

DYNAMIC EQUILIBRIUM AT A CONGESTIBLE FACILITY UNDER MARKET POWER*

Erik T. Verhoef** and Hugo E. Silva
Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Phone: +31-20-5986094
Fax: +31-20-5986004
Email: e.t.verhoef@vu.nl

This version: 19/05/2015

Key words: Congestion pricing, dynamic congestion, market power, internalization

JEL codes: R41, R48, D62

Abstract

Various contributions to the recent literature on congestion pricing have demonstrated that when services at a congestible facility are provided by operators with market power, the case in point often being a few airlines jointly using a congested airport, optimal congestion pricing rules deviate from the familiar Pigouvian rule that tolls be equal to the marginal external costs. The reason is that an operator with market power has an incentive to internalize the congestion effects that its customers and vehicles impose upon one-another, so that Pigouvian tolling would lead to overpricing of congestion. More recent contributions to this literature, however, have brought to the fore that when congestion at the facility takes on the form of dynamic bottleneck congestion à la Vickrey (1969), where trip scheduling is the key behavioural margin, there may exist no Nash equilibrium in arrival schedules for oligopolistic operators also under rather plausible assumptions on parameters. This paper investigates whether in such cases, an equilibrium does exist for another congestion technology, namely the Henderson-Chu dynamic model of flow congestion. We find that a stable and unique equilibrium exists also in cases where it fails to exist under bottleneck congestion (notably when the value of schedule late exceeds the value of travel delays). Our results suggest that self-internalization with only two firms leads to a considerable efficiency gain compared to the atomistic equilibrium (83% or more of the gain from first-best pricing in our numerical exercises).

* Financial support from ERC (AdG Grant #246969 OPTION) is gratefully acknowledged. We are also grateful to Robin Lindsey for helpful suggestions.

** Affiliated to the Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam.

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 would 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 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 that one would expect 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, as 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 other player keeps his fare fixed, his quantity will rise in response to a reduction in the player’s own fare. 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, and Verhoef (2012) studies the design of self-financing mechanisms for congestible facilities with market power, motivated by the observation that the lowering of tolls under Cournot behaviour, to levels below the marginal external cost, makes the well-known Mohring-Harwitz (1962) theorem on self-financing of optimally priced and dimensioned infrastructure break down.

More recently, the dynamics of congestion have been subject to analysis. Silva, Verhoef and Van den Berg (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, and developed in various direction later on by Arnott, De Palma and Lindsey (1993). While Silva, Verhoef and Van den Berg (2014) could use this model to describe monopoly and leader-

fringe cases, with the latter showing that the leader is forced to schedule arrivals according to atomistic patterns in the peak center, they also reported a rather discomfoting result, namely that the model seems to have no pure-strategies equilibrium in arrival schedules for the case of a Cournot oligopoly.¹

Silva, Lindsey, De Palma and Van den Berg (2014) confirmed the non-existence of equilibrium in the said setting for a duopoly, albeit that they demonstrate that it applies only when the so-called value of schedule delay late (γ) exceeds the value of travel delays (α). 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 in any sense, 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.

The proof is, not surprisingly, lengthy, but the main idea is that an equilibrium should be robust against both marginal and non-marginal changes that a player can make in his arrival schedule. The former condition implies that the firm-internal marginal cost be equal across the arrival interval(s) that the airline actually uses. Equilibrium breaks down because the only arrival pattern that secures this condition to be met for all players, is not robust against a unilateral non-marginal deviation in a player’s arrival schedule. In particular, starting from that equilibrium, the player finds it profitable to reschedule some of its flights. Keeping the other players’ departure schedules fixed, their arrivals will have been completed before the end of the original candidate peak equilibrium, giving the deviating player the opportunity to use the freed up interval for unqueued arrivals, which will lower this player’s cost compared to the candidate equilibrium.

Given the widespread prevalence of congestion on 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 probably not just an intellectual challenge to identify and possibly remedy the characteristic of the model that is responsible for the non-existence of equilibrium. It also seems to be a task with a clear societal relevance. 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 (β for early arrivals, γ for late ones), and the existence of a disutility of travel delays (with unit value α) – 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)

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

homogeneity, no equilibrium would exist. The same is true for uncertainty in travel times. Hence, the specific congestion technology seems a first candidate to consider.

In this paper, we build upon the framework set out in Silva, Verhoef and Van den Berg (2014), but replace the bottleneck congestion technology with a model of flow congestion, and focus on simultaneous scheduling by 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 more accurate description of dynamic congestion patterns is when the facility operator has the possibility of opening additional but lower quality capacity at busier times, such as more remote terminals or runways, which bring additional on-the-ground travel delays.

We investigate whether under this alternative congestion technology, there is an equilibrium in pure strategies in terms of arrival flows, and if it does, we assess its efficiency for the case of duopoly relative to that of two natural benchmarks: the purely atomistic equilibrium, and the social optimum, which 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 can be expected to 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) model²

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

² The exposition of the basic model in this sub-section 2.1 draws heavily from the one in Verhoef (2014). Literal citations are not marked as such and are taken to be acknowledged through this footnote.

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 α , and values of schedule delay of β for early arrivals and γ 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 χ determines the curvature of $T(\cdot)$. Note that the regular BPR function pre-multiplies the second term with $T_f b$ 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 $c_{SD}(t)$:

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

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 equation (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. To keep the analysis tractable, we will be considering the case of two operators (i and j). Our focus on the internalization of self-imposed congestion, and on the existence of a dynamic equilibrium, allows us to ignore firm interactions resulting from competition for the same passengers. We thus assume that the two operators serve different markets (which is why we will not refer to a duopoly).

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. 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 firms i and j :

$$ac_i(t) = ac_j(t) = c_T(f_i(t) + f_j(t); K) + c_{SD}(t) \quad (11a)$$

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:

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

(t^* is again set at 0).

Under these assumptions, both firms face the incentive to fully internalize firm-internal congestion externalities, just as in the static model Brueckner's (2002). Also, and as a result, both firms will find it optimal to equalize over time, as long as it schedules arrivals, its firm-internal marginal cost mc_x , defined as:

$$mc_x(t) = c_T(f_i(t) + f_j(t); K) + f_x(t) \cdot c_T'(\cdot) + c_{SD}(t) \quad x = \{i, j\} \quad (12a)$$

or, for the specific model:

$$mc_x(t) = \alpha \cdot \left(\frac{f_i(t) + f_j(t)}{K} \right)^\chi + f_x(t) \cdot \chi \cdot \frac{1}{K} \cdot \left(\frac{f_i(t) + f_j(t)}{K} \right)^{\chi-1} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \quad x = \{i, j\} \quad (12b)$$

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 x to equalize mc_x as long as the firm schedules arrivals, and to choose not to use particular time intervals when mc_x – then equal to ac – exceed mc_x in used time intervals. For most of the paper we will therefore focus attention on the question of whether given a pair of chosen quantities N_i and N_j , a dynamic equilibrium exists, and how its efficiency compares to that of the two benchmarks of the atomistic no-toll equilibrium and the social optimum, without modelling explicitly how the firms choose N_i and N_j by equating marginal revenue to marginal cost.

To determine whether a Nash equilibrium in arrival flows exists, and if so what it looks like, it is convenient to first establish a number of features that characterize an equilibrium.

F1: Equilibrium cannot entail two disjoint arrival intervals without any overlap between each firm's arrival periods

Proof This type of equilibrium 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. But $ac_{i,min} > mc_j$

implies $ac_{i,min} > ac_j(t)$ for all $t \in S_j$, so that we also have $mc_i > ac_j(t)$ for all $t \in S_j$. But then firm i would find it profitable to invade j 's interval S_j . ■

F2: 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 both operators schedule services)*

Proof Label the (larger) firm that is present in the two shoulder periods firm i , the shoulders where only i schedules arrivals S_i , and the interval where firm j (and hence both firms) has arrivals S_j . The reasoning is then as above, $ac_{i,min} > mc_j$ is needed so that firm j does not have an incentive to reschedule arrivals to S_i , which again implies $mc_i > ac_j(t \in S_j)$ in the interval S_j where also j schedules arrivals. This can only occur if firm i schedules arrivals in S_j . Then, $mc_i(t \in S_i) = mc_i(t \in S_j) > ac_j(t \in S_j)$ is still possible and both firms can have a constant mc_x over time when they schedule arrivals, so that they do not have any incentives to reschedule arrivals. Note that if firm i has an exclusive interval in the early shoulder, it also has one in the late shoulder. To see why, first assume that both firms have f equal to zero at their relevant t_q and t_e (proof follows below). Firm i 's arrival interval stretches from t_{qi} to t_{ei} , at which in both cases f_i has dropped to zero so that mc_i equals c_{SD} at both moments. Firm j faces the same $mc_j = c_{SD}$ at t_{qi} and t_{ei} , and because it sets $t_{qj} > t_{qi}$ it has $mc_j(t_{qj}) = ac(t_{qj}) < mc_i(t_{qj}) = ac(t_{qi})$. Firm j will therefore also not schedule arrivals at t_{ei} but instead terminates arrivals earlier at some t_{ej} for which $mc_j(t_{ej}) = ac(t_{ej}) < mc_i(t_{ej}) = ac(t_{ei})$.

The above reasoning rests on both firms having $f_x = 0$ at the relevant t_q and t_e , and this is the final part to be proven. For the large firm i , this is clearly profitable as otherwise $mc_i(t_{qi}-\varepsilon) = ac(t_{qi}-\varepsilon) < mc_i(t_{qi})$; and similarly $mc_i(t_{ei}+\varepsilon) = ac(t_{ei}+\varepsilon) < mc_i(t_{ei})$ (ε is a small positive constant). Intuitively, if firm i had $f_i > 0$ at t_{qi} and t_{ei} , it could reduce its costs by shifting the arrival to just before t_{qi} or just after t_{ei} , completely eliminating travel delays and only increasing the schedule delay costs infinitesimally. For firm j , a non-zero f_j at t_{qj} and t_{ej} can only be consistent with equilibrium for firm i if also the latter has a discontinuity in f_i at t_{qj} and t_{ej} . Specifically, f_i should jump down at t_{qj} and up at t_{ej} . Because firm i will make the jump such that mc_i is equalized around t_{qj} and around t_{ej} , and because firm i internalizes all congestion only outside S_j , the total flow right after t_{qj} will be higher than right before it; and similarly right before t_{ej} it will be higher than right after it. But then it must be true that $mc_j(t_{qj}-\varepsilon) = ac(t_{qj}-\varepsilon) < ac(t_{qj}) < mc_j(t_{qj})$, where the final inequality results from the assumed discontinuity in f_j at t_{qj} . Therefore, this cannot be an equilibrium for firm j . A similar reasoning of course applies around t_{ej} . ■

The two features, *F1* and *F2*, greatly narrow down the set of candidate Nash equilibria, and simplify the task of deriving them. Either the two firms' 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 equation (8) for the specific model, since the firm internalizes all congestion. Formally, this means that we have for the general model:

$$m\dot{c}_i(t) = \dot{c}_{SD}(t) + 2 \cdot \dot{f}_i(t) \cdot c_T'(\cdot) + f_i(t) \cdot \dot{f}_i(t) \cdot c_T''(\cdot) = 0 \quad \forall t: t_{qi} \leq t < t_{qj} \wedge t_{ej} < t \leq t_{ei} \quad (13)$$

where a dot denotes a time derivative and a (double) prime for c_T represent 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:

$$f_i(t) = \begin{cases} K \cdot \left(\frac{1}{1+\chi} \cdot \frac{\beta}{\alpha} \cdot (t - t_{qi}) \right)^{\frac{1}{\chi}} & \forall t: t_{qi} \leq t \leq 0 \\ K \cdot \left(\frac{1}{1+\chi} \cdot \frac{\gamma}{\alpha} \cdot (t_{ei} - t) \right)^{\frac{1}{\chi}} & \forall t: 0 < t \leq t_{ei} \end{cases} \quad (14)$$

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, if it exists, follows as the solution of a system of two differential equations, that for the general case read:

$$m\dot{c}_i(t) = \dot{c}_{SD}(t) + \dot{f}(t) \cdot c_T'(\cdot) + \dot{f}_i(t) \cdot c_T'(\cdot) + f_i(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) = 0 \quad \forall t: t_{qi} \leq t \leq t_{ej} \quad (15)$$

$$m\dot{c}_j(t) = \dot{c}_{SD}(t) + \dot{f}(t) \cdot c_T'(\cdot) + \dot{f}_j(t) \cdot c_T'(\cdot) + f_j(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) = 0 \quad \forall t: t_{qj} \leq t \leq t_{ej}$$

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

Although there is no general closed-form solution for the system of equations in (15), 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:

$$\dot{f}_i(t) \cdot c_T'(\cdot) + f_i(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) = \dot{f}_j(t) \cdot c_T'(\cdot) + f_j(t) \cdot \dot{f}(t) \cdot c_T''(\cdot) \quad \forall t: t_{qj} \leq t \leq t_{ej} \quad (16)$$

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 :

$$\frac{c_T''(\cdot)}{c_T'(\cdot)} = \frac{\chi - 1}{f} \quad (17)$$

we can rework (16), for the BPR function, into:

$$\dot{f}_j(t) = \dot{f}_i(t) + \frac{f_i(t) - f_j(t)}{f_i(t) + f_j(t)} \cdot (\dot{f}_i(t) + \dot{f}_j(t)) \cdot (\chi - 1) \quad \forall t: t_{qj} \leq t \leq t_{ej} \quad (18)$$

Equation (18) 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 $\chi > 1$, 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:

$$\dot{f}_i(t) = \dot{f}_j(t) \cdot \frac{\chi}{2 - \chi} \quad \text{for } t = \{t_{qj}, t_{ej}\} \quad (19)$$

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 (15) does not seem to produce manageable closed-form solutions for the general case, and also not for the specific model in its most general form where χ 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 χ as an undetermined parameter.

Before considering these cases, we address the important question of whether there will be a Nash-Cournot equilibrium in terms of departure time schedules for the most general version of the model. Because the firm sets a continuous time profile of $f_x(t)$, it is not straightforward to prove existence and uniqueness formally; for this version of the paper we only provide a sketch of proof. Consider the firm-internal marginal cost of (12a). It is, at every instant, strictly increasing in the firm's own arrival rate. This implies that when mc_x is equalized over S_x , and ac outside S_x exceeds the value of mc_x inside S_x , there is indeed only one value of $f_x(t)$ that equates $mc_x(t)$ to that equilibrium value of mc_x for any t in S_x , and none outside S_x . Moving any service, or set of services, to any other moment, be it inside or outside S_x , can only increase the firm-internal cost since it would imply an mc_x above the equilibrium level at moments to which additional schedules are shifted, and reduce mc_x below the equilibrium level at moments from which the shifted schedules were taken. Undoing the shift thus reduces the firm's total cost. Furthermore, (12a) implies that the two outputs at any given instant are strategic substitutes, where the firm-internal marginal cost is more responsive (in absolute terms) to the own output than to the competitor's output. This means that at any given instant t , there can be only one stable intersection of the two reaction functions (depicting for either firm the choice of $f_x(t)$ given the other firm's flow at that instant, and given the firm's instantaneous objective to equate $mc_x(t)$ to the equilibrium value of mc_x). And finally, a firm's arrival interval and therefore its firm-internal marginal cost is strictly increasing in its own output and in the other firm's output, securing that also with price-sensitive demands, there will be only one equilibrium in terms of total outputs when these would be price-sensitive.

As discussed previously, in the periods where only one firm operates alone, it internalizes all congestion. Therefore the Nash equilibrium, if it exists, will have inefficient scheduling only in the period where both firm schedule arrivals. An inspection of the firm-internal marginal cost in equation (12) 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 firm and its 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 Bruckner 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 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):

$$\begin{aligned}
 mc_i(t) &= \alpha \cdot \left(\frac{f_i + f_j}{K} \right)^\chi + f_i \cdot \alpha \cdot \chi \cdot \frac{1}{K} \left(\frac{f_i + f_j}{K} \right)^{\chi-1} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \\
 m\dot{c}_i(t) &= \alpha \cdot \chi \cdot \frac{1}{K} \cdot \left(\frac{f_i + f_j}{K} \right)^{\chi-1} \cdot (2 \cdot \dot{f}_i + \dot{f}_j) + \\
 &\quad f_i \cdot \alpha \cdot \chi \cdot (\chi-1) \cdot \frac{1}{K^2} \left(\frac{f_i + f_j}{K} \right)^{\chi-2} \cdot (\dot{f}_i + \dot{f}_j) + \begin{cases} -\beta & \text{if } t \leq 0 \\ \gamma & \text{if } t > 0 \end{cases}
 \end{aligned} \tag{20}$$

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:

$$mc_i(t) = \frac{\alpha}{K} \cdot (f_i + f_j) + f_i \cdot \frac{\alpha}{K} + \begin{cases} -\beta \cdot t & \text{if } t \leq 0 \\ \gamma \cdot t & \text{if } t > 0 \end{cases} \tag{21}$$

so that:

$$m\dot{c}_i(t) = 2 \cdot \frac{\alpha}{K} \cdot \dot{f}_i + \frac{\alpha}{K} \cdot \dot{f}_j + \begin{cases} -\beta & \text{if } t \leq 0 \\ \gamma & \text{if } t > 0 \end{cases} \tag{22a}$$

and:

$$m\dot{c}_j(t) = 2 \cdot \frac{\alpha}{K} \cdot \dot{f}_j + \frac{\alpha}{K} \cdot \dot{f}_i + \begin{cases} -\beta & \text{if } t \leq 0 \\ \gamma & \text{if } t > 0 \end{cases} \tag{22b}$$

The system of equations (22ab) can be solved to yield:

$$\left\{ \begin{array}{l} \dot{f}_i = \frac{1}{2} \cdot \frac{\beta}{\alpha} \cdot K \quad \wedge \quad \dot{f}_j = 0 \quad \forall t: t_{qi} \leq t < t_{qj} \\ \dot{f}_i = \dot{f}_j = \frac{1}{3} \cdot \frac{\beta}{\alpha} \cdot K \quad \forall t: t_{qj} \leq t < 0 \\ \dot{f}_i = \dot{f}_j = -\frac{1}{3} \cdot \frac{\gamma}{\alpha} \cdot K \quad \forall t: 0 \leq t < t_{ej} \\ \dot{f}_i = -\frac{1}{2} \cdot \frac{\gamma}{\alpha} \cdot K \quad \wedge \quad \dot{f}_j = 0 \quad \forall t: t_{ej} \leq t < t_{ei} \end{array} \right. \quad (23)$$

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

$$\left\{ \begin{array}{l} \dot{f}_A = \frac{\beta}{\alpha} \cdot K \quad \forall t: t_{qA} \leq t < 0 \\ \dot{f}_O = \frac{1}{2} \cdot \frac{\beta}{\alpha} \cdot K \quad \forall t: t_{qO} \leq t < 0 \\ \dot{f}_A = -\frac{\gamma}{\alpha} \cdot K \quad \forall t: 0 \leq t < t_{eA} \\ \dot{f}_O = -\frac{1}{2} \cdot \frac{\gamma}{\alpha} \cdot K \quad \forall t: 0 \leq t < t_{eO} \end{array} \right. \quad (23)$$

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:

$$\left\{ \begin{array}{l} t_{qi} = -\frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\gamma} \cdot \sqrt{2 \cdot N_i + N_j}}{\sqrt{\beta} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{qj} = -\frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\gamma} \cdot \sqrt{3 \cdot N_j}}{\sqrt{\beta} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{ej} = \frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{3 \cdot N_j}}{\sqrt{\gamma} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{ei} = \frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{2 \cdot N_i + N_j}}{\sqrt{\gamma} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \end{array} \right. \quad (24)$$

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

$$\left\{ \begin{array}{l} t_{qO} = -\frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\gamma} \cdot \sqrt{2 \cdot N_i + 2 \cdot N_j}}{\sqrt{\beta} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{qA} = -\frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\gamma} \cdot \sqrt{N_i + N_j}}{\sqrt{\beta} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{eA} = \frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{N_i + N_j}}{\sqrt{\gamma} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ t_{eO} = \frac{\sqrt{2} \cdot \sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{2 \cdot N_i + 2 \cdot N_j}}{\sqrt{\gamma} \cdot \sqrt{\beta + \gamma} \cdot \sqrt{K}} \end{array} \right. \quad (25)$$

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:

$$\left\{ \begin{array}{l} mc_i = \frac{\sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{\gamma} \cdot \frac{1}{\sqrt{2}} \cdot (2 \cdot \sqrt{2 \cdot N_i + N_j})}{\sqrt{\beta + \gamma} \cdot \sqrt{K}} \\ mc_j = \frac{\sqrt{\alpha} \cdot \sqrt{\beta} \cdot \sqrt{\gamma} \cdot \frac{1}{\sqrt{2}} \cdot (\sqrt{3 \cdot N_j} + \sqrt{2 \cdot N_i + N_j})}{\sqrt{\beta + \gamma} \cdot \sqrt{K}} \end{array} \right. \quad (26)$$

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 Figure 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 travellers. 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).

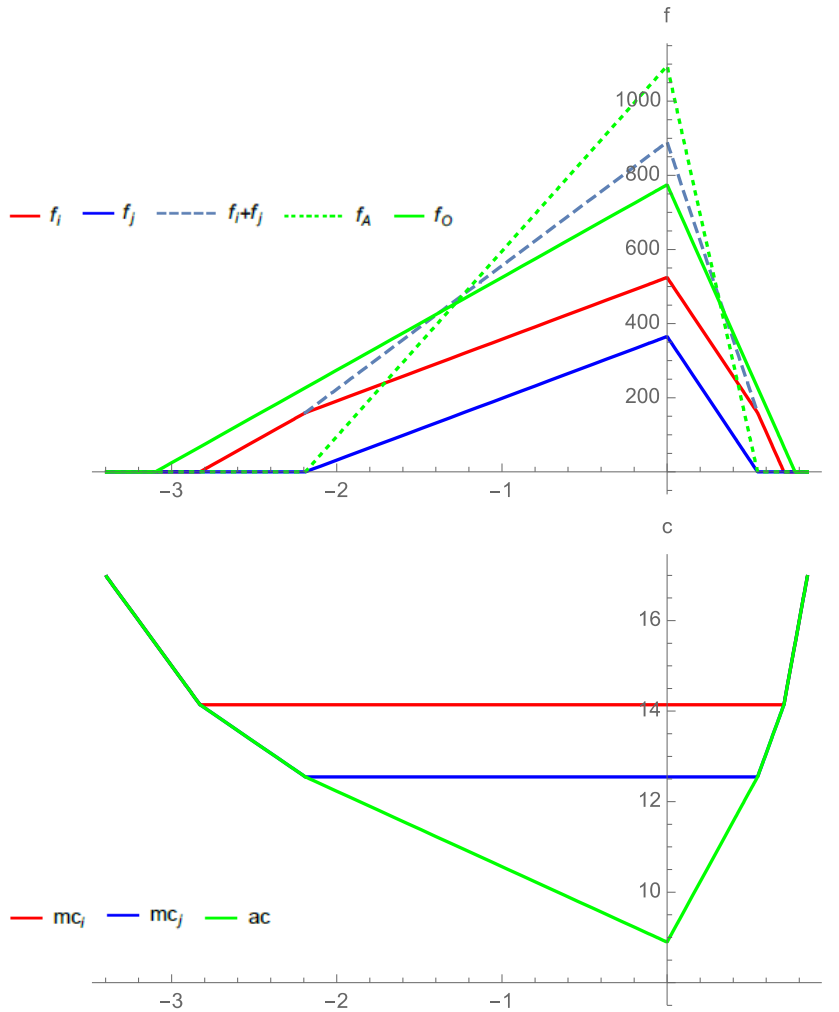


Figure 1. Arrival flows (upper panel) and firm-internal marginal costs (lower panel) with $\chi=1$

Figure 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.

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 (20). Still, we succeed to find a numerical solution, and we will briefly present it here. Apart from χ which is set at 4, the parameters remain the same compared to the linear model of Section 3.2. Figure 2 shows the results.

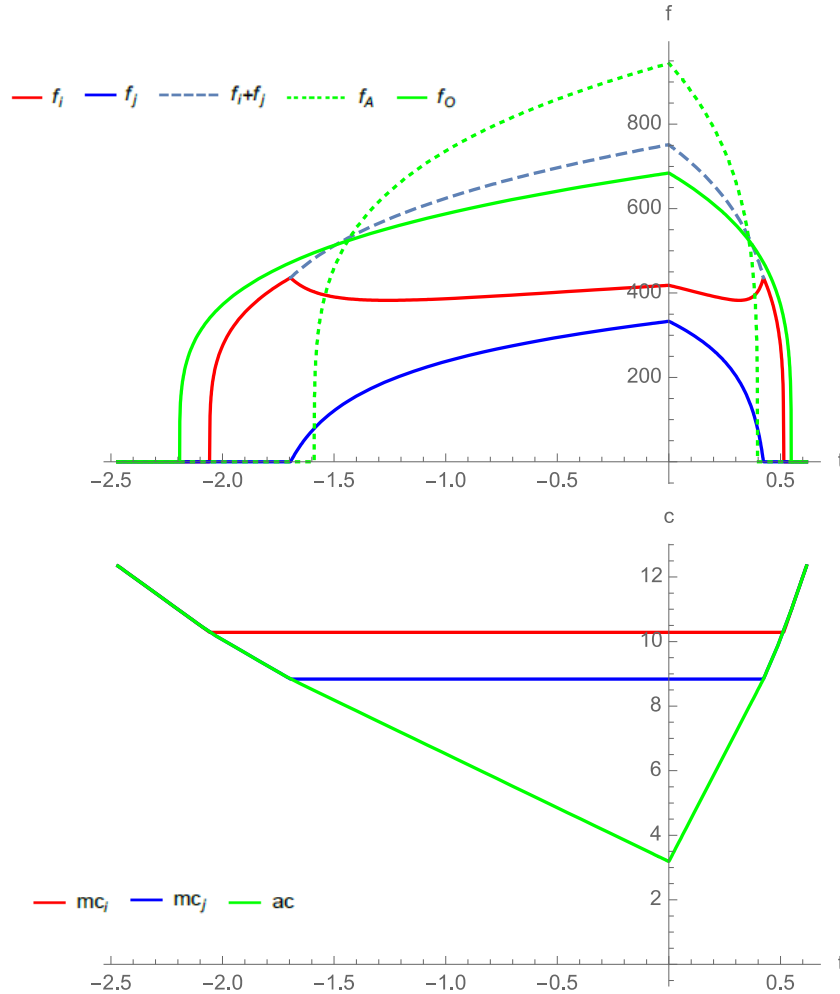


Figure 2. Arrival flows (upper panel) and firm-internal marginal costs (lower panel) with $\chi=4$

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 (19). 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%. And 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 atomic equilibrium and what applies in the optimum.

5. A third special case: a possibly non-linear travel delay function (χ undetermined) with 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 two firms. When firms are of equal size, the only possible equilibrium has both 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 both firms have arrival (see (15)), which is inconsistent with symmetry in firm size. In a perfectly overlapping period of arrivals, both firms have $f(t)=0$ at the beginning and end of the peak (see the proof in F2), which implies that the firm-internal marginal cost for both firms is the same and equal to the average cost in the borders. Therefore, the solution has $f_i(t)=f_j(t)=f_S(t)$, and $t_{qi}=t_{qj}=t_{qS}$ and $t_{ei}=t_{ej}=t_{eS}$, where subscript S stands for symmetric. Solving the system of equations yields:

$$f_S(t) = \begin{cases} \frac{K}{2} \cdot \left(\frac{1}{1+\frac{1}{2}\cdot\chi} \cdot \frac{\beta}{\alpha} \cdot (t-t_{qS}) \right)^{\frac{1}{\chi}} & \forall t: t_{qS} \leq t \leq 0 \\ \frac{K}{2} \cdot \left(\frac{1}{1+\frac{1}{2}\cdot\chi} \cdot \frac{\gamma}{\alpha} \cdot (t_{eS}-t) \right)^{\frac{1}{\chi}} & \forall t: 0 < t \leq t_{eS} \end{cases} \quad (27)$$

$$t_{qS} = -\left(1+\frac{1}{2}\cdot\chi\right)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\beta} \quad (28)$$

$$t_{eS} = \left(1+\frac{1}{2}\cdot\chi\right)^{\frac{1}{1+\chi}} \cdot \Psi \cdot \frac{\alpha}{\gamma} \quad (29)$$

As expected, (27)-(29) reveal that the peak again 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).

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

$$TVC_A = \Psi \cdot \alpha \cdot N$$

$$TVC_S = \Psi \cdot \alpha \cdot N \cdot \frac{1}{1+2\cdot\chi} \cdot \left((1+\chi) \cdot \left(\frac{2}{2+\chi} \right)^{\frac{\chi}{1+\chi}} + \chi \cdot \left(\frac{2+\chi}{2} \right)^{\frac{1}{1+\chi}} \right) \quad (30)$$

$$TVC_O = \Psi \cdot \alpha \cdot N \cdot \frac{1}{1+2\cdot\chi} \cdot (1+\chi)^{\frac{2+\chi}{1+\chi}}$$

Although especially the second and third expression are not easily interpreted, it is clear from (30) that the ratios of total variable cost are a function of χ alone. We exploit this in Figure 3 where we plot the relative efficiency of self-internalization by two symmetric firms, defined as $\omega_S = (TVCA - TVCS)/(TVCA - TVCO)$, as a function of χ alone. The figure shows that as

the curvature of the travel delay function becomes stronger, this relative efficiency increases, but that even at the lowest value of $\chi = 1$ that we consider, ω_S is already around 0.83. This confirms our earlier numerical results in the sense that again, the relative efficiency is substantial, and it increases with χ .

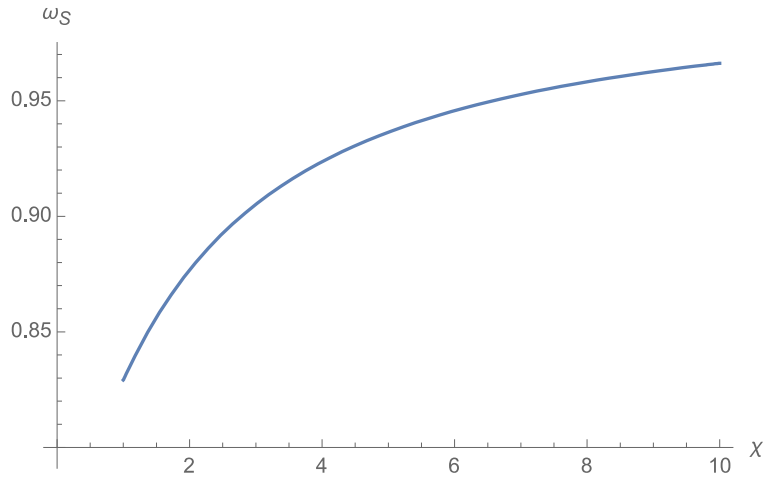


Figure 3. The relative efficiency of self-internalization by two symmetric firms

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 two firms are present; above 83% in our numerical exercises. Presumably, when more firms are present, and a larger share of congestion remains uninternalized, this relative efficiency will drop. This is one of the issues we will address in follow-up research.

Our first research effort, however, will be to consider a second alternative congestion technology, in which interactions between travellers 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. 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.

References

- Agnew, C.E. (1977) "The theory of congestion tolls" *Journal of Regional Science* **17** (3) 381-393.
- Arnott, R., A. de Palma and R. Lindsey (1993) "A structural model of peak-period congestion: a traffic bottleneck with elastic demand" *American Economic Review* **83** (1) 161-179.
- Basso, L. J. and A. Zhang (2007) "Congestible facility rivalry in vertical structures" *Journal of Urban Economics* **61** (2) 218-237.
- Brueckner, J.K. (2002) "Airport congestion when carriers have market power" *American Economic Review* **92** (5) 1357-1375.
- Brueckner, J.K. and K. van Dender (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 E.T. Verhoef (2010) "Manipulable congestion tolls" *Journal of Urban Economics* **67** (3) 315-321.
- Chu, X. (1995) "Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach" *Journal of Urban Economics* **37** (3) 324-343.
- Daniel, J.I. (1995) "Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues" *Econometrica* **63** (2) 327-370.
- 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 K.T. Harback (2008) "(When) do hub airlines internalize their self-imposed congestion delays?" *Journal of Urban Economics* **63** (2), 583-612.
- Henderson, J.V. (1974) "Road congestion: a reconsideration of pricing theory" *Journal of Urban Economics* **1** (3) 346-365.
- Henderson, J.V. (1981) "The economics of staggered work hours" *Journal of Urban Economics* **9** (3) 349-364.
- Lindsey, C.R. and E.T. Verhoef (2008) "Congestion modelling". In: D.A. Hensher and K.J. Button (eds.) (2008) *Handbook of Transport Modelling, Handbooks in Transport (Second Edition)* Elsevier / Pergamon, Amsterdam, pp. 353-373.
- Mayer, C. and T. Sinai (2003) "Network effects, congestion externalities, and air traffic delays: Or why not all delays are evil" *American Economic Review* **93** (4) 1194-1215.
- Mohring, H. and M. Harwitz (1962) "Highway Benefits: An Analytical Framework" Northwestern University Press, Evanston, IL
- Pels, E. and E.T. Verhoef (2004) "The economics of airport congestion pricing" *Journal of Urban Economics* **55** (2) 257-277.
- Silva, H.E. and E.T. Verhoef (2013) "Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges" *Journal of Public Economics* **106** 1-13.
- Silva, H.E., E.T. Verhoef and V.A.C. van den Berg (2014) "Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion" *Journal of Urban Economics* **80** 13-27.
- Silva, H.E., R. Lindsey, A. De Palma and V.A.C. van den Berg (2014) "On the existence and uniqueness of equilibrium in the bottleneck model with atomic users" TI Discussion Paper 14-077, Tinbergen Institute Amsterdam/Rotterdam.
- Verhoef, E.T. (2012) "Cost recovery of congested infrastructure under market power" TI Discussion Paper 12-064, Tinbergen Institute Amsterdam/Rotterdam.

- Verhoef, E.T. (2014) “Optimal congestion pricing with diverging long-run and short-run scheduling preferences” Unpublished manuscript, VU University Amsterdam.
- Vickrey, W.S. (1969) “Congestion theory and transport investment” *American Economic Review* **59** (2) 251-260.

Appendix A. 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.³ 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 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:

$$ac_i(t) = ac_j(t) = \int_t^{t+T} c_T(Q_i(y) + Q_j(y); K) + c_{SD}(y) dy \quad (31)$$

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:⁴

$$Q_x(t) = \int_{t-T}^t r_x(y) dy \quad x = \{i, j\} \quad (32)$$

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 :

³ 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.

⁴ As every visit lasts T , the rate of exit at time t is simply given by the entry T units of time earlier, $r(t-T)$.

$$m_x(t) = \int_t^{t+T} c_T(Q_i(y) + Q_j(y); K) + Q_x(y) \cdot c_T'(\cdot) + c_{SD}(y) dy = \int_t^{t+T} mc_x(y) dy \quad x = \{i, j\} \quad (33)$$

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-duopoly 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. (27) 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. (32):

$$\dot{Q}_s(t) = r_s(t) - r_s(t-T) \Rightarrow r_s(t) = \dot{Q}_s(t) + r_s(t-T) \quad (34)$$

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. (27) with respect to time. Using that $\chi=1$, we get:

$$\dot{Q}_s(t) = \begin{cases} \frac{K}{3} \cdot \frac{\beta}{\alpha} & \forall t: t_{qS} \leq t \leq 0 \\ -\frac{K}{3} \cdot \frac{\gamma}{\alpha} & \forall t: 0 < t \leq t_{eS} \end{cases} \quad (35)$$

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. (34). The resulting candidate equilibrium inflow rate is the following piece-wise function:

$$r_s(t) = \begin{cases} \frac{K}{3} \cdot \frac{\beta}{\alpha} & \forall t: t_{qS} \leq t \leq t_{qS} + T \\ \frac{2K}{3} \cdot \frac{\beta}{\alpha} & \forall t: t_{qS} + T \leq t \leq t_{qS} + 2T \\ \vdots & \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} & \forall t: t_{qS} + (\eta_e - 1)T \leq t \leq t_{qS} + \eta_e T = 0 \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{K}{3} \cdot \frac{\gamma}{\alpha} & \forall t: 0 < t \leq T \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{2K}{3} \cdot \frac{\gamma}{\alpha} & \forall t: T < t \leq 2T \\ \vdots & \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{(\eta_l - 1)K}{3} \cdot \frac{\gamma}{\alpha} & \forall t: (\eta_l - 2)T < t \leq (\eta_e - 1)T = t_{eS} - T \\ \frac{\eta_e \cdot K}{3} \cdot \frac{\beta}{\alpha} - \frac{\eta_l \cdot K}{3} \cdot \frac{\gamma}{\alpha} & \forall t: (\eta_l - 1)T < t \leq \eta_e T = t_{eS} \end{cases} \quad (36)$$

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. (36) 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. (36), 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 α , β , γ , χ 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 η_e and η_l are integers. The upper panel of Figure 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 Figure 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 Figure 1.

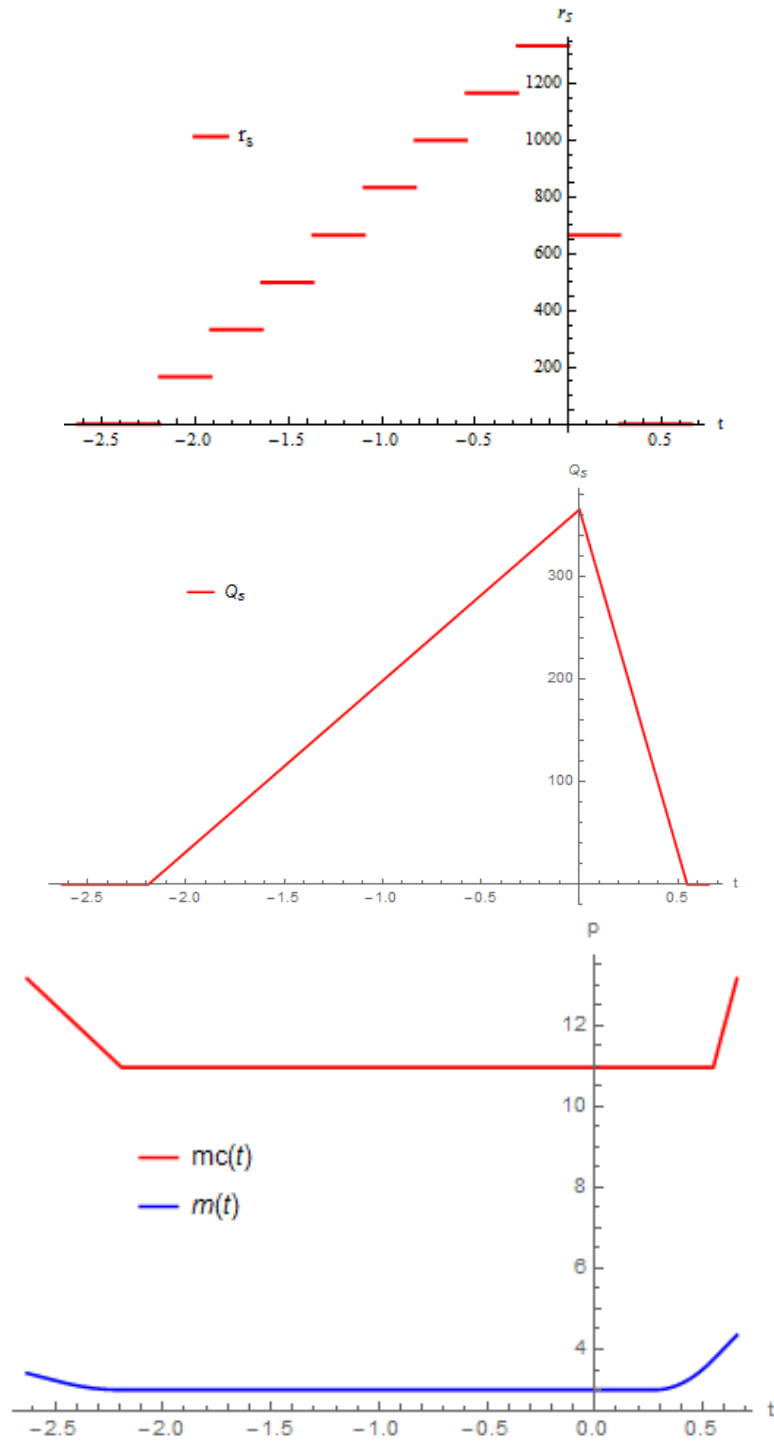


Figure 4. Inflow rate (upper panel), implied presence pattern (middle panel) and firm-internal marginal costs (lower panel) with $\chi=1$.

Figure 4 confirms that the r_s in equation (36) 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_{q_s} and t_{e_s} 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.