



# Dynamic equilibrium at a congestible facility under market power<sup>☆</sup>



Erik T. Verhoef<sup>a,b</sup>, Hugo E. Silva<sup>c,d,\*</sup>

<sup>a</sup> Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

<sup>b</sup> Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

<sup>c</sup> Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile

<sup>d</sup> Instituto de Economía, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile

## ARTICLE INFO

### Article history:

Received 3 March 2017

Revised 20 August 2017

Accepted 22 August 2017

### JEL classification:

R41

R48

D62

### Keywords:

Congestion pricing

Dynamic congestion

Market power

Internalization

## ABSTRACT

This paper studies equilibrium and optimum at a congested facility when firms have market power; e.g., when a few airlines jointly use a congested airport. Unlike most of the previous literature, we characterize the equilibrium in terms of timing of arrivals in a continuous-time congestion model when firms simultaneously schedule services. Using the Henderson-Chu dynamic model of flow congestion in a multiple-firm setting, we find that a stable and unique Nash equilibrium in pure strategies always exists. Importantly, it also exists in cases where it fails to exist under bottleneck congestion (notably when the value of schedule late exceeds the value of travel delays). We find that symmetric firms schedule arrivals inefficiently, and strongly concentrated around the desired arrival time so that the peak is shorter and delays are higher than socially optimal. We show that when firms are asymmetric in terms of output, all firms schedule vehicles in the peak center, around the desired arrival time, with arrival windows increasing with firm size such that a smaller firm's window is always fully contained in a larger firm's window and only the largest firm operates in the early and late shoulders. Furthermore, for any pair of asymmetric firms, the larger firm has a higher instantaneous arrival rate at any moment where both firms schedule arrivals. Our results also show that even though self-internalization can be substantial, there is scope for decentralizing the first-best outcome through time-varying tolls.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

Congestion at airports has recently grown into an important theme in the economics literature. Early contributions by Daniel (1995), Brueckner (2002) and Pels and Verhoef (2004) have brought to the fore that models of road traffic congestion are not directly applicable to the economic analysis of airport congestion. In contrast to atomistic road users, airlines with market power face an incentive to internalize self-imposed congestion. As a result, a traditional Pigouvian toll equal to the marginal external cost would lead to overcharging of congestion, and would have to be corrected by a certain term – one

<sup>☆</sup> Financial support from ERC (AdG Grant #246969 OPTION) is gratefully acknowledged. This research was partially funded by the Center for Sustainable Urban Development CEDEUS (CONICYT/FONDAP 15110020) and the Complex Engineering Systems Institute (CONICYT – PIA – FB0816). We are also grateful to the editor, Robin Lindsey, and three anonymous referees for helpful suggestions.

\* Corresponding author.

E-mail addresses: [e.t.verhoef@vu.nl](mailto:e.t.verhoef@vu.nl) (E.T. Verhoef), [husilva@uc.cl](mailto:husilva@uc.cl) (H.E. Silva).

minus the airline's market share in a basic Nash–Cournot setting – to secure a socially optimal outcome. Moreover, when airlines with market power have an incentive to also apply demand-related mark-ups in their pricing policies, a further downward adjustment compared to the Pigouvian rule is in order for efficiency reasons.

There is mixed evidence on the degree to which internalization of self-imposed congestion indeed occurs in reality. Mayer and Sinai (2003), for example, provide evidence supporting such “self-internalization”, while Daniel and Harback (2008) argue that traffic patterns at airports mostly follow patterns consistent with atomistic behaviour. Several theoretical contributions shed further light on this seemingly contradictory evidence. Notably, Brueckner and Van Dender (2008) show that a Stackelberg leader competing with a competitive fringe of atomistic players would not internalize self-imposed congestion if the products are perfect substitutes. This is because the leader would realize that unused capacity will be filled-up with aircraft from the fringe, leaving congestion unaltered but reducing the leader's profit. Silva and Verhoef (2013) consider Bertrand (rather than Cournot) behaviour of non-atomistic airlines offering imperfect substitutes, and find that also in this case self-internalization is limited, and more so when the products of the competing airlines become closer substitutes. Quite intuitively, a Bertrand player realizes that even when the rival player keeps his fare fixed, the rival's quantity responds to changes in the player's own fare. In particular, increasing its own fare causes an increase in the rival's quantity and therefore increased congestion costs. This makes self-internalization appear less attractive than what it seems under a Cournot assumption of a fixed quantity supplied by the competitor.

Apart from the nature of the game, also some other aspects of the problem have received attention. For example, Basso and Zhang (2007) consider the role of airport capacity choice, Czerny and Zhang (2011, 2014a) study the role of business and leisure passengers for the incentives to internalize self-imposed congestion, and Verhoef (2017a) studies the design of self-financing mechanisms for congestible facilities with market power. Another branch of the literature has focused on peak-load airport pricing using a model of discrete time with two periods, peak and off-peak, where only the peak period can be congested. In this line, Brueckner (2002), Basso and Zhang (2008) and Czerny and Zhang (2014b) have shown that firms competing in Cournot fashion internalize the self-imposed delays also in the two period framework.

More recently, the dynamics of congestion have been subject to analysis. Silva et al. (2014) proposed a dynamic model of airport congestion, combining the game-theoretic set-up of most of the earlier work with the bottleneck congestion technology proposed by Vickrey (1969) for the analysis of road pricing. Silva et al. (2014) describe monopoly and leader-fringe cases and show that the leader is forced to schedule arrivals according to atomistic patterns in the peak center. They also reported a rather discomfiting result, namely that the model seems to have no pure-strategies Nash equilibrium in arrival schedules for the case of a Cournot oligopoly.<sup>1</sup>

Silva et al. (2017) confirmed the non-existence of equilibrium in the described setting for a duopoly, albeit that they demonstrate that it applies only when the so-called value of schedule delay late ( $\gamma$ ) exceeds the value of travel delays ( $\alpha$ ). When the opposite applies, there is an equilibrium; however, it is one in which the strategies are such that no queuing occurs. Although this does not overturn the relevance of the particular case, it does suggest that this would make the model still unable to describe equilibrium on congested facilities with “visible” (queued) congestion. As a consequence, the model would also be unable to provide insights on optimal congestion pricing policies.

Given the widespread prevalence of congestion at airports and the presumed relevance of its dynamic “peak-hour” nature in many instances, combined with the rather extreme nature of leader-fringe competition – the only type of competition that does have an equilibrium in pure strategies with dynamic bottleneck congestion – it is not just an intellectual challenge to characterize a congested equilibrium with firms making decisions simultaneously in a dynamic model of congestion. It is also a task with a clear societal relevance, as the efficiency of regulation (or a lack thereof) has not been assessed in a model with continuous time and travel and schedule delays.

Against that background, this paper investigates whether equilibrium is restored under alternative assumptions on the congestion technology and, if so, whether there is room for implementing optimal congestion pricing. In particular, we believe that the assumptions on the demand side – basically entailing the existence of a most-desired arrival moment and shadow costs of deviations from this moment ( $\beta$  for early arrivals,  $\gamma$  for late ones), and the existence of a disutility of travel delays (with unit value  $\alpha$ ) – are too reasonable to drop. Although heterogeneity in consumer preferences or in airlines' cost structures may also help to establish equilibrium, it remains a bit awkward that under (too much) homogeneity, no equilibrium would exist. The same is true for uncertainty in travel times. Hence, the specific congestion technology seems a reasonable first candidate alternative assumption to consider.

We build upon the framework set out in Silva et al. (2014), but use a model of flow congestion instead of the bottleneck congestion technology, and focus on simultaneous scheduling by multiple firms. Again borrowing from classics in road congestion modelling, we consider the model originally proposed by Henderson (1974, 1981), and later refined by Chu (1995). The essential feature of the model is that it makes the travel delay associated with an arrival at a certain moment a function of the instantaneous arrival flow at that moment. The model may thus seem to give a reasonable description and approximation of dynamic equilibrium in terms of scheduling behaviour and travel delays when there is no strict and predictable FIFO (first-in-first-out) queuing discipline, and players instead have an unbiased but somewhat rough expectation of the travel delay that an arrival at a certain moment will bring. Another reason why the congestion technology may give a reasonable description of dynamic congestion patterns is when the facility operator has the possibility of opening additional but lower

<sup>1</sup> Daniel (2009) studies in a similar setting the leader-fringe case finding also that the leader is forced to behave atomistically in the center.

quality capacity at busier times, such as more remote terminals or runways, which bring additional on-the-ground travel delays. Nevertheless, as is true for other dynamic congestion technologies used for analytical modelling, including the pure bottleneck model, the congestion technology remains a simplification, which is the price to pay for analytical tractability. Arguably the most important weakness of the model in its purest form is the complete lack of direct congestion interactions between individuals travelling at different moments, no matter how close these moments are. To address this concern, we develop, in the [Appendix](#) of this paper, an extension of the model that remedies this shortcoming, and find that the important conclusions that we draw from the basic model survive this extension. We believe the models are well-suited to serve the main aim of this paper, namely to establish in a tractable model whether a revision of assumptions on the congestion technology restores dynamic equilibrium under market power.

We show that under this alternative congestion technology there is an equilibrium in pure strategies in terms of arrival flows, and we assess its efficiency relative to that of two natural benchmarks: the purely atomistic equilibrium, and the social optimum. When firms are fully symmetric, the unique equilibrium has all firms scheduling arrivals in a perfectly overlapping interval around the desired arrival time. We show that the peak period starts later and ends earlier than the socially optimal peak period, and that the growth rate of the aggregate arrival rate is larger than in the optimum. In other words, firms schedule arrivals inefficiently concentrated around the desired arrival time. Due to the internalization, however, this concentration is not as high as it would be if firms were atomistic. We also find that when firms differ in terms of output, all firms schedule arrivals in the peak center around the desired arrival time with arrival windows increasing with firm size such that a smaller firm's window is always fully contained in a larger firm's window and only the largest firm operates in the early and late shoulders. Furthermore, the larger a firm is, the higher is its instantaneous arrival rate at any moment of the peak center, thus dominating the most preferred arrival times. Finally, we find that the social optimum can be supported as an equilibrium by applying time-varying congestion charges.

The paper proceeds as follows. [Section 2](#) presents the model and the equilibrium conditions for the most general case in which none of the parameters is pinned down. We show that, in general, equilibrium always exist with this congestion technology. However, because this equilibrium cannot be given in closed-form, we proceed by discussing a number of special cases. A first case, in [Section 3](#), uses a special congestion function, namely one that is linear in the aggregate arrival flow. For this case, we can present the equilibrium in closed form, independent of whether the two firms are symmetric. We next turn to a more general case in [Section 4](#), in which the power of the congestion function is larger than one. By staying as close as possible to the numerical version of the linear model, we provide some further insights, even though there is no closed-form solution. Finally, in [Section 5](#), we simplify by imposing symmetry upon the two firms, and again obtain closed-form solutions but now also for non-linear congestion functions. [Section 6](#) concludes.

## 2. A dynamic model of operators with market power using the same facility with flow congestion

### 2.1. The original [Chu \(1995\)](#) model<sup>2</sup>

The congestion technology that we use in our analysis was originally proposed by [Chu \(1995\)](#), building on earlier work by [Henderson \(1974, 1981\)](#). What distinguishes this congestion technology from others is that it assumes that a traveller's speed will be constant over time throughout the trip, depending only on the arrival rate at the road's exit at the instant that the trip is completed (in Chu's version). The model thus ignores interactions between travellers who have departed at different instants, no matter how close. [Lindsey and Verhoef \(2008\)](#) therefore refer to this assumption as "no propagation", to distinguish it from models where shock waves travel at finite speeds along a road.

The basic model considers  $N$  identical travellers who use a single road for their trip in the morning commute. They have perfectly inelastic demand, a desired arrival time denoted  $t^*$ , a value of time  $\alpha$ , and values of schedule delay of  $\beta$  for early arrivals and  $\gamma$  for late ones. As is customary, we define  $\delta \equiv (\beta \cdot \gamma) / (\beta + \gamma)$  as a composite schedule delay cost coefficient. The capacity of the road is given, and is denoted  $K$  (there are some changes in notation compared to Chu's), and the travel time  $T(t)$  associated with an arrival at time  $t$  depends on both  $K$  and on the instantaneous arrival rate  $f(t)$ . To obtain closed-form solutions, a functional form for the travel time function  $T(f(t); K)$  needs to be specified. Chu uses a power-law or BPR (Bureau of Public Roads) type of function:

$$T(f(t); K) = T_f + \left( \frac{f(t)}{K} \right)^\chi \quad (1)$$

where  $\chi$  determines the curvature of  $T(\cdot)$ . Note that the regular BPR function pre-multiplies the second term with  $T_f \cdot b$  with  $b$  being a second parameter, but this can be dropped by choosing the appropriate units for  $K$ . Letting  $\tau(t)$  denote a possibly time-varying toll, the generalized price for an arrival at  $t$  can be written as the sum of  $\tau(t)$ , the travel time cost  $c_T(t)$ , and the schedule delay cost  $c_{SD}(t)$ :

$$p(t) = \tau(t) + c_T(t) + c_{SD}(t) = \tau(t) + \alpha \cdot T(f(t); K) + \begin{cases} \beta \cdot (t^* - t) & \text{if } t \leq t^* \\ \gamma \cdot (t - t^*) & \text{if } t > t^* \end{cases} \quad (2)$$

<sup>2</sup> The exposition of the basic model in this [Section 2.1](#) draws heavily from the one in [Verhoef \(2017b\)](#). Literal citations are not marked as such and are taken to be acknowledged through this footnote.

In the dynamic equilibrium, arrival rates for early (before  $t^*$ ) and late (after  $t^*$ ) arrivals should be such that  $p(t)$  remains constant over time. The timing of the peak then follows from the conditions that (i) the schedule delay cost for the very first driver, arriving at  $t_q$ , and the very last driver, arriving at  $t_e$ , should be the same, and (ii) between  $t_q$  and  $t_e$ , exactly  $N$  drivers should have arrived. For the travel time function of Eq. (1), the no-toll equilibrium is then characterized by (the subscript A stands for “atomistic no-toll equilibrium”):

$$f_A(t) = \begin{cases} K \cdot \left( \frac{\beta}{\alpha} \cdot (t - t_{qA}) \right)^{\frac{1}{\chi}} & \forall t : t_{qA} \leq t \leq t^* \\ K \cdot \left( \frac{\gamma}{\alpha} \cdot (t_{eA} - t) \right)^{\frac{1}{\chi}} & \forall t : t^* < t \leq t_{eA} \end{cases} \quad (3)$$

Using the short-hand parameter

$$\Psi = \left( \frac{N}{K} \cdot \frac{1 + \chi}{\chi} \cdot \frac{\delta}{\alpha} \right)^{\frac{\chi}{1+\chi}} \quad (4)$$

and setting  $t^* = 0$  without loss of generality, the peak’s start and end times can be written as:

$$t_{qA} = -\Psi \cdot \frac{\alpha}{\beta} \quad (5)$$

$$t_{eA} = \Psi \cdot \frac{\alpha}{\gamma} \quad (6)$$

The equilibrium in (3)–(6) is not efficient due to the uninternalized congestion externality, and Chu (1995) shows that the first-best optimum can be attained by setting a time-varying toll that each instant takes on the familiar Pigouvian form:

$$\tau(t) = f(t) \cdot \frac{\partial c_T(f(t))}{\partial f(t)} \quad (7)$$

Applying this toll for the fixed-demand case widens the peak and flattens the arrival rate pattern, as is shown by the optimal counterparts of (3)–(6) (where O stands for “optimum”):

$$f_O(t) = \begin{cases} K \cdot \left( \frac{1}{1+\chi} \cdot \frac{\beta}{\alpha} \cdot (t - t_{qO}) \right)^{\frac{1}{\chi}} & \forall t : t_{qO} \leq t \leq t^* \\ K \cdot \left( \frac{1}{1+\chi} \cdot \frac{\gamma}{\alpha} \cdot (t_{eO} - t) \right)^{\frac{1}{\chi}} & \forall t : t^* < t \leq t_{eO} \end{cases} \quad (8)$$

$$t_{qO} = -(1 + \chi)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\beta} \quad (9)$$

$$t_{eO} = (1 + \chi)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\gamma} \quad (10)$$

Integrating the equilibrium arrival rates confirms that the proportions of early and late drivers are such that a fraction  $\gamma/(\beta + \gamma)$  of the  $N$  drivers arrive early, and a fraction  $\beta/(\beta + \gamma)$  late. This is true both in the no-toll equilibrium and in the first-best optimum.

## 2.2. Operators with market power

Now let us turn to the case of actual interest in this paper, where operators with market power provide services using the congestible facility. While there are many studies that analyze the properties of the bottleneck model with general travel delay functions (e.g., Astarita, 1996; Heydecker and Addison, 2005; Chow, 2009a, 2009b), they all focus on users that decide the timing of only one trip (their own). Our paper is different in that we are concerned with few users (firms in our setting) that control many vehicles, and therefore take into account the impact of the timing of a single vehicle on the rest of its vehicles. As pointed out in the introduction, Silva et al. (2017) considered this problem in the context of the Vickrey bottleneck model and found that a pure strategy Nash equilibrium in which there is congestion does not exist.

Following Silva et al. (2014), in our framework there is one facility that is prone to congestion (a bottleneck) with fixed capacity  $K$ . To maintain the focus on congestion dynamics, on the internalization of self-imposed congestion, and on the existence of a dynamic equilibrium, we ignore firm interactions resulting from competition for the same passengers. We thus assume that all operators serve different markets (which is why we will not refer to an oligopoly). Each firm  $i$  faces an inverse demand  $P_i(N_i)$ , where  $N_i$  is the firm’s total quantity (number of trips) during the peak that we consider, and schedules arrival times of its passengers to the facility to maximize profit. These arrival times are implied by the choice of arrival rates of passengers,  $f_i$ . We assume that the arrival rate of flights is perfectly proportional to  $f_i$  with a fixed factor  $F$ , which is determined by aircraft size and load factor. We will treat both arrival rates as continuous. We can therefore directly use aggregate passengers flows as the argument in the congestion function and do not have to model flows of aircraft explicitly.

Thus, the only strategic interaction between firms comes from the negative congestion externality. Congestion costs could accrue to passengers (their valuations of delays and scheduling disutilities), to the operators (e.g. fuel costs and costs related

to crew costs), or to both. Silva and Verhoef (2013), among others, have argued that the two types of congestion costs would essentially enter a firm's profit optimization problem in the same way when congestion costs incurred by passengers translate on a dollar-by-dollar basis into a lower willingness to pay fares, as long as firms compete in Cournot fashion. Under our assumptions, with travellers who are identical in terms of preferences, and with fares that may be freely differentiated over clock time to support a firm's preferred arrival time pattern, this is indeed the case. This means that costs incurred by the passengers and by the firm enter the firm's optimization problem symmetrically. More specifically, because we keep load factors fixed, we need not distinguish between the airline's costs per flight, the airline's costs per passenger, and costs incurred by the airline's passengers. Dropping all time-independent costs per-passenger as these provide no useful insight into the issues we study, and assuming symmetry in costs, we may define the following average (per passenger) costs for any firm  $i$ :

$$ac_i(t) = c_T(f_i(t) + f_{-i}(t); K) + c_{SD}(t) \quad (11a)$$

where  $f_i(t)$  is the instantaneous arrival rate of firm  $i$  and  $f_{-i}(t)$  is the aggregate instantaneous arrival rate of all other firms. Note that the average cost for any given moment is the same for all firms, which is why the index  $i$  will be dropped from  $ac$  from now onwards.

For the more specific assumptions of " $\alpha\beta\gamma$ -preferences" (linear schedule delay costs and a constant value of time) and a BPR travel time function, the case that we will henceforth refer to as the "specific model" for the sake of brevity, this becomes:

$$ac(t) = \alpha \cdot \left( \frac{f_i(t) + f_{-i}(t)}{K} \right)^X + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad (11b)$$

( $t^*$  is again set at 0).

Under these assumptions, there is no loss of generality in normalizing the fixed factor  $F$  to be equal to 1, and the firm  $i$  is able to charge a price for the trip equal to  $P_i(N_i) - ac(t)$  and solves the following profit maximization problem:

$$\begin{aligned} \max_{N_i, f_i(t), t_{qi}, t_{ei}} \quad & \pi_i = \int_{t_{qi}}^{t_{ei}} f_i(t) \cdot [P_i(N_i) - ac(t)] dt \\ \text{s.t.} \quad & \int_{t_{qi}}^{t_{ei}} f_i(t) dt = N_i \\ & f_i(t) \geq 0 \end{aligned} \quad (12a)$$

As shown by Silva et al. (2014), this can be studied without loss of generality as a two-step maximization. First, for any  $N_i$  the firm minimizes its total delay cost  $TC(N_i)$ :

$$\begin{aligned} \min_{f_i(t), t_{qi}, t_{ei}} \quad & TC_i(N_i) = \int_{t_{qi}}^{t_{ei}} f_i(t) \cdot ac(t) dt \\ \text{s.t.} \quad & \int_{t_{qi}}^{t_{ei}} f_i(t) dt = N_i \\ & f_i(t) \geq 0 \end{aligned} \quad (12b)$$

This yields the cost minimizing trip timing,  $f_i^*(t)$ , for a given total output  $N_i$ . The minimization problem above already highlights the main difference between the behavior of firms and that of small users in scheduling trips.<sup>3</sup> Each firm minimizes its aggregate travel cost and in doing so takes into account the delays imposed by any of its arrivals on the delay experienced by all the others. In contrast, a small user chooses the time of arrival to minimize the experienced delay by its single vehicle. Next, in the second step the firm maximizes:

$$\max_{N_i} \pi_i = N_i \cdot P_i(N_i) - TC_i^*(N_i) \quad (12c)$$

where  $TC_i^*(N_i)$  is the optimal value of the objective function as found in (12b). The insights that can be obtained from the quantity choice of the second step are not novel. It repeats the traditional monopoly solution in which the firm equalizes marginal revenue with marginal cost.

Again, with the purpose of maintaining the focus on the congestion dynamics, for most of the paper we will therefore focus attention on trip timing decisions for given quantities, without modelling explicitly how the firms choose quantities by equating marginal revenue to marginal cost. So the main focus is on, first, whether a dynamic equilibrium exists for a given set of chosen quantities, and, second, how its efficiency compares to that of the two benchmarks of the atomistic no-toll equilibrium and the social optimum.

Under these assumptions, all firms face the incentive to fully internalize firm-internal congestion externalities, just as in the static model of Brueckner (2002). Also, and as a result, each firm will find it optimal to equalize over time, as long as

<sup>3</sup> In the game theoretical literature, small users are non-atomic and large users such as firms are atomic. So, in that terminology, our paper deals with atomic games.

it schedules arrivals, its firm-internal marginal cost  $mc_i$  (see [Proposition 1](#) below), which we define as the derivative of its total cost,  $TC_i^*$ , with respect to the firm's own instantaneous arrival rate:

$$mc_i(t) = c_T(f_i(t) + f_{-i}(t); K) + f_i(t) \cdot c_T'(\cdot) + c_{SD}(t) \quad \forall i \quad (13a)$$

or, for the specific model:

$$mc_i(t) = \alpha \cdot \left( \frac{f_i(t) + f_{-i}(t)}{K} \right)^x + f_i(t) \cdot \chi \cdot \frac{1}{K} \cdot \left( \frac{f_i(t) + f_{-i}(t)}{K} \right)^{x-1} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad \forall i \quad (13b)$$

If a supplier's firm-internal marginal cost were not constant over time, transferring a passenger to a moment with a lower  $mc_i$  would increase profit. Note that this is true independent of whether the lower  $mc_i$  would arise from lower costs incurred by the firm while the fare is kept constant, or from lower costs incurred by the passenger and the fare for that passenger is raised accordingly. Also note that this is independent of any demand-related mark-up that the firm may apply in its pricing policy. Given that mark-up and given the number of passengers it chooses to serve, it remains profit maximizing for firm  $i$  to equalize  $mc_i$  as long as the firm schedules arrivals, and to choose not to use particular time intervals when  $mc_i$  – then equal to  $ac$  – exceeds  $mc_i$  in used time intervals.

### 2.3. Analytical results on equilibrium

In what follows we prove that a Nash equilibrium in arrival flows exists and that it is unique and stable. We reiterate that we do this for the case where firms are identical in all respects apart from, possibly, the equilibrium output. We proceed in several steps. First, we formally prove that the best response of a firm, for any given pattern of arrivals of the other firms, is to schedule its arrivals in such a way that the firm-internal marginal cost is equalized over time.

**Proposition 1.** *The best response of a firm, for any given pattern of arrivals of the other firms, is to schedule arrivals in such a way that the firm-internal marginal cost is equalized over time when it schedules arrivals. At times when the firm does not schedule arrivals, the average cost can, in equilibrium, not be smaller than the equalized firm-internal marginal cost of the moments when it schedules arrivals.*

**Proof.** Suppose firm  $i$  uses two arrival moments  $t_1$  and  $t_2$  with unequal firm-internal marginal costs  $mc_i(t_1) < mc_i(t_2)$  (without loss of generality). A marginal shift of arrivals from  $t_2$  to  $t_1$  then reduces costs by  $mc_i(t_2) - mc_i(t_1) > 0$ . Only when  $mc_i$  is equalized over times for used arrival moments are such profitable shifts exhausted, and are costs minimized. At unused moments,  $mc_i$  equals  $ac$ , so that whenever  $ac$  at some  $t_2$  is lower than  $mc_i$  at a used moment  $t_1$ , the firm would find it profitable to shift a marginal arrival from  $t_1$  to  $t_2$  and  $t_2$  becomes a used moment.

Suppose the competitors' aggregate arrival schedule is continuous. Then, the firm's arrival interval is defined by the moments where  $mc_i(t)$  for  $f_i(t) = 0$ , which equals  $ac(t)$ , is exactly equal to the equilibrium (equalized) firm-internal costs  $mc_i$ . This is because, as arrivals are non-negative, the firm-internal marginal cost is continuous in  $f$ , and the schedule delay costs are continuous in arrival moment. In other words, denoting the timing of firm  $i$ 's earliest arrival  $t_{qi}$  and its latest arrival  $t_{ei}$ ,  $f_i(t_{qi}) = f_i(t_{ei}) = 0$  and  $mc_i(t_{qi}) = ac(t_{qi}) = ac(t_{ei}) = mc_i(t_{ei}) = mc_i(t | f_i(t) > 0)$  determines the firm's arrival interval.  $\square$

The equalization of firm-internal marginal cost over time as a best-response implies that the dynamic equilibrium will deviate from both the atomistic no-toll equilibrium (equalizing average cost over time) and the first-best optimum (equalizing marginal social cost) that are known from the road congestion model of [Chu \(1995\)](#) and, also, the bottleneck model ([Arnott et al., 1993](#)).

To show that a unique Nash equilibrium exists, it is convenient to first establish two important features that characterize an equilibrium.

**Lemma 1.** *Equilibrium cannot entail for any pair of firms two disjoint arrival intervals without any overlap between those firm's arrival periods.*

**Proof.** Denote by  $i$  and  $j$  two firms that both schedule trips in two disjoint arrival intervals. We prove [Lemma 1](#) by assuming that there is no overlap between the firms' arrivals in those disjoint periods and show that this cannot be part of an equilibrium. Denote  $S_i$  the interval in which only firm  $i$  schedules arrivals and  $S_j$  analogously for firm  $j$ . To be supported as an equilibrium, by [Proposition 1](#), either firm would have to equalize  $mc$  in its interval. This would require the minimum of  $ac$  in  $i$ 's interval  $S_i$ ,  $ac_{i,\min}$ , to exceed the (equalized) maximum of  $mc_j$  in  $j$ 's interval  $S_j$  to keep  $j$  out of  $i$ 's interval, by [Proposition 1](#). But  $ac_{i,\min} > mc_j$  implies, by  $mc_j > ac$  in  $S_j$ , that  $ac_{i,\min} > ac(t)$  for all  $t \in S_j$ , so that  $mc_i > ac(t)$  holds for all  $t \in S_j$ . This cannot occur as a best-response for firm  $i$ , as firm  $i$  would then find it profitable to invade  $j$ 's interval  $S_j$ .  $\square$

In other words, [Lemma 1](#) shows that in equilibrium there cannot be two periods for which it is true that only one firm schedules arrivals in the one period, and only one other firm schedules arrivals in the other period. If firms' arrival windows are of unequal size, then the period in which the firm with the smaller window schedules arrivals has to be a subset of the period in which the other firm schedules arrivals. Note that nothing in the proof of [Lemma 1](#) precludes the simultaneous presence of other firms than  $i$  and  $j$ :  $ac$  can include travel delays caused by third firms, on top of schedule delay costs.



**Lemma 2.** *In equilibrium there can be at most one early time interval and one late time interval in which only one operator schedules arrivals, and these intervals, if they exist, occur at both shoulder periods of the peak (i.e. earlier and later than a period around  $t^*$  during which also other operators schedule services). This can only be an equilibrium if the one firm scheduling arrivals in these shoulders has a larger output  $N$  than each of the other firms. More generally, for any pair of firms with unequal outputs  $N$ , the arrival window of the smaller firm is a subset of the arrival window of the larger firm, with  $t^*$  in that subset, and with the larger firm having a higher instantaneous arrival rate at any moment where both firms schedule arrivals.*

**Proof.** Consider a pair of firms  $i$  and  $j$ , with unequal outputs. We start by proving by contradiction that these firms cannot have perfectly overlapping arrival windows. Suppose that both firms schedule arrivals in the same window and denote it  $S_{ij} = [t_q, t_e]$ . By Proposition 1,  $f_i(t_q) = f_i(t_e) = 0 = f_j(t_q) = f_j(t_e)$  must hold and therefore  $ac(t_q) = ac(t_e)$  also holds. This implies that to support it as an equilibrium, both firms must have equal firm-internal marginal cost defined by the average cost at the borders:  $mc_i(t \in S_{ij}) = mc_j(t \in S_{ij}) = ac(t_q) = ac(t_e)$ . To see that this cannot be consistent with unequal outputs, observe that equalization of firm-internal marginal cost implies  $c_{SD} + c_T(f_i + f_j + f_{-i-j}) + f_i \cdot c'_T = c_{SD} + c_T(f_i + f_j + f_{-i-j}) + f_j \cdot c'_T \forall t \in S_{ij}$ , where  $f_{-i-j}$  represents the aggregate flow of all firms other than  $i$  and  $j$ . As it is obvious from the expression, this can only be an equilibrium if firms have exactly the same instantaneous arrival rate in the period  $S_{ij}$ , which is inconsistent with firms having unequal outputs.

Label the firm with the larger window  $i$ , and the interval where it schedules arrivals and  $j$  does not as  $S_i$ . By Lemma 1 there is also an interval where both firms have arrivals, and denote that interval as  $S$ . For this to occur in equilibrium, by Proposition 1, the firm-internal marginal cost for firm  $i$   $mc_i$  must be equal throughout (and between)  $S_i$  and  $S$ . Also, for firm  $j$  the firm-internal marginal cost has to be constant in  $S$ , while it should be higher elsewhere including in  $S_i$ . We now prove that the only pattern of arrivals that can be an equilibrium is as described in Lemma 2.

The first step is to show that unequal arrival windows are possible only if firm  $i$  has a higher firm-internal marginal cost than firm  $j$ , and a higher arrival rate at any moment in  $S$ . The reasoning is as in the proof of Lemma 1: the average cost at any time in  $S_i$ ,  $ac(t \in S_i)$  must be higher than firm  $j$ 's equalized  $mc_j$  in  $S$ . As  $mc_i > ac$  holds by definition, then it must be true that  $mc_i > mc_j$ . Therefore  $mc_i(t \in S_i) = mc_i(t \in S) > mc_j(t \in S) > ac(t \in S)$  holds, and a period in which only firm  $i$  schedules arrivals is feasible in equilibrium. The next part is to prove that firm  $i$  has to be larger in terms of output than firm  $j$ . As the firm-internal marginal cost in  $S$  has to be larger for firm  $i$  than for firm  $j$  to support it as an equilibrium, i.e.:  $c_{SD} + c_T(f_i + f_j + f_{-i-j}) + f_i \cdot c'_T > c_{SD} + c_T(f_i + f_j + f_{-i-j}) + f_j \cdot c'_T \forall t \in S_j$ , it follows that  $f_i > f_j$  in  $S_j$ . As  $f_j = 0$  in  $S_i$ , then it is easily established that  $N_i > N_j$ .

The second step is to show that arrivals are scheduled around  $t^*$ . First consider the firm with the largest output of all, which by Lemma 1 defines the window for the entire equilibrium peak and schedules arrivals throughout that period. Denote that firm as  $i$  and denote the timing of its earliest arrival  $t_{qi}$  and its latest arrival  $t_{ei}$ . By Proposition 1,  $f_i(t_{qi}) = f_i(t_{ei}) = 0$  and  $ac(t_{qi}) = ac(t_{ei})$ . This implies that the schedule delay cost of the earliest and latest arrival of firm  $i$  has to be the same. As the schedule delay cost of an arrival decreases as the arrival is closer to  $t^*$ , this can only be true if the earliest arrival is in the early period and the latest in late period, i.e.  $t_{qi} < t^* < t_{ei}$ . As at times earlier than  $t_{qi}$  and later than  $t_{ei}$  the average cost is higher than  $mc_i$  in  $S_i$  and therewith definitely higher than  $mc_j$  in  $S_j$  for all  $j$ , it follows that  $S_j$  is a subset of  $[t_{qi}, t_{ei}]$ .

The third step is to show that the period of joint arrivals,  $S$ , of this largest firm paired with any other firm is around  $t^*$ , and  $S_i$  consists of an early shoulder and a late shoulder compared to the joint interval  $S$ . Denote the timing of firm  $j$ 's earliest arrival  $t_{qj}$  and its latest arrival  $t_{ej}$ . Also by Proposition 1,  $f_j(t_{qj}) = f_j(t_{ej}) = 0$  and  $ac(t_{qj}) = ac(t_{ej})$ , which combined with the fact that  $ac(t_{qi}) = ac(t_{ei}) = mc_i(t \in S_j) > ac(t \in S_j)$  holds, implies that  $t_{qi} < t_{qj}$  and that  $t_{ej} < t_{ei}$ .

Next, we show that also the period  $S_j$  is around  $t^*$ . Suppose that  $t_{qj}$  and  $t_{ej}$  are either both early or both late. As  $f_j$  is zero at  $t_{qj}$  and at  $t_{ej}$ , the average cost can be equal at those times only if the aggregate arrival rate of the other firms  $f_{-j}$  is different at both times. Because the schedule delay cost is different, the travel delay has to compensate this difference and, as it only depends on the firms' aggregate rate, these must be different. On the other hand, the firm-internal marginal cost for firm  $i$  at those times is  $[ac(t) + f_i(t) \cdot c'_T(f_i(t))]$ . Given that  $ac$  is equal at  $t_{qj}$  and at  $t_{ej}$ ,  $mc_i$  cannot be equal unless  $f_i$  is equal at both times. So the only possibility to explain equal average cost at  $t_{qj}$  and  $t_{ej}$  that is consistent with an equilibrium is that there is another firm,  $k$ , with  $f_k(t_{qj}) \neq f_k(t_{ej})$ . But, again, this is not consistent with equilibrium as if they are both positive, they must be equal to equalize firm-internal marginal cost and if one is zero, average cost would need to be different to be consistent with equilibrium for firm  $k$ , which leads to the contradiction. Thus,  $t_{qj} < t^* < t_{ej}$  must hold and the period in which multiple firms schedule arrivals is around  $t^*$  for each firm.

The final step is to show that firm  $i$  schedules arrivals at all times in  $[t_{qi}, t_{qj}]$  and in  $[t_{ej}, t_{ei}]$  when  $j$  has the one-but-largest arrival interval, so that  $S_i$  for the largest firm is effectively the early and late shoulders of the peak with only one firm ( $i$ ) active. As the firm-internal marginal cost in  $S_i$  must be equal to the schedule delay cost of the earliest and latest arrivals, the arrival rate of firm  $i$  has to be positive in  $[t_{qi}, t_{qj}]$  and in  $[t_{ej}, t_{ei}]$ . This is because the firm-internal marginal cost in  $S_i$  is  $c_{SD} + c_T(f_i) + f_i \cdot c'_T$ . As  $c_{SD}$  is lower in  $(t_{qi}, t_{qj})$  and in  $(t_{ej}, t_{ei})$  than at the borders  $t_{qi}$  and  $t_{ei}$ , then the only possibility for equalized firm-internal marginal cost, is  $f_i > 0$ . This completes the proof.  $\square$

We now turn to the existence and uniqueness of equilibrium for fully symmetric firms, a first main result of this section.

**Proposition 2.** *If firms are fully symmetric, a unique and stable pure-strategy Nash equilibrium (PSNE) exists. This unique equilibrium has all firms scheduling arrivals symmetrically in the peak center, which is a period around  $t^*$  with equal average cost at the beginning and end.*

**Proof.** By Proposition 1, firms must equalize the firm-internal marginal cost in the period with arrivals and, as they are symmetric in output, by Lemmas 1 and 2, this can only be possible if all firms schedule arrivals in a perfectly overlapping period. Denote this period  $S$ , the time of earliest arrival  $t_q$  and of the latest arrival  $t_e$ .

The firm-internal marginal costs for any firm  $i$  is, omitting the time argument,  $mc_i = c_{SD} + c_T(f_i + f_{-i}) + f_i \cdot c'_T$ . If  $mc$  is larger for one firm, then the same must be true for that firm's arrival rate  $f$  at every moment with positive arrivals, which contradicts symmetry of firms. Thus  $mc$  must be equal for all firms, and this can only happen if  $f_i = f_j$  for all  $i \neq j$  and for all  $t \in S$ . By Proposition 1, each firm's arrival rate is zero at the borders and the equalized marginal cost has to be equal to the schedule delay cost at the borders. That is,  $f_i(t_q) = f_i(t_e) = 0$  for all  $i$  and  $ac(t_q) = ac(t_e)$ . Moreover, the average cost is higher elsewhere (earlier than  $t_q$  and later than  $t_e$ ) so there is no profitable deviation, by Proposition 1. The next step is to show that the solution is unique.

Let  $C$  be the equilibrium firm-internal marginal cost, which satisfies  $C = c_{SD}(t_q) = c_{SD}(t_e)$ . Using symmetry in the first-order conditions, the equilibrium arrival rate of each firm,  $f$ , is found solving the following equation:

$$mc = c_{SD} + c_T(n \cdot f) + f \cdot c'_T = C \quad (14)$$

with  $n$  denoting the number of firms. For any given  $C$ , there is a unique arrival rate  $f$  that solves the equation. This is because  $mc(f=0) < C$  for all  $t$  in the interior of  $S$ , because the schedule delay cost is lower than  $C$  by construction. Then, as  $mc$  is a strictly increasing function of  $f$  because  $c_T$  is strictly increasing in  $f$ , convex, and  $c_{SD}$  does not depend on  $f$ , the solution is unique.

The last step to show uniqueness of the equilibrium is to show that for a given output  $N_i$  of each firm,  $C_i$  is unique. Because  $mc_i$  in (13a) is strictly increasing in  $f_i$ , the inverse function  $f_i(t; C_i)$  is strictly increasing in  $C_i$ , and is therewith single-valued for a given level of  $C_i$ . Because  $C_i$  is strictly decreasing in  $t_{qi}$  (provided smaller than  $t^*$ ) and strictly increasing in  $t_{ei}$  (provided larger than  $t^*$ ), we can define monotonous functions  $t_{qi}(C_i)$  and  $t_{ei}(C_i)$ , and next define a function  $N_i(C_i)$  as follows:

$$N_i(C_i) = \int_{t_{qi}(C_i)}^{t_{ei}(C_i)} f_i(t; C_i) dt \quad (15)$$

Because the integrand is non-negative and strictly increasing in  $C_i$ , the lower limit is strictly decreasing in  $C_i$ , and the upper limit strictly increasing, the function  $N_i(C_i)$  is strictly increasing in  $C_i$ . The inverse,  $C_i(N_i)$ , is therefore also strictly increasing, and therefore it gives the unique firm-internal marginal cost consistent with an output  $N_i$ .

Finally, we need to establish stability. Given that firm-internal marginal cost  $mc_i$  is increasing in the instantaneous arrival rate  $f_i$ , a downward perturbation of  $f_i$  requires an upward adjustment of  $f_i$  to return to the firm's equilibrium, and reversely for an upward perturbation. In other words, the unique equilibrium pattern of  $f_i$  that equalizes  $mc_i$  over time for a given  $N_i$  is stable. Because  $C_i$  increases in  $N_i$ , it is sufficient to have non-increasing marginal revenues, and therefore a non-increasing inverse demand, to also have uniqueness and stability of  $N_i$ .  $\square$

The closed-form solution of the equilibrium arrival rates is not possible to obtain at this stage, as we have not assumed any functional forms for the essential functions, but we come back to this later.

Next, Proposition 3 provides the second main result of this section, and generalizes Proposition 2 to the case of asymmetric firms in terms of output.

**Proposition 3.** *If firms differ only in terms of output and are otherwise equal, a unique and stable PSNE exists. In equilibrium, all firms schedule arrivals around the peak center, for each firm in a single closed interval that starts before  $t^*$  and ends after it. The larger a firm's equilibrium output, the earlier its interval starts and the later it ends. A larger firm schedules more vehicles in the jointly used interval than a smaller firm.*

**Proof.** Lemma 2 and its proof already establish that for any pair of asymmetric firms with  $N_i > N_j$ , equilibrium will entail  $t_{qi} < t_{qj} < t^* < t_{ej} < t_{ei}$ ;  $mc_j < mc_i$ ; and that whenever both firms schedule arrivals,  $f_i > f_j$ . What remains to be proven is existence, uniqueness and stability of equilibrium. We do so in two steps. First, we establish that given the temporal structure of equilibrium established in Lemma 2, at every instant there is a unique and stable vector of instantaneous arrival rates. Then, we establish that there is a unique and stable equilibrium in terms of arrival windows and firm-internal marginal costs.

For the first step, we order the firms by decreasing size, so that  $i=1$  is the largest firm,  $i=2$  the one-but-largest, etc. Let  $I(t)$  be the number of firms that schedules arrivals at moment  $t$ . By Lemma 2,  $I(t)$  will thus increase in discrete steps when  $t$  gets closer to  $t^*$ . In this first part of the proof we take as given all arrival windows and hence  $I(t)$  for each  $t$ , as well as the firms' equilibrium values of firm-internal marginal costs,  $C_i$ . Uniqueness of equilibrium then requires that at any moment  $t$  at which there are arrivals, there is only one vector of instantaneous arrival flows for which it is true that:

$$\begin{aligned} mc_{i=1} &= c_{SD} + c_T(f_{i=1} + \dots + f_{i=I(t)}) + f_{i=1} \cdot c'_T = C_{i=1} \\ &\vdots \\ mc_{i=I(t)} &= c_{SD} + c_T(f_{i=1} + \dots + f_{i=I(t)}) + f_{i=I(t)} \cdot c'_T = C_{i=I(t)} \end{aligned} \quad (16)$$



To prove uniqueness we first show that for given values of firm-internal marginal costs,  $C_i$ , there is a unique aggregate instantaneous arrival flow at any instant that satisfies (16), and then that this implies that there is only one vector of firm-specific instantaneous arrival flows.

Denote  $F(t)$  the aggregate arrival flow at time  $t$ ,  $F(t) = f_{i=1} + \dots + f_{i=I(t)}$ . Summing the  $I(t)$  equations in (14), for any instant  $t$ , we obtain:

$$I(t) \cdot c_{SD}(t) + I(t) \cdot c_T(F(t)) + F(t) \cdot c'_T(F(t)) = C_{i=1} + \dots + C_{i=I(t)} \quad (17)$$

The firm-internal marginal costs are given at this point, so the right-hand side of Eq. (17) is constant. As the function in the left-hand side is a strictly increasing continuous function of  $F(t)$ , because  $c_T$  is continuous and strictly increasing in  $F(t)$ , convex, and  $c_{SD}$  does not depend on  $F(t)$ , the solution for  $F(t)$  is unique. To see that this implies uniqueness of the vector of instantaneous arrival flows that satisfy (16), take the equilibrium condition for any firm  $i$  with positive arrival rate at instant  $t$ :

$$c_{SD} + c_T(f_1 + \dots + f_{I(t)}) + f_i \cdot c'_T = C_i \quad (18)$$

Rewriting, we obtain the following expression for  $f_i$ :

$$f_i(t) = \frac{C_i - c_{SD} - c_T(F(t))}{c'_T(F(t))} \quad (19)$$

As the right-hand side is a strictly decreasing continuous function of the aggregate arrival flow, for every firm there is a unique instantaneous arrival rate consistent with system (16). In particular, if there were multiple solutions, the one with a lower value for  $f_i$  would require a higher value of  $F$ . But a higher value of  $F$  implies that all  $f_j$  become smaller, which is of course inconsistent with having a higher  $F$ .

Given the arrival time windows defined by the vectors  $t_q$  and  $t_e$ , and given the assumed equilibrium values of firm-internal marginal costs  $C_i$ , unique time patterns of instantaneous flows are thus found. These are stable, because  $mc_i$  is strictly increasing in  $f_i$ , so that after a perturbation of  $f_i(t)$  firm  $i$  will find it profitable to return to the proposed equilibrium.

The second main step is to prove that given the behavior just displayed, a unique and stable pattern of arrival windows and firm-internal marginal cost levels arises. Lemma 2 did most of the job, and already establishes that given the arrival pattern of other firms, there is a unique and stable cost-minimizing arrival pattern and implied arrival window for firm  $i$  for which the aggregate flow adds up to  $N_j$ . Firm-internal marginal cost are equalized over used arrival moments, and smaller firms' arrival windows are contained in firm  $i$ 's arrival window, which itself is contained in larger firm's windows. Because any perturbation to the equilibrium can be interpreted as a perturbation of instantaneous flow, to establish stability it is sufficient to see that because a firm's firm-internal marginal cost is more sensitive with respect to own flow than to other firms' flows, other firms' reactions to a perturbation cannot be so strong that a return to the equilibrium for the perturbed firm would be discouraged.  $\square$

#### 2.4. Towards closed-form and numerical solutions for specific functional forms

We now turn to more specific closed-form analytical and numerical solutions. Many of these already occur for the simple case with two firms, and much of further discussion will consider that situation. We have already shown that with two firms  $i$  and  $j$ , either their arrival intervals perfectly overlap, or we have an equilibrium at which from some  $t_{qi}$  onwards, first firm  $i$  (the label is assigned without loss of generality) is the only to schedule arrivals; then from some  $t_{qj}$  onwards, both firms schedule arrivals up until some  $t_{ej}$ ; followed by a final period lasting until  $t_{ei}$  in which again only firm  $i$  schedules arrivals. In the periods where firm  $i$  operates alone, the equilibrium arrival pattern matches the socially optimal rates of change as given in Eq. (8) for the specific model, since the firm internalizes all congestion. Formally, this means that we have for the general model:

$$m\dot{c}_i(t) = \dot{c}_{SD}(t) + 2 \cdot \dot{f}_i(t) \cdot c'_T(\cdot) + f_i(t) \cdot \dot{f}_i(t) \cdot c''_T(\cdot) = 0 \quad \forall t : t_{qi} \leq t < t_{qj} \wedge t_{ej} < t \leq t_{ei} \quad (20)$$

where a dot denotes a time derivative and a (double) prime for  $c_T$  represents the first (second) derivative with respect to the aggregate arrival rate:  $\partial c_T / \partial f$  ( $\partial^2 c_T / \partial f^2$ ). For the specific model, this translates into arrival rates:

$$f_i(t) = \begin{cases} K \cdot \left( \frac{1}{1+\chi} \cdot \frac{\beta}{\alpha} \cdot (t - t_{qi}) \right)^{\frac{1}{\chi}} & \forall t : t_{qi} \leq t < t_{qj} \leq 0 \\ K \cdot \left( \frac{1}{1+\chi} \cdot \frac{\gamma}{\alpha} \cdot (t_{ei} - t) \right)^{\frac{1}{\chi}} & \forall t : 0 \leq t_{ej} < t \leq t_{ei} \end{cases} \quad (21)$$

It is, of course, the period in which the firms are both present for which the determination of equilibrium arrival rates is the most challenging. The Nash equilibrium in arrival patterns follows as the solution of a system of two differential equations, which for the general case read:

$$\begin{aligned} m\dot{c}_i(t) &= \dot{c}_{SD}(t) + \dot{f}(t) \cdot c'_T(\cdot) + \dot{f}_i(t) \cdot c'_T(\cdot) + f_i(t) \cdot \dot{f}(t) \cdot c''_T(\cdot) = 0 \quad \forall t : t_{qj} \leq t \leq t_{ej} \\ m\dot{c}_j(t) &= \dot{c}_{SD}(t) + \dot{f}(t) \cdot c'_T(\cdot) + \dot{f}_j(t) \cdot c'_T(\cdot) + f_j(t) \cdot \dot{f}(t) \cdot c''_T(\cdot) = 0 \quad \forall t : t_{qj} \leq t \leq t_{ej} \end{aligned} \quad (22)$$

where we introduce  $f(t) \equiv f_i(t) + f_j(t)$  as shorthand for the aggregate arrival rate in the two-firm setting.

Although there is no general closed-form solution for the system of equations in (22), the fact that the first two terms in the middle expressions are the same allows us to write down the following necessary condition for equilibrium:

$$\dot{f}_i(t) \cdot c_T'(\cdot) + f_i(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) = \dot{f}_j(t) \cdot c_T'(\cdot) + f_j(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) \quad \forall t : t_{qj} \leq t \leq t_{ej} \quad (23)$$

This means that, when  $c_T'(\cdot)$  and  $c_T''(\cdot)$  are both unequal to zero, the firm with the larger flow has a smaller time-derivative (in absolute terms) of its flow. For the cases where  $f_j$  starts and ends at zero at moments that  $f_i$  is positive, it will then display a steeper growth or decline, approaching  $f_i$  only asymptotically from below as  $t$  approaches  $t^* = 0$ .

Since for the BPR function there is a convenient expression for the ratio of the second and first derivative of  $c_T$ :

$$\frac{c_T''(\cdot)}{c_T'(\cdot)} = \frac{\chi - 1}{f} \quad (24)$$

we can rework (22), for the BPR function, into:

$$\dot{f}_j(t) = \dot{f}_i(t) + \frac{f_i(t) - f_j(t)}{f_i(t) + f_j(t)} \cdot (\dot{f}_i(t) + \dot{f}_j(t)) \cdot (\chi - 1) \quad \forall t : t_{qj} \leq t \leq t_{ej} \quad (25)$$

Eq. (25) shows that for a linear travel time function ( $\chi = 1$ ), the growth rates will be equal. Quite intuitively, also when the flows are equally large – and the second term on the right-hand side vanishes – the growth rates will be (and remain) equal. Furthermore, for  $\chi \neq 1$ , we can still relate the two growth rates at the moments  $t_{qj}$  and  $t_{ej}$  where firm  $j$  starts and stops operations as follows:

$$\dot{f}_i(t) = \dot{f}_j(t) \cdot \frac{2 - \chi}{\chi} \quad \text{for } t = \{t_{qj}, t_{ej}\} \quad (26)$$

This shows that with a sufficiently curved travel time function for which  $\chi > 2$ , the arrival rate of firm  $i$  will fall when firm  $j$  commences operations, and will rise when firm  $j$  is close to termination.

Unfortunately, the system of equations in (22) does not produce manageable closed-form solutions for the general case, and also not for the specific model in its most general form where  $\chi$  is left as an undetermined parameter and firms possibly differ in size. When making further assumptions, however, we can provide further insight into the properties of the solution. In Section 3, we will show that for a linear travel time function ( $\chi = 1$ ), we can provide a closed-form solution for the equilibrium independent of whether firms are symmetric or asymmetric in size. We will also illustrate that solution numerically. Next, in Section 4 we will show that also for  $\chi > 1$ , we can still obtain a numerical solution, even though an analytical solution is outside reach. And finally, in Section 5, we will show that for firms that are symmetric in size, we can still find an analytical closed-form solution, even when we leave  $\chi$  as an undetermined parameter.

As discussed previously, in the periods where only one firm operates alone, it internalizes all congestion. Therefore the Nash equilibrium will have inefficient scheduling only in the period where multiple firms schedule arrivals. An inspection of the firm-internal marginal cost in equation (13a) reveals that only one term is lacking to make it match the marginal social cost, and that is the instantaneous marginal external congestion cost imposed on the other firms and their customers – fully matching the insights from earlier static analysis (e.g. Brueckner, 2002). Therefore, by charging firm  $i$   $\tau_i(t) = f_j(t) \cdot c_T'(\cdot)$  during the period where firm  $j$  also schedules vehicles, firm  $i$ 's firm-internal marginal cost plus the toll it faces becomes equal to the marginal social cost (assuming firm  $i$  treats the toll as parametric; see also Brueckner and Verhoef, 2010). By charging the analogous expression to firm  $j$ , the sum of firm-internal marginal cost plus the toll faced for each firm becomes equal to the marginal social cost, and its equalization over time, the optimal strategy of each firm, thus decentralizes the social optimum. This toll is, at every instant, the fraction of the instantaneous marginal congestion cost that is not internalized. It can be interpreted as the application, at every instant  $t$ , of the standard congestion pricing result in static models of congestion with market power (e.g. Brueckner, 2002).

### 3. A first specific model: linear travel delay function ( $\chi = 1$ ) with two possibly asymmetric firms

#### 3.1. Analytical solution

It is instructive to start our exposition by giving the firm-internal marginal cost and its time derivative for the specific model (with  $\alpha\beta\gamma$ -preferences and a BPR congestion function):

$$\begin{aligned} mc_i(t) &= \alpha \cdot \left(\frac{f}{K}\right)^\chi + f_i \cdot \alpha \cdot \chi \cdot \frac{1}{K} \left(\frac{f}{K}\right)^{\chi-1} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \\ m\dot{c}_i(t) &= \alpha \cdot \chi \cdot \frac{1}{K} \cdot \left(\frac{f}{K}\right)^{\chi-1} \cdot (2 \cdot \dot{f}_i + \dot{f}_j) + f_i \cdot \alpha \cdot \chi \cdot (\chi - 1) \cdot \frac{1}{K^2} \left(\frac{f}{K}\right)^{\chi-2} \cdot (\dot{f}_i + \dot{f}_j) + \begin{cases} -\beta & \text{if } t \leq 0 \\ \gamma & \text{if } t > 0 \end{cases} \end{aligned} \quad (27)$$

The corresponding expressions for firm  $j$  are isomorphic. It is the fact that the sum of flows appears in terms raised to powers of  $(\chi - 1)$  and  $(\chi - 2)$  that prevent us from finding an analytical solution. This complication vanishes for the linear

model, since then we find:

$$mc_i(t) = \frac{\alpha}{K} \cdot (f) + f_i \cdot \frac{\alpha}{K} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad (28)$$

so that:

$$m\dot{c}_i(t) = 2 \cdot \frac{\alpha}{K} \cdot \dot{f}_i + \frac{\alpha}{K} \cdot \dot{f}_j + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad (29a)$$

and:

$$m\dot{c}_j(t) = 2 \cdot \frac{\alpha}{K} \cdot \dot{f}_j + \frac{\alpha}{K} \cdot \dot{f}_i + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad (29b)$$

The system of Eqs. (29a,b) can be solved to yield:

$$\begin{cases} \dot{f}_i = \frac{1}{2} \cdot \frac{\beta}{\alpha} \cdot K & \wedge \quad \dot{f}_j = 0 & \forall t : t_{qi} \leq t < t_{qj} \\ \dot{f}_i = \dot{f}_j = \frac{1}{3} \cdot \frac{\beta}{\alpha} \cdot K & & \forall t : t_{qj} \leq t < 0 \\ \dot{f}_i = \dot{f}_j = -\frac{1}{3} \cdot \frac{\gamma}{\alpha} \cdot K & & \forall t : 0 \leq t < t_{ej} \\ \dot{f}_i = -\frac{1}{2} \cdot \frac{\gamma}{\alpha} \cdot K & \wedge \quad \dot{f}_j = 0 & \forall t : t_{ej} \leq t < t_{ei} \end{cases} \quad (30)$$

The time-derivatives of (3) and (8) for the atomistic equilibrium and the social optimum become, for  $\chi = 1$ :

$$\begin{cases} \dot{f}_A = \frac{\beta}{\alpha} \cdot K & \forall t : t_{qA} \leq t < 0 \\ \dot{f}_O = \frac{1}{2} \cdot \frac{\beta}{\alpha} \cdot K & \forall t : t_{qO} \leq t < 0 \\ \dot{f}_A = -\frac{\gamma}{\alpha} \cdot K & \forall t : 0 \leq t < t_{eA} \\ \dot{f}_O = -\frac{1}{2} \cdot \frac{\gamma}{\alpha} \cdot K & \forall t : 0 \leq t < t_{eO} \end{cases} \quad (31)$$

It is easily checked that as long as firm  $i$  operates alone, the time derivative of aggregate arrivals equals that in the social optimum, reflecting that firm  $i$  internalizes all congestion externalities. When both firms are active, the slope of the aggregate arrivals is between that in the optimum and that in the atomistic equilibrium (2/3 is between 1/2 and 1).

After some algebra, we can derive the relevant starting and ending times of the two operators as a function of the two total quantities:

$$\begin{cases} t_{qi} = -\frac{\sqrt{2\alpha\gamma} \cdot \sqrt{2 \cdot N_i + N_j}}{\sqrt{\beta \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{qj} = -\frac{\sqrt{2\alpha\gamma} \cdot \sqrt{3 \cdot N_j}}{\sqrt{\beta \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{ej} = \frac{\sqrt{2\alpha\beta} \cdot \sqrt{3 \cdot N_j}}{\sqrt{\gamma \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{ei} = \frac{\sqrt{2\alpha\beta} \cdot \sqrt{2 \cdot N_i + N_j}}{\sqrt{\gamma \cdot (\beta + \gamma)} \cdot \sqrt{K}} \end{cases} \quad (32)$$

We may compare also these times to those in the atomistic equilibrium and in the social optimum for the same aggregate number of travelers, as we can derive them from (5), (6), (9), and (10):

$$\begin{cases} t_{qO} = -\frac{\sqrt{2\alpha\gamma} \cdot \sqrt{2 \cdot N_i + 2 \cdot N_j}}{\sqrt{\beta \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{qA} = -\frac{\sqrt{2\alpha\gamma} \cdot \sqrt{N_i + N_j}}{\sqrt{\beta \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{eA} = \frac{\sqrt{2\alpha\beta} \cdot \sqrt{N_i + N_j}}{\sqrt{\gamma \cdot (\beta + \gamma)} \cdot \sqrt{K}} \\ t_{eO} = \frac{\sqrt{2\alpha\beta} \cdot \sqrt{2 \cdot N_i + 2 \cdot N_j}}{\sqrt{\gamma \cdot (\beta + \gamma)} \cdot \sqrt{K}} \end{cases} \quad (33)$$

Quite intuitively, especially after what we found for the flows, the peak with two operators starts between the moments that would be observed in the optimum and the atomistic equilibrium; and the same is true for the ending.

Finally, we can derive the equilibrium levels of the firm-internal marginal cost where, not surprisingly, we find a higher value for  $mc_i$  than for  $mc_j$  unless the two quantities  $N_i$  and  $N_j$  are equal:

$$\begin{cases} mc_i = \frac{\sqrt{\alpha\beta\gamma} \cdot (2 \cdot \sqrt{2 \cdot N_i + N_j})}{\sqrt{\beta + \gamma} \cdot \sqrt{2K}} \\ mc_j = \frac{\sqrt{\alpha\beta\gamma} \cdot (\sqrt{3 \cdot N_j} + \sqrt{2 \cdot N_i + N_j})}{\sqrt{\beta + \gamma} \cdot \sqrt{2K}} \end{cases} \quad (34)$$

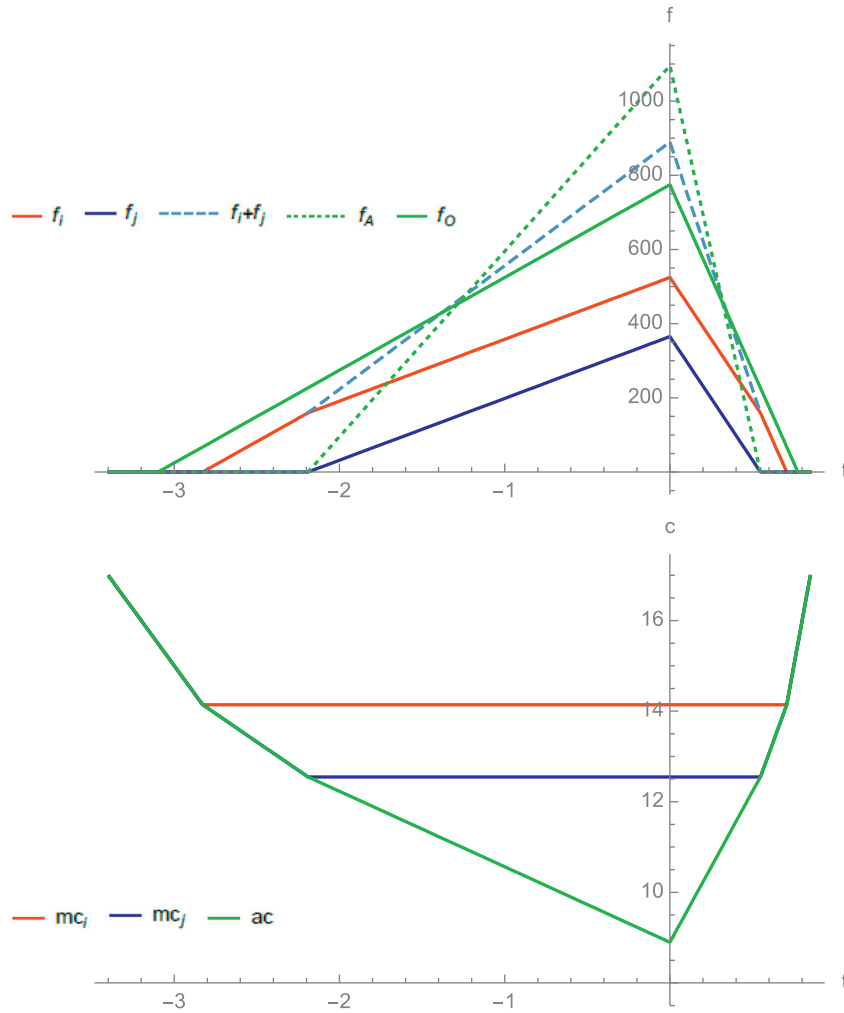


Fig. 1. Arrival flows (upper panel) and firm-internal marginal costs (lower panel) with  $\chi = 1$ .

### 3.2. A numerical illustration

To illustrate the equilibrium outlined above, we consider the specific model with  $\alpha\beta\gamma$ -preferences and the BPR congestion technology with  $\chi = 1$ . We set the parameters as follows:  $\alpha = 10$ ;  $\beta = 5$ ;  $\gamma = 20$ ;  $\chi = 1$ ;  $K = 1000$ ;  $N_i = 1000$ ;  $N_j = 500$ . The upper panel of Fig. 1 shows the equilibrium in terms of arrival rates, and contrasts these with the atomistic equilibrium and the optimum for the same total number of travelers. The lower panel shows the firm-internal marginal cost, as well as the average cost (note that the former overlap with the latter when the firm does not provide services).

Fig. 1 confirms that there is an equilibrium: the firm-internal  $mc$ 's are constant as long as the firm schedules arrivals and are higher (and equal to  $ac$ ) otherwise. The upper panel furthermore confirms the above results that the peak with two operator starts and ends between the moments that apply in the atomistic equilibrium and the peak, and that the slope of aggregate arrivals is also between the slopes in these same two benchmarks.

A visual inspection of the upper panel suggests that the arrival pattern with two operators is already pretty close to the socially optimal pattern, even though the two firms only internalize self-imposed congestion. This is confirmed by the aggregate generalized cost levels, which amount to 16 432 in the atomistic equilibrium, 15 625 with two operators, and 15 492 in the social optimum. This implies that, compared to the atomistic equilibrium, the self-internalization by two firms brings already 86% of the efficiency gain that a move to the optimum would bring. This relative efficiency is large compared to what studies have found using static models of congestion because we are ignoring the losses from monopoly pricing and considering only the inefficiencies due to the timing of departures for a given set of quantities. If the quantity-setting stage is included the relative efficiency would be smaller. Also, from Proposition 3, we obtain that in the peak shoulders the larger firm schedules vehicles efficiently. In those period the firm acts as a monopoly, so the negative externality is fully internalized. Due to this result, the relative efficiency of the timing of the vehicles increases as the share of total flights by

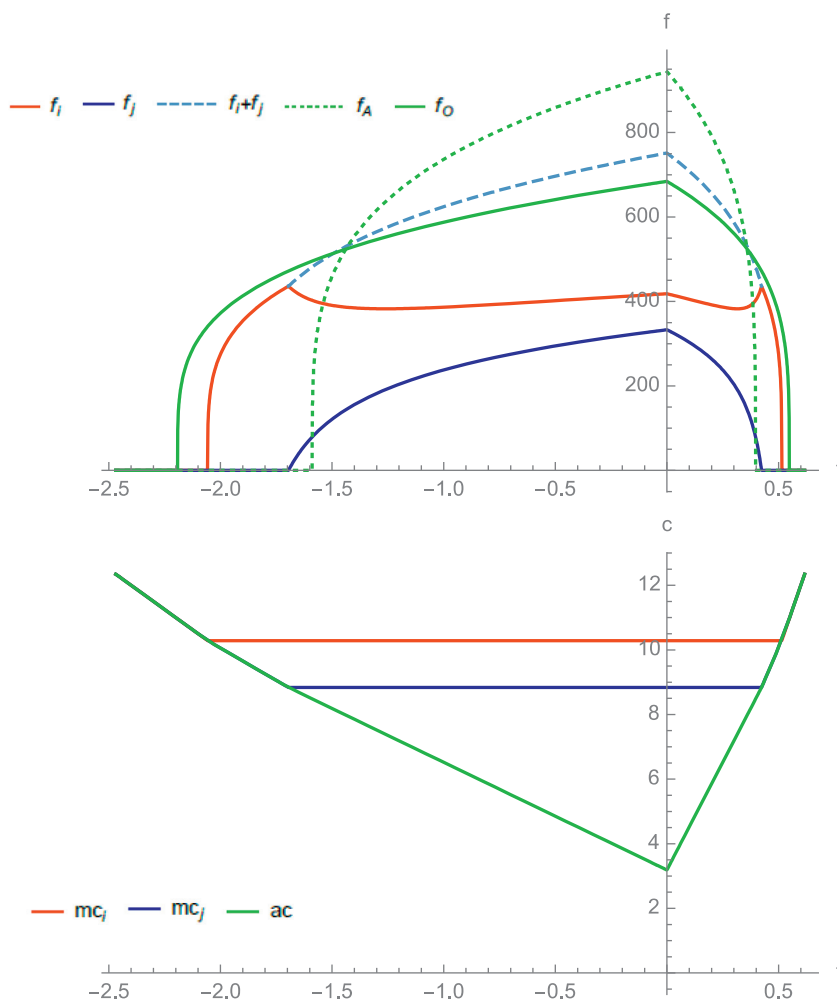


Fig. 2. Arrival flows (upper panel) and firm-internal marginal costs (lower panel) with  $\chi = 4$ .

the larger firm increases because it approaches the monopoly solution. Of course, if consumer surplus is included, then the efficiency would most likely decrease when the monopoly solution is approached as overpricing depresses demand.

#### 4. A second specific model: non-linear travel delay function ( $\chi = 4$ ) with possibly asymmetric firms

When the travel delay function is non-linear, as is the case for the conventional power of  $\chi = 4$  for the BPR function, no analytical closed-form solutions seem to exist, due to the appearance of the sum of flows in terms raised to powers of  $(\chi - 1)$  and  $(\chi - 2)$  in the firm-internal marginal cost and especially its time derivative in (27). Still, we succeed to find a numerical solution, and we will briefly present it here. Apart from  $\chi$  which is set at 4, the parameters remain the same compared to the linear model of Section 3.2. Fig. 2 shows the results.

Again, the lower panel confirms that equilibrium is reached. Note, in the upper panel, that indeed  $f_i$  drops right after  $t_{qj}$  and rises right before  $t_{ej}$ , as predicted by (26). Again, the aggregate arrival pattern with two operators appears to be relatively close to the optimal pattern, with the relative efficiency gain now being 95%. Again, this relative efficiency ignores the quantity setting stage and therefore the deadweight losses that would exist due to monopoly pricing. Finally, as it was true for the linear case, we find that the start and ending of the peak, as well as the rate of change of the arrival rate, is, for the case with two operators, between what is found for the atomistic equilibrium and what applies in the optimum.

#### 5. A third special case: a possibly non-linear travel delay function ( $\chi$ undetermined) with multiple symmetric firms

Another way to avoid the lack of closed-form analytical results is to impose symmetry in terms of also the size of the multiple firms. When firms are of equal size, the only possible equilibrium has all firms scheduling arrivals in a perfectly overlapping interval and with equal flows (i.e., fully symmetric). An equilibrium with one firm operating in the shoulders



cannot be supported because it requires a higher firm-internal marginal cost for the firm that has arrivals also in the shoulders. This can only be achieved if the arrival flow at every instant is higher for that firm in the period where all firms have arrival (see (22)), which is inconsistent with symmetry in firm size. In a perfectly overlapping period of arrivals, all firms have  $f(t)=0$  at the beginning and end of the peak (see the proof in Proposition 2), which implies that the firm-internal marginal cost for all firms is the same and equal to the average cost in the borders. Therefore, the solution has  $f_i(t)=f_S(t)$  for all  $i$ , and  $t_{qi}=t_{qS}$  and  $t_{ei}=t_{eS}$ , where subscript  $S$  stands for symmetric. The system of equations that allows for characterizing the equilibrium is equalization of firm-internal marginal cost for all firms over time, being equal also to the average cost at the beginning and end of the peak period:

$$\begin{aligned} mc_i &= c_{SD} + c_T(f_i + f_{-i}) + f_i \cdot c'_T = C \quad \forall i, \forall t \in [t_{qS}, t_{eS}] \\ C &= c_{SD}(t_{qS}) = c_{SD}(t_{eS}) \end{aligned} \quad (35)$$

Solving the system of equations for  $n$  firms and imposing symmetry yields:

$$f_S(t) = \begin{cases} \frac{K}{n} \cdot \left( \frac{1}{1 + \frac{1}{n} \cdot \chi} \cdot \frac{\beta}{\alpha} \cdot (t - t_{qS}) \right)^{\frac{1}{1+\chi}} & \forall t : t_{qS} \leq t \leq 0 \\ \frac{K}{n} \cdot \left( \frac{1}{1 + \frac{1}{n} \cdot \chi} \cdot \frac{\gamma}{\alpha} \cdot (t_{eS} - t) \right)^{\frac{1}{1+\chi}} & \forall t : 0 < t \leq t_{eS} \end{cases} \quad (36)$$

$$t_{qS} = -\left(1 + \frac{1}{n} \cdot \chi\right)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\beta} \quad (37)$$

$$t_{eS} = \left(1 + \frac{1}{n} \cdot \chi\right)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\gamma} \quad (38)$$

As expected, (36)–(38) reveal that the peak starts and ends between the moments applying in the atomistic case and in the optimum, whereas the growth rate of the aggregate arrival rate is again smaller than in the atomistic equilibrium in (3) (provided  $\chi > 1$ ), but larger than in the optimum in (8). Also as expected, (36)–(38) approach the social optimum as  $n$  approaches 1, and the atomistic equilibrium as  $n$  approaches infinity.

We can also solve for the total variable cost (TVC) of travel for the  $N$  users in this symmetric equilibrium ( $TVC_S$ ), and compare it to that in the atomistic equilibrium ( $TVC_A$ ) and that in the optimum ( $TVC_O$ ) (both derived by Chu, 1995):

$$\begin{aligned} TVC_A &= \Psi \cdot \alpha \cdot N \\ TVC_S &= \Psi \cdot \alpha \cdot N \cdot \frac{1}{1 + 2 \cdot \chi} \cdot \left( (1 + \chi) \cdot \left( \frac{n}{n + \chi} \right)^{\frac{\chi}{1+\chi}} + \chi \cdot \left( \frac{n + \chi}{n} \right)^{\frac{1}{1+\chi}} \right) \\ TVC_O &= \Psi \cdot \alpha \cdot N \cdot \frac{1}{1 + 2 \cdot \chi} \cdot (1 + \chi)^{\frac{2+\chi}{1+\chi}} \end{aligned} \quad (39)$$

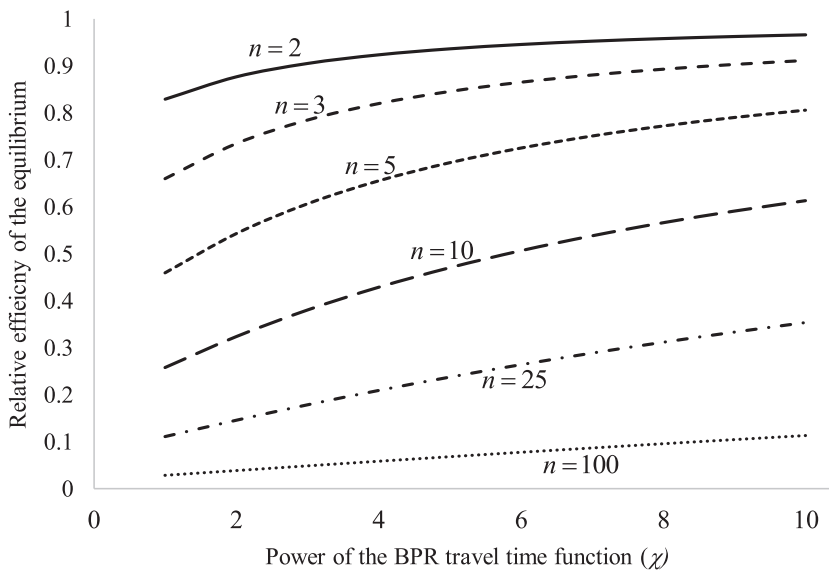


Fig. 3. The relative efficiency of self-internalization by multiple symmetric firms.

Although especially the second and third expression are not easily interpreted, it is clear from (39) that for a given  $n$ , the ratios of total variable cost are a function of  $\chi$  alone. We exploit this in Fig. 3 where we plot the relative efficiency of self-internalization by symmetric firms, defined as  $\omega_S = (TVCA - TVCS)/(TVCA - TVCO)$ , as a function of  $\chi$  alone and for different number of firms. The figure shows that as the curvature of the travel delay function becomes stronger, this relative efficiency increases. It can be easily shown that as  $\chi$  approaches infinity,  $\omega_S$  approaches 1.

Also as the number of firms increases, the relative efficiency of self-internalization decreases. For two firms, even at the lowest value of  $\chi = 1$  that we consider,  $\omega_S$  is already around 0.83.<sup>4</sup> This confirms our earlier numerical results in the sense that again, the relative efficiency with two firms is substantial, and it increases with  $\chi$ . As the number of firms grows, the relative efficiency decreases rapidly. For 5 firms, the lowest relative efficiency is 46% (when  $\chi = 1$ ) and it increases up to 80% when  $\chi = 10$  and for 10 firms both values decrease by 20 percentage points approximately.

## 6. Conclusion

We have investigated self-internalization of dynamic congestion by operators with market power. Earlier contributions using a bottleneck congestion technology found that no equilibrium may exist for Nash competitors. An important result of our analysis is that we do find a stable and unique Nash equilibrium when employing Henderson-Chu dynamic flow congestion. We also derive the first-best time-varying tolls that decentralize the social optimum.

Our results suggest that the relative efficiency of self-internalization may be rather high for a setting in which only few firms are present; above 83% in our numerical exercises for two firms. We also show that when more firms are present, and a larger share of congestion remains uninternalized, this relative efficiency drops. In the numerical example, for 10 firms the relative efficiency may thus be between 25% and 50%, depending on the curvature of the congestion function.

An important avenue for future research is to consider a second alternative congestion technology, in which interactions between travelers arriving at different instants are still present (unlike what is the case in the Henderson-Chu model), but the discontinuities of the pure bottleneck congestion technology are nevertheless avoided. This should allow us to identify better the cause of the non-existence of equilibrium in the bottleneck model; in particular whether this is an unavoidable consequence of having direct congestion interactions across arrival times, or whether it is due to some peculiarity of the bottleneck model. A possible and likely choice of congestion technology would be the model proposed by Agnew (1977), in which the speed for an individual user can vary during the trip as the instantaneous speed depends on the instantaneous total number of users present on the road, so that users who enter the facility at different instants nevertheless directly impose congestion externalities upon one another. Although Agnew's congestion technology is unrealistic for roads as this implies that vehicles can slow down if upstream density builds up, the model may be reasonable for, for example, computer or telecom networks, and it would be interesting to see whether a dynamic equilibrium exists for operators with market power who jointly use such a network. A downside of Agnew's model is its lack of analytical closed-form solutions. In the appendix of this paper, we present a simpler model with direct inter-temporal interactions between users, which is arguably the simplest possible variation of the model used above allowing for such interactions. What that model shows is that inclusion of direct inter-temporal interaction does not mean that equilibrium should necessarily break down. In other words, the non-existence of equilibrium in the Vickrey model is not exclusively due to the existence of direct inter-temporal interactions between users.

## Appendix. Constant duration model

This section briefly presents a model of dynamic congestion in which there is inter-temporal interaction between users. The purpose is to show that also with a congestion technology that allows for interactions between travellers who depart at different moments, equilibrium can exist. This, in turn, reveals that the complete lack of inter-temporal interaction in our model is not a decisive aspect for the existence of equilibrium.

The main features of the model are as follow: (i) users visit the congestible facility for an exogenously fixed amount of time,  $T$ , and choose the time of entry; (ii) the per-unit-of-time valuation of time spent at the facility varies over time and is highest at  $t^*$ . We represent this as users incurring zero schedule delay costs at instant  $t^*$ , while at other instants the schedule delay cost increases linearly in the same way as it does in the Chu model (see Eq. (2)); and (iii) there is a negative consumption externality in that at any moment of presence in the facility,  $t'$ , there is a cost,  $c_T(t')$ , that depends both on  $K$ , the capacity of the facility, and on the amount of users present at the facility.<sup>5</sup> We assume that the negative externality has the same functional form as the travel time in the Chu model. Under these assumptions, the instantaneous generalized price incurred by a user visiting the facility mirrors that in our main model. Yet, the user experiences costs during the entire visit and the generalized cost of the visit will, therefore, be the integral of the instantaneous generalized costs over the relevant duration  $T$ . The resulting model is arguably the smallest possible variation of the model in the main text that

<sup>4</sup> This 83% may seem contradictory with the 86% reported in Section 3 for two firms and linear delays. The difference is that in this section we are considering symmetric firms and in Section 3 firms were asymmetric. Intuitively, the relative efficiency is larger in that case as there is a period, the peak shoulders, in which the larger firm schedules vehicles efficiently.

<sup>5</sup> Note that the duration of the visit is constant so that this negative externality cannot be interpreted as increased travel time. It can be discomfort due to crowding, decreased quality of the visit, etc.

allows for inter-temporal interactions between users, which is why we chose it. It could describe a congestible facility such as a telecom network on which users want to spend some pre-determined time, where the desirability of using it varies over the day.

Let  $t$  be the moment of entry to the facility, or the beginning of the visit (which therefore ends at  $t+T$ ). The average (per passenger) costs for firms  $i$  and  $j$  is then simply:

$$ac_i(t) = ac_j(t) = \int_t^{t+T} c_T(Q_i(y) + Q_j(y); K) + c_{SD}(y) dy \quad (40)$$

where  $Q_x(t)$  is the number of users of firm  $x$  present at the facility at time  $t$ , analogous to the instantaneous arrival rate  $f_x(t)$  of the main model, and it is defined by the firm's pattern of inflow rates  $r_x$  in the following way.<sup>6</sup>

$$Q_x(t) = \int_{t-T}^t r_x(y) dy \quad x = \{i, j\} \quad (41)$$

Also under these assumptions, both firms face the incentive to fully internalize firm-internal negative externalities, and to equalize over time their firm-internal marginal cost  $m_x$ :

$$m_x(t) = \int_t^{t+T} c_T(Q_i(y) + Q_j(y); K) + Q_x(y) \cdot c_T'(\cdot) + c_{SD}(y) dy = \int_t^{t+T} mc_x(y) dy \quad x = \{i, j\} \quad (42)$$

A simple inspection of the firm-internal marginal cost reveals that if a firm equalizes over time its instantaneous firm-internal marginal cost  $mc_x$  during the period where its presence is positive, its firm-internal marginal cost  $m_x$  will be constant over the period where its inflow is positive. Therefore, any inflow pattern  $r_x(\cdot)$  that implies a presence pattern  $Q_x(\cdot)$  that makes  $mc_x(\cdot)$  constant over time is a best-response to any inflow pattern of the other firm. As a consequence, the inflow patterns that produce the presence pattern in terms of  $Q_x(\cdot)$  that is equivalent to an equilibrium in terms of  $f_x(\cdot)$  in the model from the main text, form an equilibrium also for the present model. However, as a firm controls the inflow over a period that is  $T$  units of time shorter than the period where it has presence on the facility (also the last user to enter stays in the facility  $T$  units of time), it is not straightforward that this candidate equilibrium presence pattern in terms of  $Q_x(\cdot)$  can be replicated by an inflow rate pattern in terms of  $r_x(\cdot)$  over a shorter period of time. Nevertheless, we study a simple case where this can be achieved and, therefore, we show that a dynamic equilibrium at a congestible facility under market power can exist also when there is inter-temporal interaction between users.

For this purpose, we turn to the symmetric-two-firms model of Section 5 with a linear travel delay function ( $\chi = 1$ ). The inflow rate pattern  $r_s$  that replicates the presence pattern  $f_s$  defined in Eq. (36) in the period  $[t_{qS}; t_{eS}]$  will ensure that the firm-internal marginal cost  $m_x$  is constant over  $[t_{qS}; t_{eS}-T]$ , which is the period where firms have users entering the facility. That inflow rate pattern will also make  $m_x$  to be higher outside  $[t_{qS}; t_{eS}-T]$ , as the instantaneous firm-internal marginal cost  $mc_x$  is, by construction of the equilibrium in the original model, higher outside  $[t_{qS}; t_{eS}]$ . Therefore, the firms will have a positive inflow rate  $r_s$  only in  $[t_{qS}; t_{eS}-T]$  and it will be an equilibrium if the presence pattern of the original model is replicated. Assume that  $T$  is such that both  $t_{qS}$  and  $t_{eS}$  are multiples of  $T$  (again  $t^*$  is normalized to zero) and let  $\eta_e = t_{qS}/T$  and  $\eta_l = t_{eS}/T$ . The candidate equilibrium inflow rate can be obtained by differentiating Eq. (35):

$$\dot{Q}_s(t) = r_s(t) - r_s(t-T) \Rightarrow r_s(t) = \dot{Q}_s(t) + r_s(t-T) \quad (43)$$

where  $\dot{Q}_s(t)$  is the time-derivative of the equilibrium presence pattern of the original model, and it is thus obtained by taking the derivative of  $f_s$  in Eq. (36) with respect to time. Using that  $\chi = 1$ , we get:

$$\dot{Q}_s(t) = \begin{cases} \frac{K}{3} \cdot \frac{\beta}{\alpha} & \forall t : t_{qS} \leq t \leq 0 \\ -\frac{K}{3} \cdot \frac{\gamma}{\alpha} & \forall t : 0 < t \leq t_{eS} \end{cases} \quad (44)$$

As the inflow rate should be zero at earlier moments than  $t_{qS}$ ,  $r_s(t-T)$  is zero for times  $t$  in  $[t_{qS}-T; t_{qS}]$  and the inflow rate in that period is equal to  $\dot{Q}_s(t)$ . For later moments, the inflow rate can be determined recursively by using Eq. (43). The

<sup>6</sup> As every visit lasts  $T$ , the rate of exit at time  $t$  is simply given by the entry  $T$  units of time earlier,  $r(t-T)$ .

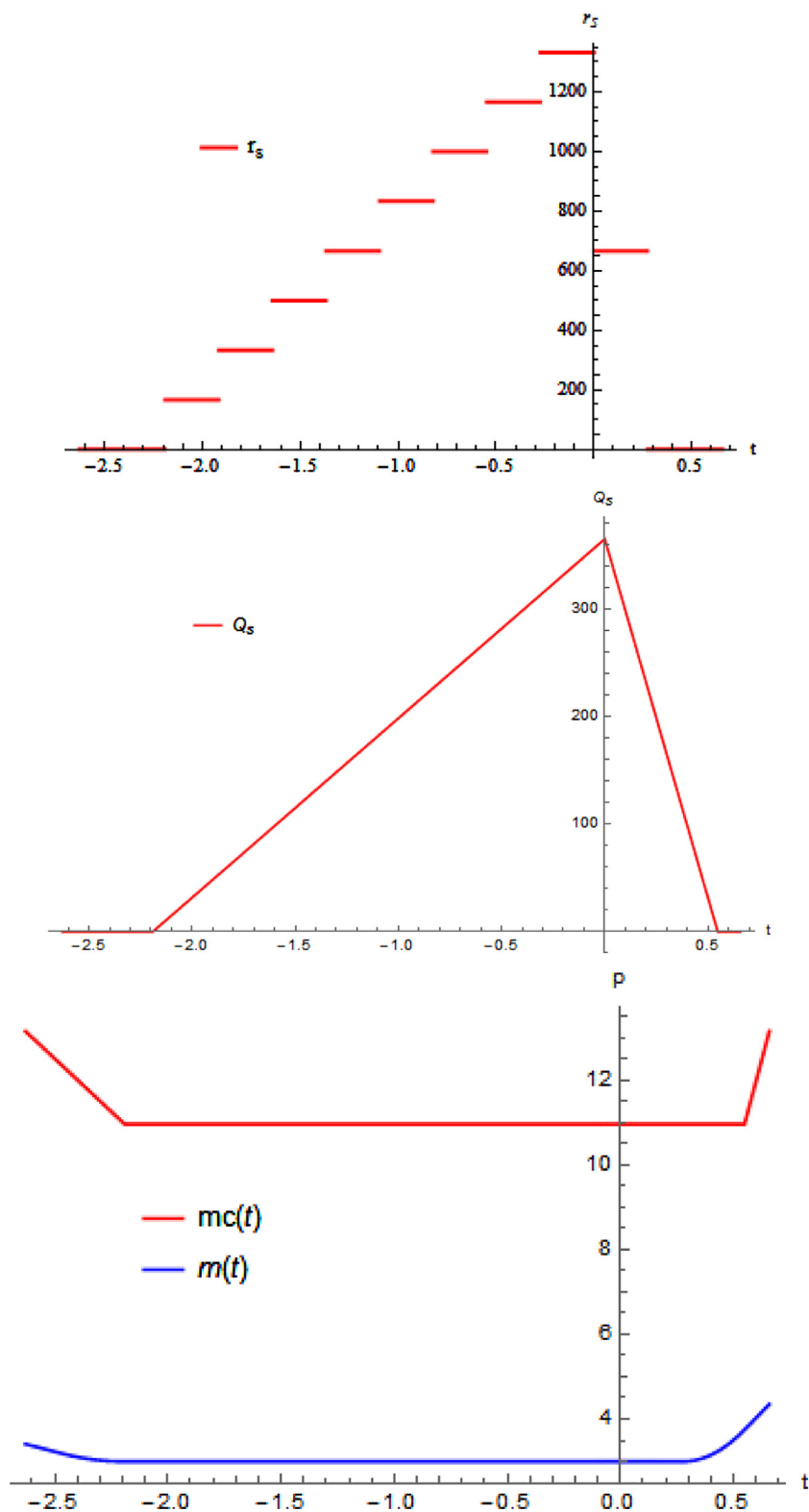


Fig. 4. Inflow rate (upper panel), implied presence pattern (middle panel) and firm-internal marginal costs (lower panel) with  $\chi = 1$ .

resulting candidate equilibrium inflow rate is the following piece-wise function:

$$r_s(t) = \begin{cases} \frac{K}{3} \cdot \frac{\beta}{\alpha} & \forall t : t_{qS} \leq t \leq t_{qS} + T \\ \frac{2K}{3} \cdot \frac{\beta}{\alpha} & \forall t : t_{qS} + T \leq t \leq t_{qS} + 2T \\ \vdots \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} & \forall t : t_{qS} + (\eta_e - 1)T \leq t \leq t_{qS} + \eta_e T = 0 \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{K}{3} \cdot \frac{\gamma}{\alpha} & \forall t : 0 < t \leq T \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{2K}{3} \cdot \frac{\gamma}{\alpha} & \forall t : T < t \leq 2T \\ \vdots \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{(\eta_l - 1)K}{3} \cdot \frac{\gamma}{\alpha} & \forall t : (\eta_l - 2)T < t \leq (\eta_e - 1)T = t_{eS} - T \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{\eta_l \cdot K}{3} \cdot \frac{\gamma}{\alpha} & \forall t : (\eta_l - 1)T < t \leq \eta_e T = t_{eS} \end{cases} \quad (45)$$

Note that to replicate the equilibrium presence pattern of the original model, the inflow rate is uniquely determined in the period where it is supposed to be positive  $[t_{qS}; t_{eS}-T]$ , but also for the period  $[t_{eS}-T; t_{eS}]$ , where it has to be zero to satisfy the equilibrium conditions. If it were positive in that last interval of duration  $T$ , there would be users present at the facility after  $t_{eS}$ , which is inconsistent with the equalization of firm-internal marginal cost. Nonetheless, in this case, we can show that the rate in the possibly conflicting interval  $[t_{eS}-T; t_{eS}]$  is zero and, therefore, the rate in Eq. (45) is an equilibrium. This is because  $t_{qS}$  and  $t_{eS}$  are such that the (instantaneous) schedule delay cost is the same, so that  $-\beta \cdot t_{qS} = \gamma \cdot t_{eS}$  holds. As  $\eta_e = t_{qS}/T$  and  $\eta_l = t_{eS}/T$  hold by definition,  $t_{qS}/t_{eS} = -\gamma/\beta = \eta_e/\eta_l$  also holds, implying  $\eta_e \cdot \beta - \eta_l \cdot \gamma = 0$ . Therefore, the inflow rate in Eq. (45), which replicates the equilibrium presence pattern is positive in  $[t_{qS}; t_{eS}-T]$  and zero elsewhere.

To illustrate the equilibrium described above, we use a numerical example with the following parameters:  $\alpha = 10$ ;  $\beta = 5$ ;  $\gamma = 20$ ;  $\chi = 1$ ;  $K = 1000$ ;  $N_i = 1825.74$ ;  $N_j = 1825.74$ ;  $T = 0.273861$ . The parameters  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\chi$  and  $K$  are the same as in the numerical illustration in Section 3.2,  $N_i$  and  $N_j$  are chosen to replicate the presence pattern of the small firm (firm  $j$ ) in the numerical illustration of Section 3.2 and  $T$  is chosen such that  $\eta_e$  and  $\eta_l$  are integers. The upper panel of Fig. 4 shows the equilibrium in terms of inflow rates. The middle panel shows the implied (equilibrium) presence pattern, which is the same as the inflow rate of firm  $j$  in Fig. 1. The lower panel shows the firm-internal marginal cost ( $m$ ) for entries, as well as the firm-internal instantaneous marginal cost ( $mc$ ), which again replicates the firm-internal marginal cost of the original model for firm  $j$  in Fig. 1.

Fig. 4 confirms that the  $r_s$  in Eq. (45) is an equilibrium: the firm-internal  $m$  is constant for the period where it has a positive inflow and higher otherwise.

Note that the requirement that  $t_{qS}$  and  $t_{eS}$  are multiples of  $T$  is highly restrictive. It is needed to make the equilibrium in terms of presence in this model isomorphic to an equilibrium in terms of arrival rates as discussed in the main body of this paper. That does not mean that if the condition is not fulfilled, the model would have no equilibrium in terms of inflows and presence levels; it does mean that the instantaneous firm-internal marginal costs as a function of presence is no longer constant over time. The firm-internal marginal costs as a function of inflows will be constant, as long as inflows occur. Deriving this equilibrium, which appears to be a problem to be addressed numerically, is considered to be outside the scope of this paper and is postponed to future work. Also the case with possibly asymmetric firms and possibly non-linear delays is under study, but outside the scope of this paper.

## References

- Agnew, C.E., 1977. The theory of congestion tolls. *J. Reg. Sci.* 17 (3), 381–393.
- Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *Am. Econ. Rev.* 83 (1), 161–179.
- Astarita, A., 1996. A continuous time link model for dynamic network loading based on travel time functions. In: Lesort, J.B. (Ed.), *Transportation and Traffic Theory*. Pergamon, Oxford, pp. 79–102.
- Basso, L.J., Zhang, A., 2007. Congestible facility rivalry in vertical structures. *J. Urban Econ.* 61 (2), 218–237.
- Basso, L.J., Zhang, A., 2008. Sequential peak-load pricing: the case of airports and airlines. *Can. J. Econ.* 41 (3), 1087–1119.
- Brueckner, J.K., 2002. Airport congestion when carriers have market power. *Am. Econ. Rev.* 92 (5), 1357–1375.
- Brueckner, J.K., van Dender, K., 2008. Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate. *J. Urban Econ.* 64 (2), 288–295.
- Brueckner, J.K., Verhoef, E.T., 2010. Manipulable congestion tolls. *J. Urban Econ.* 67 (3), 315–321.
- Chow, A.H.F., 2009a. Properties of dynamic system optimal assignment and its solution method. *Transport. Res. Part B* 43 (3), 325–344.
- Chow, A.H.F., 2009b. Dynamic system optimal traffic assignment – a state-dependent control theoretic approach. *Transportmetrica* 5 (2), 85–106.
- Czerny, Achiml., Zhang, A., 2011. Airport congestion pricing and passenger types. *Transport. Res. Part B* 45 (3), 595–604.
- Czerny, Achiml., Zhang, A., 2014a. Airport congestion pricing when airlines price discriminate. *Transport. Res. Part B* 65, 77–89.
- Czerny, Achiml., Zhang, A., 2014b. Airport peak-load pricing revisited: the case of peak and uniform tolls. *Econ. Transport.* 3 (1), 90–101.
- Chu, X., 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. *J. Urban Econ.* 37 (3), 324–343.
- Daniel, J.L., 1995. Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues. *Econometrica* 63 (2), 327–370.
- Daniel, J.L., (2009). The deterministic bottleneck model with non-atomistic traffic. Working Papers 09-08, University of Delaware, Department of Economics.

- Daniel, J.I., Harback, K.T., 2008. (When) do hub airlines internalize their self-imposed congestion delays? *J. Urban Econ.* 63 (2), 583–612.
- Henderson, J.V., 1974. Road congestion: a reconsideration of pricing theory. *J. Urban Econ.* 1 (3), 346–365.
- Henderson, J.V., 1981. The economics of staggered work hours. *J. Urban Econ.* 9 (3), 349–364.
- Heydecker, B.G., Addison, J.D., 2005. Analysis of dynamic traffic equilibrium with departure time choice. *Transport. Sci.* 39 (1), 39–57.
- Lindsey, C.R., Verhoef E.T., 2008. Congestion modelling. In: Hensher D.A., Button K.J. (eds.) (2008) *Handbook of Transport Modelling*, Handbooks in Transport (Second Edition) Elsevier / Pergamon, Amsterdam, pp. 353–373.
- Mayer, C., Sinai, T., 2003. Network effects, congestion externalities, and air traffic delays: or why not all delays are evil. *Am. Econ. Rev.* 93 (4), 1194–1215.
- Pels, E., Verhoef, E.T., 2004. The economics of airport congestion pricing. *J. Urban Econ.* 55 (2), 257–277.
- Silva, H.E., Verhoef, E.T., 2013. Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges. *J. Pub. Econ.* 106, 1–13.
- Silva, H.E., Verhoef, E.T., van den Berg, V.A.C., 2014. Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion. *J. Urban Econ.* 80, 13–27.
- Silva, H.E., Lindsey, R., De Palma, A., van den Berg, V.A.C., 2017. On the existence and uniqueness of equilibrium in the bottleneck model with atomic users. *Transport. Sci.* 51 (3), 863–881.
- Verhoef, E.T., 2017a. Cost recovery of congested infrastructure under market power. *J. Urban Econ.* 101, 45–56.
- Verhoef, E.T., 2017b. Optimal Congestion Pricing with Diverging Long-Run and Short-Run Scheduling Preferences. Discussion paper TI 17-077/VIII, Tinbergen Institute, Amsterdam-Rotterdam.
- Vickrey, W.S., 1969. Congestion theory and transport investment. *Am. Econ. Rev.* 59 (2), 251–260.