TI 2013-189/VIII
Tinbergen Institute Discussion Paper

Airline Route Structure Competition and Network Policy

Hugo Emilo Silva
Erik Verhoef
Vincent van den Berg

Faculty of Economics and Business Adminstration, VU University Amsterdam, and Tinbergen Institute, The Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Airline route structure competition and network policy

Hugo E. Silva*, Erik T. Verhoef, Vincent A.C. van den Berg

Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS, Amsterdam, The Netherlands.

Abstract

This paper studies whether a regulator needs to correct the route structure choice by carriers with market power in the presence of congestion externalities, in addition to correct their pricing. We account for passenger benefits from increased frequency, passenger connecting costs, airline endogenous hub location and route structure strategic competition. We find that, for some parameters, 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 when the regulator is constrained to set non-negative tolls, but also for the case of unconstrained tolling.

Keywords: Route structure competition, Aviation policy, Hub-and-spoke networks, Fully-connected networks

---

1. Introduction

Following the deregulation of the airline industry, several changes in aviation markets were observed (see Morrison and Winston (1995) for an overview of the changes in the US industry, and Burghouwt and Hakfoort (2001) for Europe). In addition to changes in fares, the most notorious change was in the way markets were served: the adoption of hub-and-spoke 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 hub-and-spoke structures. In a monopoly framework, Hendricks et al. (1995) show that economies of density alone can induce an airline to adopt a hub-and-spoke route structure; Pels et al. (2000) analyze the choice of route structure under linear marginal costs and the effect

*Corresponding author

Email addresses: `h.silvamontalva@vu.nl` (Hugo E. Silva), `e.t.verhoef@vu.nl` (Erik T. Verhoef), `vberg@feweb.vu.nl` (Vincent A.C. van den Berg)

of fixed costs in the decision; and Brueckner (2004) shows how frequency effects favors the adoption of hub-and-spoke. The third argument, strategic advantages, refers to the fact that adopting hub-and-spoke route structures may bring, in oligopolistic competition, further advantages because of the effect it may have on competitors. For instance, Oum et al. (1995) show that using a hub-and-spoke structure can be a top-dog strategy, in the sense of Fudenberg and Tirole (1984), allowing the carrier to be more aggressive in output market competition; it can deter entry in hub markets if the complementarities among hub markets are large, or the number of complementary hub markets is large (Hendricks et al., 1997); and it can prevent competition in local markets between two hub carriers because invading the competitor's local market may reduce own profit in all connecting markets due to a more aggressive competition in the trans-hub market (Zhang, 1996).

The outcomes of a deregulated environment where carriers can choose how to serve markets, and the conditions that give rise to hub-and-spoke route structures as an equilibrium of unregulated competition, have been well studied in the literature. Recent contributions include Hendricks et al. (1999), Alderighi et al. (2005), Barla and Constantatos (2005), and Flores-Fillol (2009, 2010). These studies, however, ignore the endogenous nature of the hub location, do not study the socially optimal route structure, and most of them also ignore congestion effects. 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 with fixed route structures, hence ignoring its endogenous nature and its effect on optimal policy.

The objective of this paper is to extend the congestion pricing and regulation analysis by elaborating on policy instruments that can decentralize the socially 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 (e.g. Brueckner (2002); Pels and Verhoef (2004)). 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).[1] 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 what we believe is 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.

[1] See Zhang and Czerny (2012) for an up-to-date review on airport pricing.

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, given the networks chosen. We find that social welfare can be increased by using an additional policy instrument when the regulator is restrained from subsidizing airlines, but also when it does not face such constraint on tolling. Specifically, the first-best optimal route structures and output levels cannot always be decentralized 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. This is because these tolls, especially the market-specific per-passenger subsidies, which are required to counteract market power exertion, provide incentives to airlines to adopt the route structure that better exploits the subsidies, and which may differ from the one that maximizes welfare.

First-best pricing as just discussed typically requires a regulator to give asymmetric per-passenger subsidies to airlines, a policy that is arguably, in practice, impossible to implement. To address this limitation, we also study the case in which the regulator is constrained to charge non-negative tolls. We show that when the second-best optimal route structure is different to the outcome of unregulated competition, second-best tolls may not induce the second-best optimal equilibrium. That is, the route structure and output levels that maximize welfare in absence of market power corrections possibly cannot be decentralized through tolls alone. Thus, also in the absence of subsidization, tolls designed to correct output inefficiencies are not always a sufficient instrument for (constrained) welfare maximization.

Our results may have important policy implications. In some cases an instrument directly aimed at regulating route structure choice is needed for welfare maximization, and in the cases where the pricing instruments are sufficient, the rationale for the charges is not always the same. In some cases they are required only to correct output choices, in other cases the tolls are needed to correct simultaneously output and route structure choices, and finally they can also be needed in order to change the market structure in terms of suppliers present in the network, in addition to correct output and route structure. It is therefore evident from the analysis that a regulator designing a pricing policy for the aviation industry has to take into account its effect on the long-run route structure equilibrium, and assess the optimality of the observed setting before deciding on toll levels. In addition, more information is needed to achieve the first-best: if only "local variables" (notably marginal external cost and marginal benefit) are available, adaptive pricing would "guide" the regulator to the (second-best) optimum. But, with discrete changes in route structures, and multiple local optima, this is no longer true. A "naive" regulator, who observes a sub-optimal route structure configuration, and set the tolls based on this, may not always achieve welfare maximization, neither first-best nor second-best. Finally, the extent to which each of the different situations described above hold is a matter of empirical research.

In Section 2 we introduce the model and the configuration of nodes between which endogenous route structures can be offered. Section 3 analyzes the equilibria of the untolled competition for a duopoly, while Section 4 derives the welfare maximizing equilibria. Section 5 studies whether pricing is enough to achieve the welfare maximizing setting as the result of tolled competition, and extends the analysis to the second-best case where the regulator cannot subsidize airlines. Finally, Section 6 concludes.

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); direction-specific return-trips (e.g. people making round-trips both from A to B and from B to A) may be introduced without altering the analysis. 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); or hub-and-spoke (henceforth H), 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.[2] We let each airline's hub to be endogenous, therefore asymmetric settings with hub-and-spoke structures may arise.

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,

[2]Equilibria where an airline chooses to serve less than the 3 origin-destination pairs will be considered in the numerical model.

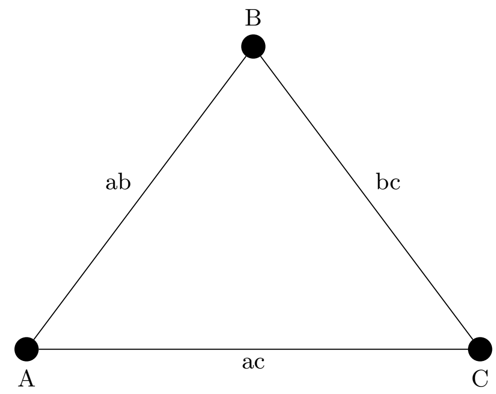


Figure 1: Network

as given.[3]

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 the sum of the fare, p_i^m, the congestion delay cost, D_i^m, and schedule delay cost, g_i^m.[4] 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:

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

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 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.

[3]The assumption that airlines compete in a Cournot fashion is common in the airline literature and is supported by empirical evidence by Brander and Zhang (1990) and Oum et al. (1993). For a discussion of the implications of leadership behavior see Brueckner and Van Dender (2008), and for Bertrand competition with differentiated airlines see Silva and Verhoef (2013).

[4]The schedule delay is the time difference between a passengers desired departure time and the actual departure time. As we do not explicitly model trip-timing decisions, this represents an average measure of the schedule delay cost, that, at least, provides the right intuition.

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

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

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% that allows for analytical tractability. A more realistic model would have endogenous aircraft size and load factors, with stochastic demand, but this would prevent us from providing analytical transparent understanding of route structure equilibria.

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 sets 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).[5] In the second and third stage, airlines choose route structure and output respectively.[6] We look at sub-game perfect equilibria through backward induction, so we first analyze, in the following section, the airlines' Nash equilibria.

3. Airlines equilibrium

In the second stage, airlines have a discrete choice between alternative route structures (F or H at any of the airports). Hence, we first look at their profits while 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 chooses 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 flying through 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.

[5] This formulation is a reduced form of a system of welfare maximizing airports, or a regulator, setting charges per passenger and charges per flight at each airport. The difference would matter if each airport is controlled by a different authority, that may, for instance, maximize local welfare.

[6] We assume that airlines treat the tolls as parametric, i.e. they do not believe that their actions may change the way a regulator chooses instruments. See Brueckner and Verhoef (2010) for a detailed discussion on how to account for such behavior of agents with market power in presence of externalities.

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 airline 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 full price then equals the sum of the fare; the congestion at the origin, A, which depends on the total number of flights operating at that airport ($F^{ab} + F^{ac}$) and the airport's capacity (K_A); the congestion at the destination airport, B; and the schedule delay cost.[7] We look at the particular case where all airports have the same capacity, but this could easily be extended. The airline's profit, using (3) and $s_i^l = q_i^l / f_i^l$, is:

$$\pi_i^F = \sum_{m \in M} q_i^m \cdot (p_i^m - c_q - \tau_i^m) - \sum_{l \in L} f_i^l \cdot (c_f + \tau_i^l) \;, \tag{5}$$

where the superscript F refers to the fully connected structure. Using this, together with (2), we can rewrite profit as:

$$\pi_i^F = \sum_{m \in M} q_i^m \cdot (A - B \cdot q_i^m - E \cdot q_j^m - g_i^m - D_i^m - c_q - \tau_i^m) - \sum_{l \in L} f_i^l \cdot (c_f + \tau_i^l) \;. \tag{6}$$

Airline profit indirectly depends also on the rival's route structure. That structure will be reflected in the rival's number of passengers and in 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 which route structure underlies 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:

$$\frac{\partial \pi_i^F}{\partial q_i^m} = 0 \Rightarrow \; p_i^m = c_q + \tau_i^m + B \cdot q_i^m \;, \tag{7}$$

$$\frac{\partial \pi_i^F}{\partial f_i^l} = 0 \Rightarrow \; - \sum_{m \in M} q_i^m \cdot \left(\frac{\partial D_i^m}{\partial f_i^l} + \frac{\partial g_i^m}{\partial f_i^l} \right) = c_f + \tau_i^l \;. \tag{8}$$

Equation (7) 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 (8) states that airline's marginal cost per flight (right-hand side of (8)) equals marginal revenue (left-hand side, marginal congestion costs plus marginal schedule delay benefits).

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

Therefore, airlines internalize own-passenger congestion. These rules are analogous to the rules obtained previously in Cournot competition (e.g. Pels and Verhoef, 2004).

The airline's profit using a fully connected route structure, in sub-game equilibrium, is:

$$\Pi_i^F = \sum_{m \in M} B \cdot (q_i^m)^2 - \sum_{l \in L} f_i^l \cdot (c_f + \tau_i^l) \,, \tag{9}$$

which is obtained by replacing Eq. (7) into Eq. (6). This is the revenues from the markup ($B \cdot q_i^m$ per passenger) minus the costs that are not directly charged to passengers.

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 purpose, we analyze in this section the case where the airline chooses airport B as its 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, in addition to the passenger from hub markets, the passengers from the spoke market are also traveling through links ab and bc. As a consequence, in this setting, aircraft sizes will satisfy:

$$\begin{aligned} s_i^{ab} &= (q_i^{AB} + q_i^{AC})/f_i^{ab} \,, \\ s_i^{bc} &= (q_i^{BC} + q_i^{AC})/f_i^{bc} \,. \end{aligned} \tag{10}$$

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 gives:

$$\begin{aligned} \widetilde{D}_i^{AC} &= D_i^{AB} + D_i^{BC} \,, \\ \widetilde{g}_i^{AC} &= g_i^{AB} + g_i^{BC} \,. \end{aligned} \tag{11}$$

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—leaving congestion and travel delays aside—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.[8] These assumptions shape profit in the following way:

$$\pi_i^H = \sum_{m\in\{AB,BC\}} q_i^m \cdot (A - B\cdot q_i^m - E\cdot q_j^m - g_i^m - D_i^m - c_q - \tau_i^m) +$$
$$q_i^{AC}\cdot(A - B\cdot q_i^{AC} - E\cdot q_j^{AC} - \widetilde{g}_i^{AC} - \widetilde{D}_i^{AC} - 2\cdot c_q - T_i^{AC}) - \sum_{l\in\{ab,bc\}} f_i^l\cdot(c_f+\tau_i^l)\,, \quad (12)$$

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 lead to the following pricing and frequency setting rules:

$$\frac{\partial\pi_i^H}{\partial q_i^m} = 0 \Rightarrow\; p_i^m = c_q + \tau_i^m + B\cdot q_i^m \quad \forall\, m\in\{AB,BC\}\,, \quad (13)$$

$$\frac{\partial\pi_i^H}{\partial q_i^{AC}} = 0 \Rightarrow\; p_i^{AC} = 2\cdot c_q + T_i^{AC} + B\cdot q_i^{AC}\,, \quad (14)$$

$$\frac{\partial\pi_i^H}{\partial f_i^l} = 0 \Rightarrow\; -\sum_{m\in\{AB,BC\}}(q_i^m + q_i^{AC})\cdot\left(\frac{\partial D_i^m}{\partial f_i^l} + \frac{\partial g_i^m}{\partial f_i^l}\right) = c_f + \tau_i^l \;\;\forall\, l\in\{ab,bc\}\,. \quad (15)$$

These equations 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, using the first-order conditions (Eqs. (13) and (14) in Eq. (12)), can be written as:

$$\Pi_i^H = \sum_{m\in M} B\cdot(q_i^m)^2 - \sum_{l\in\{ab,bc\}} f_i^l\cdot(c_f+\tau_i^l)\,. \quad (16)$$

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

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 either serve the markets with a fully connected route structure or with a hub-and-spoke route structure. As we ignore the possibility of serving only one market in the

[8] 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.

analytical model, no other configurations are possible. 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:

$$\Pi_i(r, r) \geq \Pi_i(u, r) \quad \forall\, u \in RS\ . \tag{17}$$

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 ($\neq u$), will arise if and only if the following holds:

$$\Pi_i(u, v) \geq \Pi_i(w, v) \ \wedge\ \Pi_i(v, u) \geq \Pi_i(y, u) \quad \forall\, w \in RS\ , \forall\, y \in RS\ \ u \neq v\ . \tag{18}$$

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, (17) and (18), are not necessarily mutually exclusive, therefore multiple 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. (9) and (16), it is given by:

$$\Gamma_i = \sum_{m \in M} B \cdot \left[(q^m_{i|(F,F)})^2 - (q^m_{i|(H_B,F)})^2 \right] + c_f \cdot \left[\sum_{l \in \{ab,bc\}} f^l_{i|(H_B,F)} - \sum_{l \in L} f^l_{i|(F,F)} \right] , \quad (19)$$

where the variables in Eq. (19) are evaluated at the untolled equilibrium with route structures indicated in parentheses. Eq. (19) 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. (11)). 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.

For a monopoly and in absence of congestion, Brueckner (2004) already showed that this expression 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 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 hub-and-spoke over fully connected route structures. In contrast to previous studies that analyze route structure choice, an asymmetric equilibrium with both airlines having a hub-and-spoke route structure may arise because the hub location is endogenous.

Figure 2 shows the equilibria for a 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 Appendix 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 to describe the unregulated equilibrium.

Figure 2 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). Figure 2 also suggests that 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 $E/B = 0$.

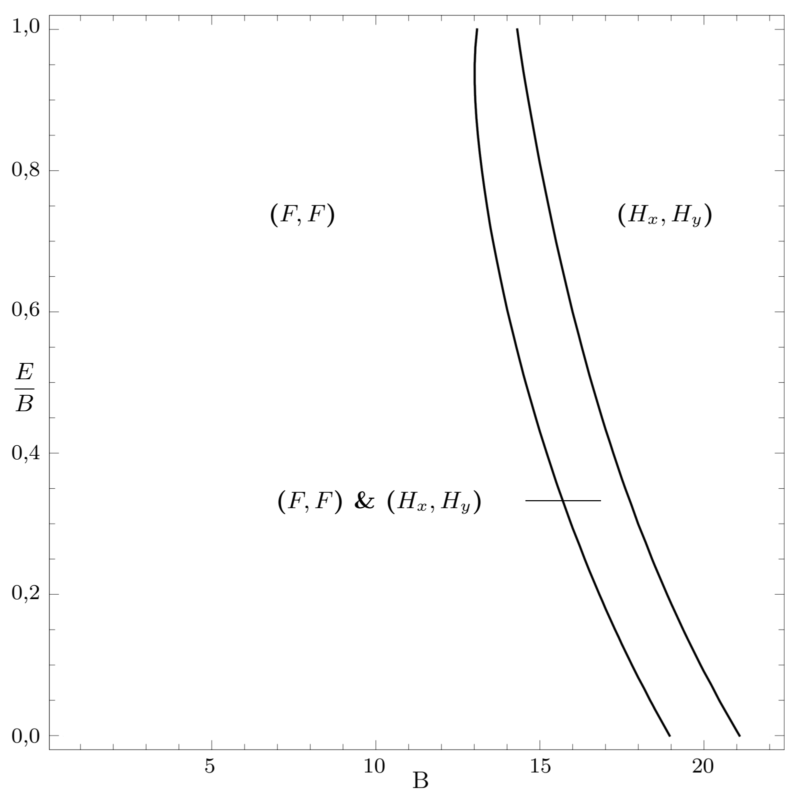


Figure 2: Untolled equilibrium in route structures. Main parameterization (see Appendix A).

Our results in Fig. 2 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 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 (H_x, H_y), arise instead of symmetric hub-and-spoke equilibria (H, H). As the cost advantages that adopting hub-and-spoke brings can be exploited in a similar way under symmetric and under asymmetric hub-and-spoke 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; 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, which explains why, given that the rival is using a hub-and-spoke structure, choosing a different airport as the hub dominates choosing the same when considering adopting a hub-and-spoke structure.[9]

As the purpose of the paper is to compare the untolled equilibria with the welfare maximizing situation, and analyze how to enforce it, the next section studies the first-best combination of route structure and output.

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. Thereafter, we study the socially efficient route structure equilibrium and whether these tolls are sufficient to achieve it as an equilibrium of the full game.

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. Straightforward calculations yield the following expression for social welfare:

$$SW^{(F,F)} = \left[\sum_{m \in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m\right] + \left[\pi_i^F + \pi_j^F\right] + \left[\sum_{m \in M} \tau_i^m \cdot q_i^m + T_j^m \cdot q_j^m\right] + \left[\sum_{l \in L} \tau_i^l \cdot f_i^l + \tau_j^l \cdot f_j^l\right], \tag{20}$$

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

[9]We numerically check that choosing a different airport as a hub dominates choosing the same airport as the rival in the whole parameter range, i.e. not only in the region where (H_x, H_y) is an equilibrium.

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:

$$\frac{\partial SW^{(F,F)}}{\partial q_i^m} = 0 \Rightarrow \; p_i^m = c_q \;\; \forall \, m \in M, \tag{21}$$

$$\frac{\partial SW^{(F,F)}}{\partial f_i^l} = 0 \Rightarrow \; -\sum_{m \in M} (q_i^m + q_j^m) \cdot \frac{\partial D^m}{\partial f_i^l} + q_i^m \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f \; . \tag{22}$$

Note that the tolls cancel out, because they are only monetary transfers, and 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 (21) states that the fare should equal the marginal cost of a seat in all markets, and (22) states that, in every link, frequency should be such that the airline's marginal cost per flight equals marginal benefits for all passengers. Comparing (7) with (21), and (8) with (22), we derive the tolls that maximize social welfare under (F, F):

$$\tau_i^m = -q_i^m \cdot B \;\; \forall \, m \in M \; , \tag{23}$$

$$\tau_i^l = \sum_{m \in M} q_j^m \cdot \frac{\partial D^m}{\partial f_i^l} \;\; \forall \, l \in L \; . \tag{24}$$

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. This is a traditional result in the airport pricing literature (e.g. Brueckner, 2005).

The two instruments above (Eqs. (23) and (24)) attain the social optimum, if both airlines exogenously choose fully connected structures. The optimal value for social welfare, in this fixed route structure equilibrium, is:

$$SW^{(F,F)} = \sum_{m \in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m - \sum_{l \in L} (f_i^l + f_j^l) \cdot c_f \; , \tag{25}$$

with quantities and frequencies satisfying (21) and (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):

$$\frac{\partial SW^{(H,H)}}{\partial q_i^m} = 0 \Rightarrow \; p_i^m = c_q \;\; \forall m \in \{AB, BC\} \; , \tag{26}$$

$$\frac{\partial SW^{(H,H)}}{\partial q_i^{AC}} = 0 \Rightarrow \; p_i^{AC} = 2 \cdot c_q \; , \tag{27}$$

$$\frac{\partial SW^{(H,H)}}{\partial f_i^l} = 0 \Rightarrow -\sum_{m\in\{AB,BC\}} (q_i^m + q_i^{AC} + q_j^m + q_j^{AC}) \cdot \frac{\partial D^m}{\partial f_i^l} + (q_i^m + q_i^{AC}) \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f \,. \quad (28)$$

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:

$$\tau_i^m = -q_i^m \cdot B \quad \forall\, m \in M \,, \quad (29)$$

$$\tau_i^l = \sum_{m\in\{AB,BC\}} (q_j^m + q_j^{AC}) \cdot \frac{\partial D^m}{\partial f_i^l} \quad \forall\, l \in \{ab, bc\} \,. \quad (30)$$

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:

$$SW^{(H,H)} = \sum_{m\in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m - \sum_{l\in\{ab,bc\}} (f_i^l + f_j^l) \cdot c_f \,, \quad (31)$$

with variables satisfying (26)-(28).

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 of the rules described above ((21), (22), and (26)–(28)). As a result, the tolls that maximize welfare for asymmetric equilibria will also be a combination of the tolls above; for an airline with fully connected route structure, (23) and (24) should be charged, and for an airline using hub-and-spoke, (29) and (30). The difference will be that the tolling rules will be evaluated at different outputs, and that Eqs. (29) and (30) have to be adjusted if the hub is not B.

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. (19), is sufficient for (F,F) to be an equilibrium. Using (25), and (31), we get:

$$\Delta = \sum_{m\in M} \frac{B}{2}\left[(q^m_{i|(F,F)})^2 - (q^m_{i|(H_A,H_B)})^2\right] + \sum_{m\in M} \frac{B}{2}\left[(q^m_{j|(F,F)})^2 - (q^m_{j|(H_A,H_B)})^2\right]$$
$$+E\cdot[q^m_{i|(F,F)}\cdot q^m_{j|(F,F)} - q^m_{i|(H_A,H_B)}\cdot q^m_{j|(H_A,H_B)}]$$
$$+c_f\left(\sum_{l\in\{ab,ac\}} f^l_{i|(H_A,H_B)} - \sum_{l\in L} f^l_{i|(F,F)}\right) + c_f\left(\sum_{l\in\{ab,bc\}} f^l_{j|(H_A,H_B)} - \sum_{l\in L} f^l_{j|(F,F)}\right), \quad (32)$$

where the variables in (32) are evaluated at the social optimum for the given route structure setting in parentheses. The comparison between Γ_i in Eq. (19), and Δ in Eq. (32) 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.

In a monopoly context, Brueckner (2004) 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 because frequency setting—for a given number of passengers—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. (32) reveals that what drives which route structure composition maximizes welfare, at least 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. (32)), versus the changes in the number of passengers in each market. Figure 3 summarizes, for the same parameter region used in Section 3 (see Appendix A for details), the welfare maximizing route structure (when evaluated at the optimal output for those parameters). Recall that the first-best setting in Figure 3 is the one that maximizes welfare, and it is not necessarily an equilibrium of the tolled competition. We analyze whether and when the first-best setting and the (tolled) equilibrium coincide in the following section.

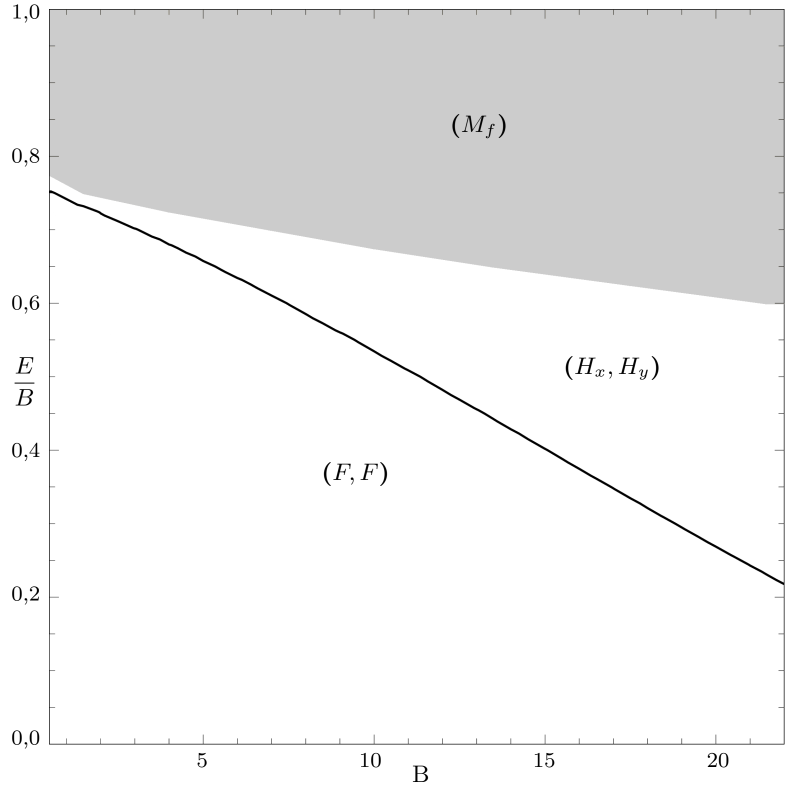


Figure 3: Welfare maximizing route structures. Main parameterization (see Appendix A).

Figure 3 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 2 and 3 confirms that the outcome of the unregulated competition may lead to route structure equilibria that are different from the efficient one. Figure 3 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 3), 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 full 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 that, for a single market, higher welfare is achieved under the full regulation of a monopoly than of perfect substitute competing airlines. This is because we are looking at a regulator who solves the congestion inefficiency and the market power exertion through tolls; therefore, there is no dead-weight loss regardless of the number of firms. In addition, the frequency set by a fully regulated monopoly airline is higher than for a fully regulated airline in oligopoly, as demand is divided between firms, thus schedule delay costs will be lower. This is what favors a monopoly. When differentiation is weak ($B \gg E$), the expansion of demand generated by a new firm may overweight the schedule delay cost advantages of a monopoly, and competition will bring higher welfare.[10]

This may also hold when airlines are close substitutes, as Figure 3 reveals: in the parameter region M_f, it is welfare maximizing to have a regulated monopoly airline serving the three markets with a fully connected route structure. 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.[11]

[10]Despite having a different model, the intuition is the same as in Basso (2008).

[11]For the parameters of our main case, the value of the own-demand sensitivity parameter that makes a hub-and-spoke route structure more efficient than a fully connected structure under a regulated monopoly is $B = 30.4$ (not shown in Figure 3).

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?

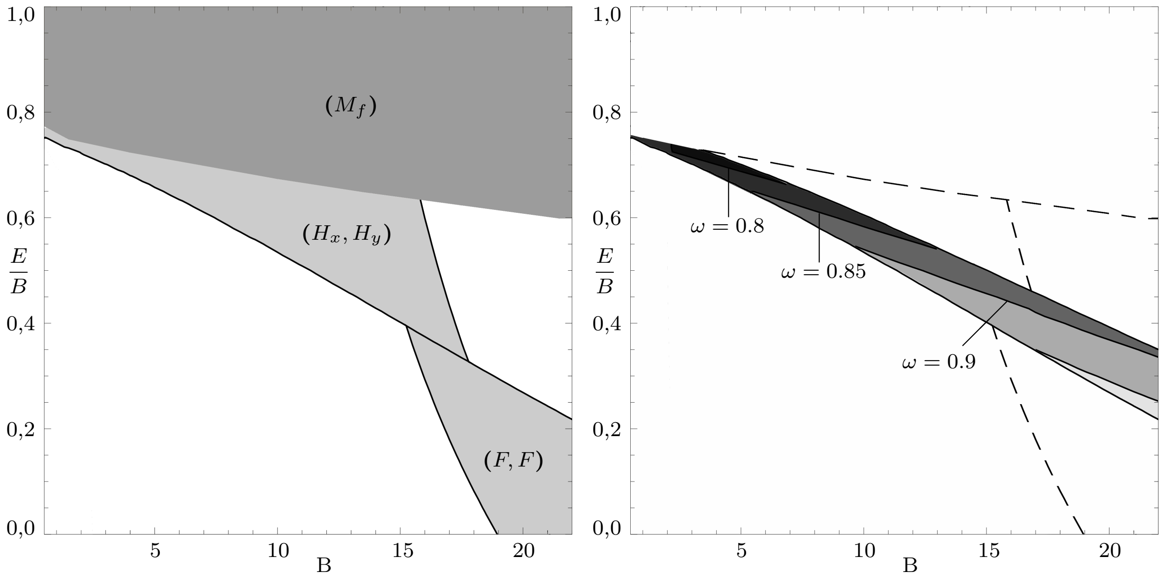
5. Sufficient instruments for social welfare maximization

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 3, 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 3, the regulator set the tolls according to rules (21) and (22); 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 4a compares the untolled equilibrium in Fig. 2 with the welfare maximizing setting in Fig 3, 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 airline 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.

Let us first discuss the difference between the route structure in the untolled equilibrium versus that in the optimum (later we will discuss whether tolls alone can decentralize the optimum). Figure 4a 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



(a) Tolls' rationale.

(b) Tolls' relative efficiency.

Figure 4: First-best tolls' rationale and relative efficiency of output-based tolling. Main parameterization (see Appendix A).

not too low (roughly, in the triangle where B is lower than 18 and E/B between 0.4 and 0.8 in Figure 4a). The reason why in that area we find (F, F) in the untolled equilibrium is that, when changing from a fully connected to a hub-and-spoke route structure, the full price in the connecting market increases. Therefore, when demand is relatively sensitive to price changes (low B), the profit loss in the connecting market is larger than the cost benefits and the equilibrium is with (F, F). However, this demand reduction in the airlines' connecting market does not necessarily harm welfare. In this area, we find (H_x, H_y) in the optimum because the cost advantages that hub-and-spoke structures bring can be exploited without having a large number of connecting passengers. Under (H_x, H_y), in each airline's connecting market the competitor offers a direct service with a lower full price. As demand is relatively sensitive to price changes and substitutability is not low, the number of connecting passengers is low. This is only possible with airlines adopting different airports as their hub, as 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 4a) and own-demand sensitivity with respect to price changes is low (high B).

The results also indicate that a "naive" regulator, who observes the unregulated equilibrium and set the tolls based on the then observed route structure, may not always achieve the first-best. The regulator should 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.

A 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 between airlines and when it has no budget constraints. The shaded area in Figure 4b represents the parameter range where the toll instruments described in the previous sections cannot decentralize the first-best. 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 a subsidy in each market that increases with the firm's number of passengers, 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, because the intended connecting market can be served directly without paying a per-flight toll. Doing this may increase revenue due to an increased number of passengers, because the full price in that market is lower when served directly. In the shaded region of Figure 4b, this effect dominates the potential cost advantages that the firm would have from adopting a hub-and-spoke structure, in presence of tolls intended for a hub-and-spoke setting.[12] To reach the first-best outcome in these cases, an additional instrument is therefore required.[13] This also implies that more information is needed to attain the first-best, as the variables usually needed for first-best tolling (e.g. marginal external cost) are not enough to assess discrete changes in welfare under different route structures.

A natural question that follows from the above results is how big the loss in welfare is, when charging the output-based tolls and not achieving the first-best route structure. Figure 4b also shows the relative efficiency of such tolls (ω), as the percentage of the maximum welfare gain that can be obtained (the first-best setting in Section 4) using the untolled equilibrium as the reference scenario for the welfare level. The relative efficiency ranges from 1 in the white areas of Fig. 4b where the tolls are sufficient instruments to achieve the first-best; to 0.77, the lowest possible relative efficiency of the output-based tolls, in the darkest region of the figure.

Although, in our model, the regulator does not have direct control on the number of airlines, he can price an airline out of one market through the tolls. To do this, one airline

[12] We do not find in the numerical analyses that the equilibrium with tolls has airlines adopting hub-and-spoke route structure in a parameter region where fully connected is the welfare maximizing structure.

[13] For example a arbitrarily high per-flight toll in the connecting market of each airline can act as a barrier to fully connected structures. As this charge is not paid in equilibrium, because the airlines adopt a hub-and-spoke structure, it does not induce any inefficiency in output, but only the incentives to adopt the optimal route structure. A direct restriction to an airline to fly from those airports is also a sufficient additional instrument for these cases.

should receive the per-passenger (market power) subsidy that corresponds to the monopoly output, and face no congestion toll. This is because a monopolist perfectly internalizes congestion externalities. The other airline should face a prohibitive toll that removes the incentives to become active. Note that the latter may not always be needed; in some cases, a large output choice by the subsidized airline may be enough to shift the rival's demand to levels that are not profitable or non-positive, given the congestion delay costs.

Finally, additional numerical analyses (not shown) show that the boundaries indicated in Figure 4 change when parameters are changed, but the main qualitative results hold.[14]

5.2. Second-best analysis

The results in Section 5.1 require the regulator to be able to give subsidies to airlines, and, in some cases, to exclude one airline out of the market. This is arguably close to impossible to carry out in real networks, and, instead, a regulator will most likely be constrained to charge non-negative tolls. This section considers the realistic case where only non-negative tolls can be charged, and compares the second-best optimal route structure equilibrium with the first-best equilibrium. An important question is, again, whether the second-best tolls are able to induce the desired (second-best) outcome also in terms of route structure.

As we normalize airport costs to zero, the first-best toll per passenger we found above is equal to a market power subsidy, and therefore it is always negative (see Eqs. (23) and (29)). The first-best per-flight toll is always non-negative (un-internalized congestion, see Eqs. (24) and (30)). It is, therefore, expected that in the second-best case we now study, the per-flight tolls are adjusted downwards to compensate for the lack of subsidies, and that in some parameter region they may even become (constrained to) zero. This would be the case if the congestion externality is sufficiently low compared to the justified market power subsidy.[15]

Figure 5 shows the second-best optimal setting for the same parameter constellation and functional forms as in Sections 3 and 4. The solid line divides the different second-best optimal regions, which are again (F, F) or (H_x, H_y), and the dashed line is the analogous divisional line for the first-best case. This allows for a comparison between the second- and first-best equilibria.

[14] We assess how the airlines' untolled equilibrium in route structure changes and whether the first-best setting can be decentralized with tolls for different parameters of the delay function, the airline cost function, and the passenger values of time. The changes in route structure equilibria follow the intuition provided by previous studies (e.g., Brueckner, 2004): hub-and-spoke route structures are favored by higher values of the per-flight fixed cost c_f, lower fixed marginal cost per seat c_q, higher disutility of schedule delay γ, and lower congestion.

[15] Pels and Verhoef (2004) analyze this using a single origin-destination pair. They find that in some cases the second-best policy is to set the tolls equal to zero.

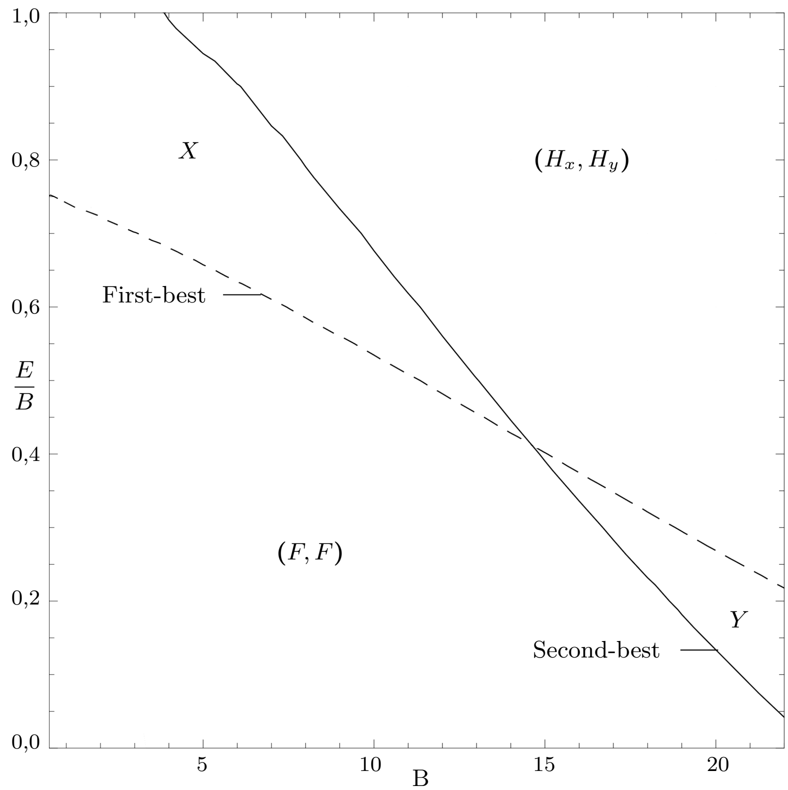


Figure 5: Second-best optimal route structures and comparison with the first-best case. Main parameterization (see Appendix A).

Figure 5 shows that, when subsidies are not feasible, fully connected route structures are favored when the degree of substitutability (E/B) is low and own-demand sensitivity is high (low B), and asymmetric hub-and-spoke route structures are favored when the opposite applies, in a similar manner as in the first-best case. However, there is a region—denoted X—where (F, F) is second-best optimal whereas (H_x, H_y) is first-best optimal, and *vice versa* (the Y region). The difference is caused by the fact that in the second-best cases, there are dead-weight losses due to the lack of subsidies. Competition under both airlines using fully connected is direct in all three markets, whereas in the asymmetric hub-and-spoke equilibrium there is one market where one airline has an advantage over the other (offering a direct flight). In the area X, where substitutability (E/B) is high and demand is more price sensitive (low B), the direct competition that a fully connected route structure brings may decrease dead-weight losses by more than under an asymmetric hub-and-spoke structure, and it is more favored than the case where all inefficiencies are dealt with through pricing. In the area Y, where demand is less price sensitive (higher B) and substitutability (E/B) is low, this effect is decreased and the difference in number of passengers becomes smaller; as a result, the cost advantages of hub-and-spoke structures become more attractive.

The results also show that the monopoly region does not exist for the considered parameters, as the dead-weight losses from monopoly pricing are no longer being corrected through the subsidy; it is better to regulate a duopoly when tolls are constrained to be non-negative.

Figure 6 compares the second-best optimal route structure equilibria with the untolled equilibria, and it reveals that also the rationale for the second-best charges may include the desire to change the route structure. The white areas again indicate the cases where the route structure in the untolled equilibrium is the same as the second-best optimum, and tolls only correct for output inefficiencies. The light gray and dark gray regions show where the inefficiency occurs both in route structures and in outputs, and therefore tolls may be required to correct both.

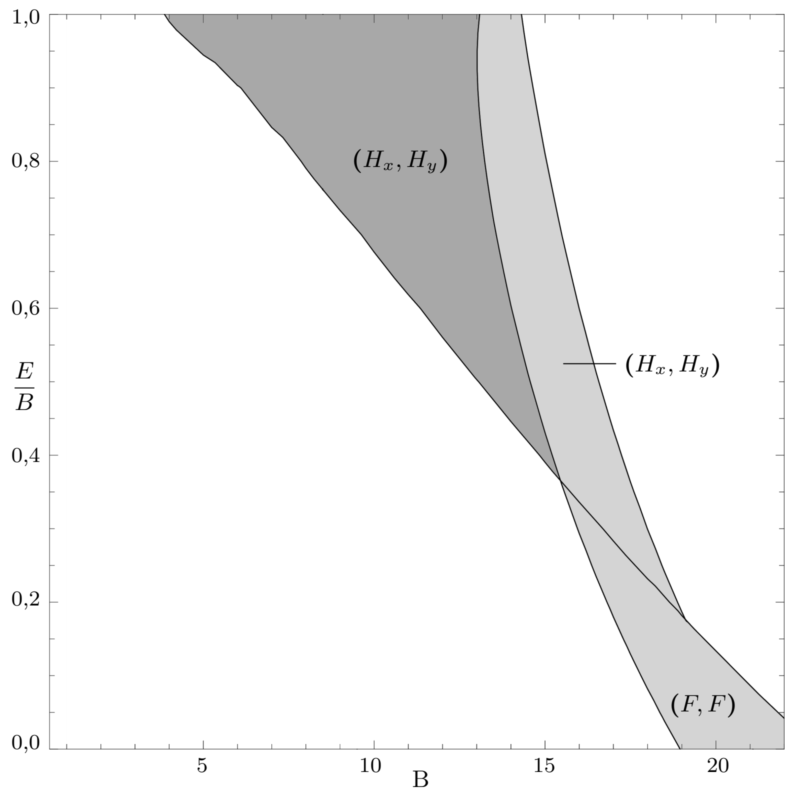


Figure 6: Second-best tolls' rationale and decentralization of the second-best optimum. Main parameterization (see Appendix A).

Finally, where the second-best optimal route structures in Figure 6 are assumed to apply, we also study whether they are decentralized by second-best tolls, in order to study whether the per-flight tolls are sufficient instruments to achieve the second-best outcome in route structures shown in Figure 5. As explained above, these tolls might be zero if the congestion externality is sufficiently low compared to the market power exertion by airlines. Therefore, it follows that in the regions where second-best tolls are zero and the route structure in the untolled equilibrium is not the same as the second-best optimal route structure, an instrument aimed at regulating route structure choices will be needed.

The main result of this exercise is that a regulatory instrument designed to correct route structure may be needed in addition to the second-best per-flight tolls: there are different regions where the second-best optimal route structures and outputs differ from the airlines' equilibrium with second-best charges. Figure 6 summarizes these results. Both the dark gray and the light gray regions show the cases where the second-best optimal equilibrium (in labels) cannot be enforced by a regulator constrained to non-negative tolls, and without additional instruments. In the light gray regions, the second-best optimal tolls are zero and the route structure in the untolled equilibrium differs from the second-best optimal. Therefore, an instrument that directly regulates route structure is required in these parameter areas. In the dark gray region, the constrained per-flight tolls are positive, but they are not able to decentralize the second-best optimal equilibrium. In this dark gray region, the airlines' equilibrium, when facing the tolls that would induce the optimal output under asymmetric hub-and-spoke route structure (H_x, H_y), is both using fully connected (F, F). This is because in this region, the loss in the connecting market of an airline moving from fully connected to hub-and-spoke is larger than the cost advantages of adopting a hub-and-spoke structure in an untolled setting, and, in absence of per-passenger subsidies, this cannot be overturned.

6. Conclusions

In the present paper, we have compared the unregulated equilibrium of route structure competition between airlines with market power, in the presence of congestion, with the welfare maximizing setting. Our analysis shows that the unregulated equilibrium in route structure may be different than the welfare maximizing one, and that a regulator may not always be able to decentralize the first-best route structures using the per-passenger tolls and the per-flight tolls that correct output inefficiency, even if these tolls are allowed to be market and airline specific. We also study the case where a regulator is constrained to use non-negative tolls. Again, the second-best optimal route structure may be different from the outcome of unregulated competition, and again, when the tolls are constrained to be non-negative, the second-best optimal equilibrium may not be decentralized with second-best tolls.

It follows from our analysis that, in some cases, a regulator may benefit from using an additional regulatory instrument, different from the tolls that are only concerned with the congestion externality and the market power exertion. The additional instrument has to align the airline choice of route structure with the welfare maximizing one, without affecting the choice of output. It also imposes additional requirements on the information needed by the regulator, as discrete changes in welfare due to discrete changes in route structures have to be assessed. For example, a direct restriction on the links that can be flown by each airline can solve the problem, or an arbitrarily high toll in the links that are not supposed to be used in equilibrium can also induce the first-best setting.

We also show that the result that a profit maximizing airline may have a bias towards the use of hub-and-spoke does not fully extend to an oligopolistic framework. We find that it is not the case that whenever hub-and-spoke route structures are welfare maximizing, they are also the equilibrium of the unregulated competition.

All of the results above, as additional sensitivity analyses show, seem to be robust to parameter values. Still, for our analysis we have used the simplest possible model that allows us to study optimal pricing in networks. We see extending the model as a natural avenue for future research. For example, the consideration of a larger network (more nodes) and the interaction between several airlines are logical extensions; 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 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.

Appendix A. Functional forms and parameters of the main case

We use the following functional forms for the schedule delay cost (g_i) and passengers' congestion cost (D):

$$g_i(f_i) = \gamma \cdot \frac{1}{f_i}\,, \tag{A.1}$$

$$D(F_i + F_j, K) = \alpha \cdot \left(\frac{F_i + F_j}{K}\right)^{\beta}\,, \tag{A.2}$$

where the schedule delay cost in (A.1) is inversely proportional to the airline frequency, and γ is a constant representing the monetary value of a unit of schedule delay time. The congestion delay at each airport in (A.2) is a function of the volume capacity ratio, with α being proportional to the passengers' value of travel time, K the capacity, and β the power of the function.

Table A.1: Parameter values.

Parameter	Value
A	1750
B	$\in$ [0.5,22]
E/B	$\in$ [0,1]
α	180
β	1
γ	80
K	4.85
c_q	50
c_f	33240

References

Alderighi, M., Cento, A., Nijkamp, P. and Rietveld, P. (2005), 'Network competitionthe coexistence of hub-and-spoke and point-to-point systems', *Journal of Air Transport Management* **11**(5), 328–334.

Barla, P. and Constantatos, C. (2005), 'Strategic interactions and airline network morphology under demand uncertainty', *European Economic Review* **49**(3), 703–716.

Basso, L. J. (2008), 'Airport deregulation: Effects on pricing and capacity', *International Journal of Industrial Organization* **26**(4), 1015–1031.

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

Brueckner, J. K. (2002), 'Airport congestion when carriers have market power', *The American Economic Review* **92**(5), 1357–1375.

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.

Brueckner, J. K. and Van Dender, K. (2008), 'Atomistic congestion tolls at concentrated airports? seeking a unified view in the internalization debate', *Journal of Urban Economics* **64**(2), 288–295.

Brueckner, J. K. and Verhoef, E. T. (2010), 'Manipulable congestion tolls', *Journal of Urban Economics* **67**(3), 315–321.

Burghouwt, G. and Hakfoort, J. (2001), 'The evolution of the european aviation network, 1990–1998', *Journal of Air Transport Management* **7**(5), 311–318.

Flores-Fillol, R. (2009), 'Airline competition and network structure', *Transportation Research Part B: Methodological* **43**(10), 966–983.

Flores-Fillol, R. (2010), 'Congested hubs', *Transportation Research Part B: Methodological* **44**(3), 358–370.

Fudenberg, D. and Tirole, J. (1984), 'The fat-cat effect, the puppy-dog ploy, and the lean and hungry look', *The American Economic Review* **74**(2), 361–366.

Hendricks, K., Piccione, M. and Tan, G. (1995), 'The economics of hubs: The case of monopoly', *The Review of Economic Studies* **62**(1), 83–99.

Hendricks, K., Piccione, M. and Tan, G. (1997), 'Entry and exit in hub-spoke networks', *The RAND Journal of Economics* **28**(2), 291–303.

Hendricks, K., Piccione, M. and Tan, G. (1999), 'Equilibria in networks', *Econometrica* **67**(6), 1407–1434.

Morrison, S. A. and Winston, C. (1995), *The evolution of the airline industry*, Brookings Institution Press, Washington, DC.

Oum, T. H., Zhang, A. and 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.

Oum, T. H., Zhang, A. and Zhang, Y. (1995), 'Airline network rivalry', *The Canadian Journal of Economics/Revue canadienne d'Economique* **28**(4a), 836–857.

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

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

Silva, H. E. and Verhoef, E. T. (2013), 'Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges', *Journal of Public Economics* **106**, 1–13.

Zhang, A. (1996), 'An analysis of fortress hubs in airline networks', *Journal of Transport Economics and Policy* **30**(3), 293–307.

Zhang, A. and Czerny, A. I. (2012), 'Airports and airlines economics and policy: An interpretive review of recent research', *Economics of Transportation* **1**(1–2).