fully connected route structure (F,F), and both using hub-and-spoke but in different hubs (H_x,H_y); it also reveals that there are regions where both sets of equilibria may arise (the region between the two lines). Moreover, a higher substitutability between airlines favors the choice of hub-and-spoke route structures, as the indifference lines are to the left of the indifference point for $E/B = 0$.

Our results show that, when markets and airlines are symmetric and for the considered parameters, the best response to the rival using a hub-and-spoke route structure, is either adopting a fully



(a) Untolled equilibrium . (b) First-best.

Figure 2: Untolled equilibrium and welfare maximizing route structure. Main case (see A).

connected or a hub-and-spoke route structure, but using a different airport as the hub. That is the reason why asymmetric hub-and-spoke equilibria arise instead of symmetric hub-and-spoke equilibria (H,H) in Fig. 2a. As the cost advantages that adopting hub-and-spoke brings can be exploited under symmetric and under asymmetric settings, the main difference between both structures comes from the change in number of passengers in the connecting market. In a symmetric hub-and-spoke setting (H,H), the competition is direct in all markets, while in an asymmetric hub-and-spoke setting (H_x,H_y) the connecting market of the rival is dominated (as it is served point to point). This gain from dominating the rival's connecting market seems to be higher than the loss of being dominated in the own connecting market.

4 Welfare analysis

We first look at the welfare maximizing output for a given choice of route structure by airlines, deriving the tolls that induce that output choice. Therefore, we study the socially efficient route structure and whether these tolls are sufficient to achieve it as an equilibrium.

4.1 The symmetric fully connected case

In this case, denoted (F,F), both airlines serve the markets with a fully connected route structure. We look at a regulator that maximizes unweighted social surplus, with the toll per-passenger in each market (τ_i^m) and the toll per-flight in each link (τ_i^l) as instruments. Social welfare is:

$$SW^{(F,F)} = \left[\sum_{m\in M}\frac{B}{2}\cdot((q_i^m)^2+(q_j^m)^2)+E\cdot q_i^m\cdot q_j^m\right]+\left[\pi_i^F+\pi_j^F\right]+ \left[\sum_{m\in M}\tau_i^m\cdot q_i^m+T_j^m\cdot q_j^m\right]+\left[\sum_{l\in L}\tau_i^l\cdot f_i^l+\tau_j^l\cdot f_j^l\right], \quad (19)$$

where the first term in brackets is the consumer surplus, the second term is the airlines' profit, the third term is the revenue from per-passenger tolls, and the fourth term is the revenue from per-flight tolls. The first-order conditions for welfare maximization under fully connected route structures imply the following pricing and frequency setting rules:

$$\frac{\partial SW^{(F,F)}}{\partial q_i^m} = 0 \Rightarrow \; p_i^m = c_q \;\; \forall m \in M \; \wedge \; \frac{\partial SW^{(F,F)}}{\partial f_i^l} = 0 \Rightarrow \; -\sum_{m \in M} (q_i^m + q_j^m) \cdot \frac{\partial D^m}{\partial f_i^l} + q_i^m \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f \,. \tag{20}$$

Note that we drop the index on the delay cost function as it is the same for both airlines, because we are looking at a symmetric route structure setting. Equation (20) states that the fare should equal the marginal cost of a seat in all markets, and that, in every link, frequency should be such that the airline's marginal cost per flight equals marginal benefits for all passengers. Comparing (6) with (20), and (7) with (20), we can derive the tolls that maximize social welfare under (F,F):

$$\tau_i^m = -q_i^m \cdot B \;\; \forall m \in M \; \wedge \; \tau_i^l = \sum_{m \in M} q_j^m \cdot \frac{\partial D^m}{\partial f_i^l} \;\; \forall l \in L \,. \tag{21}$$

This is simply a per-passenger subsidy equal to the markup for each market, to eliminate the dead-weight loss in all markets, and a per-flight toll equal to the uninternalized congestion for each link, a traditional result in the airport pricing literature (e.g. Brueckner 2005). The two instruments above (Eq. (21)) attain the social optimum, if both airlines exogenously choose fully connected. The optimal value for social welfare (with variables satisfying (20)) is:

$$SW^{(F,F)} = \sum_{m \in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m - \sum_{l \in L} (f_i^l + f_j^l) \cdot c_f \,, \tag{22}$$

4.2 The symmetric hub-and-spoke case

Following the same procedure as in Section 4.1, straightforward calculations yield the following rules for welfare maximizing pricing and frequency setting under (H,H) route structure (with B as the hub airport in this case):

$$\frac{\partial SW^{(H,H)}}{\partial q_i^m} = 0 \Rightarrow \; p_i^m = c_q \; \forall m \in \{AB, BC\} \; \wedge \; \frac{\partial SW^{(H,H)}}{\partial q_i^{AC}} = 0 \Rightarrow \; p_i^{AC} = 2 \cdot c_q \,, \tag{23}$$

$$\frac{\partial SW^{(H,H)}}{\partial f_i^l} = 0 \Rightarrow \; -\sum_{m \in \{AB,BC\}} (q_i^m + q_i^{AC} + q_j^m + q_j^{AC}) \cdot \frac{\partial D^m}{\partial f_i^l} + (q_i^m + q_i^{AC}) \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f \,. \tag{24}$$

Again, in the optimum, the fare should equal the marginal cost in all markets, and frequencies should be such that the airline's marginal cost per flight equals marginal benefits for all passengers. Comparing first-order conditions, we obtain the tolls that maximize social welfare under symmetric hub-and-spoke route structures:

$$\tau_i^m = -q_i^m \cdot B \;\; \forall m \in M \; \wedge \; \tau_i^l = \sum_{m \in \{AB,BC\}} (q_j^m + q_j^{AC}) \cdot \frac{\partial D^m}{\partial f_i^l} \;\; \forall l \in \{ab, bc\} \,. \tag{25}$$

These are the sufficient instruments when route structure is fixed to be hub-and-spoke for both airlines. In equilibrium, the optimal value for social welfare under (H,H) can be written as:

$$SW^{(H,H)} = \sum_{m \in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m - \sum_{l \in \{ab,bc\}} (f_i^l + f_j^l) \cdot c_f \,, \tag{26}$$

with variables satisfying (23)-(24).

4.3 The asymmetric cases

We have shown in Section 3.3 that also asymmetric equilibria may arise, in particular (F,H) and (H_x,H_y). It is straightforward to show that optimal pricing and frequency setting rules will be a combination. The difference will be that the tolling rules will be evaluated at different outputs.

4.4 The optimal route structure

We now look at the combination of route structure and output that maximizes welfare. Again, complexity prevents us from fully comparing social welfare values analytically. As in Section 3, we combine analytical results with numerical examples to identify the equilibria and provide intuition.

First, we compare the choice of route structure by unregulated firms with the social welfare maximizing choice, to identify the sources of potential inefficiency. Consider the comparison only between the following route structure equilibria: (F,F) and (H_A,H_B). We focus on the comparison between these particular structures because the numerical analysis suggests that, for the considered symmetry, those are the settings that can be first-best optimal. To compare the social optimum with the untolled equilibrium, let the difference between social welfare in both settings be $\Delta \equiv SW^{(F,F)} - SW^{(H_A,H_B)}$. The condition $\Delta > 0$ is necessary for (F,F) to be the welfare maximizing route structure setting, while the condition $\Gamma_i > 0$ in Eq. (18), is sufficient for (F,F) to be an equilibrium. Using (22), and (26), we get:

$$\Delta = \sum_{m\in M}\frac{B}{2}\left[(q^m_{i|(F,F)})^2-(q^m_{i|(H_A,H_B)})^2\right]+\sum_{m\in M}\frac{B}{2}\left[(q^m_{j|(F,F)})^2-(q^m_{j|(H_A,H_B)})^2\right]$$
$$+E\cdot[q^m_{i|(F,F)}\cdot q^m_{j|(F,F)}-q^m_{i|(H_A,H_B)}\cdot q^m_{j|(H_A,H_B)}]$$
$$+c_f\left(\sum_{l\in\{ab,ac\}}f^l_{i|(H_A,H_B)}-\sum_{l\in L}f^l_{i|(F,F)}\right)+c_f\left(\sum_{l\in\{ab,bc\}}f^l_{j|(H_A,H_B)}-\sum_{l\in L}f^l_{j|(F,F)}\right), \tag{27}$$

where the variables in (27) are evaluated at the social optimum for the given route structure setting in parentheses. The comparison between Γ_i in Eq. (18), and Δ in Eq. (27) sheds light on the inefficiency of route structures choice by profit maximizing agents and makes implausible that unregulated competition between airlines will always lead to the first-best route structure. First, recall that there is a difference in output between the two cases, as discussed above, due to two effects: market power exertion and the presence of congestion externalities. This clearly makes the variables of Δ and Γ_i differ. Even if the outputs were the same, in the social welfare comparison there is a term involving the cross sensitivity parameter (E) that is absent in profit comparison; i.e. a firm ignores the effect of its choices on the consumer surplus derived by the competitor's passengers. Moreover, the airlines' relevant comparison is for a given route structure of the rival, which again makes Δ and Γ_i diverge.

Brueckner (2004), in a monopoly context, shows that the choice of route structure by a monopoly airline will be biased towards hub-and-spoke. In our problem, the divergence between settings is complicated further because frequency setting is distorted by congestion effects, and, when airlines are substitutes, both are distorted by strategic effects. Therefore, whether competing airlines are biased towards hub-and-spoke cannot be assessed analytically.

A look at Δ in Eq. (27) reveals that what drives which route structure composition maximizes welfare, between the symmetric fully connected setting (F,F) and the asymmetric hub-and-spoke settings (H_x,H_y), is the cost advantages that hub-and-spoke may bring (last two terms on the right-hand side of Eq. (27)), versus the changes in the number of passengers in each market. Figure 2b summarizes, for the same parameter region used in Section 3, the welfare maximizing route structure (when evaluated at the optimal output for those parameters). We analyze whether the first-best setting and the (tolled) equilibrium coincide in the following section.

Figure 2b suggests that, for the chosen parametrization, the route structure configuration that maximizes welfare, when all markets are served by both airlines, is either symmetric fully connected (F,F) or asymmetric hub-and-spoke (H_x,H_y). A simple comparison between Figures 2a and 2b confirms that the outcome of the unregulated competition may lead to route structure equilibria that are different from the efficient one. Figure 2b also suggests that the higher the substitutability between airlines, the more likely is (H_x,H_y) to be a more efficient route structure equilibrium than (F,F). It still brings cost savings and frequency benefits, but it is less essential that all airlines are present on all routes. One of the effects that favors the fully connected symmetric equilibrium over the asymmetric hub-and-spoke structures, is the absence of connecting passengers, because they have a higher marginal cost per seat and face a higher full price. In an asymmetric hub-and-spoke setting (H_x,H_y), the number of connecting passengers decreases as the airlines are perceived as closer substitutes, because in every connecting market of one airline, the rival provides a direct service priced at marginal cost (because we are looking at the first-best setting). Therefore, when products are close substitutes, the number of connecting passengers is low, and the gains from lower total costs due to reduced total frequency dominate. Note that this is not possible to achieve under symmetric hub-and-spoke (H,H).

It is also worth noting that when products are independent, asymmetric settings may be welfare maximizing. This occurs when demand is low (right end of the horizontal axis in Figure 2b), where (H_x,H_y) is the most efficient setting. The intuition behind this is that cost advantages from reduced total number of flights that hub-and-spoke brings in this region dominates, and that total congestion costs may be lower under (H_x,H_y) than under a symmetric hub-and-spoke setting, where all flights either take off or land at the hub airport. Asymmetric hub-and-spoke structures can benefit non-connecting passengers, as one of the airports will be less congested compared to when it is the hub of both airlines. Additional numerical examples, not shown here, reveal that only when airlines are perceived as independent and there is no congestion, symmetric and asymmetric hub-and-spoke structures yield the same welfare, for the considered parameters.

On the other hand, when products are perfect substitutes, the social optimum cannot have two airlines using different route structures ((H_x,H_y) and (F,H)). This is because generalized prices must be the same for all airlines that are serving the market, but also should be set at marginal social cost. As marginal social costs, under different route structures, are different in at least one market, these two constraints make an asymmetric route structure setting incompatible with welfare maximization when products are perfect substitutes ($E/B=1$). What is optimal, instead, is to have regulated monopolized markets. This result is driven by the fact that, for a single market, higher welfare is achieved under the full regulation of a monopoly than of perfect substitute competing airlines. This may also hold when airlines are close substitutes, as Figure 2b reveals: in the parameter region M_f, this is true. As already shown by Brueckner (2004), depending on parameters, it may be more efficient from a social welfare point of view, to have a regulated monopoly using fully connected route structure (M_f) or serving the market with hub-and-spoke.

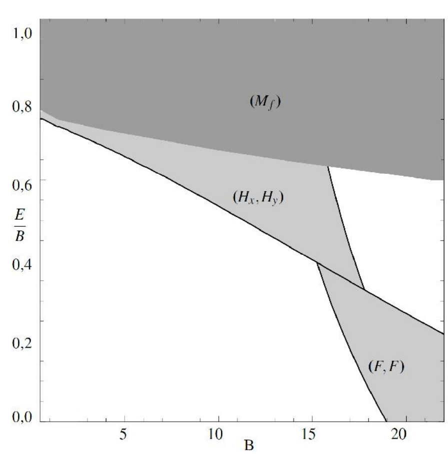
We now turn to the analysis of how to enforce the first-best described in this section. This is, can the first-best setting be a toll-decentralized equilibrium?

5 Sufficient instruments for social welfare maximization

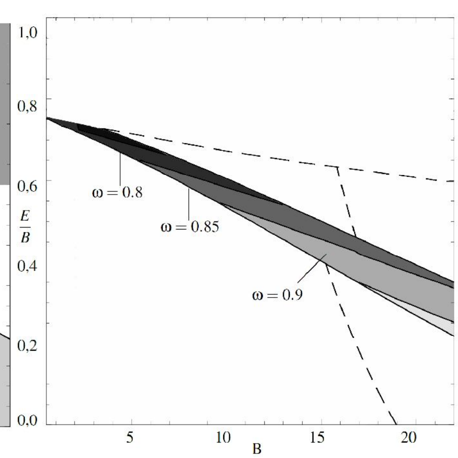
5.1 First-best analysis

In order to study whether the two pricing instruments described in Section 4 align airline choices with welfare maximization, we numerically examine the equilibrium of the game when the regulator charges the optimal tolls conditional on the first-best route structure. In other words, we derive the outcome of the game in each of the regions of Figure 2b, when the regulator charges the tolls that induce the optimal output for the given welfare maximizing route structure. For example, in the parameter region denoted by (F,F) in Figure 2b, the regulator set the tolls according to rules (20); if the equilibrium that results from charging these tolls is with both airlines choosing fully connected route structure, the first-best is achieved as the charges ensure optimal outputs. Conversely, if the equilibrium with the optimal charges in the parameter region where (F,F) maximizes welfare is not with both airlines choosing fully connected route structure, we can conclude that the two pricing instruments are not sufficient in this case. This is because any other charge, that may induce the optimal route structure equilibrium, will not induce the optimal output.

Figure 3a compares the untolled equilibrium in Fig. 2a with the welfare maximizing setting in Fig 2b, in terms of route structure. It reveals that the rationale for the charges is not always the same: they can be required only to correct output setting, to correct simultaneously output and route structure choice, and in order to correct market structure as well. The white areas represent the cases where the airlines' route structure equilibrium is the same as the first-best, and only output corrections are needed. The light gray areas are where the route structure equilibrium from untolled competition is different from the efficient one, and the tolls are required to induce airlines to choose both the welfare maximizing outputs and route structures. The M region, in dark gray, indicates that welfare maximization requires tolls that exclude one of the two airlines from the market, as a fully regulated monopoly would be optimal. We use the labels to indicate the efficient route structure when it differs from the one adopted by unregulated competing airlines.



(a) Tolls' rationale.



(b) Tolls' relative efficiency.

Figure 3: Tolls' rationale and relative efficiency. Main case (see A).

Figure 3a shows that the result that a monopoly airline exhibits a bias towards hub-and-spoke route structure does not fully carry on to competing airlines. This is, no longer whenever (H_x, H_y) is welfare maximizing, it is also an equilibrium of the untolled competition. For the chosen parametrization, (H_x, H_y) is optimal but the untolled equilibrium is (F, F) when own-demand sensitivity to price changes and substitutability are not too low (B lower than 18 and E/B above 0.4 in Figure 3a). The intuition for this is that, when demand is relatively sensitive to price changes (low B), the loss in the connecting market, when changing from a fully connected to a hub-and-spoke route structure, is larger than the cost benefits, which drives the untolled equilibrium; however, this same effect increases welfare if substitutability is high, as the cost advantages can be exploited with an asymmetric hub-and-spoke route structure configuration, without having a large number of connecting passengers. This positive effect on welfare is only possible with airlines adopting different airports as their hub, otherwise the number of connecting passengers would not necessarily decrease. The opposite, (F, F) being efficient but untolled airlines choosing (H_x, H_y), occurs when substitutability is not high (below 0.4 in Figure 3a) and own-demand sensitivity with respect to price changes is low (high B).

These results also indicate that a “naive” regulator, who observes the unregulated equilibrium and set the tolls based on the observed route structure, may not always achieve the first-best. The regulator must realize whenever the observed equilibrium is not efficient in terms of route structure (the gray regions), and induce airlines, via tolls, to change the way they serve the markets.

One of the main result of our numerical analyses is that the first-best cannot always be enforced by using the airline- and market-specific per-passenger tolls together with the airline- and link-specific per-flight tolls designed to induce the optimal outputs of the first-best route structures. Thus, the sub game-perfect equilibrium is not always efficient, even when the regulator can perfectly discriminate airlines and has no budget constraints. The colored area in Figure 3b displays when these instruments are not sufficient for welfare maximization. This region has airlines adopting fully connected route structures instead of the first-best setting (H_x, H_y). This is due to the asymmetry of the toll structure: to enforce (H_x, H_y), the regulator must give per-passenger subsidies in each market that are proportional to the number of passengers in the first-best setting, and a per-flight toll on the two routes that are flown by each airline. Given this toll configuration, adopting a hub-and-spoke route structure may be strictly dominated by the fully connected strategy if the subsidies in the market that is intended to be a connecting market are large enough, because serving it directly increases the number of passengers and hence the revenue. The latter effect dominates the potential cost advantages from a hub-and-spoke structure, in presence of tolls intended for a hub-and-spoke setting. To reach the first-best outcome in these cases, an additional instrument is required.

A natural question that follows from the results above is how big is the loss in welfare from charging the output-based tolls and not achieving the first-best route structure. Figure 3b also shows the relative efficiency of such tolls (ω), as the percentage of the maximum welfare gain that can be obtained using the untolled equilibrium as the reference scenario. The relative efficiency ranges from 1, the white areas in Fig. 3b where the tolls are sufficient instruments, to 0.77, the lowest possible relative efficiency of the output-based tolls.

Although, in our model, the regulator does not have direct control on the number of airlines, he can indirectly leave an airline out of one market through the tolls. To do this, it needs to give to one airline the per-passenger (market power) subsidy that corresponds to the monopoly output, and charge no congestion tolls. This is because a monopolist perfectly internalizes congestion externalities. On the other hand, the other airline must be charged a "barrier toll" that removes the incentives to participate.

6 Conclusions

We have extended the airport pricing literature by analyzing how to enforce the social optimum, in terms of output and route structure, in a network with endogenous hub location. We show that per-flight and per-passengers charges are not always sufficient to attain the first-best, and that the rationale behind the charges is not always the same. In some cases, a regulator needs an additional regulatory instrument aimed at route structure choice to maximize welfare.

We see extending the model as a natural avenue for future research. For example, the consideration of a larger network and the interaction between several airlines is a logical extension; the role of the endogenous hub location is likely to be important in those settings. Considering asymmetry of markets and airlines is also an important topic for future research, as the competition between regional, national, and low-cost carriers is one of the driving forces of route structure adoption in the aviation networks. Finally, this framework can be extended to analyze how airports regulated by different authorities interact and affect route structure equilibrium, as well as to the welfare implications of alliances and merges.

Acknowledgments

Financial support from ERC Advanced Grant OPTION (#246969) is gratefully acknowledged.

References

Brander, J. A. & Zhang, A. (1990), 'Market conduct in the airline industry: An empirical investigation', *The RAND Journal of Economics* **21**(4), 567–583.

Brueckner, J. K. (2004), 'Network structure and airline scheduling', *The Journal of Industrial Economics* **52**(2), 291–312.

Brueckner, J. K. (2005), 'Internalization of airport congestion: A network analysis', *International Journal of Industrial Organization* **23**(7-8), 599–614.

Oum, T. H., Zhang, A. & Zhang, Y. (1993), 'Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets', *Journal of Transport Economics and Policy* **27**(2), 171–192.

Pels, E. & Verhoef, E. T. (2004), 'The economics of airport congestion pricing', *Journal of Urban Economics* **55**(2), 257–277.

A Functional forms and parameters of the main case

Table 1: Functional forms and parameter values.

Function	Functional form	Parameter	Value	Parameter	Value	Parameter	Value
Congestion delay	D = 180 F / K	A	1750	E/B	[0;1]	cf	33240
Schedule delay cost	$g_i = 80/f_i$	B	[0.5;22]	K	4.85	cq	50