Airlines' Strategic Interactions and Airport Pricing in a Dynamic Bottleneck Model of Congestion

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

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion

Hugo E. Silva[*], Erik T. Verhoef[1], Vincent A.C. van den Berg

Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Version of 22 May, 2012

Abstract

This paper analyzes airlines' strategic interactions and airport efficient pricing, with a deterministic bottleneck model of congestion, in Cournot-Nash competition and in sequential competition where a Stackelberg leader interacts with perfectly competitive airlines. We show that the internalization of self-imposed congestion by non-atomistic carriers is consistent with earlier literature based on static models of congestion, but the congestion tolls are not. The tolls derived for fully atomistic airlines achieve the social optimum, when charged to all carriers, in the simultanous setting as well as in the sequential setting. We also find that alternative efficient pricing schemes exist for the sequential competition between a dominant airline and a competitive follower. The analysis suggests that airport congestion pricing has a more significant role than what previous studies have suggested. Moreover, the financial deficit under optimal pricing may be less severe than what earlier studies suggest, as congestion toll revenues may cover optimal capacity investments. Political feasibility would be enhanced as efficient congestion charges do not depend on market shares and therefore may not be perceived as inequitable.

Keywords: Airport pricing, Congestion, Bottleneck model
JEL codes: H23, L50, L93, R48

1. Introduction

As congestion at major airports worldwide continue to increase and traffic approaches existing capacity, the implementation of policies aimed at reducing delays effectively are becoming essential. For example, in the first half of 2007, 30 percent of commercial flights in U.S. arrived more than 15 minutes late, and similar figures hold for European airports (Rupp, 2009; Santos and Robin, 2010). Policies to solve the congestion problem have been extensively discussed during the last decades; one alternative is capacity enlargements, but they have the drawback of bringing benefits after a long period of time and at a relatively high monetary cost (see Jorge and de Rus (2004) for a cost-benefit analysis). Another option is congestion

---

[*]Corresponding author
Email addresses: `h.silvamontalva@vu.nl` (Hugo E. Silva), `e.t.verhoef@vu.nl` (Erik T. Verhoef), `vberg@feweb.vu.nl` (Vincent A.C. van den Berg)
[1]Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam, The Netherlands.

pricing, perhaps the most discussed policy in the academic literature, often following the road pricing literature.[2] However, governments, regulators and airports have not followed this path; the current practice at many airports is to levy weight-based landing fees, a rule that has been criticized since early contributions by Levine (1969) and Carlin and Park (1970), who were the first to argue that these charges provide wrong incentives and lead to inefficiencies. Despite of four decades of theoretical and empirical contributions calling for implementation of efficient landing and takeoff charges based on economic principles, airport pricing schemes have been kept remarkably unchanged. But, as delays are reaching critical values and other negative externalities, such as pollution and noise, are becoming more important, congestion pricing is likely to turn into a serious option for governments and regulators.[3] This policy may be specially appealing because landing fees are already in place, and only changes in the way that they are charged are needed. Moreover, in some countries, such as the U.S., landing fees are allowed to vary by time of the day, a fundamental feature of an efficient congestion pricing scheme.

It is now agreed that the vast literature on road congestion pricing may not be directly applicable to airports, because airlines are non-atomistic players, in contrast to road drivers. Carriers have market power and have non-negligible shares of the overall traffic and, as a consequence, they can be expected to internalize the congestion imposed on themselves. Daniel (1995) was the first to recognize this, and Brueckner (2002) and Pels and Verhoef (2004) analyzed the problem assessing the internalization of congestion with theoretical models. Subsequent works by Brueckner (2005), Zhang and Zhang (2006), and Basso and Zhang (2007) extend the analysis. The main conclusion, regarding congestion pricing, is that carriers competing in a Cournot-Nash fashion internalize self-imposed congestion and, therefore, should be charged for the fraction of congestion that they impose on others. This leads to a congestion charge that depends on the rivals' market share at the congested airport. This may be perceived as inequitable, as dominant airlines should face lower charges than small carriers.

These results have led to a debate on congestion internalization that has not yet fully reached consensus. On the one hand, Mayer and Sinai (2003) and Santos and Robin (2010) provide empirical evidence supporting the internalization hypothesis. On the other, Daniel (1995) and Daniel and Harback (2008) argue empirically that dominant airlines do not internalize self-imposed congestion because they act as Stackelberg leaders, facing a group of perfectly competitive carriers (hereafter, the fringe). They suggest that the optimal congestion charge should be the so-called atomistic toll that ignores any internalization. Subsequent theoretical works have aimed at bringing unity to the debate, using analytical models that follow previous literature. Brueckner and Van Dender (2008) show that, indeed, an interaction be-

---

[2]Quantity-based approaches to congestion management are also being discussed as an alternative in the literature. See Brueckner (2009), Basso and Zhang (2010) and Verhoef (2010) for analyses.

[3]Congestion pricing can be a second-best solution for environmental externalities. See, for example, Carlsson (2003) for an analysis of airport pricing with congestion and emissions, and Brueckner and Girvin (2008) for an investigation of airport noise regulation.

tween a Stackelberg leader with fringe carriers can induce airlines to ignore the self-imposed congestion, and that, in general, they only partially internalize own congestion. Silva and Verhoef (2011) consider a differentiated Bertrand duopoly, and show that airlines internalize only part of the self-imposed congestion, because they realize that frequency increases affect rivals' output through increased congestion. Under both arguments, the first-best charge is in between the congestion imposed on the rivals' and the total marginal congestion costs that results from completely ignoring internalization.

Theoretical analyses that have contributed to this debate have often been based on static models of congestion. This means, basically, that delays do not vary over time, and airlines are not concerned with the time of arrival or departure of aircrafts. However, a dynamic model of congestion is more appropriate when scheduling preferences underly observed peaks in travel, and when airports' operational conditions follow the first-come first-served discipline. In fact, this queuing discipline is the dominant mechanism in U.S. airports; for example, Daniel and Harback (2008) show that the observed traffic pattern in most of the major U.S. airports is consistent with the dynamic bottleneck model of congestion. Studies to date, have not provided transparent analytical insights, in a context of dynamic congestion, for the strategic interactions of airlines, and for the first-best charges that result from these interactions.

Previous exercises such as Daniel (1995, 2001) and Daniel and Harback (2008, 2009) consider dynamic congestion at airports, but focus on cost minimization of scheduling flights, hence ignoring the passengers' role in the problem, or at least treat that role only implicitly. Moreover, most of these papers aim at testing whether the observed patterns of arrivals and departures of flights support the internalization hypothesis. Furthermore, congestion tolls are derived for an atomistic equilibrium only, ignoring airlines' strategic interactions. Daniel (2009) analytically studies the conditions under which dominant airlines internalize self-imposed congestion with a deterministic bottleneck model, but focuses only on a Stackelberg-fringe competition and omits the passengers in the model, hence ignoring the fact that airlines use the airport as an input to sell an output in a downstream market. As a result, congestion charges are not directly applicable and comparable to previous findings.

The contribution of this paper is to provide clear-cut insights and understanding of airlines' strategic interactions and airport congestion pricing in a model of dynamic congestion. We recognize the vertical nature of aviation markets, thus explicitly including the role of airport's tolls on airlines' behavior, and incorporating that airlines compete taking these into account while facing the passengers' demand for trips. We use the dynamic model of congestion used in most economic analyses in the literature, namely the deterministic bottleneck model of congestion developed by Vickrey (1969) and Arnott et al. (1990, 1993). This allows for an analysis that balances analytical tractability and the inclusion of behavioral decisions: airlines endogenously adjust departure or arrival rates trading off queuing delays and schedule delays, and passengers dislike queuing and schedule delays in a different manner than airlines. Furthermore, we derive the equilibrium and first-best pricing for the case of a Cournot-oligopoly of airlines competing simultaneously, and the sequential competition between a Stackelberg leader and a competitive fringe.

Our main result is that, while the (untolled) equilibrium is fully consistent to what previous literature with static congestion suggests, first-best congestion pricing is not. In particular, non-atomistic carriers competing in a Cournot-Nash fashion do internalize the self-imposed congestion, but the first-best congestion charge is not altered by this internalization result and resembles the charges with fully atomistic carriers. Likewise, when a Stackelberg leader faces a fringe, the equilibrium is fully consistent with static models regarding internalization of congestion (e.g. Brueckner and Van Dender, 2008), but the first-best congestion toll is not. We find that one way of inducing the first-best outcome is by charging the atomistic toll, in the dynamic context, to the fringe carriers and a flat toll to the leader that accounts only for the congestion imposed on the fringe. The anticipation by the leader of the fringe carriers' response to the toll provides incentives to internalize the self-imposed congestion. We further show that charging the toll that is derived for fully atomistic carriers to both leader and fringe carriers also yields the first-best outcome. This is independent of the assumed pattern of substitution between airlines, and holds with imperfectly elastic demand if a market-power correcting subsidy is given. These results confirm Brueckner's (2002) claim that the allocation of number of flights in peak hours may be not as severe as atomistic models predicts and that delays are smaller at more concentrated airports (in absence of time-variant tolls), because airlines itnernalize congestion based on own market shares. But, at the same time, suggests that the first-best congestion charges have the same structure as in the case with atomistic airlines that do not internalize congestion.

The results of this paper suggest that optimal congestion pricing has a more significant role than what has been suggested in the literature before. The congestion pricing scheme that is obtained for fully atomistic carriers, that is independent of airlines' market shares, induces the first-best outcome and results in a revenue for the airport that restores the well known self-financing result for congested facilities. This also suggests that the political feasibility of congestion pricing would be enhanced as congestion charges do not vary across airlines and therefore may not be perceived as inequitable. Finally, we also show in the analyses how the market-power exertion has to be corrected, finding insights that are consistent with those in the previous literature, and that this overturns the self-financing result if market-specific subsidies are drawn from the airport budget.

Our results have to be qualified according to our assumptions. Naturally, the dynamic bottleneck model is not directly applicable when queuing is not necessary or helpful in order to obtain a certain arrival time, as in fully slot-constrained airports. For this case, more common in European airports, the quantity-based approach is more pertinent (see Brueckner 2009). We also assume that airlines and passengers share the desired time of arrival or departure and that airlines are homogeneous in values of time. The model can be straightforwardly extended following the road pricing literature.[4] Lastly, we use the deterministic version of the bottleneck model for analytical simplicity. A stochastic version that does not require operation

---

[4]The original model by Vickrey (1969) analyses heterogeneity in desired arrival time. For heterogeneity in values of time see e.g. Vickrey (1973), Arnott et al. (1994) and Van den Berg and Verhoef (2011).

at capacity to yield (non-linear) queues would be more realistic. However, as the trade off between expected queuing and expected schedule delays will be driving airlines' interactions, general results may not change significantly, while detailed results such as equilibrium delays, traffic rates and queue lenghts will change.

The paper is organized as follows. Section 2 introduces the model and the assumptions that are necessary for the analysis. We illustrate the main features of the model by characterizing the untolled equilibrium and deriving first-best and time-invariant second-best tolls for perfectly competitive airlines. We then study a monopoly carrier in the market and Cournot-Nash competition. Section 3 extends the analysis to competition where a Stackelberg leader faces a group of competitive carriers, focusing on the untolled equilibrium and on first-best tolling. We study the case of imperfectly elastic demand and imperfectly substitutable airlines, and also look at the special cases of perfect substitution, independent markets and perfectly elastic demand. Finally, Section 4 concludes.

2. The model and basic analysis

We use a vertical setting, recognizing that airports provide an input for the output sold by airlines in a downstream market, and explicitly incorporate the passengers' role. The demand for an airline i follows a linear inverse demand function:

$$D_i(q_i, q_j) = A_i - B_i \cdot q_i - E_i \cdot q_{-i} \tag{1}$$

which gives the marginal willingness to pay for travelling with airline i; q_i is the number of passengers travelling with airline i, and q_{-i} the aggregate of the rivals' passengers; A_i represents the maximum reservation price, B_i is the own-demand sensitivity parameter, and E_i is the cross-demand sensitivity parameter. We assume $B_i \geq E_i$ in general, and $B_i > E_i$ usually so that outputs are imperfect substitutes. Perfect substitutability is a special case of our specification ($E_i = B_i = B \; \forall \; i$), while the special case where airlines serve independent markets is represented by $E_i = 0$. We use the linear specification for analytical simplicity, but our results do not depend crucially on this particular functional form.

The full price p_i for a passenger travelling with airline i is the sum of the fare (ρ_i) and the generalized cost experienced by the passenger (C_i). As we consider dynamic congestion, the various components of the generalized cost are generally not constant over time. We use the deterministic bottleneck model of Vickrey (1969) and follow Small's (1982) model of scheduling behavior, so that the passengers generalized cost is the sum of travel delay cost and schedule delay cost. In a nutshell, passengers face travel delays that include queuing delays to land or depart, and have a preferred arrival or departure time from which any scheduled deviation (early or late) induces a schedule delay cost. For simplicity we consider one congested airport and focus on arrivals, so the bottleneck is the airport's runway. However, the analytical results would apply for departures as well, and can in principle be extended to a network setting with multiple airports and delays in both arrivals and departures. Our bottleneck model is relevant when the airport's operational conditions for arrivals (or departures) follow the first-in first-

out (FIFO) discipline, and is not directly applicable when the airport is managed with slots, so that queuing is not needed or helpful to obtain a certain arrival (or departure) moment.

We further assume that passengers have the same desired arrival time (t^*), that free-flow travel time is zero,[5] and that the generalized cost, for passengers in a flight arriving at the destination at t, is:

$$C(t) = \alpha \cdot T(t) + \begin{cases} \beta \cdot (t^* - t) & \text{if } t \leq t^* \\ \gamma \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \quad (2)$$

where $T(t)$ is the travel time for a flight arriving at t, in our case consisting only of queuing delay; α is the passengers' value of travel time, β is the value of early schedule delay and γ is the value of late schedule delay. We drop the subindex referring to the airline in (2), because the generalized cost experienced by the passenger does not depend on the identity of the airline; preferences over airlines are captured in the demand function (1).

As it is usual in the airport pricing literature, we assume that the product of the load factor and the seat capacity is constant, so that the number of passengers per flight is given, and that airlines compete in a Cournot fashion, so they choose the number of flights.[6] They also choose when to schedule each flight, which is represented by t. Consequently, the equilibrium condition for any airline i is that the full price of a trip arriving at t, equals the inverse demand $D_i(q_i, q_{-i})$:

$$\rho_i + C(t) = A_i - B_i \cdot q_i - E_i \cdot q_{-i} \quad (3)$$

This implies that an airline may have to charge passengers a time-dependent fare $\rho_i(t)$. This is a condition for an interior equilibrium, where all flights are used by passengers and where passengers are indifferent between all the flights of a single airline. Flights with the same generalized cost must have the same fare for user equilibrium to hold. Forbes (2008) provides empirical evidence that airlines indeed charge lower fare when they face higher delays. Note that this does not mean that different airlines charge the same fare, even for flights with the same passenger generalized costs; fares may differ between airlines because of the imperfect substitution assumption.

Airline's cost consists of a time-invariant operating cost per flight c_1, a time-invariant operating cost per passenger c_2, and a time-variant cost $C_a(t)$. Furthermore, we assume that the same scheduling model holds for airlines as for passengers; therefore, the airline's delay cost differs from (2) only in the values of time.[7] Let α_a, β_a and γ_a the values of travel time,

---

[5]In a single origin-destination pair, we can assume zero free-flow travel time without loss of generality, but this is generally different with multiple origin-destination pairs.

[6]This is equivalent to set the number of passengers (quantity) because the fixed-proportion assumption implies $q_i = f_i \cdot s$, where s is the product between seat capacity and load factor.

[7]This also means that airlines share the desired departure time t^* with the passengers. Although the preferred arrival time for airlines may be endogenous, following from desired arrival times for passengers, the analysis of this issue is beyond the scope of this paper. With endogenous t^* it can be expected that airlines' preference is significantly affected by passenger's preferred arrival time and will be close in practice. For example, in hub-and-spoke networks, airlines coordinate arrivals and departures to facilitate passenger

early schedule delay and late schedule delay for the airline, respectively. Denoting the constant product between seat capacity and load factor as s, and $c_p = c_1 + s \cdot c_2$ as the time-invariant cost per flight,[8] an airline's cost per flight can be written as:

$$c_1 + s \cdot c_2 + C_a(t) = c_p + \alpha_a \cdot T(t) + \begin{cases} \beta_a \cdot (t^* - t) & \text{if } t \leq t^* \\ \gamma_a \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \tag{4}$$

With the cost structure defined, we can analyze the equilibrium in the airline market and then study the regulator's problem. This section looks at three market structures to illustrate the main features of the model, namely the perfectly competitive case, the monopoly case, and an oligopoly interacting in Cournot-Nash fashion.

2.1. The competitive case

In the competitive case, airlines are atomistic and each carrier operates a single flight with s passengers. Because of free entry, fares just cover operating cots; therefore, airlines face a zero-profit condition. We assume that carriers are homogeneous, i.e. airlines are perfect substitutes. This implies that $E_i = B_i = B$ in the inverse demand function in (1). The profit of an airline, that schedules its only flight at t, is revenues minus costs:

$$\pi(t) = s \cdot \rho_i(t) - C_a(t) - c_p - \tau(t) \tag{5}$$

where $\tau(t)$ is the time-variant per-flight toll (in this case, landing fee) that the regulator might charge to airlines. Denoting f as the aggregate number of flights, the total number of passengers is sf, and using (3) we can write profit for a flight arriving at t as:

$$\pi(t) = s \cdot \left[A - B \cdot sf - \alpha \cdot T(t) - \begin{cases} \beta \cdot (t^* - t) & \text{if } t \leq t^* \\ \gamma \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \right] - C_a(t) - c_p - \tau(t) \tag{6}$$

where the term between square brackets is the fare. Using (4) and defining $\overline{\alpha} = s\alpha + \alpha_a$, $\overline{\beta} = s\beta + \beta_a$ and $\overline{\gamma} = s\gamma + \gamma_a$, the profit of an airline whose flight arrives at time t at the bottleneck can be simplified as:

$$\pi(t) = s \left[A - B \cdot sf \right] - c_p - \tau(t) - \overline{\alpha} \cdot T(t) - \begin{cases} \overline{\beta} \cdot (t^* - t) & \text{if } t \leq t^* \\ \overline{\gamma} \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \tag{7}$$

This reduced form shows that airlines take into account the generalized cost of its own passengers, because the lower the passengers' generalized cost is, the higher the fare can be (see (3)).

---

connections. Cost advantages because of high passenger density may also drive airlines to adopt the passengers' preferred arrival time.

[8]Because of the fixed-proportions assumption, constant costs per passenger and per flight have the same effect and can be aggregated. The same occurs with a landing fee; it does not matter if it is a per-passenger fee or a per-flight fee.

Therefore, we can interpret the airline's problem as if they face a *generalized cost per flight* that is the sum of its own delay costs, $C_a(t)$, and the generalized cost of all the passengers on its flight, $s \cdot C(t)$.

The dynamic equilibrium is such that an airline cannot improve its benefits by changing the schedule of its single flight. By looking at (7), this can only be achieved when every airline, i.e. every flight, faces the same generalized cost per flight (last two terms on the right-hand side of equation (7)), because all other terms are time-invariant. This generalized cost per flight from the airlines' perspective, is similar to the generalized costs typically found in bottleneck road pricing literature for individual drivers (e.g., Arnott et al., 1990, 1993). A difference is that the values of time considered by the airline, for a single flight, are its own values of time plus the summed passengers' values of time in that flight. But, through the use of the composite shadow prices $\overline{\alpha}$, $\overline{\beta}$, and $\overline{\gamma}$, this difference disappears from the formal model. This enables us to describe the equilibrium in schedules following the road pricing literature, and keep the discussion concise.

We first characterize the untolled equilibrium. In this case, the generalized cost per flight must be constant over time:

$$C_a(t) + s \cdot C(t) = \frac{\overline{\delta} \cdot f}{K} \quad \forall \quad t \in [t_s, t_e] \tag{8}$$

where K is the runway capacity in terms of flights per unit of time, t_s the start of the operation period, t_e the end of the operation period, and $\overline{\delta} = (\overline{\beta} \cdot \overline{\gamma})/(\overline{\beta} + \overline{\gamma})$.[9] This yields an equilibrium profit (superscript e) for any airline of:

$$\pi^e = s \left[A - B \cdot sf \right] - c_p - \frac{\overline{\delta} \cdot f}{K} \tag{9}$$

Recall that airlines are indifferent between any arrival time t and passengers are indifferent between any flight, because the full price of all flights is constant, equal to $A - B \cdot sf$, and given by:

$$p_i = \rho_i(t) + C(t) = A - B \cdot sf = \frac{1}{s} \cdot \left(c_p + \frac{\overline{\delta} \cdot f}{K} \right) \tag{10}$$

where the last equality comes from the zero-profit condition ($\pi^e = 0$). The passengers' full price in the no-toll equilibrium equals the airlines' constant operating cost per passenger (c_p/s) plus the generalized cost per flight divided by the number of passengers. The total generalized costs (or total delay costs) are the generalized costs per flight times the number of flights, $\overline{\delta} \cdot f^2/K$, as in the road case.

We can also study how fare (ρ) changes over time by using (2), (10), and taking the derivative with respect to t:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial C}{\partial t} = -\alpha \frac{\partial T}{\partial t} - \begin{cases} \beta \frac{\partial(t^* - t)}{\partial t} \\ \gamma \frac{\partial(t - t^*)}{\partial t} \end{cases} = \begin{cases} -\left[\alpha \cdot \frac{\overline{\beta}}{\overline{\alpha}} - \beta \right] = \beta \left[1 - \frac{\alpha/\beta}{\overline{\alpha}/\overline{\beta}} \right] & \text{if } t \leq t^* \\ \left[\alpha \cdot \frac{\overline{\gamma}}{\overline{\alpha}} - \gamma \right] = \gamma \left[\frac{\alpha/\gamma}{\overline{\alpha}/\overline{\gamma}} - 1 \right] & \text{if } t \geq t^* \end{cases} \tag{11}$$

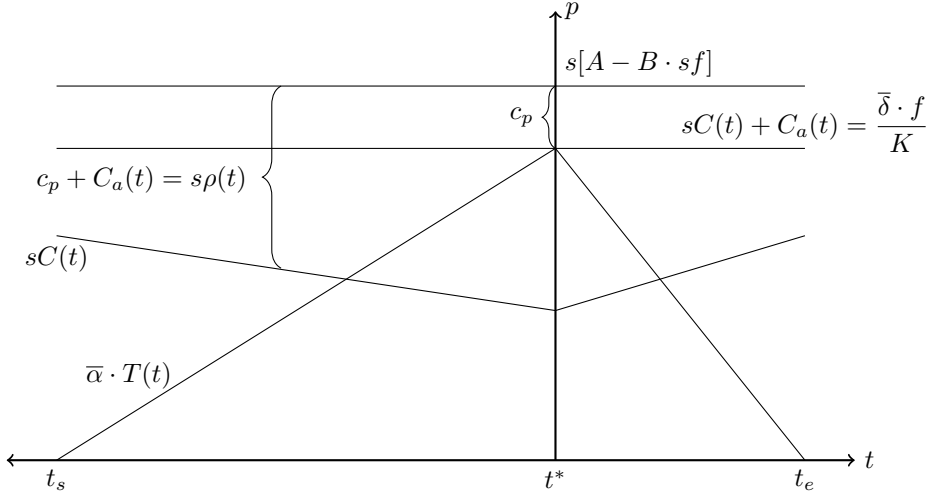[9]See Arnott et al. (1990, 1993) for a derivation of this result.

Figure 1: Competitive no-toll equilibrium ($\alpha/\beta < \overline{\alpha}/\overline{\beta} < \alpha_a/\beta_a$ and $\alpha/\gamma < \overline{\alpha}/\overline{\gamma} < \alpha_a/\gamma_a$).

where we use that, in equilibrium, queuing delays, $T(t)$, have a slope of $\overline{\beta}/\overline{\alpha}$ for early arrivals and $-\overline{\gamma}/\overline{\alpha}$ for late arrivals. This reveals that only when the ratios α/β and α/γ equals $\overline{\alpha}/\overline{\beta}$ and $\overline{\alpha}/\overline{\gamma}$ respectively, the fare (and generalized cost) is constant over time. This is when the passengers' willingness to accept schedule delays in order to reduce travel times, as represented by the ratios α/β and α/γ, equals the equilibrium rates of change between schedule delays and travel time. On the other hand, when the passengers' ratios α/β and α/γ are lower (higher) than the equilibrium rates of change, the fare will be higher (lower) for passengers travelling closer to t^*.

Figure 1 illustrates the no-toll equilibrium for the competitive case. The equilibrium is represented by the constant generalized costs per flight (see (8)), and the depiction of $s[A - B \cdot sf]$ satisfying (10). Figure 1 also shows the fare structure for the parametrization where values of time are such that $\alpha_a/\beta_a > \alpha/\beta$ and $\alpha_a/\gamma_a > \alpha/\gamma$.

With the untolled equilibrium characterized, we analyze the regulator's problem of maximizing social welfare through a per-flight toll. First, consider the case of a time-invariant toll. As it does not vary over time, the airlines treat it as a constant operating cost and, for a given number of flights, it does not alter the scheduling decisions. The regulator's optimization problem follows:

$$\max SW = \int_0^{sf} (A - Bx)dx - \int_{t_s}^{t_e} (K \cdot s \cdot C(t))dt - \int_{t_s}^{t_e} (K \cdot c_p + K \cdot C_a(t))dt \qquad (12)$$

where the first term is gross benefits for sf travelers, the second is total passengers' generalized costs (at t, a flow of K flights will serve s passengers each), and the third term is total airlines' costs that includes constant and generalized costs (fares and tolls cancel out). Rewriting,

$$SW = \int_0^{sf} (A - Bx)dx - K \cdot \int_{t_s}^{t_e} (s \cdot C(t) + C_a(t))dt - K \cdot c_p \int_{t_s}^{t_e} dt$$

$$= \int_0^{sf} (A - Bx)dx - \frac{\overline{\delta} \cdot f^2}{K} - f \cdot c_p \qquad (13)$$

9

where the second equality uses that the duration of the peak is f/K, and that, in equilibrium, $s \cdot C(t) + C_a(t)$ is constant (condition (8)).

Let $\widehat{\tau}$ be the time-invariant toll. Comparing the first-order conditions for welfare maximization and the airline zero-profit condition, we then obtain:

$$\frac{\partial SW}{\partial f} - \pi^e = s(A - B \cdot sf) - 2\frac{\overline{\delta} \cdot f}{K} - c_p - \left[s(A - B \cdot sf) - c_p - \frac{\overline{\delta} \cdot f}{K} - \widehat{\tau} \right] \tag{14}$$

As a consequence, the welfare maximizing time-invariant toll per flight is:

$$\widehat{\tau} = \frac{\overline{\delta} \cdot f}{K} \tag{15}$$

This toll matches the flat toll for the road bottleneck (Arnott et al., 1993), because without altering the flights' schedule, generalized costs per flight are $\overline{\delta} \cdot f/K$, and marginal social generalized costs are $2 \cdot \overline{\delta} \cdot f/K$, which is fully consistent with the road case. As a consequence, it is straightforward that the second-best flat toll is the difference between both.

Note that—given that airlines are queuing such that generalized costs are constant over time—the aggregate number of flights is not optimal, and the flat-toll (15) equal to the marginal delay cost that a flight imposes on all airlines' flights (including their passengers) is necessary. This is analogous to the atomistic toll derived in static congestion models (e.g., Brueckner, 2002), because—as showed by Arnott et al. (1993)—the untolled equilibrium in the bottleneck model is characterized by a supply function (8) that depends (linearly) on the aggregate number of flights. This time-invariant toll induces an aggregate number of flights f', which is second-best optimal, given that queuing is not eliminated. The fares will remain with the same dynamic structure that they also have in the no-toll equilibrium.

As queuing delay is a pure loss in this model, welfare can be improved further. The reason is that, any number of flights in an equilibrium with queues can be served in the same time interval, without queuing and without increasing schedule delay costs. This requires an arrival rate equal to the capacity throughout the peak, and cannot be achieved spontaneously in equilibrium, as the flights closer to t^* face a lower generalized cost. The first-best charge is the time-variant toll $\tau(t)$ that fully eliminates queuing. It is equal to the value of queuing delay per flight of the no-toll equilibrium, $\overline{\alpha} \cdot T(t)$ in Figure 1, and induces that the generalized cost of each flight equals marginal social generalized cost. Denoting f^* as the optimal aggregate number of flights, the optimal toll is,

$$\tau(t) = \frac{\overline{\delta} f^*}{K} - \begin{cases} \overline{\beta} \cdot (t^* - t) & \text{if } t \leq t^* \\ \overline{\gamma} \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \tag{16}$$

We call this toll structure the *dynamic atomistic toll*, in contrast to the atomistic toll derived in static models of congestion and the second-best flat toll of this problem (15). In the first-best equilibrium, as there is no queuing, total generalized costs are $\overline{\delta} \cdot f^2/2K$, hence marginal social generalized costs are $\overline{\delta} \cdot f/K$. With $\tau(t)$, the flight that arrives at t^* faces no generalized costs and pays a toll equal to the marginal social cost. Moreover, the first and last flight face

a schedule delay equal to the marginal social generalized cost and therefore do not pay any toll. The equilibrium full price does not differ from the no-toll equilibrium, but generalized cost does. The queuing costs completely disappear and are replaced by the toll, hence total revenue is equal to the total queuing costs of the untolled equilibrium. The part of the toll that reflects passenger valuation of delays is transferred to them through the fare to maintain passenger equilibrium. In other words, the fare will show a stronger time variation than in the no-toll equilibrium.

2.2. The monopoly case

Here, we consider a market with a single airline facing a linear inverse demand (as in (1) with $E = 0$). The monopoly carrier chooses the number of flights F and how to schedule them. Since the airline faces no competition,[10] the flights will be scheduled such that there is no queue at all. The airline realizes that by choosing a departure rate equal to the runway capacity, it will achieve the minimum possible delay, i.e. only schedule delay costs. Let t_s be the start of the period and t_e the end, then the airline's profit is:

$$\pi = \int_{t_s}^{t_e} K \cdot s \cdot \rho(t) - K \cdot c_p - K \cdot C_a(t)dt = K \int_{t_s}^{t_e} s[A - B \cdot sF] - sC(t) - c_p - C_a(t)dt$$

$$= s \cdot F \cdot [A - B \cdot sF] - F \cdot c_p - K \int_{t_s}^{t_e} sC(t) + C_a(t)dt \tag{17}$$

Where the second equality uses the appropriate version of (3), and the third equality that the peak lasts F/K. We have shown that the last term on the right hand side of (17) reflects the road case with $\overline{\alpha}$, $\overline{\beta}$, and $\overline{\gamma}$. This allows us to write the generalized costs per flight given that the monopoly does not queue, i.e. these are schedule delay costs that diminish linearly from $\overline{\delta}F/K$ at t_s to zero at t^* and then grow to $\overline{\delta}F/K$ at t_e. Taking this into account and considering a per-flight time-invariant toll $\widehat{\tau}$ (that is seen as parametric by the airline), the profit in (17) can be expressed as:

$$\pi = s \cdot F \cdot [A - B \cdot sF] - F \cdot c_p - \frac{\overline{\delta} \cdot F^2}{2K} - F \cdot \widehat{\tau} \tag{18}$$

The airline first-order condition for profit maximization is,

$$\frac{\partial \pi}{\partial F} = s[A - B \cdot sF] - B \cdot s^2F - c_p - \frac{\overline{\delta}F}{K} - \widehat{\tau} = 0 \tag{19}$$

which means that the (constant) full price paid by passengers is:

$$p = \rho(t) + C(t) = A - B \cdot sF = \frac{1}{s} \cdot \left(c_p + \frac{\overline{\delta} \cdot F}{K}\right) + B \cdot sF + \frac{\widehat{\tau}}{s} \tag{20}$$

implying that the fare, that maintains equilibrium, is:

$$\rho(t) = \frac{1}{s} \cdot \left(c_p + \frac{\overline{\delta} \cdot F}{K}\right) + B \cdot sF + \frac{\widehat{\tau}}{s} - C(t) \tag{21}$$

---

[10]We are abstracting from potential entry in this setting, but we address this question in Section 3.
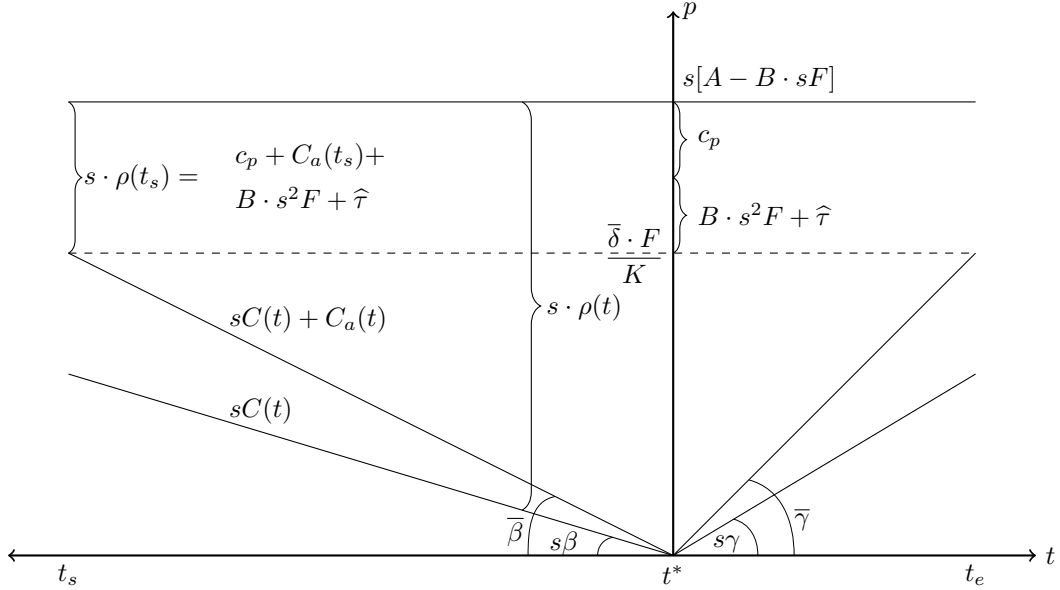
Figure 2: Monopoly time-invariant-toll equilibrium.

In contrast to the competitive case, this condition shows that the monopoly carrier charges to the passengers a markup of $B \cdot sf$. This is simply the number of passengers times the own-demand price sensitivity, the market power effect first described by Pels and Verhoef (2004). Figure 2 depicts the time-invariant-toll equilibrium for a monopoly. There is no queue, and the first and last flight (at t_s and t_e, respectively) experience a generalized cost of $\bar{\delta}F/K$. The fulfillment of the first-order condition for profit maximization is represented in the vertical axis, where $s[A - B \cdot sF] = c_p + \bar{\delta}F/K + B \cdot sF + \hat{\tau}$ by construction. The time-variant per-flight fare, $s \cdot \rho(t)$ in (21), is also depicted in Figure 2. The slopes of passengers' generalized cost ($C(t)$) and airline's delay costs ($C_a(t)$) are the same as in the optimum of the competitive case, and therefore the slope of the per-flight fare is also the same. Finally, note that the fare at the first and last flight is just the sum of the constant operating cost, the airline's delay cost, the toll and the market power markup. Now, the regulator's maximization problem is:

$$SW = \int_0^{sF} (A - Bx)dx - K \cdot \int_{t_s}^{t_e} (s \cdot C(t) + C_a(t))dt - f \cdot c_p \tag{22}$$

but, in contrast with the competitive case, the airline is scheduling the flights in such a way that there is no queue. Hence, the second term on the right-hand side of (22) is the same as derived in (18), shaping social welfare in the following way:

$$SW = \int_0^{sF} (A - Bx)dx - \frac{\bar{\delta} \cdot F^2}{2K} - F \cdot c_p \tag{23}$$

This is gross benefits of sF passengers minus total social costs; when there are no queuing delays, total generalized costs equals $\bar{\delta} \cdot F^2/2K$. Taking the derivative with respect to F, we get the first-best condition:

$$s[A - B \cdot sF] = c_p + \frac{\bar{\delta} \cdot F}{K} \tag{24}$$

12

This is simply that, at the optimum, full price equals marginal social cost, which is the sum of the marginal operating cost plus marginal total generalized costs (including airlines and passengers through $\overline{\delta}$).[11] Comparing the monopolist's first-order condition (19) and the first-order condition for welfare maximization (24), it is straightforward that the first-best toll is:

$$\widehat{\tau} = -B \cdot s^2 F \tag{25}$$

The regulator corrects the market power exertion by subsidizing the airlines, and does not have to give an incentive to the monopolist to internalize congestion. This subsidy ($-B \cdot sF$ per passenger) induces the optimal number of passengers, and is analogous to the one obtained in the static model (Pels and Verhoef, 2004).

A monopoly airline internalizes all the congestion costs by scheduling the flights efficiently: there is no queuing and therefore there is no need for congestion pricing. In Figure 2, when the optimal subsidy is applied, the term $B \cdot s^2 F + \widehat{\tau}$ disappears and the first-best condition (24) is satisfied. Moreover, the per-flight fare (s multiplied by the per-passenger fare) at the first and last flight is simply the airline's costs per flight, as Figure 2 shows.

2.3. Cournot oligopoly

In this section we analyze the case where a small number of airlines compete for passengers and face horizontally differentiated demands as in (1). We look at the case where they make their decisions simultaneously, and believe that they are not able to influence the competitors' traffic volume. This problem has been studied in various settings in the literature, but mainly in the context of static congestion models. In this section we derive the equilibrium and first- and second-best tolls with the dynamic deterministic bottleneck model of congestion.

The equilibrium that mirrors the road case with the adjusted values of time ($\overline{\alpha}, \overline{\beta}$ and $\overline{\gamma}$), where every airline faces the same generalized cost per flight, is the unique Cournot-Nash equilibrium. Airlines cannot increase profit by changing the schedule of flights, and, even though they have market power and the potential to internalize congestion imposed on their own flights, they cannot reach the equilibrium without queuing delays. An equilibrium without queues cannot be supported because one flight will always face higher schedule delays than other; therefore, one airline will have an incentive to reschedule flights. This is obviously not the case when all the flights face the same generalized cost. It is worth stressing that the equilibrium pattern of queuing implies that there is no particular unique "order" of flights over operators.

Despite the fact that in the untolled equilibrium there is queuing, airlines do internalize the generalized costs imposed on themselves—given that they are queuing. To see this, let f_i be the flights of the airline i and f_{-i} the rivals' aggregate number of flights. Then, the constant generalized cost per flight is $\overline{\delta} \cdot (f_i + f_{-i})/K$ and profit and first-order conditions for

---

[11]We look at the full price of a flight (the full price of a trip $A - B \cdot sF$ times the number of passengers in a flight s) and the marginal social cost of a flight, but there is no loss of generality. The condition also implies that the full price of a trip equals the marginal social cost of a seat.

the airline i are:

$$\pi_i = s \cdot f_i[A_i - B_i \cdot sf_i - E_i \cdot sf_{-i}] - f_i \cdot c_p - f_i \cdot \frac{\overline{\delta} \cdot (f_i + f_{-i})}{K}$$

$$\frac{\partial \pi_i}{\partial f_i} = 0 \Rightarrow s[A_i - B_i \cdot sf_i - E_i \cdot sf_{-i}] = c_p + \frac{\overline{\delta}(f_i + f_{-i})}{K} + B_i \cdot s^2 f_i + \frac{\overline{\delta} f_i}{K} \qquad (26)$$

Recall that in this equilibrium marginal social costs are $c_p + 2 \cdot \overline{\delta} \cdot (f_i + f_{-i})/K$. Therefore, the first-order condition (26) shows that airlines apply the conventional monopolistic markup ($B \cdot sf_i$ per passenger) and fail to internalize the congestion imposed on the rivals ($\overline{\delta} \cdot f_{-i}/K$). As a result, the second-best optimal flat toll is fully consistent with the earlier literature with static congestion in Cournot competition, and is given by:

$$\widehat{\tau} = \overline{\delta} \cdot f_{-i}/K - B \cdot s^2 f_i \qquad (27)$$

Airlines are still not capable of reaching the equilibrium without queues, but they internalize self-imposed congestion. This toll in (27) is not the first-best toll. As queuing costs remain a pure loss in this model, welfare can be improved further with a time-variant toll that fully eliminates queuing. This is the dynamic atomistic toll that reflects the road case, and is given by:

$$\tau(t) = \frac{\overline{\delta}(f_i^* + f_{-i}^*)}{K} - \begin{cases} \overline{\beta} \cdot (t^* - t) & \text{if } t \le t^* \\ \overline{\gamma} \cdot (t - t^*) & \text{if } t \ge t^* \end{cases} \qquad (28)$$

where $f_i^* + f_{-i}^*$ is the first-best number of flights. This toll is completely analogous to the toll derived in the perfect competitive case, and transfers queuing costs into toll revenues. With (28), there is no need for flat congestion tolls because there are no queues and marginal social cost is reduced to $\overline{\delta} \cdot (f_i + f_{-i})/K$; hence, congestion costs are at the efficient level when airlines pay (28). In order to achieve the first-best equilibrium, the market power exertion has to be corrected through the optimal per-passenger subsidy of $B_i \cdot sf_i^*$. The first-best congestion toll, in contrast to previous findings and to the second-best flat toll, does not have to be different for airlines with different market shares of flights; however, the market-power correcting subsidy does. Moreover, the first-best congestion toll increases consumer surplus, as the the full price of a trip decreases with respect to the untolled equilibrium (the last term of (26) is eliminated). However, the effect on profits is ambiguous; it can increase or decrease depending on the relative magnitude of the market power effect.

This result for a Cournot-oligopoly of airlines has an important implication for the financial situation of the airport. As Arnott et al. (1993) demonstrate, the self-financing results of Mohring and Harwitz (1962) for capacity investments hold for the bottleneck model with elastic demand. As we have disclosed, the results of our perfect competitive case parallel results for the road case; therefore, the self-financing result also holds when (28) is charged to airlines. If there are constant returns to scale in capacity provision, the revenues from the first-best toll exactly cover the cost of providing the optimal capacity.[12] This differs

---

[12]In general, "the ratio of the revenue collected from the optimal toll to the costs of constructing optimal capacity equals the elasticity of construction cost with respect to capacity" (Arnott et al., 1993).

from earlier results because first-best tolls are not discounted by the fraction of congestion that is internalized by carriers; therefore, the self-financing result is not overturned by the internalization of congestion. However, the market-power subsidy that corrects dead-weight losses can overturn this self-financing. As suggested by Brueckner (2005), this should be corrected at a city-pair market level instead of at an airport level, but even then the breakdown of self-financing, of course, remains.

The results of this section show that the dynamic atomistic toll is the only toll that leads to the first-best when carriers are atomistic, but also when competing airlines have market power and internalize self-imposed congestion. In fact, when non-atomistic airlines offer perfect substitutes, the first-best toll is exactly the same as in the perfect competitive case (see (16) and (28)). Conversely, for a monopoly, only a subsidy that decreases price is needed, however, it is important to emphasize that the dynamic atomistic toll *could* also be charged to the monopoly airline without altering social welfare, but transferring part of the monopoly carrier profits to the regulator. This is the result of the congestion technology: the monopoly airline cannot do better than setting the arrival rate equal to capacity in time windows where it has arrivals, so as to face only schedule delay costs, regardless of the dynamic toll schedule it faces. To see this, it is enough to add the time-variant toll to the monopoly profit in (17):

$$\pi = s \cdot F[A - B \cdot sF] - F \cdot c_p - K \int_{t_s}^{t_e} sC(t) + C_a(t) + \tau(t) \, dt \qquad (29)$$

By charging the dynamic atomistic toll (28) to the monopoly, it is straightforward that $sC(t) + C_a(t)$ in (29) cancels out with the time-variant part of the toll, and the airline will set full price equal to c_p plus the constant part of the toll per flight ($\bar{\delta}F^*/K$). By including the subsidy, it is clear that the outcome will be the first-best.

These results suggest that a simple congestion pricing scheme, the dynamic atomistic toll at an airport level, that is independent of airlines' market shares and degree of competition (it only depends on the aggregate optimal traffic) will induce the first-best outcome, if the market power exertion is corrected with a subsidy at a market level. Moreover, it is the unique first-best congestion toll when there is, at least, some degree of competition.

3. A Stackelberg leader with a competitive fringe

We now turn to the case of a Stackelberg leader with a follower that behaves competitively to assess the internalization of congestion by the leader, and derive the optimal congestion tolls for this situation. This issue was raised by Daniel (1995) and studied further by Brueckner and Van Dender (2008), Daniel and Harback (2008) and Daniel (2009). On the one hand, Brueckner and Van Dender (2008)—with a static congestion model—show that the internalization of self-imposed congestion by a Stackelberg leader facing a competitive follower depends on the assumed substitution pattern. When all carriers are perceived as perfect substitutes, the leader does not internalize congestion and, as a result, the optimal congestion charge for all carriers would be the so-called atomistic toll that accounts for the marginal congestion

damage from an extra flight imposed on all flights. When leader and follower offer imperfect substitutes, and when they offer independent goods, the leader partially internalizes the self-imposed congestion. Consequently, the optimal toll lies in between the congestion imposed on the fringe and the atomistic toll. On the other hand, Daniel (1995) and Daniel and Harback (2008) find empirically that traffic patterns at most U.S. airports are not consistent with the internalization hypothesis and argue that atomistic tolls are needed for all carriers. Daniel (2009) also argues the need for atomistic tolls with an analytical model.

The competitive follower can be interpreted as one airline behaving competitively, or a group of competitive airlines, as in Section 2.1 with a free-entry condition. Following the aviation literature, we use the term "fringe" for the group of airlines that behaves competitively, regardless of the temporal location of its flights; note, in particular, that the fringe may operate in the temporal center of the peak. To avoid confusion, we will therefore not use the term fringe for indicating the shoulder periods of the peak. We assume that the leader and the fringe treat the toll that the regulator sets as parametric, and that when the Stackelberg leader makes the decisions, it is aware of the toll that the regulator applies to the fringe.[13]

3.1. Untolled equilibrium

To study the airlines' interactions, we first look at the no-toll equilibrium. In this case, the fringe carriers are characterized by a zero-profit condition and satisfy the dynamic equilibrium of constant generalized cost per flight, as in equation (8). This can only be possible by queuing in the center of the peak. On the other hand, the Stackelberg leader has the choice of scheduling flights joining the queue of the fringe operators, and to schedule flights outside this congested period, in the peak shoulders, bearing only schedule delay costs.[14] Let f be the number of flights that the fringe operates, l_c the number of flights that the leader schedules in the peak center along with the fringe, and l_s the leader's number of flights in the peak shoulders. We follow the framework proposed in Section 2, with fringe airlines being pure substitutes and inverse demands given by (1), but with symmetric parameters ($B_i = B \wedge E_i = E \, \forall \, i$); both carriers and passengers are homogeneous in values of time and in preferred arrival time, and the airlines have the same constant operating cost.

The fringe equilibrium condition is the same as in the competitive case (see (9)), but including the fact that the leader has $s(l_c + l_s)$ passengers, affecting its inverse demand. Moreover, the generalized cost per flight has to be constant in the peak center, and will equal $\overline{\delta} \cdot (f + l_c)/K$. This is because the leader can only operate jointly with the fringe in the same period if the aggregate queue pattern satisfies the competitive conditions. If the leader causes

---

[13]Brueckner and Verhoef (2010) point out that assuming that agents are large enough to exert market power and to recognize the impact of their decisions on overall congestion, but that they do not take into account the impact of their actions on the tolls, is a strong assumption. We maintain this assumption to focus on the first-order effects and comparison with earlier literature, but discuss how the solution proposed by Brueckner and Verhoef (2010) applies to our case in Section 3.3.

[14]The leader can set the departure rate equal to the capacity of the bottleneck and achieve the minimum time costs, because it does not face competition or potential entry in the peak shoulders.

steeper queue rates of change, the fringe will be pushed to the outsides, and the leader will harm itself unnecessarily at central times. Conversely, if the leader causes flatter queue rates of change, the fringe will be pushed to the inside and the leader will harm itself unnecessarily at shoulder times. The fringe zero-profit equilibrium condition is given by:

$$s\left[A - B \cdot sf - E \cdot s(l_c + l_s)\right] - c_p - \frac{\overline{\delta} \cdot (f + l_c)}{K} = 0 \tag{30}$$

Note that this condition defines f as a function of l_c and l_s. Solving for f, we obtain:

$$f = \frac{s\left[A - E \cdot s(l_c + l_s)\right] - c_p - \overline{\delta} \cdot l_c/K}{Bs^2 + \overline{\delta}/K} \tag{31}$$

First, note that the fringe's number of flights depends only on the number of flights set by the leader in the peak center (l_c) and not those in the shoulder (l_s) if, and only if, $E = 0$. This is an important point to stress, because allows us to identify the condition that makes the fringe care only about what happens in the center. The latter is the assumption made by Daniel (2009), thus the full independence case of our model is the case where results may be comparable with Daniel's (2009) findings.

By taking derivatives of (31) we can obtain the response of the fringe to a change in l_c, the number of flights that the leader schedules in the peak center, yielding the following:

$$\frac{\partial f}{\partial l_c} = -\frac{Es^2}{Bs^2 + \overline{\delta}/K} - \frac{\overline{\delta}/K}{Bs^2 + \overline{\delta}/K} = -\frac{Es^2 + \overline{\delta}/K}{Bs^2 + \overline{\delta}/K} \equiv \phi \tag{32}$$

Since $E < B$, it follows that $-1 < \phi < 0$. That is, a frequency change by the leader in the peak center, yields an opposite change in number of flights by the fringe, but that is not equal in magnitude because of the assumed substitution pattern in demand. As (32) shows, two effects are present in the fringe response. The first one is because of the substitutability; the leader will anticipate the fact that any reduction in the number of passengers (through a reduction in frequency, because of the fixed proportions assumption) will induce a shift in the inverse demand of the fringe, that will be met by an increase in the fringe's number of passengers, or, equivalently, new entry until the fringe profit is again zero. Any reduction in the number of passengers by the leader will thus be partially offset by an increase from the fringe. The second effect is because both carriers are imposing congestion on each other; the leader predicts that any frequency reduction is partially offset by an increase of the number of flights set by the follower in response to reduced queuing.

Note that in the case where outputs are perfect substitutes ($E = B$), $\phi = -1$, which means that any frequency reduction by the leader in the congested period is fully offset by an increase in number of flights by the competitive fringe. Note also that when outputs are independent ($E = 0$), the response of the follower still partially offsets a leader's frequency change, because of the congestion effect.

We can also derive the response of the fringe to a change in the number of flights scheduled in the peak shoulders, by differentiating (31) with respect to l_s, yielding:

$$\frac{\partial f}{\partial l_s} = -\frac{Es^2}{Bs^2 + \overline{\delta}/K} \equiv \lambda > \phi \tag{33}$$

17

This is the substitutability effect discussed above. Since the flights scheduled in the peak shoulders do not impose congestion on the fringe, there is no congestion effect. When $0 \leq E < B$, the response of the fringe, to an increase of the leader number of flights scheduled in the peak shoulders, satisfies $-1 < \lambda \leq 0$. Note that $\lambda > \phi$ means that the response ϕ is stronger than the response λ, because both are negative.

With the response of the fringe defined, we can look at the first-order conditions for the Stackelberg leader and derive the equilibrium. In this untolled equilibrium, the leader's profit can be separated into two terms, the profit from the operations in the peak center and the profit from the peak shoulders. In the center, because of the fringe's presence, the generalized cost per flight must be constant and equal to $\bar{\delta} \cdot (f + l_c)/K$. In the shoulders, the leader sets the arrival rate equal to the bottleneck capacity and experiences only schedule delay costs. Since the duration of the entire peak has to be $(f + l_c + l_s)/K$, the schedule delay of the first and last flight is $\bar{\delta} \cdot (f + l_c + l_s)/K$. The flights at the interior borders of the shoulders still experience only schedule delay costs and, as the duration of the peak center is $(f + l_c)/K$, they experience a schedule delay cost of $\bar{\delta}(f + l_c)/K$. Therefore, with linear delays, the average generalized cost per flight in the peak shoulders will be $\bar{\delta} \cdot (f + l_c)/K + \bar{\delta} \cdot l_s/(2K)$. This shapes the profit in the following way:

$$\Pi = \quad l_c \left(s[A - B \cdot s(l_c + l_s) - E \cdot sf] - c_p - \frac{\bar{\delta}(f+l_c)}{K} \right) +$$
$$l_s \left(s[A - B \cdot s(l_c + l_s) - E \cdot sf] - c_p - \frac{\bar{\delta}(f+l_c)}{K} - \frac{\bar{\delta}l_s}{2K} \right) \tag{34}$$

The first-order conditions are:

$$\frac{\partial \Pi}{\partial l_c} = 0 = \quad s[A - B \cdot s(l_c + l_s) - E \cdot sf] - c_p - \frac{\bar{\delta}(f+l_c)}{K}$$
$$- [(B + \phi E) \cdot s^2(l_c + l_s)] - \left[\frac{\bar{\delta}(l_c+l_s)}{K}(1 + \phi) \right] \tag{35}$$

$$\frac{\partial \Pi}{\partial l_s} = 0 = \quad s[A - B \cdot s(l_c + l_s) - E \cdot sf] - c_p - \frac{\bar{\delta}(f+l_c+l_s)}{K}$$
$$- [(B + \lambda E) \cdot s^2(l_c + l_s)] - \left[\frac{\bar{\delta}(l_c+l_s)}{K} \cdot \lambda \right] \tag{36}$$

In both first-order conditions, the last two terms in square brackets on the right-hand side are the market power effect and the congestion internalization, respectively. By subtracting (35) and (36), we can explicitly write the fraction of flights that the leader schedules in the shoulders as:

$$\frac{\partial \Pi}{\partial l_c} - \frac{\partial \Pi}{\partial l_s} = \frac{\bar{\delta}l_s}{K} - \left[\phi E \cdot s^2(l_c + l_s) \right] + \left[\lambda E \cdot s^2(l_c + l_s) \right] - \frac{\bar{\delta}(l_c + l_s)}{K}(1 + \phi - \lambda) = 0$$

$$\Rightarrow \frac{l_s}{l_s + l_c} = \frac{E \cdot s^2(\phi - \lambda) + \bar{\delta}/K(1 + (\phi - \lambda))}{\bar{\delta}/K} = 1 - \frac{Es^2 + \bar{\delta}/K}{Bs^2 + \bar{\delta}/K} \tag{37}$$

This shows that in the untolled equilibrium, when demand is imperfectly elastic and airlines are imperfect substitutes (i.e. $0 < E < B$), the leader schedules flights in both the peak center as well as in the peak shoulders ($l_s/(l_s + l_c) > 0$). This is also the case when the outputs

are independent ($E = 0$). In the case of perfect substitution ($E = B$) and when demand is perfectly elastic ($B = E = 0$), the leader sets all of its flights in the peak center, queuing along with the fringe. The reason is that the leader knows that the fringe reacts to increases in l_c by offsetting them, so that the fringe will make room for the leader's flights; conversely, if the leader decreases the number of flights in the peak center by shifting to the shoulders, the fringe will increase the number of flights raising the generalized cost. When the fringe fully offsets the changes in the leader's number of flights, the leader is better off setting all the flights in the peak center along with the fringe. When this effect is partial, the leader is better off setting part of the flights in the center.

Recall that the leader schedules flights in the peak center, queuing in the same way as the fringe and partially internalizes self-imposed congestion (as shown by the last term on the right-hand side of (35)). Conversely, in the peak shoulders it is scheduling the flights such that there is no queue, therefore being efficient in this sense, but still failing to fully internalize self-imposed congestion because of anticipation of the fringe response (captured by the last term on the right-hand side of (36)). This means that the leader does not internalize any congestion when demands are perfectly elastic or when demand is imperfectly elastic and products are perfect substitutes. In the case of full independence and imperfect substitutability, the leader internalizes a fraction of the self-imposed congestion. These results reproduce previous findings, regarding internalization of self-imposed congestion, by Brueckner and Van Dender (2008), but now in a dynamic congestion model.

Our result for full independence is also similar to the result by Daniel (2009) who finds that the leader sets a fraction of the flights in the peak center that ranges from 0 to 1 in the untolled equilibrium (Daniel's proposition 1). He argues that the leader sets all of the flights in the peak shoulders when the number of flights by the fringe is fixed. This is also true in our model, and is obtained when market are independents and the fringe faces a perfectly inelastic demand.

In the next section, we analyze the regulator's problem, deriving the first-best tolls for the three cases separately.

3.2. First-best tolls

- Perfect substitution

When outputs are perfect substitutes ($E = B$), it is straightforward from (37) that the leader sets all the flights in the peak center, queuing along with the fringe. In this case, (35) is the only relevant first-order condition, and using $\phi = -1$, the last two terms in square brackets become zero. Hence, the leader behaves atomistically in the sense that it does not exert market power (consistent with its own demand becoming, in practice, perfectly elastic), and does not take into account the fact that scheduling flights imposes congestion on its own flights (because elastic supply from the fringe frustrates any attempt to reduce travel delays). Let l be the number of flights scheduled by the leader;[15] in this case, both the zero-profit condition

---

[15]There is no need, in this case, to differentiate the leader's flights as it schedules all of them in the congested

of the fringe (30) as well as the first-order condition of the leader (35) are:

$$s\left[A - B \cdot s(f + l)\right] - c_p - \frac{\overline{\delta} \cdot (f + l)}{K} = 0 \tag{38}$$

As a consequence, the leader's profit will also be zero and only the aggregate number of flights can be determined. The fact that airlines are perfect substitutes makes the leader behave atomistically, because the fringe zero-profit condition requires a constant value for the overall number of flights, as (38) shows. This implies that the fringe fully offsets any change of leader's number of flights, and that the equilibrium is the same as Figure 1 depicts.

Since queuing delay is a pure loss in the model, the first-best toll scheme has to fully eliminate queuing. The fringe, by definition, operates atomistically; therefore, the only way to induce the first-best is by charging the dynamic atomistic toll described in Section 2.1 to the fringe: a per-flight toll equal to the queuing costs (per flight) of the untolled equilibrium:

$$\tau(t) = \frac{\overline{\delta} F^*}{K} - \begin{cases} \overline{\beta} \cdot (t^* - t) & \text{if } t \leq t^* \\ \overline{\gamma} \cdot (t - t^*) & \text{if } t \geq t^* \end{cases} \tag{39}$$

where F^* denotes the aggregate number of flights of the no-toll equilibrium, that results from (38); and $\overline{\delta} F^*/K$ is the marginal social cost when there is no queue.

Conversely, the Stackelberg leader has the potential to schedule its flights without incurring queuing delays, as we discussed in Section 2.2 for a monopoly, but it has no incentive to do so in the no-toll equilibrium because of the fringe's presence. However, when the regulator imposes the dynamic atomistic toll (39) only to the fringe, the leader realizes that can schedule F^* flights efficiently (without queuing and operating at capacity), earning profit (because it is not paying any toll), and keeping the fringe completely out. This is because whenever the fringe schedules a flight, it unavoidably experiences at least the constant generalized cost per flight of the untolled equilibrium $(\overline{\delta} F^*/K)$, as it incurs schedule delay cost and pays the toll that mirrors the queuing costs. Anticipating this, the leader—when not facing any toll—schedules F^* flights from t_s to t_e, leaving no room for the fringe. Figure 3 shows the first-best equilibrium.

As there are no queuing delays, there is no congestion inefficiency. Moreover, the leader does not exert market power, as can be seen by noting that the first-best condition is being satisfied: the full price of a flight $(s[A - B \cdot sF^*])$ equals marginal social costs $(c_p + \overline{\delta} F^*/K)$. Finally, the leader's profit is positive and equal to the queuing costs savings. To see this, note that at any time t, the revenue per-flight equals $s \cdot \rho(t)$ and the costs are $c_p + C_a(t)$. The subtraction of both equals $\overline{\delta} F^*/K - (sC(t) + C_a(t))$, exactly the queuing costs savings with respect to the no-toll equilibrium. The first and last flights have a marginal profit of zero, ensuring that the fringe does not have incentive to enter.

Even though the fringe is not operating along with the leader, the latter has no incentive to exert market power, because any reduction of number of flights from F^* results in a profit

---

period along with the fringe.

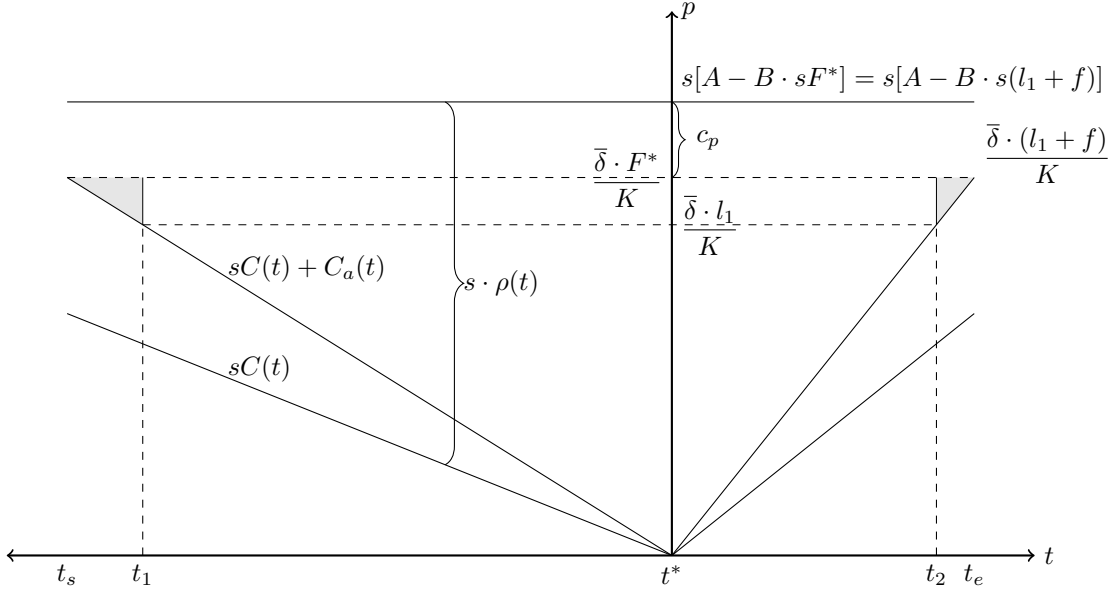Figure 3: First-best equilibrium with perfect substitution.

loss. To see this, graphically, suppose the leader schedules l_1 flights from t_1 to t_2, where schedule delay costs per flight equal $\bar{\delta}l_1/K$ at the borders. As a response to this, the fringe will schedule flights from t_s to t_s and from t_2 to t_e, because of the perfect substitution, keeping constant the total number of flights $(l_1 + f = F^*)$. The leader's outcome from reducing the number of flights is a profit loss equal to the shaded area.

This results in a first-best equilibrium where the dynamic atomistic toll is being only requested to the fringe. This toll works by inhibiting its operation and, at the same time, maintaining the leader's lack of capability to exert market power, but providing incentives to schedule the flights efficiently. However, this toll regime that exempts the leader from paying, and does not bring any revenue to the regulator, is not the only one that induces the first-best equilibrium. For example, the regulator can charge the dynamic atomistic toll (39) to both the fringe and to the leader, and the equilibrium will have the same scheduling pattern (only schedule delay costs). The difference is that all the leader's profit will be transferred to the regulator, making all carriers' profit zero. This is because of the congestion technology and the perfect substitution; in the deterministic bottleneck model, the leader is always better off setting the arrival rate equal to capacity so as to face only schedule delay, minimizing costs, but the fringe response prevents the market power exertion.

Brueckner and Van Dender (2008) find that the atomistic toll must be charged to both the leader and the fringe to obtain the first-best outcome. In our case, the dynamic atomistic toll—analogous to the first-best of the road case—for both agents leads to the first-best outcome, but it is not the only way, because the airline's interactions are different with dynamic congestion. The regulator can charge to the leader the dynamic atomistic toll (39), with the only effect of transferring leader's profit to toll revenues, but it may also charge a zero toll to the leader. The former case makes the outcome consistent with Brueckner and Van Dender (2008); the latter corresponds to the elastic case where the congestion toll approaches zero as the firm's

21

market share approaches unity.

This has an important implication for the financial situation of the regulator, since it can finance airport's capacity investment. As we discuss in Section 2.3, the self-financing results of capacity investments of Mohring and Harwitz (1962) hold in our problem when the dynamic atomistic toll is charged. Therefore, if the costs of providing capacity exhibit constant returns to scale, the regulator can finance the optimal capacity investments by charging the dynamic atomistic tolls to both the leader and the fringe.

- Full independence

In this setting, the fringe and the leader serve two different markets, they face independent demands ($E = 0$), and we let parameters A and B vary among airlines (subscripts f and l). As a result, the fringe is not affected by the operations in the peak shoulders ($\lambda = 0$), but still partially offsets peak center frequency reductions ($-1 < \phi < 0$), implying that the leader schedules flights in both periods (from (37), $l_s/(l_s + l_c) > 0$). The full independence also implies that, in both the peak center and the peak shoulders, the leader is exerting market-power with a markup equal to $B_l \cdot s^2(l_c + l_s)$, as a result of the full price being independent of the fringe's number of flights. This is exactly the markup that is present in the monopoly case, related to the own-price sensitivity and the total number of passengers (see the second term in square brackets (35) and (36) using $E = \lambda = 0$).

Consistent with Brueckner and Van Dender (2008), the internalization in the no-toll equilibrium is in between the Cournot case and the atomistic behavior, and the leader exerts market power. This result is comparable to the results obtained by Daniel (2009), as he finds that the dominant airline schedules a fraction between 0 and 1 with the fringe in the center. However, we find that the leader exerts market power raising the full price of a trip, that its number of flights cannot be represented as a function of average generalized cost per flight, and that it partially internalizes self-imposed congestion in the center (in contrast to Daniel (2009)).

Note that in the first-best for full independence, it cannot be socially optimal to fully deter fringe operations because the fringe serves a different market than the leader, where optimal consumption is not generally zero. Moreover, queuing delays have to be fully eliminated and, for f flights of the fringe and l flights of the leader, marginal social costs equals $\overline{\delta}(f + l)/K$ plus the airlines' constant operating costs. Therefore, first-best conditions for both markets are:

$$
s[A_l - B_l \cdot sl] = c_p + \frac{\overline{\delta} \cdot (l + f)}{K}
$$
$$
s[A_f - B_f \cdot sf] = c_p + \frac{\overline{\delta} \cdot (l + f)}{K} \tag{40}
$$

This indicates that the full price should equal marginal social costs. Note also that we do not restrict the demands to be the same, i.e. markets are asymmetric, and therefore optimal number of flights are different. Let f^* and l^* be the optimal number of flights of the fringe

and the leader respectively, that result from (40). Following the arguments of the previous section, the only way to eliminate queuing delays caused by the fringe is by charging it the following dynamic atomistic toll:

$$\tau(t) = \frac{\overline{\delta}(f^* + l^*)}{K} - \begin{cases} \overline{\beta} \cdot (t^* - t) & \text{if } t \le t^* \\ \overline{\gamma} \cdot (t - t^*) & \text{if } t \ge t^* \end{cases} \tag{41}$$

This toll makes the fringe indifferent between scheduling flights at any time when the aggregate arrival rate does not exceed capacity. The equilibrium interval of time where the fringe operates depends on decisions by the leader, and as a consequence, on the toll that the regulator charges to the leader.

The leader, that anticipates the fringe's reaction to the toll in (41), realizes that it can schedule flights efficiently (without queuing) because by setting the arrival rate equal to capacity is enough to keep the fringe outside its operating period. In doing so, the leader earns all the saved queuing costs as profit, and it is better off operating in the peak center where the schedule delays are lower. In this case, the leader's profit is:[16]

$$\Pi = l \cdot \left(s[A_l - B_l \cdot sl] - c_p - \frac{\overline{\delta} l}{2K} \right) \tag{42}$$

This is the monopoly profit described in Section 2.2 (see (18)). The first-order condition yields:

$$s[A_l - B_l \cdot sl'] = c_p + \frac{\overline{\delta} \cdot l'}{K} + B_l \cdot s^2 l' \tag{43}$$

where l' denotes the number of flights set by the leader when it does not face any toll and the regulator charges the dynamic atomistic toll to the fringe. Comparing (43) with the first-best condition for the leader (40), we see that the leader exerts market power by charging a per-flight markup equal to $B_l \cdot s^2 l'$, and does not internalize congestion imposed on the fringe. Hence, the leader sets frequency taking into account its own marginal generalized cost $\overline{\delta} l'/K$, instead of the social marginal cost $\overline{\delta}(l' + f)/K$.

As a consequence, the regulator has to correct both effects, but this is possible to achieve in a number of ways. We look first at the case where the regulator induces the first-best outcome with a flat toll to the leader, whereas the other socially optimal configurations are discussed below. The flat-toll that corrects these two effects is the sum of the uninternalized congestion and a subsidy equal to the markup:

$$\hat{\tau} = \frac{\overline{\delta} f^*}{K} - B_l \cdot s^2 l^* \tag{44}$$

---

[16]This is when the toll is defined as in (41) and not confined to the optimal period of operation $[t_s, t_e]$; this means that outside that period the toll is negative. If the toll is not negative outside the first-best interval of operation, the analysis holds if and only if parameters are such that l' is smaller than l^*. This is when the number of flights set by the leader when it does not face any toll and the regulator charges the atomistic toll to the fringe, is smaller than the first-best number of flights for the leader.
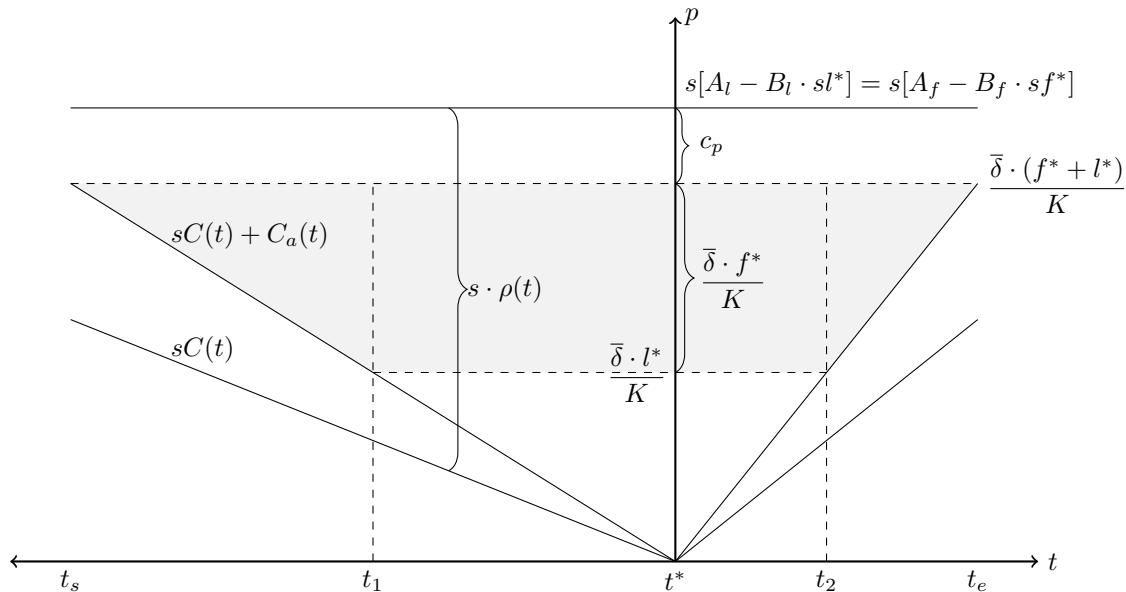
Figure 4: First-best equilibrium with full independence.

Note that depending on the parameters, the toll might be negative or even zero. Figure 4 shows the unique (first-best) equilibrium that results from charging $\widehat{\tau}$ in (44) to the leader, and the dynamic atomistic toll in (41) to the fringe. The leader schedules its flights to arrive, in the center, between $[t_1, t_2]$, and the fringe operates outside. As Figure 4 shows, the first-best conditions (40) are satisfied, and there are no queuing delays. The leader's profit is equal to the saved queuing costs $\overline{\delta}l^{*2}/2K$ and the revenues from the market power effect $(l^* \cdot B \cdot s^2 l^*$, not shown graphically). The congestion toll revenues (before subtracting the subsidy) are equal to the shaded area in Figure 4: the sum of the revenues from the leader (the rectangle in the center) and from the fringe (the two triangles at the shoulders).

This first-best toll regime is again different from the results by Brueckner and Van Dender (2008), because the leader is not failing to internalize self-imposed congestion. Moreover, it also differs from the first-best toll proposed by Daniel (2009), which was the dynamic atomistic toll (41) for both the leader and the fringe. In our model, the leader anticipates the fringe's decisions and realizes that—when the regulator charges the dynamic atomistic toll to the fringe—it is enough, and profit maximizing, to schedule flights efficiently around the preferred arrival time, keeping the fringe out of this period. As a consequence, the leader only fails to internalize the congestion imposed on the fringe through the expansion of the peak period.

Nevertheless, as in the perfect substitution case, this toll regime is not the only one that yields the first-best outcome. Although the market power has to be corrected in any case, the regulator can induce the leader to set the full price of its flights equal to the marginal social cost with regimes that correct the congestion effect differently.

First, note that if the leader is forced to give up the peak center, and operate outside the fringe's period, the first-best can again be attained (given that the market power is being corrected). In this case the fringe will operate around t^*, between $[t_1', t_2']$, paying the dynamic atomistic toll, ensuring that full price equals marginal social cost. The generalized cost per

flight at t_1' and t_2' for the leader is equal to $\overline{\delta} f^*/K$ and it schedules from t_s to t_1' and from t_2' to t_e, without queuing and earning the saved queuing costs as profit. The leader's first and last flight face a generalized cost (per flight) of $\overline{\delta}(f^* + l^*)/K$, hence satisfying first-best condition (40).[17] The toll regime that induces this outcome is the dynamic atomistic toll (41) for the fringe, and for the leader the per-flight market power subsidy to correct dead-weight losses $B_l \cdot s^2 l^*$, and a toll equal or higher than $\overline{\delta}(f^* + l^*)/K$ only during the period $[t_1', t_2']$. The latter works as a barrier for the leader to operate in the peak center, as it makes him better off operating outside and not paying the toll. In fact, it is equivalent to restrict the interval of time where the leader can operate.

These two configurations are similar in the sense that full price does not change, because the gain in costs by the fringe (resulting from operating closer to t^*) is offset by higher tolls, and the cost increase of the leader is offset by the absence of congestion tolls. This makes these settings identical in social welfare, consumer surplus, profit per firm (hence total profit) and total revenue. The difference, besides the times of operation for each firm, is that the tax revenues are not the same for each individual firm, but total tax revenues remain unchanged. In fact, there is a continuum of configurations that follow these properties, defined by more elaborate patterns of temporal separation of leader and fringe operations, but the congestion tolls become a more complicated matter. The leader can operate outside any restricted time-window around t^*, and it will be an efficient outcome as long as the regulator charges the dynamic atomistic toll to the fringe, corrects the market power effect with a subsidy, and charges for the congestion imposed by the leader on the fringe's flights that did not fit in the central time-window, and are outside the leader's time of operation. Any time restriction (or barrier-toll) together with an efficient flat congestion toll designed for that time restriction gives a unique first-best equilibrium pattern of arrivals. The two outcomes discussed above are special cases of a combination of time-restriction and efficient congestion flat-toll: (i) no restriction together with a toll equal to the congestion imposed on all the fringe's flights, and (ii) a period where all of the fringe's flights can operate and hence no congestion tolls. All the other possible equivalent configurations are in between these two.

Finally, if the dynamic atomistic toll (41) is applied to both the leader and the fringe, and the market power subsidy is given to the leader, the first-best outcome is also attained. In this case, the equilibrium pattern of queuing implies that there is no particular order of flights. The intuition is that, with the bottleneck model, the leader cannot be better than when it sets the arrival rate equal to capacity so as to face only schedule delay costs, because it is cost minimizing and regardless of paying (41) since profit is still positive. With the dynamic atomistic toll and the subsidy, full price will equal marginal social cost and the part of the leader's profit that corresponded to queuing costs savings in previous regimes, is transferred to the regulator. The leader will still have a positive profit because of the market power. This is the first-best regime that yields the highest revenues for the regulator, and the revenue gain

---

[17]This equilibrium is not shown graphically, but it is enough to see Figure 4 and change l^* for f^* (and vice versa), t_1 for t_1' and t_2 for t_2'. The duration of the center is $t_2' - t_1' = f^*/K$.

with respect to the toll regimes discussed above is equal to $\bar{\delta}l^{*2}/2K$. The total revenue from congestion tolls allows the Mohring-Harwitz self-financing result to hold, and it will only be overturned by the market-power subsidy.

- General case

As discussed in Section 3.1, in the untolled equilibrium of the general case of imperfect substitution ($0 < E < B$), the leader schedules flights in the peak center along with the fringe and also in the peak shoulders. The fact that the full price of the leader's trips depends also on the fringe's number of flights, implies that the offsetting behavior (see (32) and (33), $0 < \lambda < \phi < -1$) partially restricts the leader's ability to exert market power, because the fringe partially offsets any leader's reduction of flights.

The derivation of the first-best tolls is similar to the case of full independence, therefore we keep discussion concise and begin with the flat toll case. In the first-best equilibrium, queuing again has to be eliminated and full prices must equal marginal social cost. Denote f^* and l^* the optimal number of flights of the fringe and the leader. The regulator then charges the dynamic atomistic toll (41) to the fringe, and a flat toll $\hat{\tau}$ to the leader. The leader is better off when scheduling flights efficiently around t^*, therefore the leader's profit and first-order conditions are:

$$\Pi = l \cdot \left(s[A - B \cdot sl - E \cdot sf] - c_p - \frac{\bar{\delta}l}{2K} \right)$$

$$\frac{\partial \Pi}{\partial l} = 0 \Rightarrow s[A - B \cdot sl - E \cdot sf] = c_p + \frac{\bar{\delta} \cdot l}{K} + (B + E \cdot \frac{\partial f}{\partial l}) \cdot s^2 l + \hat{\tau} \qquad (45)$$

This is analogous to the result of the full independence case (see (43)), where the leader does not take into account the delays imposed on the fringe and exerts market power, in this case reduced because of the (partial) offsetting behavior of the fringe. The fringe's full price depends on l and hence, any reduction of frequency by the leader will result in a lower full price for the fringe, that—because of the free-entry (zero-profit condition)—translates into an output increase by the fringe. The first-best flat-toll is simply the flat toll that corrects market power and congestion effects:

$$\hat{\tau} = \frac{\bar{\delta}f^*}{K} - (B + E \cdot \frac{\partial f}{\partial l}) \cdot s^2 l^* \qquad (46)$$

The equilibrium is very similar to the one depicted in Figure 4. The difference in optimal number of flights for the leader and the fringe, makes the actual values different, but the structure is the same (e.g. t_s and t_e are different, but slopes do not vary).

Likewise, the flat-toll higher or equal than $\bar{\delta}(f^* + l^*)/K$ applied to the leader in the center peak (of duration f^*/K around t^*), together with the subsidy in (46) also yields the first-best outcome. The intuition is the same: this toll works as a time-restriction that keeps the leader in the shoulders, where it does not impose congestion on the fringe and is better of scheduling efficiently; on the other hand, the fringe pays the dynamic atomistic toll that ensures full price equal marginal social cost, and keeps revenues unchanged with respect to

the equilibrium where the leader pays (46). Moreover, again there is a continuum of first-best configurations that yield the same social welfare and revenue, each of them obtained with a unique combination of time-restriction and efficient congestion toll. Finally, the regulator again can charge the dynamic atomistic toll (41) to both the leader and the fringe, to induce the first-best outcome and maximize revenue, as long as the market power subsidy is given to the leader. The intuition is that the leader is better off scheduling flights efficiently because of the congestion technology, and the subsidy ensures the optimal number of flights and positive profit to the leader. The profits that the leader was earning (apart from the market power) are transferred to the regulator.

The first-best toll regimes for the case of imperfect substitutability always includes the subsidy that corrects the market power exertion by the leader, but the congestion side of the toll appears to be different from what has been found in the literature before. Brueckner and Van Dender (2008) find that the leader should pay a congestion toll that lies in between the congestion imposed on the fringe and total marginal congestion costs. We find that, when the regulator charges the dynamic atomistic toll to the fringe, because of the sequential nature of the game and dynamic congestion, the leader does not fail to internalize self-imposed congestion. As a consequence, the regulator can induce the first-best outcome by charging the congestion imposed by the leader on the fringe (analogous to the so-called "Cournot" toll), or charge him a barrier-toll that induces him to operate in the shoulders, hence not imposing congestion on the fringe. The regulator can also increase revenues by charging the dynamic atomistic toll to both the leader and the fringe, and (partially) finance capacity investments, because with deterministic bottleneck congestion this does not induce scheduling inefficiencies, but only induces monetary transfers. In addition, consumer surplus increases with the implementation of efficient congestion pricing, as the the full price of a trip decreases with respect to the untolled equilibrium.

Although the untolled equilibria derived in this Section are consistent to previous literature, including static congestion models, the first-best congestion tolls are not. Results show, in general, that the first-best can be induced with a dynamic atomistic toll for the group of competitive carriers and a flat toll to the leader that accounts for market power exertion and only the congestion imposed on the fringe. These tolls yield an equilibrium where leader and fringe operate in different times. Moreover, results suggest that the simple congestion pricing scheme of levying the dynamic atomistic toll to all airlines (at an airport level), also yield the first-best equilibrium (market power being corrected at a market level). This congestion pricing scheme increases revenues without altering social welfare and is easier to implement in the sense that it may not be perceived as inequitable. The non-uniqueness of optimal toll schedules that we find has not been reported before. Note, however, that the resulting equilibria are identical in terms of welfare, and only the order of operation is different.

These results also imply that, in a sequential leader-fringe competition, when there is more than one non-atomistic carrier in the first-stage, the first-best congestion toll is unique and fully atomistic (as in (28)), because, as derived in Section 2.3, a duopoly cannot reach the efficient outcome, and only market power subsidies have to be airline and market specific.

3.3. Manipulable tolls

We assume in the analysis above that airlines that are large enough to exert market power and to recognize the impact of their decisions on overall congestion (and followers), do not take into account the impact of their actions on the tolls. We are aware that this is a strong assumption, but it is common to most previous works. Brueckner and Verhoef (2010) propose a manipulable toll rule designed to induce the social optimum when carriers predict the impact of their decisions on tolls that can also be applied to our problem. They propose an adjustment such that the carriers' profit plus the (manipulable) toll liability varies perfectly in parallel with social surplus. In our problem, market power correction can be straighforwardly adjusted with their methodology: the subsidy for an airline is the surplus from own-passengers consumption. This makes the airlines pricing strategy maximize social welfare. On the other hand, congestion tolls have to be corrected differently as per-flight tolls that vary over time are required. The efficient time-variant toll has to be corrected by modifying the intercept and keeping the slopes unchanged. This is because the slopes provide the incentives to schedule flights efficiently and the intercept to set the efficient number of flights. The manipulable toll has to be such that the derivative of the number of flights times the intercept yields marginal social cost; in the case of the Stackelberg leader this also needs to take into account that the leader does not view the follower's number of flights as parametric.

4. Conclusion

This paper studies airlines' interactions and scheduling behavior together with airport efficient pricing with a deterministic bottleneck model of congestion. We confirm that non-atomistic airlines internalize self-imposed congestion in the sense that, without facing tolls, they schedule less flights than perfectly competitive carriers would, achieving lower social congestion costs. Consistent with earlier literature with static models of congestion (e.g., Brueckner, 2002; Brueckner and Van Dender, 2008), the degree of internalization of self-imposed congestion depends critically on the market structure. Nevertheless, our results suggests that social welfare maximizing congestion tolls do not depend crucially on the degree of internalization, and that the time-variant tolls derived for perfectly competitive carriers apply also to a Cournot oligopoly and to a setting where a Stackelberg leader interacts with a group of competitive carriers as followers. On the other hand, time-invariant tolls—that are second-best optimal—depend on market structure and are consistent with earlier literature. This is because time-variant tolls aim at changing the schedule of flights so as to reduce inefficient queuing delays, whereas time-invariant tolls aim at reducing the number of flights that are being scheduled inefficiently.

Our analyses suggest that optimal congestion pricing may have a more significant role than what has been suggested in the earlier literature based on static models. Moreover, the efficient fully time-variant congestion toll regime results in a revenue for the airport that restores the well known self-financing result for congested facilities.[18] Results also suggest

---

[18]The market-power subsidy that corrects dead-weight losses can overturn this, but, as suggested by Brueck-

that the political feasibility of congestion pricing would be enhanced as efficient congestion charges do not vary with market shares, and therefore may not be perceived as inequitable.

We also find agreement with Daniel (1995) and Daniel and Harback (2008) in that dynamic atomistic tolls are efficient in markets well represented by an interaction between a leader and a competitive follower, but we show that this is not the only efficient solution. The non-uniqueness of social welfare maximizing congestion tolls in this setting allows for other pricing schemes that achieve the social optimum. As the dominant airline behaves inefficiently because of the competitive follower's presence, it is enough to toll fringe carriers out from the leader's operation time-window to provide the right incentives (although further time-invariant corrections are needed).

We see incorporating heterogeneity and studying step-tolling as natural extensions of the present analysis, to complement Daniel's (2009) work. Our model allows to include heterogeneity in values of time and preferences for both airlines and passengers. Certainly, the equilibrium and optimal toll will depend on the type of heterogeneity considered. Finally, step-tolling, a relevant alternative in practice, may bring important benefits compared with the social optimum; as the number of steps is increased, it approaches the dynamic atomistic congestion toll, and, consequently, its efficieny (and consumer surplus) increases approaching the optimum values.

5. Acknowledgements

References

Arnott, R., De Palma, A. and Lindsey, R. (1990), 'Economics of a bottleneck', *Journal of Urban Economics* **27**(1), 111–130.

Arnott, R., De Palma, A. and Lindsey, R. (1993), 'A structural model of peak-period congestion: A traffic bottleneck with elastic demand', *The American Economic Review* **83**(1), 161–179.

Arnott, R., De Palma, A. and Lindsey, R. (1994), 'The welfare effects of congestion tolls with heterogeneous commuters', *Journal of Transport Economics and Policy* **28**(2), 139–161.

Basso, L. J. and Zhang, A. (2007), 'Congestible facility rivalry in vertical structures', *Journal of Urban Economics* **61**(2), 218–237.

Basso, L. J. and Zhang, A. (2010), 'Pricing vs. slot policies when airport profits matter', *Transportation Research Part B: Methodological* **44**(3), 381–391.

---

ner (2005), this should be corrected at a city-pair market level instead of at an airport level. But even then the break-down of self-financing remains.

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

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

Brueckner, J. K. (2009), 'Price vs. quantity-based approaches to airport congestion management', *Journal of Public Economics* **93**(5-6), 681–690.

Brueckner, J. K. and Girvin, R. (2008), 'Airport noise regulation, airline service quality, and social welfare', *Transportation Research Part B: Methodological* **42**(1), 19–37.

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

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

Carlin, A. and Park, R. (1970), 'Marginal cost pricing of airport runway capacity', *The American Economic Review* **60**(3), 310–319.

Carlsson, F. (2003), 'Airport marginal cost pricing: Discussion and an application to swedish airports', *International Journal of Transport Economics* **30**(3), 283–303.

Daniel, J. I. (1995), 'Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues', *Econometrica: Journal of the Econometric Society* **63**(2), 327–370.

Daniel, J. I. (2001), 'Distributional consequences of airport congestion pricing', *Journal of Urban Economics* **50**(2), 230–258.

Daniel, J. I. (2009), 'The deterministic bottleneck model with non-atomistic traffic', *Working Papers 09-08, University of Delaware, Department of Economics* .

Daniel, J. I. and Harback, K. T. (2008), '(When) do hub airlines internalize their self-imposed congestion delays?', *Journal of Urban Economics* **63**(2), 583–612.

Daniel, J. I. and Harback, K. T. (2009), 'Pricing the major us hub airports', *Journal of Urban Economics* **66**(1), 33–56.

Forbes, S. J. (2008), 'The effect of air traffic delays on airline prices', *International Journal of Industrial Organization* **26**(5), 1218–1232.

Jorge, J. D. and de Rus, G. (2004), 'Cost-benefit analysis of investments in airport infrastructure: a practical approach', *Journal of Air Transport Management* **10**(5), 311–326.

Levine, M. E. (1969), 'Landing fees and the airport congestion problem', *Journal of Law and economics* **12**(1), 79–108.

Mayer, C. and Sinai, T. (2003), 'Network effects, congestion externalities, and air traffic delays: Or why not all delays are evil', *The American Economic Review* **93**(4), 1194–1215.

Mohring, H. and Harwitz, M. (1962), *Highway benefits: An analytical framework*, Evanston, IL: Northwestern University Press.

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

Rupp, N. (2009), 'Do carriers internalize congestion costs? empirical evidence on the internalization question', *Journal of Urban Economics* **65**(1), 24–37.

Santos, G. and Robin, M. (2010), 'Determinants of delays at european airports', *Transportation Research Part B: Methodological* **44**(3), 392–403.

Silva, H. E. and Verhoef, E. T. (2011), 'Optimal pricing of flights and passengers at congested airports: The efficiency of atomistic charges', *Tinbergen Institute Discussion Paper No. 11-179/3* .

Small, K. A. (1982), 'The scheduling of consumer activities: work trips', *The American Economic Review* **72**(3), 467–479.

van den Berg, V. A. C. and Verhoef, E. T. (2011), 'Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay', *Journal of Public Economics* **95**(7–8), 983–992.

Verhoef, E. T. (2010), 'Congestion pricing, slot sales and slot trading in aviation', *Transportation Research Part B: Methodological* **44**(3), 320–329.

Vickrey, W. (1973), 'Pricing, metering, and efficiently using urban transportation facilities', *Highway Research Record* **476**, 36–48.

Vickrey, W. S. (1969), 'Congestion theory and transport investment', *The American Economic Review* **59**(2), 251–260.

Zhang, A. and Zhang, Y. (2006), 'Airport capacity and congestion when carriers have market power', *Journal of Urban Economics* **60**(2), 229–247.