Transportation Research Part B 67 (2014) 320–343

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

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

ARTICLE INFO

Article history:
Received 8 October 2013
Received in revised form 20 May 2014
Accepted 21 May 2014

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

ABSTRACT

We analyze the behavior of airlines in terms of route structure choice using a differentiated duopoly model that accounts for congestion externalities, passenger benefits from increased frequency, passenger connecting costs and airline endogenous hub location. We also examine the route structure configuration that maximizes welfare and whether it can arise as an equilibrium when a regulator implements optimal airport pricing, but does not regulate directly the route structure choice. We find that this is not always the case and that, therefore, 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 unconstrained tolling. Finally, we also study the relative efficiency of airport pricing when the optimal route structure configuration cannot be decentralized by tolling.

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. Also since the deregulation of the markets, the costs caused by congestion at airports have grown significantly and managing the increasing congestion has become one the main concerns of governments in countries with a large aviation market. For example, Ball et al. (2010) estimate that the cost of US air transportation delay in 2007 was $16.7 billions to passengers, $8.3 billions to air carriers and that it reduced the 2007 GDP by $4 billion. Not surprisingly, optimal airport pricing has gained increased attention as a measure aimed at reducing congestion costs.

The objective of this paper is to analyze optimal airport pricing in a network setting and in the presence of congestion externalities, where carriers with market power have the route structure choice as a strategic instrument. 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 deadweight 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

* Corresponding author. Tel.: +31 20 59 86106.
E-mail addresses: h.silvamontalva@vu.nl (H.E. Silva), e.t.verhoef@vu.nl (E.T. Verhoef), vberg@feweb.vu.nl (V.A.C. van den Berg).

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

http://dx.doi.org/10.1016/j.trb.2014.05.012
0191-2615/

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 conditions that give rise to hub-and-spoke route structures as an equilibrium of unregulated competition have often been explained with three arguments: economies of density, frequency effects and strategic advantages. The first refers to that average cost in a direct route may decrease with the number of passengers, and the second, the frequency effect, 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; Brueckner (2004) shows how frequency effects favor the adoption of hub-and-spoke. The third argument, strategic advantages, reflects that adopting hub-and-spoke route structures may bring, in oligopolistic competition, further advantages because of the effect it has on competitors. For instance, Oum et al. (1995) show that using a hub-and-spoke structure may allow the carrier to be more aggressive in output market competition. Employing hub-and-spoke 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). Finally, it may 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). Recent theoretical contributions to this topic include Hendricks et al. (1999), Alderighi et al. (2005), Barla and Constantatos (2005), and Flores-Fillol (2009, 2010).[2]

The above 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. This paper contributes to the literature by including all these together. For example, we find that airlines choosing hub-and-spoke structures using different cities as their hub may be the unique equilibrium when airports, markets and airlines are symmetric. The consideration of asymmetric hub-and-spoke networks, commonly observed in real markets, is an important contribution of this paper. In addition, we show that the result that a monopolistic airline is biased towards hub-and-spoke configurations (Brueckner, 2004) does not necessarily carry over to competing airlines under our assumptions. We find that airlines adopting fully connected route structures can be the unique equilibrium when using hub-and-spoke structures is socially optimal.

The 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. Brueckner (2004) compares the optimal route structure in a three-node network with the one adopted by an unregulated monopoly, but does not study pricing policies. Flores-Fillol (2009) extends the analysis to a duopoly of airlines without analyzing the optimal route structure. Brueckner (2005) and Flores-Fillol (2010) study the optimal pricing policy in a duopoly setting with an exogenous route structure. Our contribution to the policy analysis is to identify the rationale for the tolls, when route structure is endogenous, and to extend the optimal pricing and regulation analysis by elaborating on the policy instruments that can decentralize the socially efficient outcome in terms of output and network configuration.

A 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 (needed to eliminate deadweight losses), 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, despite that they provide the incentives to set the output efficiently, cannot always align the effect of adopting different route structures on the firms' profit with the effect on social welfare of those different configurations. This is especially true when the optimal network configuration is asymmetric and requires one firm to have, in one of the markets, a significantly lower market share and profit than the competitor.

First-best pricing, as just discussed, typically requires a regulator to give per-passenger subsidies to airlines, a policy that is arguably impossible to implement in practice. To address this limitation, we study the case in which the regulator is constrained to charge non-negative tolls. We show that the route structures and output levels that maximize welfare in absence of subsidies to correct for market power exertion cannot always be decentralized through non-negative tolls alone. Thus, also in the absence of subsidization, using an additional regulatory instrument, on top of the tolls designed to correct output inefficiencies, may increase welfare.

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.

[2] For an empirical analysis and review of the size and shape of the networks in the airline industry from a cost perspective, see Jara-Díaz et al. (2013).

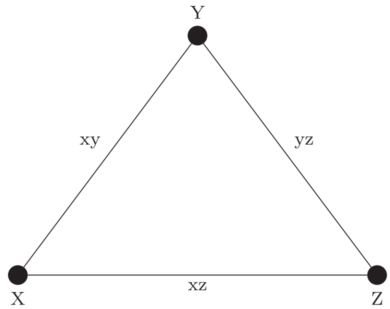

Fig. 1. Network.

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. This is particularly relevant for regions in which all airports are owned and operated by the same authority, and puts a question mark on the use of airport price-fixing systems, where all airports in a network charge identical tolls to airlines. As Bel and Fageda (2010) point out, in Spain, Greece (except Athens) and Norway there is a single operator that applies this system approach. Moreover, airlines operating in the local (national) markets in those countries operate asymmetric hub-and-spoke structures with different airports as their hubs, something that highlights our methodological contribution of allowing for these cases.

In addition to the policy implications discussed above, we show that more information is needed to achieve the regulator's objective: under fixed network structures, adaptive pricing based on "local variables" (notably marginal external cost and marginal benefit) would "guide" the regulator to the 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 sets the tolls based on this, may not always achieve welfare maximization, neither first-best nor second-best.

The remainder of the paper is structured as follows. In the next section 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 combination of output and route structure. 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 (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 Fig. 1. These markets $M = \{XY, YZ, XZ\}$ represent return-trips for simplicity (e.g. people travel from X to Y and return).[3] The links $L = \{xy, yz, xz\}$ are always available to any airline; that is, both airlines have permission to schedule flights in any city-pair. Each market m can be served by airlines either directly, flying nonstop 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.[4] As we let each airline's hub be endogenous, 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), the number of flights in each link (f_l), and the 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.[5]

We assume that the full price faced by a passenger traveling nonstop 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.[6] For a connecting trip, passengers incur an additional cost of layover time that we assume fixed and equal to μ. We further assume that airlines are perceived as

[3] Direction-specific return-trips (e.g. people making round-trips both from X to Y and from Y to X) may be introduced without altering the analysis.

[4] Equilibria where an airline chooses to serve fewer than the 3 origin–destination pairs is considered in the numerical model.

[5] 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).

[6] 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.

imperfect substitutes and that the passenger demand function for an airline i in market m, q_i^m, is linear in the own and in 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 $0 \leqslant E \leqslant B$. 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.

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

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

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%, which 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 transparent understanding of route structure equilibria. We assume that congestion affects airlines only through the demand function, in that increased congestion raises the full price faced by passengers and therefore, everything else constant, fares will be lowered by the increased congestion. We could also include that congestion affects the airlines' costs, but this would not affect the main results and conclusions in any fundamental way. In the profit function, there is no difference if congestion raises the airlines' costs or if reduces the passengers' willingness to pay.

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).[7] In the second and third stage, airlines choose route structure and output respectively.[8] We look at subgame-perfect Nash equilibria through backward induction, so we first analyze, in the following section, the airlines' output Nash equilibrium for a given network configuration and then we study the route choice Nash equilibrium (where airlines consider the effects of their network choice on the own and competitor's outputs).

3. Airlines equilibrium

In the second stage of the game, airlines have a discrete choice between alternative route structures (F or H at any of the airports). For this reason, 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. Due to our assumption that the load factor is constant and equal to 100%, 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.

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. $\partial g_i^{XY}/\partial f_i^{xy} < 0 \wedge \partial g_i^{XY}/\partial f_i^l = 0 \; \forall \; l \neq xy \wedge \partial g_i^{XY}/\partial f_j^l = 0 \; \forall \; l$). 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 at the destination, that the airport runway congestion depends on the 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 the link l, the full price faced by a passenger of market XY flying with airline i is:

$$\theta_i^{XY} = p_i^{XY} + D(F^{xy} + F^{xz}) + D(F^{xy} + F^{yz}) + g_i^{XY}(f_i^{xy}), \tag{4}$$

where D is the delay cost function, assumed common to all airports. The full price equals the sum of the following: the fare; the congestion at the origin, X, which depends on the total number of flights operating at that airport ($F^{xy} + F^{xz}$); the congestion at the destination airport, Y; and the schedule delay cost. Without loss of generality, both delay functions, D and g,

[7] 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.

[8] 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.

include the passengers valuation of time and reflect that these costs are experienced also in the return trip. We look at the case where all airports have the same delay function, but this could easily be extended.

The airline's profit, using Eq. (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 Eq. (2), we can rewrite profit as:

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

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

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

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

Eq. (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 the carrier's market power ($B \cdot q_i^m$). Eq. (8) states that the airline's marginal cost per flight (right-hand side of Eq. (8)) 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 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 constant per-flight costs.

3.2. The hub-and-spoke route structure

We now look at the case where airline i chooses to serve the markets with a hub-and-spoke route structure. For illustration purpose, we analyze the case where the airline chooses airport Y 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 change with respect to the fully connected case is that the market XZ (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 in the spoke market is set independently; this implies that the fare in market XZ is not restricted to be equal to the sum of the fares of the two hub markets (XY and YZ). The fares must, however, satisfy an arbitrage condition: the sum of the fares of the hub markets (in this case, XY and YZ) cannot be lower than the fare charged to the connecting passengers (market XZ).[9]

The number of seats per flight changes in this case because, in addition to the passenger from hub markets, the passengers from the spoke markets are also traveling through links xy and yz. As a consequence, in this setting, aircraft sizes will satisfy:

$$\begin{aligned} s_i^{xy} &= (q_i^{XY} + q_i^{XZ}) / f_i^{xy}, \\ s_i^{yz} &= (q_i^{YZ} + q_i^{XZ}) / f_i^{yz}. \end{aligned} \tag{10}$$

Full prices in the hub markets have the same structure as before (see Eq. (4)), but they change in the spoke market. We assume that the passengers' congestion delay, denoted $\widetilde{D}_i^{XZ}$, is the sum of the delays at each leg and that those passengers incur a constant transfer cost of μ, a standard assumption in the literature (e.g. Brueckner, 2004). The schedule delay of a connecting passenger is not straightforward to represent, in average terms, as a function of frequencies. Brueckner (2004) assumes that it is equal to the schedule delay of a passenger flying only one leg. In our model there is imperfect competition and endogenous hub location, therefore, that assumption would only be applicable if we impose *ex ante* symmetry. To avoid

[9] Note that we do not impose the arbitrage condition that the fare charged to connecting passengers has to be higher than the fare charged to passengers in direct markets. Instead, we assume that airlines can effectively prevent a passenger that purchases a ticket in the connecting market to make a return trip from the hub airport to one of the other two airports. As we model return trips, it is a reasonable assumption because, at least in one of the two directions, the passenger that intends to deviate will have to board a flight at the hub with a ticket that reveals the purchase of a trip originating in a different city. For the case of non-return trips, this is less likely and the additional arbitrage condition would have to be imposed.

this, and for practical purpose, we suppose that the schedule delay cost for a connecting passenger, denoted $\tilde{g}_i^{XZ}$, can be modeled as the average of the schedule delays in each of the two direct routes.[10] These assumptions give:

$$\begin{aligned}&\tilde{D}_i^{XZ} = D_i^{XY} + D_i^{YZ},\ \ \tilde{g}_i^{XZ} = (g_i^{XY} + g_i^{YZ})/2,\\ &\theta_i^{XZ} = p_i^{XZ} + \tilde{D}_i^{XZ} + \tilde{g}_i^{XZ} + \mu.\end{aligned} \tag{11}$$

The above assumptions shape profit in the following way:

$$\begin{aligned}\pi_i^H = &\sum_{m\in\{XY,YZ\}} q_i^m \cdot (A - B\cdot q_i^m - E\cdot q_j^m - g_i^m - D_i^m - c_q - \tau_i^m) + q_i^{XZ}\cdot(A - B\cdot q_i^{XZ} - E\cdot q_j^{XZ} - \tilde{g}_i^{XZ} - \tilde{D}_i^{XZ} - \mu - 2\cdot c_q - \tau_i^{XZ})\\ &- \sum_{l\in\{xy,yz\}} f_i^l \cdot (c_f + \tau_i^l),\end{aligned} \tag{12}$$

where the superscript H refers to an airline 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 XZ market requires an increase of aircraft size in both links, making the cost per connecting passenger equal to $2\cdot c_q$.

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

$$\frac{\partial \pi_i^H}{\partial q_i^m} = 0 \Rightarrow p_i^m = c_q + \tau_i^m + B\cdot q_i^m \quad \forall\, m \in \{XY, YZ\}, \tag{13}$$

$$\frac{\partial \pi_i^H}{\partial q_i^{XZ}} = 0 \Rightarrow p_i^{XZ} = 2\cdot c_q + \tau_i^{XZ} + B\cdot q_i^{XZ}, \tag{14}$$

$$\frac{\partial \pi_i^H}{\partial f_i^l} = 0 \Rightarrow -\sum_{m\in\{XY,YZ\}} (q_i^m + q_i^{XZ})\cdot\left(\frac{\partial D_i^m}{\partial f_i^l} + \frac{\partial g_i^m}{\partial f_i^l}\right) = c_f + \tau_i^l \quad \forall\, l \in \{xy, yz\}. \tag{15}$$

These equations state that airlines apply a market power markup in each market and set frequency to equalize marginal revenue with 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:

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

Just as in the previous case, it is the revenues from the markup minus frequency costs.

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 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 $r_i \in RS = \{F, H_X, H_Y, H_Z\}$, where H_C refers to a hub-and-spoke structure with airport C as the hub. The relevant comparisons are between the profits under different route structures, given the route structure of the rival. Let $\Pi_i(r, v)$ be the profit of airline i, 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 follows that a symmetric setting with both airlines using route structure r will be an equilibrium of the airlines' game if and only if:

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

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 subgame equilibrium of the airlines' competition. Also note that, whenever both airlines playing H_X is an equilibrium, both having H_Y and both having H_Z are also equilibria, because airports and 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:

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

[10] An alternative is to use the minimum frequency considering the frequency on the two routes, but this would generate unnecessary analytical complications to the model without changing results in any significant way.

This implies that u is the best response when the rival chooses v and *vice versa*, which again, because of airline, airport and market symmetry, implies that there are multiple equilibria for a particular u and v. We use (F, H) to denote 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.

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 analyze the expression that characterizes the cases in which using a fully connected structure is a best response to the rival using fully connected as well: $\Gamma_i \equiv \Pi_i(F,F) - \Pi_i(H_Y,F)$. The condition $\Gamma_i > 0$ is sufficient for (F,F) to be a subgame equilibrium of the airlines' game.[11] Denote $f_{i|(r,v)}$ the total number of flights scheduled by airline i in the third stage equilibrium where it uses r as a route structure and its rival uses v. Using Eqs. (9) and (16), Γ_i is:

$$\Gamma_i = \sum_{m \in M} B \cdot \left[(q^m_{i|(F,F)})^2 - (q^m_{i|(H_Y,F)})^2 \right] + c_f \cdot \left[f_{i|(H_Y,F)} - f_{i|(F,F)} \right], \tag{19}$$

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 types of effects driving the adoption of fully connected over hub-and-spoke: a 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 a change in costs, due to variations in the total number of flights (second bracketed term in the right-hand side).[12] Despite that it 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 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). It is, therefore, expected that the second bracketed term is negative, because the total frequency is lower under H_Y than under F, so that this favors the adoption of a hub-and-spoke route structure over a fully connected.

The change in number of passengers in each market, however, may favor the point-to-point structure. To see this, note that the connecting passengers (market XZ) face higher travel costs under H_Y than under F, because they incur higher travel delays and the additional cost of connecting (see Eq. (11)). As a result, the equilibrium number of passengers in the connecting market (for the given route structure of the rival) should be higher under a fully connected route structure.[13] This demand effect is what favors the adoption of F. On the other hand, the full price in the markets that are served directly under both route structures (markets XY and YZ) may be higher or lower due to two counteracting effects. Under a hub-and-spoke route structure, the frequency in each link is expected to be higher than under a fully connected route structure as more passengers travel between the hub and the other airports. This higher frequency implies higher congestion and decreased schedule delay costs and, therefore, its effect on the full price is, *a priori*, ambiguous. However, as markets XY and YZ are served directly under both configurations, the demand effect in these markets is expected to be significantly smaller than in the connecting market, and, therefore, it is not expected to be a main driving force of the sign of Γ_i.

The driving force behind a symmetric fully connected equilibrium (F,F) is the gain from an increased number of passengers in the connecting market when switching from a hub-and-spoke to a fully connected route structure. This demand increase has to be big enough so that the increase in revenues exceeds the increase in costs that results from operating a higher total number of flights. Therefore, it is expected that (F,F) is an equilibrium, for example, when demand is high and relatively sensitive to (full) price changes, which can be represented by a low value of B.[14] If the profit enhancing effect in the connecting market that favors the fully connected structure dominates when the rival's route structure is hub-and-spoke, but not when it is fully connected ($\Gamma_i < 0$), (F,H) is an equilibrium.

Conversely, when the cost benefits of adopting a hub-and-spoke route structure are dominant, equilibria with both firms using hub-and-spoke exist. This, as explained above, can be expected to happen when demand is relatively low and insensitive to price changes. Which of the two sets of equilibria that involve both firms adopting hub-and-spoke arises, (H,H) or (H_x,H_y), depends on the sign of $\Phi_i \equiv \Pi_i(H_X,H_Y) - \Pi_i(H_Y,H_Y)$:

$$\Phi_i = \sum_{m \in M} B \cdot \left[(q^m_{i|(H_X,H_Y)})^2 - (q^m_{i|(H_Y,H_Y)})^2 \right] + c_f \cdot \left[f_{i|(H_Y,H_Y)} - f_{i|(H_X,H_Y)} \right]. \tag{20}$$

Again, despite that it is not possible to assess the sign of Φ_i analytically, intuition can be provided. First, if demands are not interdependent, which in our model can only take place when airlines offer independent goods ($E = 0$) and there is no congestion, Φ_i becomes zero. This is because the route structure configuration and the output choice of the rival would not have any effect on own profit, and, as all markets are symmetric, the location of the hub would not affect profit either. In this extreme case, whenever the profit enhancing effects of adopting a hub-and-spoke route structure dominate the loss due

[11] Again, we use Y as the reference hub airport for illustration purpose. Because of the airlines' and the markets' symmetry, there is no need to analyze the condition $\Gamma_i > 0$ for all three airports.

[12] The decomposition of Γ_i that we use is helpful to analyze the equilibrium but it should be noted that it is not unique.

[13] With increased travel costs due to the connection, the marginal willingness to pay is shifted downwards and the equilibrium quantity is lower.

[14] Note that the direct demand system that results from Eq. (2) is: $q^m_i = A/(B+E) - \theta_i \cdot B/(B^2 - E^2) + \theta_j \cdot E/(B^2 - E^2)$. Therefore both the intercept as well as the own-price sensitivity increase as B is reduced.

to a demand reduction in the connecting market, both (H, H) or (H_x, H_y) are equilibria. It is the demand interdependencies, due to congestion or product substitution, that cause the divergence between these cases.

When Φ_i is positive, (H_x, H_y) arises instead of (H, H) if, in addition, fully connected is not the best response to the rival using hub-and-spoke. We first look at how the number of passengers in each market differs when adopting hub-and-spoke with a different city as the hub compared to using the same airport as the rival (the first bracketed term on the right-hand side of Eq. (20)). The market *XZ* is served directly under H_X and with a connection under H_Y. Given that the rival uses H_Y, choosing a different airport as the hub necessarily gives a comparative advantage in this market: by offering a direct flight where the rival offers a connecting service, the firm increases its market share and profit, compared to the case where both serve the market with a connection. The opposite holds true in the market *YZ*, which is served directly by the competitor under H_Y, but with a connection by firm *i* when adopting H_X. As a result, there is a demand reduction in the own connecting market that favors using the same city as the hub. Finally, the market *XY* is served by both firms with direct flights under both settings, so it can be expected that the demand change in this market does not play a major role in shaping the best-response. Something similar can be expected for the change in the total number of flights offered, which is the second bracketed term on the right-hand side of Eq. (20). Under both hub-and-spoke route structures, the total number of flights should not vary significantly and we can expect that the frequency costs are not a significant determinant of the sign of Φ_i.

It is therefore expected that the key determinant of the existence of the asymmetric hub-and-spoke equilibrium (H_x, H_y) should be whether the benefits from serving directly the market *XZ* (the rival's connecting market) exceed the disadvantages of serving market *YZ* with a connection (where the rival offers a direct flight). Although these two opposite effects may appear to be of the same size because of symmetry, they differ mainly because of congestion effects. Under (H, H) the hub is heavily congested because all flights either take off or land at airport *Y*, whereas under (H_x, H_y) each hub is less congested. Therefore, by serving directly the market *XZ*, the airline *i*'s passengers face a change from a heavily congested connection to a direct service. In the market *YZ*, the passengers face a change from a heavily congested direct service to a connection that is less congested. This suggests that there is more to be gained in the market *XZ* than what is to be lost in *YZ*. This is what may make using a different airport as the hub to be the best response, and therefore to give rise to the asymmetric hub-and-spoke equilibria (H_x, H_y).[15]

Finally, note that the conditions analyzed above are not necessarily mutually exclusive and multiple equilibria may arise. To complement the analysis above, we solve the equilibrium of the airlines' game numerically and provide further intuition.

3.4. Numerical example

In this section we show the equilibria in route structure of the airlines' game for a range of the own-demand sensitivity parameter, *B*, and all possible values of the ratio of substitutability, E/B, for a particular set of values of the other parameters (full details are in Appendix A). The functional forms used are: $g_i(f_i) = \gamma/f_i$ for the schedule delay cost (where γ is the disutility of a unit of schedule delay), and $D(F_i + F_j) = \alpha \cdot (F_i + F_j)/K$ for the users' congestion cost, where α is the value of travel time and *K* the airport's capacity. We choose to display our results in terms of those demand parameters, because they capture key characteristics of the markets. The parameter *B* allows us to control both the market size and the own-price sensitivity, something that, as discussed above, is expected to have a significant impact in the airlines' equilibrium: a higher value of *B* implies a lower demand and a lower own-price sensitivity (see footnote 14). The ratio E/B allows us to study the effect of different degrees of substitutability in our results, and compare them with the case where the demand interdependency is only through congestion. By varying the ratio from 0 to 1, we can assess the effect of product substitution on the results, in presence of congestion.

The remaining parameters are set according to the following criteria: (i) the second-order conditions hold in the studied parameter range (the second-order conditions involving the cross derivatives do not hold globally in our problem); (ii) the market demand price elasticities are in the range provided by InterVISTAS (2007) as representative of the airline industry (between -0.84 and -1.96); (iii) the own-demand price elasticities are in the range $[-0.1, -3.2]$, found by Brons et al. (2002) as representative, when we consider the ratios of substitutability estimated by Perloff et al. (2007) (between 0.4 and 0.8); and (iv) the airlines' operating cost parameters are the ones estimated by Swan and Adler (2006) for a long-haul aircraft configuration and a reference trip distance of 5000 km.

Fig. 2 shows the results in terms of route structure equilibria and 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 regions where both sets of equilibria may arise (the region between the two lines). The numerical analysis confirms the intuition provided above that for relatively low values of *B*, which imply higher demand levels and higher demand sensitivity, (F, F) is an equilibrium. This is where the effect of a decreased number of passengers in the connecting market dominates the cost benefits, when switching from a fully connected to a hub-and-spoke route structure. The opposite holds true for relatively large values of *B*. Our results in Fig. 2 also 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

[15] Note that we assume that the demand function for each OD pair is the same, which basically means that all origins and destinations are of the same size. In reality, cities differ in size, which makes larger cities more attractive as hubs and this would favor a symmetric network.

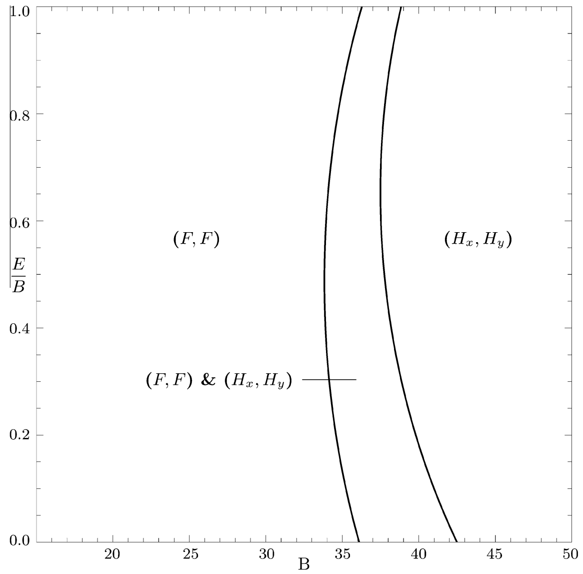

Fig. 2. Untolled equilibrium in route structures. Main parameterization (see Appendix A).

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 the number of passengers in the connecting market. The numerical example suggests, as discussed previously, that the gain from dominating the rival's connecting market is higher than the loss of being dominated in the own connecting market, when the reference point is a symmetric setting (H,H).

In addition, sensitivity analyses, in Appendix B, show that results are robust in the sense that, when different parameter values are considered, the absolute position and size of the regions in Fig. 2 change, but the relative positions and the set of equilibria that arise remain unchanged. It also confirms the previous results regarding the effect of demand and costs in the resulting equilibria (e.g. Flores-Fillol, 2009), which are summarized as:

- A higher value of the demand intercept (A), which implies higher demands, favors the adoption of fully connected route structures. This is in line with the intuition provided above; when markets are larger, the potential loss in the connecting market when adopting hub-and-spoke is increased.
- A higher value of the cost per flight (c_f) favors the hub-and-spoke route structure. This result follows directly from the analytical comparisons (see Eq. (19)), as the frequency cost advantages that using hub-and-spoke structures brings become more significant.
- A higher value of the marginal cost per seat (c_q) favors the fully connected route structure. Connecting passengers require an additional seat in two flights, so that the cost per connecting passenger is higher. This enhances the demand effect that favors the fully connected route structure.
- A higher value of the disutility of schedule delay costs (γ) favors the hub-and-spoke configuration, as the frequency benefits become more relevant. Concentrating passengers by increasing the frequency through hub-and-spoke structures increases the willingness to pay for the two direct flights and therefore the fare that can be charged.
- A higher value of travel time (α) favors the adoption of fully connected route structures. As hub-and-spoke route structures induce higher frequencies in the hub-to-spoke links and the traffic is more concentrated, it yields increased congestion. When congestion reductions are valued at a higher price, the increased congestion that hub-and-spoke structures bring shifts the passengers' willingness to pay downwards.
- A higher value of the constant transfer cost (μ) favors the fully connected route structure. When transferring is more costly for passengers, hub-and-spoke route structures become less attractive.

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 and derive the tolls that induce that output choice. Thereafter, we study the socially efficient route structure configuration 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:

$$SW^{(F,F)} = \left[\sum_{m\in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m\right] + \left[\pi_i^F + \pi_j^F\right] + \left[\sum_{m\in M} \tau_i^m \cdot q_i^m + \tau_j^m \cdot q_j^m\right] + \left[\sum_{l\in L} \tau_i^l \cdot f_i^l + \tau_j^l \cdot f_j^l\right], \tag{21}$$

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:

$$\frac{\partial SW^{(F,F)}}{\partial q_i^m} = 0 \Rightarrow p_i^m = c_q \quad \forall\, m \in M, \tag{22}$$

$$\frac{\partial SW^{(F,F)}}{\partial f_i^l} = 0 \Rightarrow -\sum_{m\in M} (q_i^m + q_j^m) \cdot \frac{\partial D^m}{\partial f_i^l} - q_i^m \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f. \tag{23}$$

We drop the index on the delay cost function, as it is the same for both airlines since we are looking at a symmetric route structure configuration. Eq. (22) states that the fare should equal the marginal cost of a seat in all markets, and Eq. (23) states that, in every link, frequency should be set such that the airline's marginal cost per flight equals the marginal benefit for all passengers. Comparing Eq. (7) with Eq. (22), and Eq. (8) with Eq. (23), we derive the tolls that maximize social welfare under (F, F):

$$\tau_i^m = -q_i^m \cdot B \quad \forall\, m \in M, \tag{24}$$

$$\tau_i^l = \sum_{m\in M} q_j^m \cdot \frac{\partial D^m}{\partial f_i^l} \quad \forall\, l \in L. \tag{25}$$

This is simply a per-passenger subsidy equal to the markup applied in each market, to eliminate the deadweight losses, and a link-specific per-flight toll equal to the uninternalized congestion. This is a traditional result in the airport pricing literature (e.g. Brueckner, 2005).

The two instruments above (Eqs. (24) and (25)) 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:

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

with quantities and frequencies satisfying Eqs. (22) and (23).

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 the (H, H) route structure (with Y as the hub airport in this case):

$$\frac{\partial SW^{(H,H)}}{\partial q_i^m} = 0 \Rightarrow p_i^m = c_q \quad \forall\, m \in \{XY, YZ\}, \tag{27}$$

$$\frac{\partial SW^{(H,H)}}{\partial q_i^{XZ}} = 0 \Rightarrow p_i^{XZ} = 2 \cdot c_q, \tag{28}$$

$$\frac{\partial SW^{(H,H)}}{\partial f_i^l} = 0 \Rightarrow -\sum_{m\in\{XY,YZ\}} (q_i^m + q_i^{XZ} + q_j^m + q_j^{XZ}) \cdot \frac{\partial D^m}{\partial f_i^l} - (q_i^m + q_i^{XZ}) \cdot \frac{\partial g_i^m}{\partial f_i^l} = c_f. \tag{29}$$

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

$$\tau_i^m = -q_i^m \cdot B \quad \forall\, m \in M, \tag{30}$$

$$\tau_i^l = \sum_{m\in\{XY,YZ\}} (q_j^m + q_j^{XZ}) \cdot \frac{\partial D^m}{\partial f_i^l} \quad \forall\, l \in \{xy, yz\}. \tag{31}$$

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:

$$SW^{(H,H)} = \sum_{m\in M} \frac{B}{2} \cdot ((q_i^m)^2 + (q_j^m)^2) + E \cdot q_i^m \cdot q_j^m - \sum_{l\in\{xy,yz\}} (f_i^l + f_j^l) \cdot c_f, \tag{32}$$

with variables satisfying Eqs. (27)–(29).

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 the optimal pricing and frequency setting rules will be a combination of the rules described above (Eqs. (22), (23) and (27)–(29)). 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, the tolls in Eqs. (24) and (25) should be charged, and for an airline using hub-and-spoke, the ones in Eqs. (30) and (31). The difference will be that the tolling rules will be evaluated at different outputs, and that Eqs. (30) and (31) have to be adjusted if the hub is not Y.

4.4. The optimal route structure

We now look at the combination of route structure and output that maximizes social 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. We also compare the choice of route structure by unregulated firms with the welfare-maximizing choice, to identify the potential sources of inefficiency.

Consider first the comparison between the following route structure configurations: (F,F) and (H,H). Let the difference between social welfare in both settings be $\Delta \equiv SW^{(F,F)} - SW^{(H_Y,H_Y)}$, where the airport Y is used as the hub for illustration purpose. The condition $\Delta > 0$ is necessary for (F,F) to be the welfare-maximizing route structure setting. Using Eq. (26), Eq. (32), and symmetry of the first-best solution we get:

$$\Delta = \sum_{m\in M} (B+E) \cdot \left[\left(q^m_{(F,F)}\right)^2 - \left(q^m_{(H_Y,H_Y)}\right)^2 \right] + 2 \cdot c_f \cdot [f_{(H_Y,H_Y)} - f_{(F,F)}], \tag{33}$$

where the variables in Eq. (33) are defined as $q^m = q_i^m = q_j^m \ \forall\, m \in M$ and $f = f_i = f_j$ because of symmetry, and they are evaluated at the social optimum for the given route structure setting in parentheses. The comparison between Γ_i in Eq. (19), which defines the sufficient condition for (F,F) to be an equilibrium ($\Gamma_i > 0$), and Δ in Eq. (33) sheds light on the inefficiency of route structures choice by profit maximizing agents. It also 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 in Δ and Γ_i to 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. In addition, in the welfare comparison (Δ), the cost difference due to a different total number of flights takes into account the savings for both firms (the factor of 2 multiplying the last term on the right-hand side of Eq. (33)). Finally, the airlines' relevant comparison is for a given route structure of the rival, which again makes Δ and Γ_i diverge.

A look at Δ in Eq. (33) reveals that the driving forces behind which route structure composition maximizes welfare, between the symmetric fully connected (F,F) and the symmetric hub-and-spoke (H,H) settings, are similar to the ones driving the airlines' choice. This is, there is a cost effect (second term on the right-hand side of Eq. (33)) that favors the hub-and-spoke over the fully connected route structure if the total number of flights is lower, which is a natural expectation. What favors the fully connected route structure is a higher consumer surplus in the market that is served with a connection under (H,H). When market XZ is served directly, the social cost (hence the full price paid) is lower because there are no connection costs and because a passenger requires an additional seat in one link instead of two (see Eqs. (27) and (28)). Therefore, demand is higher in that market when served directly and the consumer surplus increases. Again, as a result, we can expect that (F,F) yields higher welfare than (H,H) when demand is relatively high and price sensitive (low values of B). When this is not the case, and the cost benefits that hub-and-spoke route structures bring exceed the drawbacks from having connecting passengers, settings involving hub-and-spoke will be welfare maximizing. To gain insights on which of those settings could maximize welfare, let $\Lambda \equiv SW^{(H_X,H_Y)} - SW^{(H_Y,H_Y)}$ be the difference between social welfare under an asymmetric hub-and-spoke setting and a symmetric configuration involving hub-and-spoke. Straightforward calculations lead to:

$$\Lambda = \sum_{m\in M} \sum_{r\in\{i,j\}} \frac{B}{2} \left[(q^m_{r|(H_X,H_Y)})^2 - (q^m_{r|(H_Y,H_Y)})^2 \right] + \sum_{m\in M} E \left[q^m_{i|(H_X,H_Y)} \cdot q^m_{j|(H_X,H_Y)} - q^m_{i|(H_Y,H_Y)} \cdot q^m_{j|(H_Y,H_Y)} \right] + \sum_{r\in\{i,j\}} c_f \cdot \left[f_{r|(H_Y,H_Y)} - f_{r|(H_X,H_Y)} \right], \tag{34}$$

where the variables in Eq. (34) are evaluated at the social optimum for the given route structure setting in parentheses, in which the first structure refers to firm i and the second to firm j. What drives which of these two settings yields higher welfare is again related to the number of passengers who face a higher social cost and the difference in the airlines' costs. However, as both firms use a hub-and-spoke route structure in both configurations, the economies of density are exploited in similar ways and the difference in costs should not be significant. In other words, the total number of flights should not differ much between the two configurations, and, therefore, the last term on the right-hand side of Eq. (34) should not play a major role in determining the socially optimal setting between these two configurations.

Before elaborating about the sign of each of the remaining terms, note that if products are independent (i.e. $E = 0$) and there is no congestion, $\Lambda = 0$ holds and both settings yield exactly the same welfare. This is because demands are no longer interdependent and, as all markets are symmetric, the hub location would not have any effect on welfare. Consider now that the products are independent, but there is congestion. In this case, the difference between the two settings, (H_X, H_Y) and (H_Y, H_Y), is only due to differences in congestion. In the symmetric case, the hub (Y in this example) is congested by the two services of both airlines, whereas in the asymmetric case the airports that are used as a hub are congested by two services of one airline and one service of the other. This should result in a less congested situation and it should favor the asymmetric setting.

When there is also imperfect substitution, the number of connecting passengers, who are the ones facing a higher full price, is lower under (H_X, H_Y). When firms use hub-and-spoke with different airports as their hub, in each market that is served with a connection, one airline provides a direct service at marginal social cost, which is lower than the marginal social cost of the connection. Therefore, the effect of imperfect substitution also favors the asymmetric over the symmetric hub-and-spoke setting. In the limit, when products are perfect substitutes, the social optimum cannot have two airlines using different route structures (e.g. (H_x, H_y)), 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 under perfect substitution, instead, is to have regulated monopolized markets. This is one airline serving all three markets either with a fully connected (M_F hereafter) or with a hub-and-spoke route structure (M_H hereafter). The reason behind this result is that 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 deadweight 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 the schedule delay costs will be lower. This is what favors a monopoly. When differentiation is weak ($B \approx E$), this effect of lower schedule delay costs may still be dominant, and, as a result, the regulation of a monopoly may still yield higher welfare than of a duopoly. When differentiation is strong ($B \gg E$), 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.[16]

Fig. 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 Fig. 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.

Fig. 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 Figs. 2 and 3 confirms that the outcome of the unregulated competition may lead to route structure equilibria that are different from the efficient one.

Fig. 3 also suggests that the higher the substitutability between airlines (higher E/B), the more likely is (H_x, H_y) to be a more efficient route structure configuration than (F, F). It still brings cost savings and frequency benefits, but it is less essential that all airlines are present on all routes. As explained above, one of the effects that favors the fully connected symmetric setting over the hub-and-spoke structures is the lack of connecting passengers, who have a higher marginal social cost. In the 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 social 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 of (H_x, H_y) dominate the possible gains due to higher aggregate demand in the connecting markets under (F, F).

As Fig. 3 reveals, the result that monopolized markets yield higher welfare also holds when airlines are close substitutes: 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.

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

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

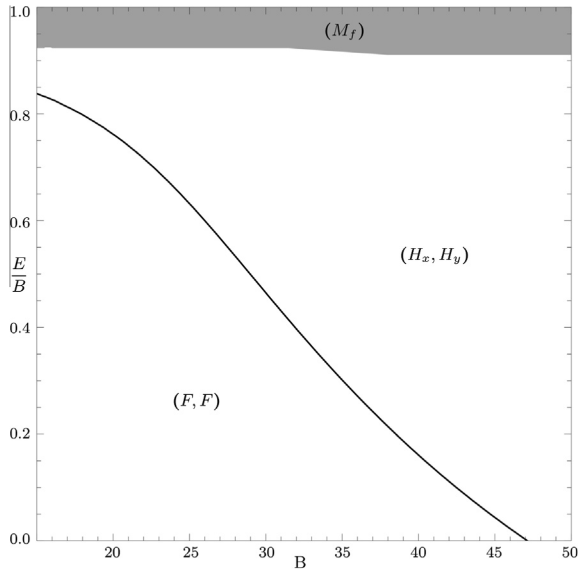

Fig. 3. Welfare-maximizing route structures. Main parameterization (see Appendix A).

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 Fig. 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 Fig. 3, the regulator set the tolls according to rules in Eqs. (22) and (23); if the equilibrium that results from charging these tolls is with both airlines choosing the fully connected route structure, the first-best is decentralized 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 the 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 configuration as an equilibrium, will not induce the optimal output.

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). Fig. 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 behind 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 unique and the same as the first-best, and only output corrections are needed. The light gray areas show where there is a route structure equilibrium of the untolled competition that 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_f 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 configuration when it is not the unique outcome of the unregulated competition.

Fig. 4a also shows that the result that a monopoly airline exhibits a bias towards hub-and-spoke route structure does not necessarily carry over to competing airlines.[17] That is, no longer whenever the welfare-maximizing setting has firms using hub-and-spoke structures, 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 the demand, the demand sensitivity to own-price changes and the degree of substitutability are not too low (roughly, in the triangle where B is lower than 40 and E/B between 0.15 and 0.9 in Fig. 4a). Recall that in that area we find (F, F) in the untolled equilibrium because, when demand is relatively high and sensitive to price changes (low B), the profit loss in the connecting market is larger than the cost benefits that may be obtained when switching from a fully connected to a hub-and-spoke route structure. However, the 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, who are the ones with a

[17] In a monopoly context, Brueckner (2004) shows that the choice of route structure by a monopoly airline is biased towards hub-and-spoke. In our problem, the divergence between settings is more complex 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.

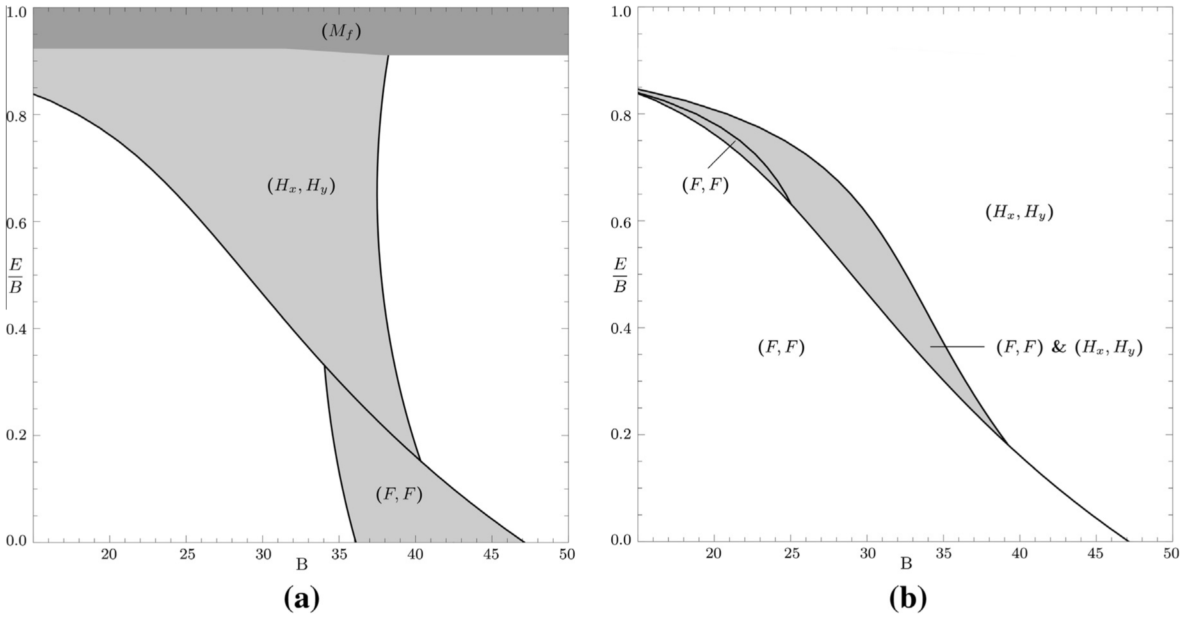

Fig. 4. First-best tolls' rationale (a) and equilibrium under optimal pricing (b). In the shaded areas of Figure (a) there are equilibria of the untolled competition that are different from the welfare-maximizing route structure configuration and labels indicate the welfare-maximizing route structure configuration. Labels in Figure (b) indicate the equilibrium in route structures under optimal airport pricing, and the shaded areas highlight the cases where tolls cannot always decentralize the first-best. Main parameterization (see Appendix A).

higher social cost. 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 the substitutability between firms is not too low, the resulting 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 results also indicate that a "naive" regulator, who observes the unregulated equilibrium and sets 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 light gray regions), and induce airlines, via tolls, to change the way they serve the markets. For example, a regulator that plans to put congestion pricing into action may need to set optimal toll levels that are far from what would be optimal if the route structure choice was exogenous.

Fig. 4b shows the subgame-perfect Nash equilibrium in terms of route structure, i.e. the equilibrium in route structure choice when the regulator applies optimal pricing designed to induce the optimal output based on the welfare-maximizing route structure configuration. 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 subgame-perfect Nash equilibrium is not always efficient, even when the regulator can perfectly discriminate between airlines and when it has no budget constraints. The shaded areas in Fig. 4b represent the parameter range where the toll instruments described in the previous sections cannot always decentralize the first-best. In these regions, airlines adopting fully connected route structures is an equilibrium whereas the first-best setting is (H_x, H_y). As Fig. 4b shows, there is a region in which (H_x, H_y) is welfare maximizing but the unique equilibrium from the tolled competition is (F, F), and there is also a region in which (H_x, H_y) is welfare maximizing but (F, F) is one of the two possible equilibria.

To provide the intuition behind this result, let us focus on firm *i*. The welfare-maximizing setting in the shaded areas in Fig. 4b is (H_x, H_y) and it has firm *i* offering a connection in a market (*YZ*) in which the rival offers a direct service with a lower full price. This maximizes welfare because it yields a low number of connecting passengers (who face a higher social cost) when the firms' products are imperfect substitutes. However, while having firm *i* offering a connection and having a small market share in the market *YZ* is welfare enhancing, it is, at the same time, profit reducing. From the airline *i* point of view, given that the competitor flies directly in that market and substitutability is not low, flying directly may be the best response, as it increases the willingness to pay of own-passengers (no layover time cost) and reduces the marginal cost (one seat instead of two). This is what occurs in the shaded regions of Fig. 4b, where the misalignment of the market *YZ*'s demand effect on welfare and profit makes the route structure equilibrium to be different than the optimal configuration. To ensure that the first-best outcome is reached in these cases, an additional instrument is therefore required.[18]

[18] For example, an arbitrarily high per-flight toll to each airline only when they fly directly in the market that is meant to be served with a connection 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.

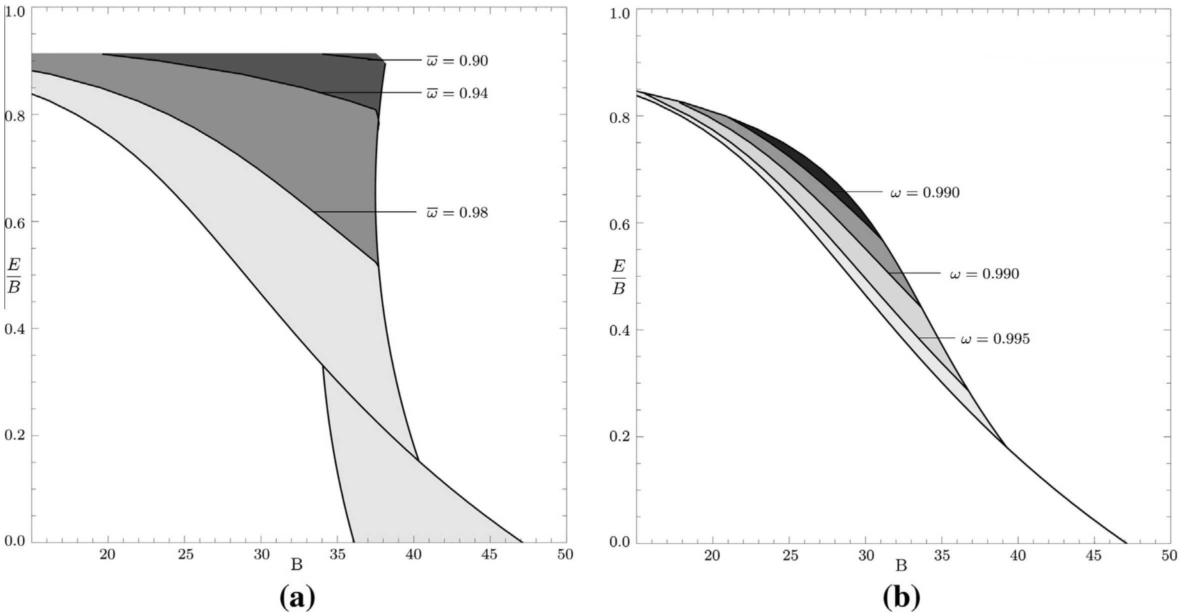

Fig. 5. Relative efficiency of naive airport pricing based on the untolled route structure equilibrium (a) and relative efficiency of optimal airport pricing (b). Main parameterization (see Appendix A).

Another policy implication of our analysis is that more information is needed to attain the first-best, as the variables usually needed for setting first-best tolls (e.g. marginal external cost) are not enough to assess discrete changes in welfare under different route structures. A question that naturally follows is how does pricing perform when the optimal network configuration is not assessed or known by the regulator. This is, what is the relative efficiency of setting the tolls based on the observed unregulated route structure equilibrium (in Fig. 2). This naive pricing is inefficient whenever the untolled equilibrium differs in terms of route structure from the welfare-maximizing setting (the light gray areas of Fig. 4a). Fig. 5a shows the relative efficiency achieved by a "naive" regulator, ϖ, as the percentage of the maximum welfare that can be gained (relative to the untolled equilibrium as the reference scenario). It reaches a minimum of 0.9 in the darkest gray regions of Fig. 5a and a maximum of 1 in the white areas. We assume that the inefficient setting arises as equilibrium in the cases where multiple equilibria may arise, in order to assess the minimum value of ϖ without using an equilibrium selection mechanism.

Another question that follows from the above results is how big the loss in welfare is, when charging the tolls that correct the output inefficiencies and not achieving the first-best route structure. Again, in this exercise, we look at the case where the inefficient route structure equilibrium arises in the region where multiple equilibria may arise (see Fig. 4b), to assess the worst-case scenario without analyzing an equilibrium selection mechanism. Therefore, there will be a welfare loss where the optimal tolls cannot always decentralize the first-best in terms of route structure, the shaded areas of Fig. 4b. There, (H_x, H_y) is the optimal route structure configuration, but (F, F) arises as an equilibrium under optimal pricing and, therefore, tolls are designed to induce the optimal output for the suboptimal (F, F) configuration. Fig. 5b shows the relative efficiency of such tolls (ω), as the percentage of the maximum welfare that can be gained. The relative efficiency ranges from 1, in the white areas of Fig. 5b where the tolls are sufficient instruments to achieve the first-best, to 0.99, in the darkest region of the figure. This suggests that although there is a theoretical possibility of decentralizing the "wrong" first-best route structure configuration, the welfare losses may be modest.

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 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 heavily subsidized airline may be enough to shift the rival's demand to levels that are not profitable or non-positive.

Finally, the sensitivity analyses in Appendix B suggest that the results are robust in that only the boundaries indicated in Fig. 4 change when parameters are changed, but the main qualitative results hold. Specifically, we find that the tolls are not always sufficient to achieve the first-best when varying the cost parameters (c_f and c_q), the demand intercept (A), the disutility of schedule delay costs (γ), and the congestion levels (e.g. through different airport capacities). Also in those cases, the relative efficiency of optimal airport pricing (ω) and naive airport pricing (ϖ) remains high.

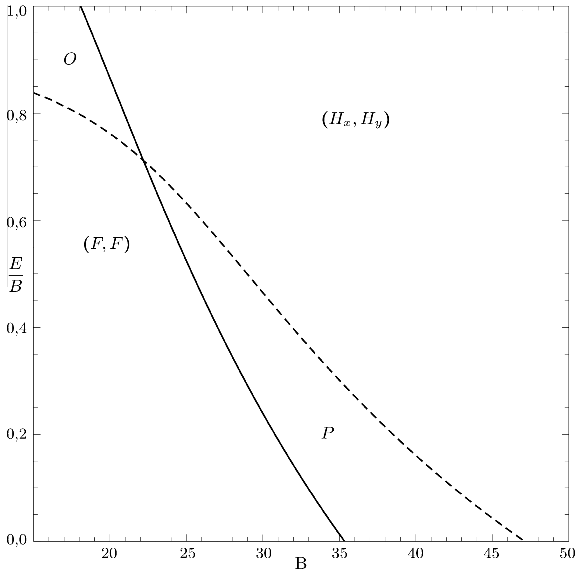

Fig. 6. Second-best optimal route structures (solid line) and comparison with the first-best case (dashed line). Main parameterization (see Appendix A).

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 more realistic case where only non-negative tolls can be charged, and compares the second-best optimal route structure configuration with the first-best setting. An important policy 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. (24) and (30)). The first-best per-flight toll is always non-negative (un-internalized congestion, see Eqs. (25) and (31)). 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 inefficiency from the congestion externality is sufficiently low compared to the inefficiency from the market power exertion.[19]

Fig. 6 shows the second-best optimal setting for the same parameter constellation and functional forms as in previous sections. 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. Note that we omit, in Fig. 6, the region in which regulating monopolized markets is optimal when the regulator can subsidize airlines (see Fig. 3). This allows for a comparison between the second- and first-best optimal route structure configuration.

Fig. 6 shows that also in the second-best situation considered, (F, F) is the welfare-maximizing setting for low values of B (relatively high demand and high price sensitivity), and (H_x, H_y) for higher values of B. In a similar way as in the first-best case, a higher ratio of substitutability favors the asymmetric hub-and-spoke over the symmetric fully connected structure.

Despite the similarity, there is a region – denoted P – where (H_x, H_y) is second-best optimal whereas (F, F) is first-best optimal, and a region – denoted O – where (F, F) is second-best optimal whereas either (H_x, H_y) or (M_f) is first-best optimal. The difference is caused by that in the second-best cases, there are deadweight losses due to the lack of subsidies and the demand is lower than in the first-best. The monopoly region does not exist for the considered parameters, as the monopolist's market power exertion is no longer being corrected through the subsidy; it is better to regulate a duopoly when tolls are constrained to be non-negative. When both airlines use fully connected, competition is direct in all three markets, whereas in the asymmetric hub-and-spoke equilibrium there are two markets where one airline has an advantage over the other. Therefore, the direct competition that a fully connected route structure brings may decrease deadweight losses by more than under an asymmetric hub-and-spoke structure. This effect, absent in the first-best case, favors (F, F). On the other hand, as demand is lower, frequencies are lower and the cost advantages of hub-and-spoke structures are more attractive. Phrased differently, there are more incentives to exploit the economies of density in the second-best that in the first-best case. In the area O, substitutability (E/B) is high and demand is high and more price sensitive (low B), so that the effect of the reduction of the deadweight losses dominates, and the more intense competition in (F, F) yields higher welfare in the second-best,

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

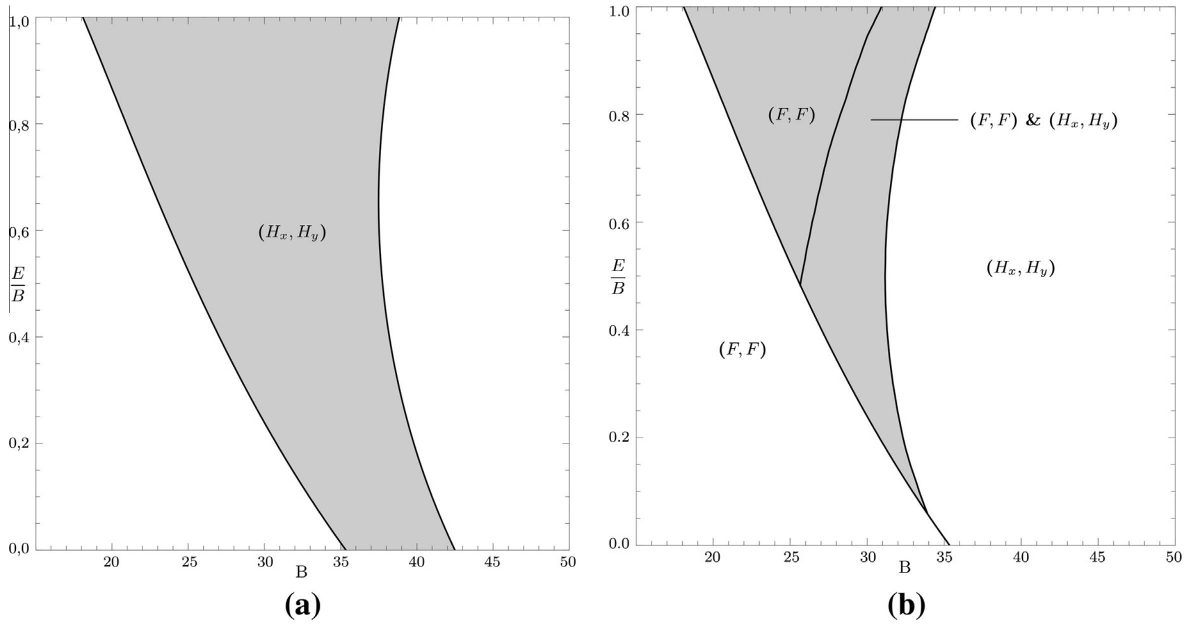

Fig. 7. Second-best tolls' rationale (a) and equilibrium under non-negative optimal pricing (b). In the shaded area of Figure (a) there is an equilibrium of the untolled competition that is different from the second-best optimal route structure configuration and the label indicates the second-best optimal route structure configuration. Labels in Figure (b) indicate the equilibrium in route structures under optimal non-negative tolls, and the shaded areas highlight the cases where non-negative tolls cannot always decentralize the second-best. Main parameterization (see Appendix A).

whereas (H_x, H_y) or (M_f), where subsidies directly stimulate demand, is first-best optimal. In the area P, where demand is lower and less price sensitive (higher B) and substitutability (E/B) is not too high, these effects are reversed, and the cost advantages of hub-and-spoke structures become more attractive, making (H_x, H_y) second-best optimal where (F, F) is first-best optimal, given the higher demand that the first-best subsidies induce.

Fig. 7a compares the second-best optimal route structure configuration 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 unique and the same as in the second-best optimal setting, and tolls are only needed to correct for output inefficiencies. The gray region shows where there is an untolled equilibrium in route structures that is different from the second-best optimal configuration, and therefore tolls ideally would be required, but sometimes are unable, to correct both. The label indicates the second-best optimal route structure setting when an inefficient untolled equilibrium exists.

Finally, where the second-best optimal route structures in Fig. 6 are assumed to apply, we also assess whether they are decentralized by non-negative tolls, in order to study whether the per-flight tolls are sufficient instruments to achieve the second-best outcome in route structures shown in Fig. 6. The main result of this exercise is that a regulatory instrument designed to correct route structure may be needed in addition to the per-flight tolls: there are different regions where the second-best optimal route structures and outputs differ from the airlines' equilibrium with non-negative charges. Fig. 7b summarizes these results by showing the equilibria of the tolled competition. The gray regions show the cases where the second-best optimal setting (H_x, H_y) cannot always be enforced by a regulator constrained to non-negative tolls, and without additional instruments. Although the constrained per-flight tolls are positive in these regions, they are not always able to decentralize the second-best optimal configuration. In these gray areas, (H_x, H_y) is the second-best optimal configuration, but both airlines using fully connected (F, F) is an equilibrium when they face the tolls designed to induce the second-best optimal output under (H_x, H_y). Again, there is a region where (F, F) is the unique equilibrium under non-negative tolls whereas (H_x, H_y) is second-best optimal, and a region where (F, F) is one of the two possible equilibria under non-negative tolls whereas only (H_x, H_y) is second-best optimal.

This result extends our finding that optimal pricing may not decentralize the choice of the optimal route structure configuration to the second-best case. However, when subsidies are not feasible, we find that the potential welfare gains from second-best tolling are limited.[20] As a consequence, the welfare losses of not having the second-best optimal route structure configuration are low. Therefore, also in the second-best case, our results suggest that there is a chance of decentralizing a sub-optimal route structure configuration through optimal pricing, but that the welfare losses may be moderate.

[20] In the studied parameter range, the potential gains from second-best tolling are not higher that 3%.

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. A first main finding from our analysis is 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. Secondly, we also study the case where a regulator is constrained to use second-best 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 setting may not be decentralized with tolls.

An important policy implication from our analysis is that 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. 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.

Another policy implication from our results is that regulators designing pricing policies need more information than what would be required if route structures were exogenous. Even if the policymaker cannot use an additional regulatory instrument and wants to, for example, put congestion pricing in place, it needs to assess discrete changes in welfare due to discrete changes in route structures. Marginal benefits and marginal costs are not always enough information to design a pricing policy. It may happen that in order to fully reap the long-run benefits from congestion pricing, the tolls should be designed for a different route structure configuration than the one that is observed. Nevertheless, our results suggest that the relative welfare losses from inducing the wrong network configuration through tolling, be it first-best or second-best, are limited. So, while the conceptual point that optimal pricing may not decentralize the choice of optimal route structures is significant, the consequent welfare losses may be limited. The latter result, in itself, is of course also an important insight for practical policy making. In any case, it appears to be so in the small numerical example we have developed and its real relevance is a matter of empirical investigation.

All of the results above, as additional sensitivity analyses show, seem 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.

Acknowledgment

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):

$$g_i(f_i) = \gamma \cdot \frac{1}{f_i}, \quad D(F_i + F_j, K) = \alpha \cdot \left(\frac{F_i + F_j}{K}\right), \tag{A.1}$$

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.1) is a function of the volume capacity ratio, with α being proportional to the passengers' value of travel time, and K the capacity. The parameter values of our main application are summarized in Table A.1.

The airlines' cost parameters are from Swan and Adler (2006) for a 5000 km. trip and a long-haul aircraft configuration. The remaining parameters are set to obtain realistic values of elasticities and aircraft size. Perloff et al. (2007) estimate linear demand functions for two Chicago-based markets in which two airlines had a joint share of 98% of passengers, so that it can be regarded as a case of duopolistic competition. They find ratios of substitutability between 0.4 and 0.8. With our parameters in the untolled competition, and with a ratio of substitutability of 0.8, the minimum elasticity in a direct market under (F, F) is -2.94 and under (H_x, H_y) is -3.12. When E/B is 0.4, we find a maximum elasticity in a direct market under (F, F) of -1.25 and under (H_x, H_y) is -1.31. All values are in the range of $[-0.1, -3.2]$ found by Brons et al. (2002) in their meta-analysis of elasticities of the airline industry.

The market elasticities that we find under (F, F) and under (H_x, H_y) are in the range $[-0.86, -2.08]$, which is roughly the same as the one found by InterVISTAS (2007): $[-0.84, -1.96]$. Finally, in the equilibrium of the untolled competition we find a minimum aircraft size of 70 seats and a maximum of 232 seats.

Table A.1
Parameter values.

Parameter	A	B	E/B	α	γ	μ	K	c_q	c_f
Value	4400	$\in$ [15,50]	$\in$ [0,1]	128.25	80	40	4.85	82.8	17,470.8

Appendix B. Sensitivity analysis

In this section, we show how our results change when different parameter values are considered. We solve, for different parameter values, the equilibrium of the airlines' untolled competition, derive the welfare-maximizing setting and study whether pricing is enough to achieve this setting as an equilibrium. For each parameter, we study an increase and a decrease of 25% of its value, and show three figures that summarize the results. The first of the three figures, to be compared with Fig. 2, shows the equilibrium in route structures of the untolled competition. If, as a result of a change in a parameter value, the lines that divide the regions in which different equilibria arise move to the left, we conclude that the hub-and-spoke route structure is favored. Phrased differently, when the region in which (H_x, H_y) is the unique equilibrium is larger, we conclude that the parameter change favors the adoption of hub-and-spoke route structures. It follows that if the lines move to the right, i.e. the region in which (F, F) is the unique equilibrium is larger, the parameter variation favors the fully connected route structure.

The second figure, to be compared with Fig. 3, shows the welfare-maximizing route structure configuration. As a result of our sensitivity analysis, we find that the effect of a change in a parameter value on the welfare-maximizing route structures has the same pattern as for the firms' choice. This is, whenever a parameter variation enlarges the region where a route structure configuration is the unique equilibrium from the untolled competition, it also enlarges the region where that configuration maximizes welfare. The only exception is the value of travel time, where the optimal route structure configuration does not vary significantly in the studied parameter region for α.

Finally, the third figure that we show for each parameter change, to be compared with Fig. 4b, displays the subgame-perfect Nash equilibrium in terms of route structure. This is, the equilibrium in route structure when the regulator applies optimal pricing based on the welfare-maximizing route structure. We find that in all cases, there is a region where (F, F) is an equilibrium of the tolled competition whereas (H_x, H_y) maximizes welfare. Therefore, also when different parameter values are assumed to apply, we find that optimal pricing cannot always decentralize the first-best also in terms of route structure, and an instrument directly aimed at regulating route structure choice may be needed to maximize welfare.

B.1. Sensitivity with respect to the demand intercept

Figs. B.8 and B.9 show that a higher value of the demand intercept favors the fully connected route structures, and that a lower value favors the hub-and-spoke route structures. This is in line with the intuition provided in the previous sections, as higher demand levels favor the fully connected route structures because the losses in connecting markets are larger.

B.2. Sensitivity with respect to the disutility of a unit of schedule delay

Figs. B.10 and B.11 reveal that when frequency benefits become more important, through a increase in the disutility of the schedule delay, hub-and-spoke route structures are favored. This is in line with the results in previous literature (e.g., Brueckner, 2004), as hub-and-spoke route structures induce higher frequency in the hub markets.

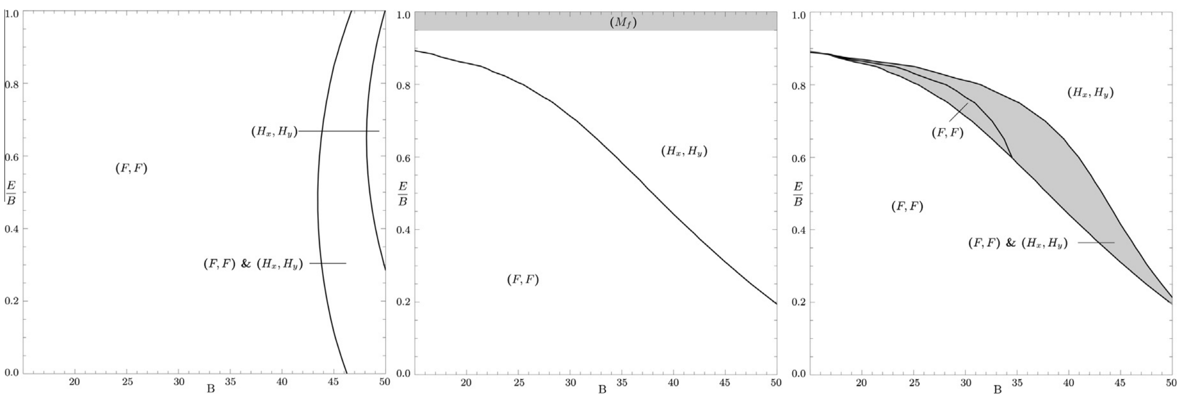

Fig. B.8. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher demand intercept ($A = 5500$).

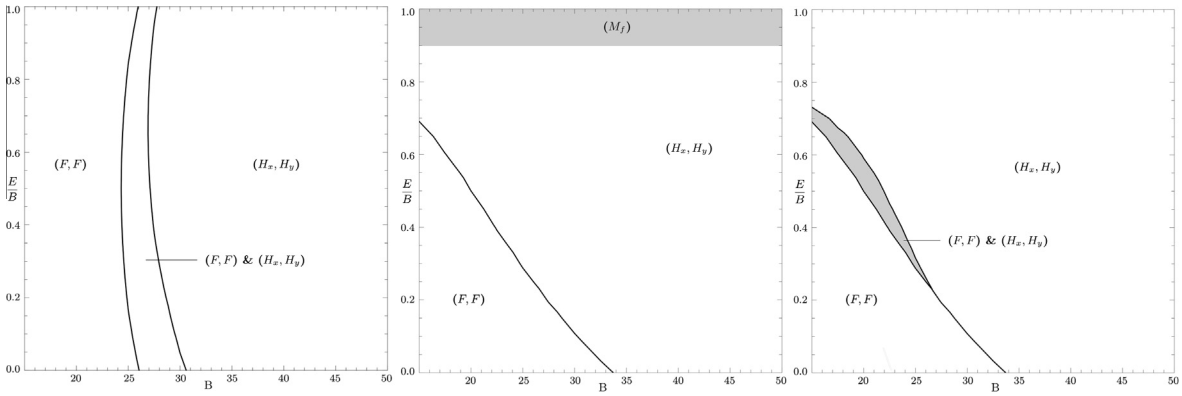

Fig. B.9. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower demand intercept ($A = 3300$).

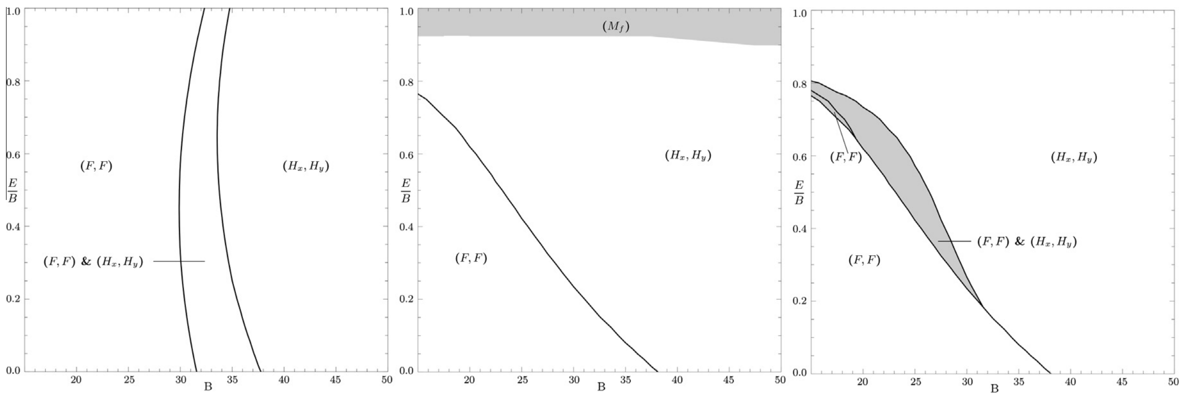

Fig. B.10. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher disutility of a unit of schedule delay ($\gamma = 100$).

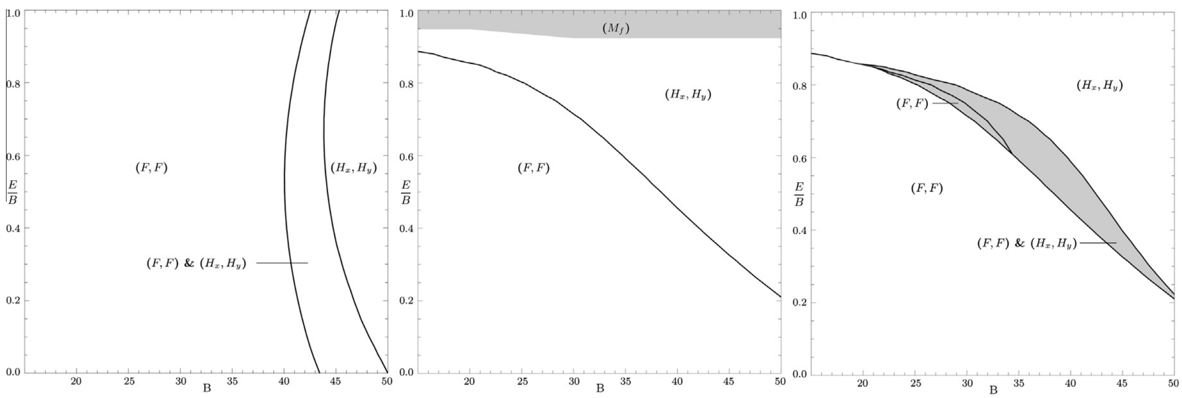

Fig. B.11. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower disutility of a unit of schedule delay ($\gamma = 60$).

B.3. Sensitivity with respect to the constant cost per flight

Figs. B.12 and B.13 show that higher values of the fixed cost per flight favor the hub-and-spoke route structures. This follows straightforwardly from that the cost benefits that hub-and-spoke networks bring, through a reduced total number of flights, are proportional to the cost per flight (see Eq. (19)).

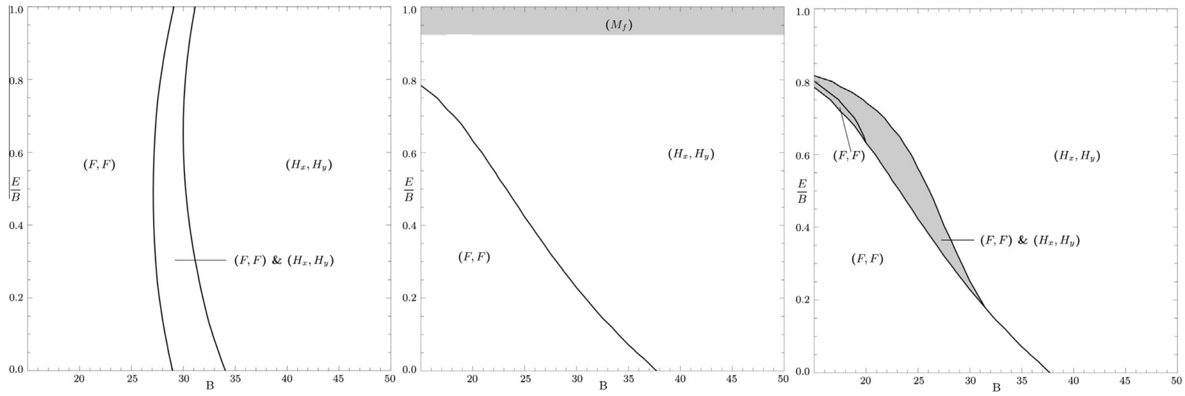

Fig. B.12. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher constant cost per flight ($c_f = 21838.5$).

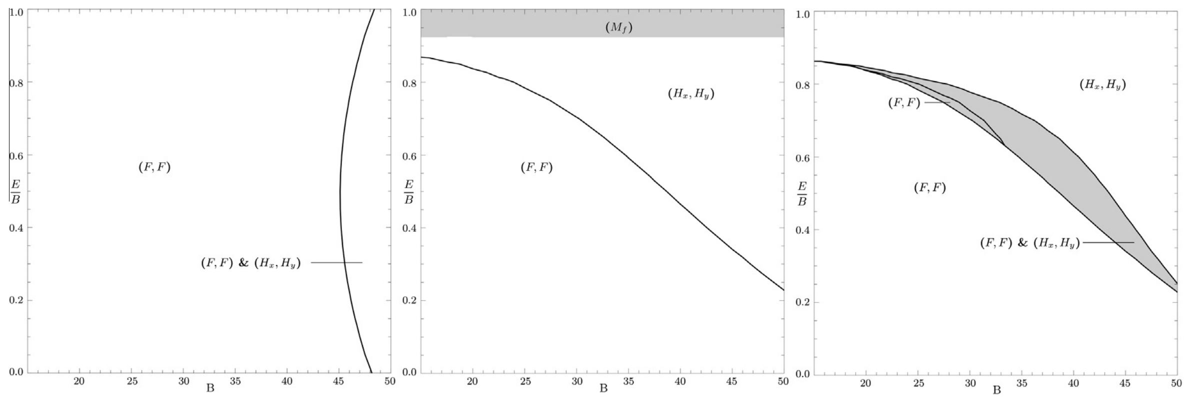

Fig. B.13. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower constant cost per flight ($c_f = 13103.1$).

B.4. Sensitivity with respect to the constant cost per seat

Figs. B.14 and B.15 display the result that a higher value of the constant cost per seat favors the fully connected route structure. When the marginal cost of a seat is higher, the connecting passengers' face higher costs as they require a seat in two legs. As a consequence, the profit that can be made in connecting markets is lower and the social cost of a connection is higher. This is the reason why a higher value of c_q favors the fully connected route structure.

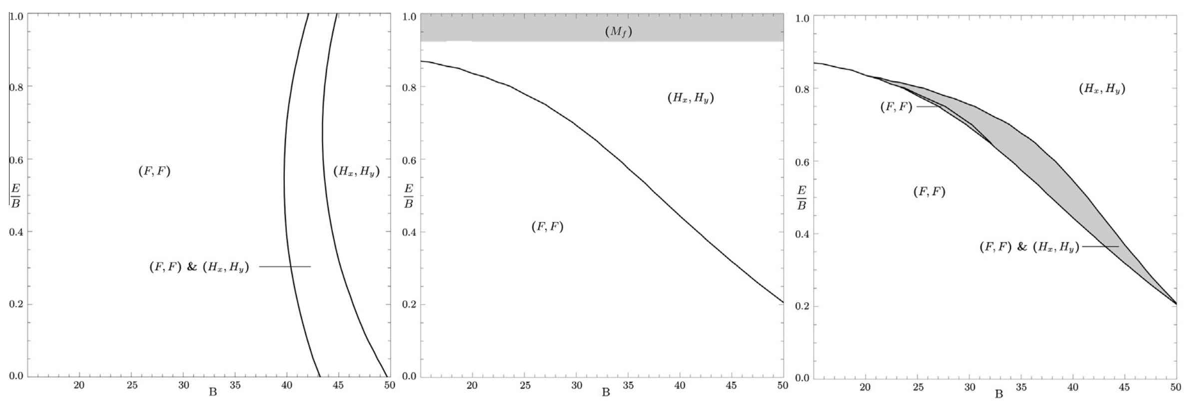

Fig. B.14. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher constant cost per seat ($c_q = 103.5$).

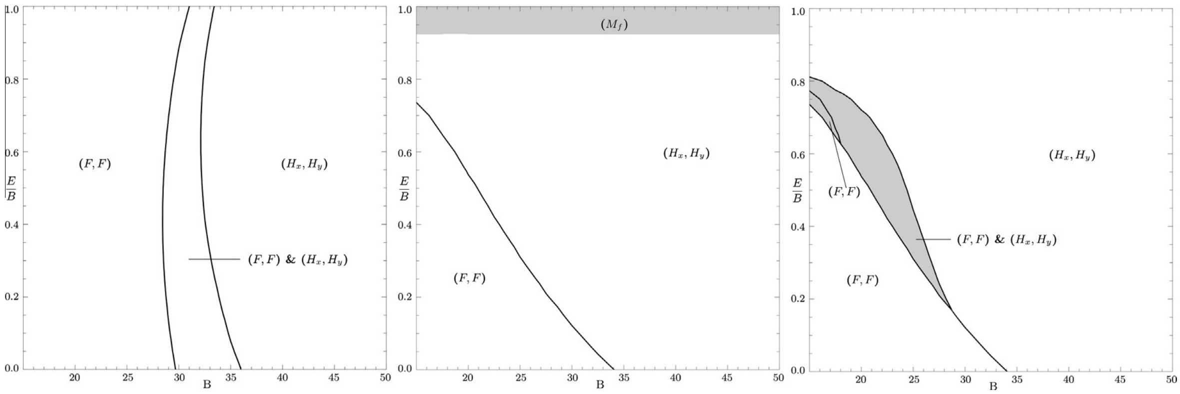

Fig. B.15. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower constant cost per seat ($c_q = 62.1$).

B.5. Sensitivity with respect to the value of travel time

The results below show that a higher value of travel time favors the adoption of fully connected route structures by (unregulated) firms. As hub-and-spoke route structures induce higher frequencies in the hub markets, they yield increased congestion, which explains why a higher value of α favors the fully connected route structures. This is the opposite of what happens when the disutility of the schedule delay increases. Figs. B.16 and B.17 show that the optimality of the different

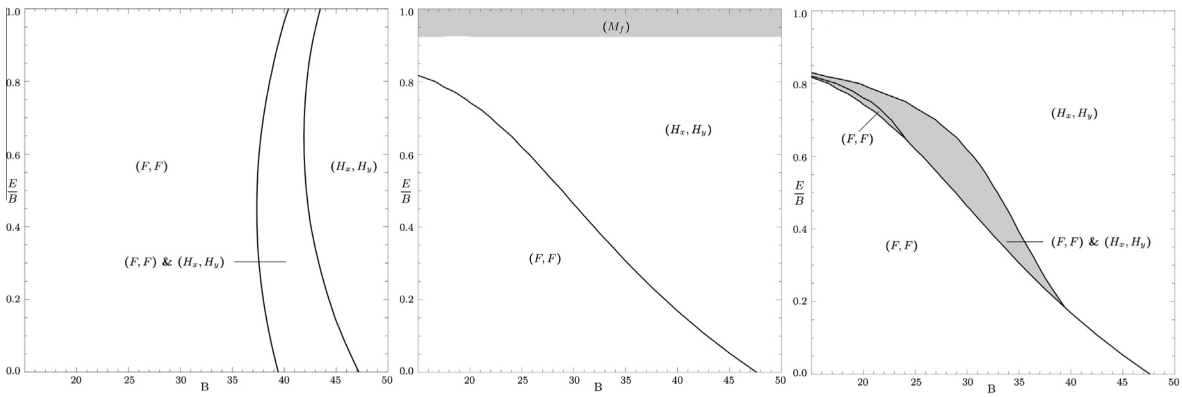

Fig. B.16. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher value of travel time ($\alpha = 160.31$).

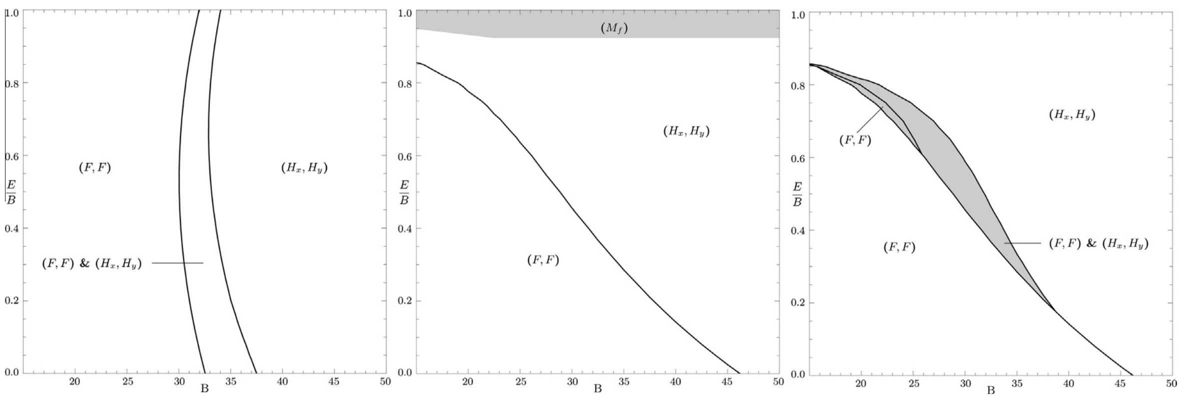

Fig. B.17. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower value of travel time ($\alpha = 96.19$).

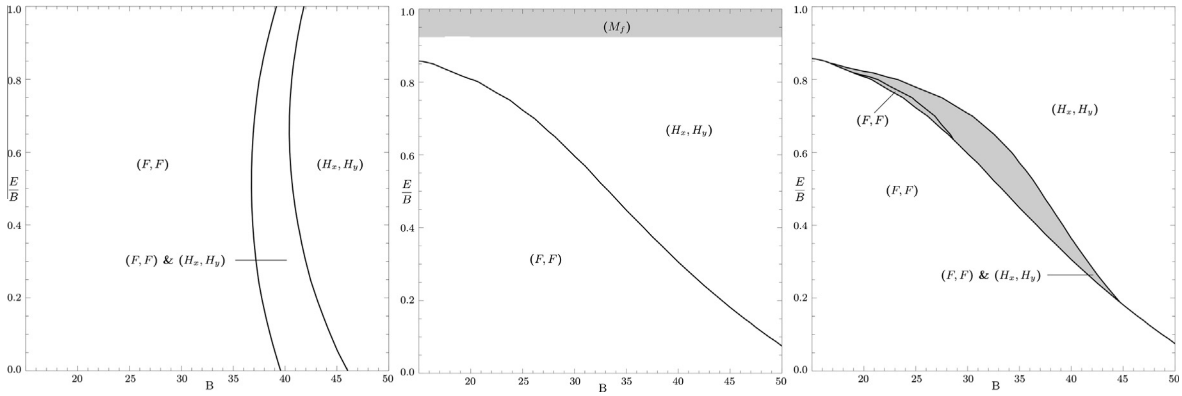

Fig. B.18. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Higher constant cost transfer cost ($\mu = 50$).

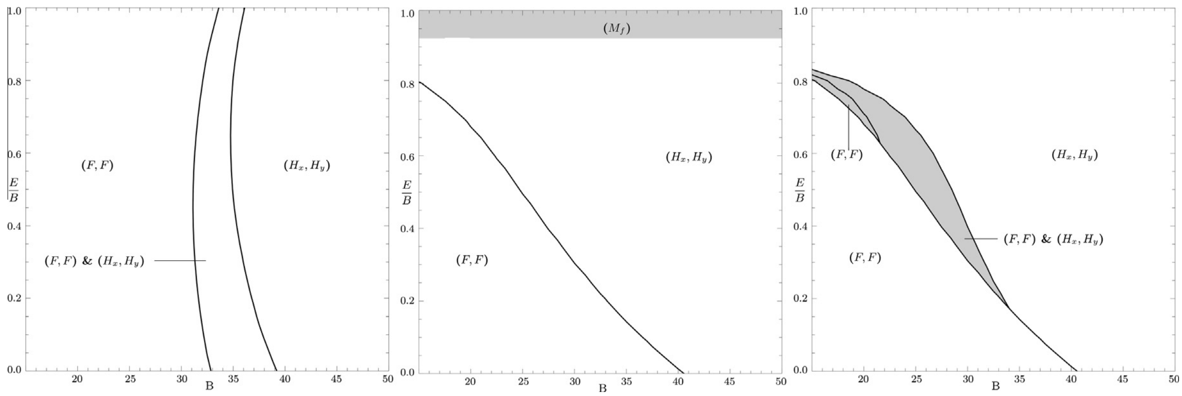

Fig. B.19. Untolled equilibrium (left), welfare-maximizing configuration (center) and equilibrium under optimal airport pricing (right) in terms of route structure. Lower constant cost transfer cost ($\mu = 30$).

route structure configurations is not very sensitive to the value of travel time, at least in the studied range. Finally, recall that a change in the airports' capacity would have the same effect as a change in the value of time because what matters is the ratio α/K (see Eq. (A.1)). Therefore, there is no need to also vary the value of K.

B.6. Sensitivity with respect to the constant transfer cost

Figs. B.18 and B.19 display the result that a higher value of the constant transfer cost favors the fully connected route structure. This is a straightforward result: when transferring is more costly for passengers, hub-and-spoke route structures become less attractive for the firm and for welfare maximization.

References

Alderighi, M., Cento, A., Nijkamp, P., Rietveld, P., 2005. Network competition—the coexistence of hub-and-spoke and point-to-point systems. Journal of Air Transport Management 11 (5), 328–334.
Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A.A., Zou, B., 2010. Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States, Technical report, National Center of Excellence for Aviation Operations Research, Final Report.
Barla, P., 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.
Bel, G., Fageda, X., 2010. Privatization, regulation and airport pricing: an empirical analysis for Europe. Journal of Regulatory Economics 37 (2), 142–161.
Brander, J.A., Zhang, A., 1990. Market conduct in the airline industry: an empirical investigation. The RAND Journal of Economics 21 (4), 567–583.
Brons, M., Pels, E., Nijkamp, P., Rietveld, P., 2002. Price elasticities of demand for passenger air travel: a meta-analysis. Journal of Air Transport Management 8 (3), 165–175.
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., 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., Verhoef, E.T., 2010. Manipulable congestion tolls. Journal of Urban Economics 67 (3), 315–321.

Burghouwt, G., 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.

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

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

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

InterVISTAS (2007), Estimating Air Travel Demand elasticities, Report for IATA, http://www.iata.org/whatwedo/Documents/economics/Intervistas_Elasticity_Study_2007.pdf (accessed September, 2013).

Jara-Díaz, S.R., Cortés, C.E., Morales, G.A., 2013. Explaining changes and trends in the airline industry: Economies of density, multiproduct scale, and spatial scope. Transportation Research Part E: Logistics and Transportation Review 60, 13–26.

Morrison, S.A., Winston, C., 1995. The Evolution of the Airline Industry. Brookings Institution Press, Washington, DC.

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.

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

Pels, E., Verhoef, E.T., 2004. The economics of airport congestion pricing. Journal of Urban Economics 55 (2), 257–277.

Perloff, J.M., Karp, L., Golan, A., 2007. Estimating Market Power and Strategies. Cambridge University Press.

Silva, H.E., 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.

Swan, W.M., Adler, N., 2006. Aircraft trip cost parameters: a function of stage length and seat capacity. Transportation Research Part E: Logistics and Transportation Review 42 (2), 105–115.

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

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