



Munich Personal RePEc Archive

Auctions for private congestible infrastructures

Vincent A.C. van den Berg

VU University Amsterdam

17. July 2012

Online at <http://mpra.ub.uni-muenchen.de/40103/>

MPRA Paper No. 40103, posted 17. July 2012 13:37 UTC

Auctions for private congestible infrastructures

Version of 17 July 2012

Vincent A.C. van den Berg
Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1081 HV, Amsterdam, The Netherlands
+31-20-598 6160
v.a.c.vanden.berg@vu.nl

Abstract

This paper investigates regulation by auctions of private supply of congestible infrastructures in two networks settings: 1) two serial facilities, where the consumer has to use both in order to consume; and 2) two parallel facilities that are imperfect substitutes. There are four market structures: a monopoly and 3 duopolies that differ in how firms interact. The effects of an auction depend on what the bidders compete. With a bid auction, the bidders compete on how much money they transfer to the government. This auction leads to the same outcome as the unregulated game (for a given market structure), since this gives the maximum profit to transfer. An auction on the capacity of a facility leads to an even lower welfare than no regulation, because firms set very high capacities and usage fees. Conversely, an auction on generalised price or number of users leads to the first-best outcome. Moreover, these two auctions are robust: they attain the first-best regardless of whether the facilities are auctioned off to a single firm or to two firms, and for all market and network structures. On the contrary, the performances (relative to the first-best) of the bid and capacity auctions strongly depend on these considerations.

Keywords: private supply, congestible facilities, auctions, serial facilities, parallel facilities, imperfect substitutes

JEL codes: D43, L13, L51, R41, R42

1. Introduction

There is a rising interest in the private supply of congestible facilities such as airports, telecommunication, roads, energy networks, public transport and waste disposal. Often named advantages of private supply are: higher efficiency than the government, stronger responsiveness to the preferences of users, and alleviating pressure on governmental budgets. But there are also disadvantages: the most important of which is that a private facility typically has substantial market power.

To limit these downsides, the government could use regulation by auctions for the right to operate a facility. The central question of this paper is: what are the effects of monopolistic or duopolistic supply of two serial or parallel facilities, and can the outcome be improved by auctions for franchises? With two serial facilities, a user has to use both facilities in order to consume. This setting is common. When flying, one first uses the origin and then the destination airport. With long-distance phone calls, there are often different operators at the origin and destination. When considering parallel facilities it is likely that they are imperfect substitutes: for example, the air or seaports of an area differ in their facilities or accessibility; when choosing between travelling by air, (high-speed) rail or car, the modes differ in flexibility, access time, travel time and comfort; and airlines are typically imperfect substitutes (if only due to frequent flyer programs).

To understand the effects of regulation, one first needs to know the outcome without regulation. Firms have two instruments: the usage fee and the capacity, which affects the usage cost for the users. I investigate four market structures: a monopoly and three duopolies that differ in how firms interact. With the first “open-loop” duopoly, a firm takes the actions of the other firm as given. With the second “closed-loop” duopoly, firms first set capacities and then fees. In each stage, the actions then are taken as given, but the capacity setting takes into account the effect on Nash equilibrium fees. This set-up seems more realistic, as capacity is a long run decision while fees can be changed more easily. In the third “Stackelberg” game, capacities are set sequentially and then fees are set in a Nash fashion. This setting seems even more realistic, since facilities are typically not all build at the same time, and if firms play a sequential game they should take this into account.

I study four perfectly-competitive auctions on: bid (i.e. transfer to the government), capacity, patronage (i.e. number of users) and generalised price (which is usage cost plus fee and henceforth referenced to as price). Such regulation seems important: private supply can lead to much lower welfare than first-best, and even lower than when the government just sets a (suboptimal) capacity financed by taxes. The auctions I study have been analysed before for a single facility or with an unpriced alternative in, for example, Ubbels and Verhoef (2008) and Verhoef (2007). Tan et al. (2010) show that it is optimal to give the franchise to the firm for the

entire lifespan of the facility, since extending the franchise allows making the contract terms more favourable for welfare by lowering the fee and/or increasing capacity while keeping profit the same. This paper will, however, look at a static setting, and hence ignore the lifespan of the facility.

The two facilities can be auctioned off to a single firm or to two separate firms (where the facilities are auctioned simultaneously),¹ and following the auction there are thus still 4 types of market structures. Since the auction is perfectly competitive, a winning firm makes a zero profit. I find that a bid auction leads to the same capacities and fees as without regulation, since this gives the highest profit to transfer to the government. The capacity auction leads to very high capacities, and typically has a much lower welfare than no intervention. The patronage and price auctions result in the first-best fees and capacities. These results are qualitatively the same as with a single link, although the exact effects differ. However, the auction types also differ in how robust their effects are to the network structure, whether there are one or two winners, and the market structure. The price and patronage auctions attain the first-best outcome regardless of these issues, whereas the effects (relative to the first-best) of the other two auction types depend on the set-up.

For the parallel setting, I find that it is also important to account for how close the substitutes are. The higher the substitutability is, the stronger duopolistic competition, which raises the welfare under a parallel duopoly without regulation and lowers the gain of regulating it.

1.1 Review of the different market structures

Before turning to the modelling, it is important to briefly discuss the different types of market structures. In the first-best case, the Pigouvian fee equals the marginal external cost that a user imposes on the other users—thereby ensuring internalisation of these external costs—and capacity minimises social-cost (i.e. capacity plus usage cost) for a given number of users. The monopolist and the firms with the “open-loop” game use the same capacity rule: firms internalise usage costs, as any decrease in usage cost can be met by an equal fee increase. Hence, firms use the same capacity rule as in the first-best, but do add a mark-up to the fee (Zhang and Zhang, 2006; Basso and Zhang, 2007; Xiao et al., 2007).

With the closed-loop game, the effects of the strategic setting of capacities depend on the network. For parallel facilities that are perfect substitutes, De Borger and Van Dender (2006) and Basso and Zhang (2007) show that firms set lower capacities (for given number of users) than with the single stage of the open-loop game, and they do this to induce higher Nash-equilibrium fees. This also means that usage cost is higher than in the first-best and welfare is further reduced.

¹ Verhoef (2008) studies a sequential-entry market structure which is related to the Stackelberg game: the auctioning of facilities is sequentially, and bidders assume that they will be the last entrant and are “surprised” when the next auction occurs. The downside of this myopic setting is that incumbents tend to make a loss when a new entry occurs, since after they entered (but before the next entry) they made a zero profit.

Conversely, I find that with serial facilities, firms typically set higher capacities, as this lowers the competitor's fee and this raises profit by attracting more users. This also means that welfare is higher than with a open-loop game.

The Stackelberg leader's capacity choice considers the effects on the follower's capacity and the Nash fee setting. I find that with serial facilities, under some common assumptions, the leader sets a higher capacity than without the sequential capacity setting, since this increases the follower's capacity and this in turn attracts more users. In the parallel setting, analytical insights are more limited. Still, in the numerical model, the result is as one would expect: the parallel leader sets a higher capacity to increase its market power (see Van den Berg and Verhoef (2012) for this setting with perfect substitutes).

The proposed model describes the situation where facilities directly serve the user—which seems accurate for roads, railways and telecommunication—but not when there is an intermediate market of firms that in turn serve the consumer—e.g. air- or sea-ports and busses on private roads—if these intermediate firms have market power. Then, the actions of the intermediate firms also need to be modelled (Zhang and Zhang, 2006; Basso and Zhang, 2007), and the distortion of ignoring the second market increases with the market power of the intermediate firms. The model could also be adapted for facilities to access a business (see, e.g., Van Dender, 2005), such as transport facilities to an (air/sea)port or register capacity at a supermarket.

The next section discusses the model set-up and assumptions, where most attention is paid to the different assumptions on the congestion function. Section 3 discusses the first-best and monopolistic equilibria, and Sections 4 and 5 discuss the analytical models for the oligopolies. Section 6 then turns to the auctions. Section 7 gives a numerical example, and 8 discusses some limitations and caveats to the research. Section 9 concludes.

2. Assumptions and model set-up

There are two congestible facilities. The price, p_i , of facility i is the sum of the cost of using it for the user, c_i , and the usage fee, f_i . To keep to model tractable, I assume that inverse demand is linear. With the imperfect substitutes i and j of the parallel case, demand follows

$$D_i = d_0 - d_1 \cdot q_i - d_2 \cdot q_j; \tag{1}$$

where constants d_0 , d_1 and d_2 are larger than zero, and q_i is the number of users of i . If $d_1=d_2$, the facilities are perfect substitutes; when $d_1>d_2$ they are imperfect substitutes. With serial facilities, demand follows

$$D = d_0 - d_1 \cdot q. \tag{2}$$

Assuming an interior solution, for user equilibrium with parallel facilities, the price of each facility has to equal its inverse demand; with serial facilities, the inverse demand has to equal the price of using both.

The cost, C^{cap} , of the capacity, s_i , of facility i is linear:

$$C^{cap} = k \cdot s_i. \quad (3)$$

The benefit to the consumers is the (line-)integral of the inverse demand, consumer surplus is benefit minus price multiplied by the number of users. Profit for a facility is total fee payment minus capacity cost. Welfare equals consumer benefit minus total usage and capacity cost. It is assumed that taxation is costless.

Facilities are assumed to be congestible. Hence, the usage cost of i increases with q_i and the second derivative is non-negative; the cost decreases in a strictly convex manner with capacity:

$$\begin{aligned} \partial c_i / \partial q_i &> 0, \\ \partial^2 c_i / \partial^2 q_i &\geq 0, \\ \partial c_i / \partial s_i &< 0, \\ \partial^2 c_i / \partial^2 s_i &> 0. \end{aligned} \quad (a.i)$$

At times I will need to add the assumption that usage cost is homogeneous of degree zero in the number of users and capacity, and thus only depends on the ratio q_i/s_i :

$$c_i(q_i, s_i) = c_i(\varepsilon q_i, \varepsilon s_i) = c_i(q_i / s_i), \quad \forall \varepsilon \in R^+. \quad (a.ii)$$

Finally, some results need the following functional form that is special case of (a.ii):

$$c_i = \chi_i + \delta_i \left(\frac{q_i}{s_i} \right)^n. \quad (a.iii)$$

Here, χ_i and δ_i are facility specific positive constants and $n \geq 1$ is the same for all. The cost function (a.iii) is still rather general and includes the widely used Bureau of Public Roads (BPR) formulation (e.g. Verhoef, 2007, 2008), costs linear in q_i/s_i (e.g. Van Dender, 2005; De Borger and Van Dender, 2006), and also Vickrey (1969) bottleneck congestion. The BPR congestion used in the numerical example is also a special case of (a.iii).

3. First-best and monopolistic outcomes

This section briefly reviews the first-best and monopoly cases for the serial and parallel settings (for a more extensive overview see, for instance, Small and Verhoef (2007)). In all cases, there are two facilities. In the *First-Best* (FB) outcome, which is defined as the one that maximises welfare, the fee of a facility equals the marginal external cost (MEC) on it, and capacity is set to

minimise social cost by equating marginal capacity cost, k , to the usage cost decrease from the marginal capacity expansion. This is reflected in these well known optimality conditions:

$$f_i^{FB} = q_i \cdot \partial c_i / \partial q_i, \quad (4)$$

$$k = -q_i \cdot \partial c_i / \partial s_i. \quad (5)$$

As showed by Mohring and Harwitz (1962), as long as assumptions (a.i) and (a.ii) hold, the first-best outcome leads to zero profit for the facilities, and thus the system is self-financing.

A monopolist also uses capacity rule (5). The intuition is as follows, for given number of users, when capacity is expanded usage cost decreases, which in turn allows the fee to be increased by this cost decrease. The firm hence internalises usage costs, because any decrease in total usage cost can be met by an equal increase in toll revenue. If usage cost follows assumption (a.ii) and thus usage cost only depend on the ratio q_i/s_i , then the facts that capacity cost are linear and the capacity rule is the same in the monopoly and first-best cases imply that monopolist will even have the same usage cost as in the first-best. The monopolistic fee, however, is much higher as it equals the MEC plus a mark-up that depends on the network and situation. Accordingly, there are fewer users and capacity is lower (Xiao et al. (2007)).

When a *parallel monopolist* (PM) controls the two facilities, its fee on i is

$$f_i^{PM} = q_i \cdot \partial c_i / \partial q_i + q_i \cdot d_1 + q_j \cdot d_2. \quad (6)$$

Here, the first term is the MEC. The second term is the monopolistic mark-up from users on i . The third term is the mark-up due to facility j : it measures the effect that a higher fee on i increases the demand for j which raises the profit from j . Note that the same considerations play in the setting of its fee on j . The closer substitutes the facilities are (i.e. d_2 is higher for given d_1 and d_0), the higher the fees, since an increase of the fee on i then pushes more users to j .

With serial facilities, users are only interested in the total price. A *serial monopolist* (SM) asks a fee for using A and B of

$$f_{A+B}^{SM} = q \cdot \partial c_A / \partial q + q \cdot \partial c_B / \partial q + q \cdot d_1. \quad (7)$$

This fee is the sum of the two marginal external costs and the monopolistic mark-up ($q \cdot d_1$).

4. Serial duopolists

In the serial duopoly, firm i always wants j 's fee to be lower and capacity higher, as both these changes lower the price and thereby attract more users and raise the users' willingness to pay i 's fee. As we will see, these strategic goals are opposite to those in the parallel case. With Nash capacity setting preceding Nash fee setting (i.e. a closed-loop game), a serial firm typically sets a

higher capacity than with a open-loop game (where fee and capacity are set at the same time). Conversely, as section 5 will find, a parallel firm typically sets a lower capacity as it wants to increase other facility's fee. In the serial Stackelberg game, the leader tends to sets a higher capacity, as it wants a higher follower's capacity since this in turn increases the number of users.

This section will start with the response of the equilibrium number of users to changes in capacities and fees. Then, the fee rule is derived, which turns out to be the same regardless of the type of duopoly. Thereafter, I turn to the capacity setting.

4.1. The response of the equilibrium number of users to changes in capacities and fees

As one would expect, the user-equilibrium number of users is higher when a fee is lower or a capacity higher, as these changes lower the (generalised) price. The derivative of the number of users to the fee is the same whether it is i 's or j 's, since users do not care about to whom they pay. The derivatives to capacity may differ, as these depend on the congestion levels.

In user equilibrium, inverse demand equals the price of using both facilities:

$$D(q) \equiv d_0 - d_1 q = C_i + f_i + C_j + f_j.$$

By differentiating this condition to fees and capacities, and solving the resulting system of equations, the responses of q to changes in these variables can be found (where superscript R indicates a response):

$$\frac{\partial q^R}{\partial f_i} = \frac{\partial q^R}{\partial f_j} = -\frac{1}{Y_1} < 0, \quad (8)$$

$$\frac{\partial q^R}{\partial s_i} = -\frac{\partial c_i / \partial s_i}{Y_1} = \frac{\partial c_i}{\partial s_i} \frac{\partial q^R}{\partial f_i} > 0; \quad (9)$$

and Y_1 follows

$$Y_1 = \partial c_i / \partial q_i + \partial c_j / \partial q_j + d_1. \quad (10)$$

As the third part of (9) shows, the response of q to capacity is just the change due to the fee multiplied by $\partial c_i / \partial s_i$, which gives the change in price due to the capacity change (whereas the derivative of the price w.r.t. the fee is 1).

4.2. Fee setting

In setting its fee, firm i takes the capacities and fee of j as given and maximises its profit:

$$\Pi_i = f_i \cdot q - k \cdot s_i. \quad (11)$$

By differencing profit to the fee, and then inserting eq. (8) for $\partial q_i / \partial f_i$ and rewriting, the profit-maximising fee rule can be found:

$$f_i = f_j = q \cdot (\partial c_i / \partial q + \partial c_j / \partial q + d_1) = q \cdot Y_1. \quad (12)$$

A firm not only asks the MEC on its own facility, $q \cdot \partial c_i / \partial q$, but also that on the other, $q \cdot \partial c_j / \partial q$: since any decrease in usage cost on i or j , for a given q , can be matched by a fee increase. The term $d_1 \cdot q_i$ in (12) is the mark-up, and it has the same form as for the monopolist, since the firm has no parallel competitors. Firm j uses the same rule, and thus has the same fee. Consequently, users face “double marginalisation”, and the total fee is higher than with a monopolist. The fee is also much higher than socially optimal.

The responses of i 's fee to j 's fee and both capacities are summarised here, for mathematical derivations please see Appendix A.1. The fee of i decreases with j 's fee for any congestion function, and the slope of the function is between $-1/2$ and 0 : a higher f_j lowers the number of users, and this lowers the MEC and mark-up parts of i 's fee. For linear in q/s_i congestion, the slope is $-1/2$; it would be 0 , if an usage cost had an infinitely large second derivative w.r.t. q ; in general, the more convex usage costs are in q , the weaker the response.

The fee of i typically decreases with its capacity by lowering the MEC part of the fee. However, there is also a counteracting indirect effect: the lower user cost (due to the higher capacity) attracts more users, and this increases the MEC and mark-up parts of the fee. Still, this second-order effect is unlikely to dominate the first effect. The $\partial f_i / \partial s_i$ is certainly non-positive when (a.iii) holds, and is negative if then usage costs increase stronger than linear in q/s_i (for linear costs, the fee is insensitive to i and j 's capacity²). Finally, i 's fee decreases with j 's capacity under the same conditions as discussed for i 's capacity.

The Nash-equilibrium (NE) fees for given capacities are at the intersection of the response functions. These NE fees generally decrease with each capacity; and the derivatives are certainly non-positive under (a.iii), and only zero when usage costs are linear.

4.3. Capacity setting under open-loop competition (serial facilities)

With the open-loop game, capacity and fee setting occur simultaneously, and firm i maximises (11) to s_i for given capacity of j and fees. The resulting rule states that the cost of a marginal capacity expansion, k , should equal the total reduction in usage cost it achieves:

$$k = -q_i \cdot \partial c_i / \partial s_i. \quad (13)$$

² Note the similarity with De Borger and Van Dender (2006) who have linear costs and perfect substitutes, and find that the fee of i is independent of i 's capacity but decreases with j 's. For imperfect substitutes and linear cost, the next section will find that the same holds.

Hence, again, the firm internalises the usage costs, since any decrease in usage cost, for given q , can be met by an equal fee increase; and the firm's capacity rule is the same as under welfare maximization,³ and, if assumption (a.ii) holds, the firm will even have the same usage costs as in the first-best.

4.4. Capacity setting under closed-loop competition (serial facilities)

With the closed-loop game, the Nash capacity setting precedes the Nash fee setting, and in its capacity setting a firm takes into account the effect on Nash-equilibrium fees. The capacity rule has the additional concern that a higher own capacity typically lowers the other firm's fee.⁴ This will lead the firms to set higher capacities than with an open-loop game for a given number of users, and hence user cost is lower.

The firm again maximises (11) but now the (Nash-equilibrium) fees are a function of s_i and, the taken as given, s_j . The first order condition is

$$\frac{\partial \Pi_i}{\partial s_i} = 0 = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q_i + \frac{dq^R}{ds_i} f_i - k = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q + \left(\frac{\partial q^R}{\partial s_i} + \frac{\partial q^R}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_i} + \frac{\partial q^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} \right) f_i - k; \quad (14)$$

where q is the number of users and $\partial q^R / \partial f_i$ the response of the number of users to a change in i 's fee. This condition can be simplified to⁵

$$k = -\frac{\partial c_i}{\partial s_i} q + \frac{\partial q^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i. \quad (15)$$

This equation for the closed-loop game differs from (13) for the open-loop game (where capacity and fee setting occur simultaneously) by the addition of the second term on the right side. The term is only non-positive when $\partial f_i^{NE} / \partial s_i$ is non-positive, and this derivative is—as discussed in Section 4.2—certainly non-negative if (a.iii) holds:

$$k = -\overbrace{\frac{\partial q}{\partial s_i}}^{>0} q + \overbrace{\frac{\partial q^R}{\partial f_j}}^{<0} \overbrace{\frac{\partial f_j^{NE}}{\partial s_i}}^{\leq 0} f_i, \quad \text{if (a.iii) holds;} \quad (15')$$

³ However, since the fee is higher than the MEC, given the fee, the welfare-maximising capacity would be higher to correct for the fee leading to too few users (Small and Verhoef, 2007, p.172).

⁴ The effect of its capacity choice via its own fee on profit drops out: when i increases its capacity, the profit increase from a higher own fee is exactly offset by of the lower number of users this fee increase induces. This occurs for any cost function (see footnote 6), and, as we will see later, it also holds for parallel facilities.

⁵ Using the f.o.c. for fee setting, one gets that $f_i \partial q^R / \partial f_i = -q$; and by using (9) and (12) one gets that $(\partial q / \partial s_i) f_i = -(\partial c_i / \partial s_i) / Y_i (q \cdot Y_i) = -\partial q / \partial s_i \cdot q > 0$. Inserting these two results into (14) and rewriting results in the simplified rule (15). A consequence of these two results is also that the effect on profit of s_i via its own fee is zero, since the direct effect, $\partial f_i^{NE} / \partial s_i \cdot q_i$, is cancelled out by the indirect effect via the number of users, $(\partial q^R / \partial f_i)(\partial f_i^{NE} / \partial s_i) f_i = -\partial f_i^{NE} / \partial s_i \cdot q$.

and only if costs are linear in q/s_i is the new term zero, since then $\partial f_i^{NE} / \partial s_i = 0$. Therefore, with linear costs the open- and closed-loop games have the same outcome. For all other functions satisfying (a.iii), the extra term is positive, and the capacity is higher with a closed-loop game.

Under the likely outcome that the new term is positive, the firm sets a higher capacity than under the open-loop rule of (13) (as this lowers the $\partial q / \partial s_j q$, thereby ensuring that the right side of (15) equals the constant k). The intuitive reason is that, if setting a higher capacity lowers the competitor's fee, this in turn attracts more users and thereby raises profit. As both firms do this, equilibrium capacities are higher and fees lower than with the single stage of the open-loop game; and this also means that welfare is higher.

4.5. Capacity setting under Stackelberg competition (serial facilities)

Under Stackelberg competition, the capacity setting is done sequentially, but the fee setting that follows is Nash. The follower j uses capacity rule (15), as it can affect the fee of leader i but takes the leader's capacity as given. The leader can also affect the capacity of the follower.

The leader's capacity rule can be shown to equal

$$k = -\frac{\partial c_i}{\partial s_i} q + \frac{\partial q^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i + \frac{\partial s_j}{\partial s_i} \left(\frac{\partial f_i^{NE}}{\partial s_j} q + \frac{\partial q^R}{\partial s_j} f_i \right); \quad (16)$$

which only differs from (15) for the closed-loop setting by the addition of the third term on the right side. This section briefly describes the effects of this new term, Appendix A.2 gives the mathematics. The new term measures the effects of the change in the follower's capacity that a change in the leader's capacity induces. The $\partial s_j / \partial s_i$ gives the induced change in s_j . The two terms between brackets measure the effects of this induced change on marginal revenue: the first item gives the profit-lowering effect that a higher s_j typically lowers i 's fee, the second item gives the profit-increasing effect that more users will be attracted.

The sign of the sum of the two terms between brackets is uncertain even with assumption (a.iii). However, when the power of the congestion function (n) is not larger than 4, the sum is positive; whereas, with $n > 4$, this seems likely, but not certain.

Similarly, $\partial s_j / \partial s_i$ is also typically positive, but it may not be. A positive slope of this reaction function is logical: an increase of s_i attracts more users and this in turn also increases congestion on j ; and both these facts make capacity expansion more attractive for j . There is, however, a third counteracting force that the increased s_i lowers j 's fee, which makes expansion less rewarding. Still under (a.iii), $\partial s_j / \partial s_i$ is always positive since the two first terms dominate (and with linear costs, the third term is even zero).

Concluding, for any congestion function following the common assumption (a.iii) with $n \leq 4$, the leader sets a higher capacity to induce the follower to also set a higher capacity; and this

higher capacity of the follower raises profit by attracting more users. For other congestion forms this is also likely. Note that even assumption (a.iii) allows the facilities to have different cost functions, and they can be very different: e.g. (short) access trip by car, and then flying (from an airport with airlines without market power).

5. Duopolists with parallel facilities

If a firm has parallel competition, it wants its competitor's fee to be higher and capacity lower, since this pushes users to its own facility. Attaining closed form solutions is more difficult in this setting, as now there are two interdependent demands.

5.1. Equilibrium number of users and capacities and fees

The effects on the equilibrium number of users of capacity and fee choices are again as expected. A higher fee or lower capacity on i lowers its number of users by increasing the price. A higher fee or lower capacity on j increases the number of i by increasing j 's price.

To prove this, I use that the prices on i and j should equal their respective inverse demands:

$$D_i(q_i, q_j) = C_i + f_i,$$

$$D_j(q_i, q_j) = C_j + f_j.$$

Then, by differentiating these conditions to fees and capacities, and solving the resulting system, one gets the responses of the number of users to these instruments:⁶

$$\frac{\partial q_i^R}{\partial f_i} = -\frac{\partial c_j / q_j + d_1}{Z_1} < 0, \quad (17)$$

$$\frac{\partial q_j^R}{\partial f_i} = \frac{\partial q_i^R}{\partial f_j} = \frac{d_2}{Z_1} > 0; \quad (18)$$

where

$$Z_1 = (\partial c_i / \partial q_i + d_1)(\partial c_j / \partial q_j + d_1) - d_2 d_2 > 0. \quad (19)$$

Similarly, the responses of q to the capacities are

$$\frac{\partial q_i^R}{\partial s_i} = -\frac{\partial c_i}{\partial s_i} \frac{\partial c_j / q_j + d_1}{Z_1} = \frac{\partial c_i}{\partial s_i} \frac{\partial q_i}{\partial f_j} > 0, \quad (20)$$

$$\frac{\partial q_j^R}{\partial s_i} = \frac{\partial c_i}{\partial s_i} \frac{d_2}{Z_1} = \frac{\partial c_i}{\partial s_i} \frac{\partial q_j}{\partial f_i} < 0. \quad (21)$$

⁶ Note that equations (4-7) in De Borger and Van Dender (2006) are special cases for perfect substitutes ($d_1=d_2$) and symmetric usage costs that are linear in q/s_i (i.e. following (a.iii) with $n=1$).

5.2. Fee setting (parallel facilities)

Firms set their fees in a Nash fashion and take capacities as given. By maximising profit to the fee in a similar manner as with serial facilities, the optimal fee rule can be found:

$$f_i = q_i \cdot \partial c_i / \partial q_i + q_i \cdot d_1 - q_i \frac{d_2 \cdot d_2}{\partial c_j / \partial q_j + d_1}. \quad (22)$$

The first term in this rule is the marginal external cost (MEC), which the firm charges as it internalises the congestion cost. The sum of the second and third term gives the mark-up: the second is the mark-up for a monopolist of a single facility, the third gives the correction due to the competition from the other facility. The closer substitutes the facilities are (i.e. the closer d_2 is to d_1), the larger this correction, the stronger the competition and the lower fees. In the special case of independent demands (i.e. $d_2=0$), the third term is zero, and the fee the highest (for given d_1); conversely, with perfect substitutes (i.e. $d_2=d_1$), the fee is the lowest. The fee is always higher than socially optimal, and thus the number of users is too low.

To find the slopes of the best responses of a firm's fee to its competitor's and the capacities, one can again write (22) in implicit form, and use the implicit function theorem. Appendix B.1 provides the mathematical details, this section summarises the results.

An increase in the fee of competitor j has two effects on i 's fee: 1) it increases the number of users on i and thus f_i ; and 2) it lowers the number of users on j , which lowers f_i : this is because, following (22), i 's mark-up is higher with a higher q_j , as this increases congestion on j , making it a less attractive substitute. The sum of these two effects is typically positive, but it may not be if firms are really asymmetric. If firms are (ex-post) symmetric, it is certain that f_i decreases with f_j , and if (a.iii) holds this is also certain with asymmetry.

A higher capacity of i typically lowers i 's fee: it lowers congestion and thus the MEC part of the fee; and it also lowers the mark-up by lowering the number of users of competitor j (and thereby lowering the usage cost of j). Still, there is a third counteracting effect from the induced increase in the number of users of i , which limits the decrease of the MEC and raises the mark-up. Hence, the fee needs not always to decrease with s_i , but it would be surprising that the third effect would dominate. Moreover, under (a.iii), $\partial f_i / \partial s_i \leq 0$ is certain, and it is only zero when usage cost is linear in the ratio.

The fee of i also typically decreases with j 's capacity, but this may not hold in extreme cases. There are three effects: 1) a higher s_j makes j more competitive and thereby decreases i 's mark-up; 2) it lowers q_i , thereby lowering the MEC and mark-up; but 3) a higher s_j also attracts more users to j , which tends to increase i 's fee by increasing the user cost on j and thus making j less attractive. Under (a.iii), i 's fee certainly decreases with j 's capacity.

Finally turning to the Nash-equilibrium (NE) fees, these generally decrease with each capacity. Moreover, they certainly decrease with the capacities under (a.iii), which corresponds with the finds of De Borger and Van Dender (2006) for linear usage costs.

5.3. Capacity setting under open-loop competition (parallel facilities)

In this setting, each firm takes the fee and capacity of the other as given in setting its capacity. Maximising profit results in the capacity rule below. This rule again is the same as in the first-best case:

$$k = -q_i \cdot \partial c_i / \partial s_i \quad (23)$$

5.4. Capacity setting under closed-loop competition (parallel facilities)

In this more realistic setting, capacity setting precedes the fee setting, and a firm recognises that its capacity affects the Nash-equilibrium (NE) fees f_i^{NE} and f_j^{NE} . Appendix B.2 shows that the capacity rule under closed-loop competition is

$$k = -q_i \cdot \partial c_i / \partial s_i + \frac{\partial q_i^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i . \quad (24)$$

Just as with serial competition, the only difference between the formula here and (23) for the open-loop game is addition of the second term on the right side: the capacity choice is adjusted for the effect on i 's number of users via the Nash-equilibrium fee of j . Under assumption (a.iii), the new term is negative,⁷ which implies that i sets a lower capacity than without the separate stages (since otherwise the right side of (24) could not equal the constant k):⁸

$$-\underbrace{q_i \cdot \partial c_i / \partial s_i}_{<0} + \frac{\partial q_i^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i = k, \quad \text{if (a.iii) holds.} \quad (24')$$

This also means that, then, welfare is lower with the closed-loop game than with the open-loop setting: with duopolistic supply, the number of users is too low from society's point of view; and the lower capacities with closed-loop competition only discourage usage further and raise costs.

⁷ Following Section 5.2 and Appendix B.1, $\partial f_i / \partial s_j < 0$ is certain under (a.iii). The other two items of the new term in (24) are positive for any congestible usage cost.

⁸ Under general congestion functions, the second-term is also typically negative, unless firms are really asymmetric.

5.5. Capacity setting under Stackelberg competition (parallel facilities)

Now, the firms set their capacities one after the other. Again, follower j has the same capacity rule as with the closed-loop game. Leader i can also affect the follower's capacity, and thus its capacity rule includes this extra consideration:

$$k = -\frac{\partial c_i}{\partial s_i} q_i + \frac{\partial q_i^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i + \frac{\partial s_j}{\partial s_i} \left(\frac{\partial f_i^{NE}}{\partial s_j} q_i + \frac{\partial q_i^R}{\partial s_j} f_i \right). \quad (25)$$

Analytically, it difficult to say what effect is of the new concern in (25). One would expect that the leader would set a larger capacity to increase its market power, just as happens in the textbook Stackelberg game without congestion; and this is also what is found in the numerical example below and in Van den Berg and Verhoef (2012).⁹ For this to happen, j 's capacity needs to decrease i 's. But even for the linear congestion of De Borger and Van Dender (2006) this is not always the case. Still, in all numerical calibrations that were tried, j 's capacity decreased with i 's

Both terms inside the brackets are typically negative, and hence the sum of the two is generally negative, thus making it profit increasing to lower the follower's capacity. The first term measures the effect of the change i 's Nash-equilibrium fee due to an induced change in the follower's capacity (s_j) and is negative under (a.iii). The second term gives the effect via i 's number of users and is negative under the most general congestion function of (a.i); the intuition is that a higher s_j lowers the congestion on j and thereby lowers i 's number of users and profit.

6. Auctions

The paper now turns to the discussion of auctions for the right to build and operate a facility. The government can auction off the facilities to a single firm or to two separate firms, where the facilities are auctioned off simultaneously. After the auction, a winner sets its fee and capacity under the constraint that the promise from the auction has to be met; otherwise, firms compete in the same way as without an auction (and, hence, there are still three duopolies and one monopoly possible). In all cases, the auction is perfectly competitive and firms attain zero profit. The auction outcome is where, given the following competition, the offer leads to zero profit.¹⁰

The four auction types are bid (highest transfer to the government wins), capacity (highest offered capacity wins), price (lowest generalised price wins), and patronage (highest number of users wins).¹¹ These auctions have been studied for a single facility, but not for two. The outcome with a bid auction is the same as without an auction, but the profit is transferred to the

⁹ See Acemoglu et al. (2009) for a related game without congestion and fixed demand unless the price exceeds the common reservation utility. Also in their set-up the leader has the higher capacity.

¹⁰ If there are multiple offers that lead to zero profit, the offer made is the one that is most likely to win. For instance, with a capacity auction, profit is zero with zero capacity and with the largest capacity that is self-financing. Naturally, only the latter offer has any change of winning.

¹¹ I do not look at an auction on fee, as with freely chosen capacities the outcome would be capacities and fees of zero; a fee auction only makes sense when capacity is fixed, constrained, or regulated.

government. With a capacity auction, the capacities are much larger than is socially optimal. The patronage and price auctions lead to the same outcome as the first-best case. These results correspond with those for a single facility in Verhoef (2007) and Ubbels and Verhoef (2008).

7. Numerical example

This section illustrates the model with a numerical example. The calibration in Table 1 builds on Verhoef (2007). For a base-case calibration without congestion pricing, there are 5000 users and the elasticity w.r.t. the own price is -0.35 , with parallel substitutes the cross-price elasticity is 0.20 . The congestion follows a Bureau of Public Roads (BPR) calibration, which is special case of (a.iii) with $n=4$. The usage-cost functions of the two facilities are the same. The marginal cost of capacity is set at 7 for an entire path,¹² and thus at 3.5 for one of two serial facilities. The calibration assumes that the facilities are ex-ante symmetric, but this assumption is not vital to the results; it only helps with the comparisons and simplifies the tables. Moreover, firms can still be asymmetric ex-post, and will be so in the unregulated Stackelberg games.

Table 1: Calibrations of the numerical models

	Serial facilities	Parallel facilities
d_0	3495/112	26795/432
d_1	233/50400	1631/118800
d_2	x	26795/432
$\delta_A=\delta_B$	9/32	9/16
$\chi_A=\chi_B$	15/8	15/4
k	3.5	7

7.1. Serial facilities

Table 2 gives the outcomes for the serial network without auctions. In the *base case*, congestion is heavy and usage cost is much higher than in the other cases. This case should not be seen as some initial situation since all private games have much lower capacities. The regime is only there for comparison.

In the *first-best* (FB) case, capacity is set following (5) to minimise social cost for a given number of users, and the fee equals total Marginal External Cost (MEC). Under some assumptions that hold here, profit is zero under these instrument rules. A *serial monopolist* uses the same capacity rule, but following (7) adds a mark-up to the fee. In accordance with the theoretical section, although the number of users and welfare are much lower, the usage cost is the same as in the first-best. Welfare is much lower than in the first-best case.

¹² See Verhoef (2007) for the calculation from the expressway construction cost of about €5 million per lane-km in the Netherlands and a road length of 60 km. As Van den Berg and Verhoef (2012) discuss, this cost seems comparable to those presented for the USA in Washington State Department of Transport (2005).

The two serial operators with the *open-loop* game both add a mark-up to the fee that has the same structure as a monopolist's, as they do not directly compete. The total fee is lower than twice the monopolist's, since the number of users is lower. Welfare under this duopoly is even lower than under a monopoly. The firms are also worse off, but unfortunately the monopolistic outcome is not a Nash-equilibrium of this duopoly.

Under my calibration, firms want their serial competitor to set a lower fee (for any level of the competitor's fee). To achieve this in the *closed-loop* game, each firm sets a higher capacity than it otherwise would. Hence, firms have a lower usage cost and fee, and attract more users than without this strategic consideration. This is opposite from what occurs with parallel facilities, where a firm typically sets a lower capacity for given number of users, since this increases fees.

Table 2: Outcome for the numerical example for the serial facilities

	Base case	First-best	Monopolist	Open-loop duopoly	Closed-loop duopoly	Stackelberg (B is the leader)
Capacity A	3000	3530.8	1765.4	856.4	1045.1	1193.1
Capacity B	3000	3530.8	1765.4	856.4	1045.1	1331.3
Number of users	5000	4430.5	2215.3	1074.6	1244.8	1403.9
Usage cost A	4.05	2.57	2.57	2.57	2.44	2.41
Usage cost B	4.05	2.57	2.57	2.57	2.44	2.22
Fee A	-	2.79 [#]	7.91 [#]	10.55	10.28	10.04
Fee B	-	2.79 [#]	7.91 [#]	10.55	10.28	10.04
Profit A	-10500	0 [#]	14455 [#]	8356	9144	9918
Profit B	-10500	0 [#]	14455 [#]	8356	9144	9434
Consumer Surplus	57788	45373	11343	2669	3582	4556
Welfare	36788	45373	34030	19341	21870	23909
Relative efficiency*	0	1	-0.32	-2.03	-1.74	-1.50

Note: [#] Here, only the total fee is defined. Hence, the given number for facility i is the system total divided by two.

* Relative efficiency is the welfare gain from the base case relative to the first best gain.

The *Stackelberg* game seems the most realistic of the three duopolies. Under BPR congestion, serial leader B wants follower A to set a higher capacity. Hence, B sets a higher capacity than it would without this extra consideration. But, since serial duopolists always have the same fees, the leader makes a lower profit than the follower. Still, the leader is better off than with the closed-loop, as fees are closer to the monopolistic ones. Nevertheless, the duopolistic fees remain above the monopolistic ones. Hence, with serial facilities, it is better to have a monopoly than a duopoly.

All private games attain a lower welfare than the arbitrary base case with very heavy congestion. Hence, it might even be better for the government to set a suboptimal capacity financed by taxes than to allow private supply. This does ignore, however, the cost of raising tax revenue and the possibility that the private sector works more efficient than the government.

7.2. Parallel facilities

This section looks at parallel facilities that are imperfect substitutes. In the parallel *base case* in Table 3, the usage cost and fee of i equal the corresponding total with serial facilities. A monopoly leads to lower welfare than a duopoly, as now the facilities directly compete.

The *closed-loop* game increases fees from the open-loop game, as by setting a lower capacity the competitor's fee is increased and this in turn makes the own facility more attractive for users. This also implies that welfare is even lower than with the single stage of the open-loop game.

Comparing the *Stackelberg* and closed-loop games, it shows that the Stackelberg game further lowers welfare, because it gives the leading firm B more market power, and this gives B a higher fee and more users. The parallel leader makes a larger profit than the follower: it has more users and can ask a higher fee (since it has the lower usage cost).

Table 3: Outcome for the numerical example for the parallel facilities

	Base case	First-best	Monopolist	Open-loop duopoly	Closed-loop duopoly	Stackelberg (B is the leader)
Capacity A	1500	1895.1	947.5	1288.9	921.7	872.5
Capacity B	1500	1895.1	947.5	1288.9	921.7	1163.9
Number of users A	2500	2378.0	1189.0	1617.3	1370.8	1319.7
Number of users B	2500	2378.0	1189.0	1617.3	1370.8	1389.5
Usage cost A	8.09	5.14	5.14	5.14	6.50	6.69
Usage cost B	8.09	5.14	5.14	5.14	6.50	4.89
Fee A	x	5.58	31.23	21.99	25.95	26.31
Fee B	x	5.58	31.23	21.99	25.95	27.70
Profit A	-10500	0	30498.8	26540.2	29120.2	28616.6
Profit B	-10500	0	30498.8	26540.2	29120.2	30346.3
Consumer surplus	134838	121995	30499	56433	40537	39595
Welfare	113838	121995	91496	109514	98777	98558
Relative efficiency*	0	1	-2.74	-0.53	-1.85	-1.87

Note: * Relative efficiency is the welfare gain from the base case relative to the first best gain.

7.3. Auctions

The *bid* auction attains the same outcome as no auction, as this leads to the highest profit to transfer. Therefore, for the outcomes of this auction, see Tables 2 (serial) and 3 (parallel). For the three other auctions, Table 4 gives the results with serial facilities and Table 5 with parallel facilities. Under these three auctions, all three duopolies lead to the same outcome. This is because both firms have to make zero profit, and thus the outcome of an auction is at the intersection of the zero-profit functions. Accordingly, with, for example, a capacity auction, the

strategic setting of a different capacity will mean that at least one firm will not be making zero profit, and thus this action is not supported by an equilibrium.¹³

An auction on *capacities* leads to very large capacities, and a user cost that is much lower than first-best. But, since to finance these large capacities the fees need to be very high, the number of users is actually low. Hence, this auction is bad for welfare. It would actually be better not to intervene. Just as without an auction, it is again better to have a serial monopoly than a serial duopoly. Still, even the monopoly auction on capacity leads to a lower welfare than an unregulated serial duopoly, which confirms for this setting the results in Verhoef (2007, 2008) on how bad an auction design this is.

Conversely, the *patronage* and *price* auctions attain the first-best outcome in all of my set-ups. They do so when the auction has a one or two winners, and for any of the analysed market or network structures. Hence, these auctions seem more robust than the other two.

Table 4: Auctions for serial facilities

	Monopoly on capacity	A duopoly on capacity	A duopoly or Monopoly on patronage	A duopoly or Monopoly on price
Capacity of i	5807.3	5162.7	3530.8	3530.8
Number of users	2949.2	1971.9	4430.5	4430.5
Usage cost of i	1.89	1.88	2.57	2.57
Fee of i	6.89	9.16	2.79	2.79
Consumer surplus	20105	8988	45373.4	45373.4
Welfare	20105	8988	45373.4	45373.4
Relative efficiency	-1.94	-3.24	1	1

Table 5: Auctions for parallel facilities

	Monopoly on capacity	A duopoly on capacity	A duopoly or Monopoly on patronage	A duopoly or Monopoly on price
Capacity of i	5621.5	4728.2	1895.1	1895.1
Number of users of i	1350.4	1887.9	2378.0	2378.0
Usage cost of i	3.75	3.76	5.14	5.14
Fee of i	29.14	17.53	5.58	5.58
Consumer surplus	39341	76894	121995	121995
Welfare	39341	76894	121995	121995
Relative efficiency	-9.13	-4.53	1	1

¹³ With the sequential-entry structure of Verhoef (2008), a different outcome would occur as the first firm to enter is myopic in its auction offer to the entry of the second firm. Hence, it will offer a higher capacity and lower fee, but it will make a loss after the second entry.

8. Discussion

When doing (numerical) research, it is important to investigate how sensitive the results are to the parameters and assumptions. The private oligopolies attain in the numerical example much lower welfares than the first-best or even the (arbitrary) base case without congestion pricing. With the serial set-up one would expect this due to the serial marginalisation, but especially with parallel facilities it is surprising. But this outcome could be different if users were more price-sensitive or the base-case capacity was smaller. Indeed, if the base case had zero capacity, then private supply always improves welfare. Also important is the cross-price sensitivity. The stronger users respond to the competitor's price, the stronger the competition with a parallel duopoly, and hence the higher welfare. When the model is re-calibrated with only the cross-price sensitivity being 50% larger, the relative efficiency (i.e. the welfare gain from the base case relative to the first-best gain) for the open-loop game would be -0.42 instead of the -0.53 found above. Hence, the stronger substitutes the parallel facilities are, the better duopolistic supply is for welfare and the lower the gain of regulation.

The form of the capacity cost function is also important. Under the assumption used here—and in much of the literature—that cost is linear, the first-best outcome has zero profit and thus the price and patronage auctions can attain the first-best. Yet, if capacity cost is not linear, this generally does not hold. Under increasing returns, the operator would need a subsidy in order to be able to attain the first-best. Such a subsidy might be difficult, because the government may not be able to provide it or have difficulty observing the cost structure of the firm.

This paper ignores costs of public funds and, hence, assumes that taxation is costless. This assumption can affect the comparison of the bid auction with the patronage and price auctions: if the cost of public funds is high enough, the revenue from a bid auction becomes so valuable that this auction actually is preferable.

For future research it seems interesting to look at larger networks where the effects of serial and (im)perfect substitute facilities interact. Here, the effect of (long-run) dynamics also seems interesting. If firms choose to invest over many years: how do they choose when to invest and are there strategic incentives such as trying to attain a monopoly in an area and preventing entry by other firms or the government? Another important issue is how to deal with risk and uncertainty. As Engel et al. (1997) argue, government bailouts are almost universal for franchises in financial trouble. This limits incentives for cost control and can lead to “opportunistic renegotiations” by firms or government. In the latter case the government, for example, demands lower fees as the franchise is making “excessive” profits.

9. Conclusion

This paper studied private supply of two congestible facilities that are either parallel or serial. It did so under four market structures: a monopoly and three duopolies that differ in how the firms interact. All private settings result in much lower welfare than the first-best. This is especially true for serial facilities, but also for parallel firms. Therefore, there is substantial scope for regulation, and this paper also investigated the effects of such regulation by four types of auctions; where the facilities can be auctioned off to one or two firms, and hence there are still 4 possible market structures after the auction.

It is well known that in a parallel setting, competition improves welfare, and hence a duopoly leads to a higher welfare than a monopoly. Conversely, with serial facilities, having multiple firms actually harms welfare, since each firm is a monopoly on its section.

This problem does not occur with the two auctions that seem preferable—those on generalised price and number of users (patronage)—since these lead to the first-best outcome regardless of whether there are one and two winners, of which market structure holds after the auction, and for both analysed network structures. Conversely, the other two auctions are sensitive to these considerations and tend to lead to lower welfare. Ignoring costs of public funds, the bid auction leads to the same outcome as no regulation, while the capacity auction actually lowers welfare. Hence, auctions need to be carefully designed: a “wrong” auction can actually be worse for welfare than no regulation.

Acknowledgements

Financial support from the ERC (AdG Grant #246969 OPTION) is gratefully acknowledged. I thank Erik Verhoef, Paul Koster, Eva Gutiérrez, and Ken Small for their suggestions. The usual disclaimer applies.

References

- Acemoglu, D., Bimpikis, K., Ozdaglar, A., 2009. Price and capacity competition. *Games and Economic Behavior* 66 (1), 1–26.
- Basso, L.J., Zhang, A., 2007. Congestible facility rivalry in vertical structures. *Journal of Urban Economics* 61(2), 218–237.
- De Borger, B., Van Dender, K., 2006. Prices, capacities and service quality in a congestible Bertrand duopoly. *Journal of Urban Economics* 60(2), 264–283.
- Engel, E., Fisher, R., Galetovic, A., 1997. Highway franchising: pitfalls and opportunities. *American Economic Review* 8(2), 68–72.
- Mohring, H., Harwitz, M., 1962. *Highway benefits: an analytical framework*. Northwestern University Press, Evanston, Illinois.
- Small, K.A., Verhoef, E.T., 2007. *The Economics of Urban Transportation*. Routledge, Abingdon, England.
- Tan, Z., Yang, H., Guo, X., 2010. Properties of Pareto-efficient contracts and regulations for road franchising. *Transportation Research Part B* 44(4), 415–433.
- Ubbels, B., Verhoef, E.T., 2008. Auctioning concessions for private roads. *Transportation Research Part A* 42(1), 155–172.
- Van Dender, K., 2005. Duopoly prices under congested access. *Journal of Regional Science* 45(2), 343–362.
- van den Berg, V.A.C., Verhoef, E.T., 2012. Is the travel time of private roads too short, too long, or just right? *Transportation Research Part B*, in press.

- Verhoef, E.T., 2007. Second-best road pricing through highway franchising. *Journal of Urban Economics* 62(2), 337–361.
- Verhoef, E.T., 2008. Private roads: auctions and competition in networks. *Journal of Transport Economics and Policy* 42(3), 463–493.
- Vickrey, W.S., 1969. Congestion theory and transport investment. *American Economic Review (Papers and Proceedings)* 59(2), 251–260.
- Washington State Department of Transport, 2005. Highway construction costs: are WSDOT’s highway construction costs in line with national experience? (version of 3 November 2005). Retrieved on 25 March 2011, from www.wsdot.wa.gov/biz/construction/CostIndex/pdf/HighwayConstructionCosts2005.pdf
- Xiao, F., Yang, H., Han, D., 2007. Competition and efficiency of private toll roads. *Transportation Research Part B* 41(3), 292–308.
- Zhang, A., Zhang, Y., 2006. Airport capacity and congestion when carriers have market power. *Journal of Urban Economics* 60(2), 229–247.

Appendix A. Mathematical discussion for the serial facilities

A.1. Reaction functions of a serial facility’s fee

This appendix mathematically discusses the reactions of i ’s fee to j ’s fee and the capacities. Section 4 gives the intuition. The fee rule (12) in text implicitly defines the reactions. By writing it in implicit form and using the implicit function theorem, one finds the slopes of the reactions:

$$w_1 = -f_i + q \cdot (\partial c_i / \partial q + \partial c_j / \partial q + d_1) = -f_i + qY_1 = 0;$$

where $Y_1 > 0$ follows (10), q is the number of users, and f_i is i ’s fee.

This gives for the reaction to f_j :

$$\begin{aligned} \frac{\partial f_i^R}{\partial f_j} &= \frac{\partial w_1 / \partial f_j}{\partial w_1 / \partial f_i} = - \frac{(\partial q^R / \partial f_j)(\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q))}{-1 + \partial q^R / \partial f_i (\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q))}, \\ &= - \frac{(-1/Y_1)(Y_1 + Y_2)}{-1 + (-1/Y_1)(Y_1 + Y_2)} = -1 + \frac{Y_1 + Y_2}{2Y_1 + Y_2} < 0; \end{aligned} \quad (26)$$

where $Y_2 = q(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q) \geq 0$. This equation not only implies that the fee of i decreases with j ’s fee, but also that $0 > \partial f_i / \partial f_j \geq -1/2$.

The slope of the response function of i ’s fee to its own capacity follows

$$\begin{aligned} \frac{\partial f_i^R}{\partial s_i} &= - \frac{\partial w_1 / \partial s_i}{\partial w_1 / \partial f_i} = - \frac{q \frac{\partial^2 c_i}{\partial q \partial s_i} + \left(\frac{\partial q^R}{\partial s_i} \right) (\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q))}{-1 + \partial q^R / \partial f_i (\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q))}, \\ &= - \frac{q \frac{\partial^2 c_i}{\partial q \partial s_i} + \left(-\frac{\partial c_i}{\partial s_i} \frac{1}{Y_1} \right) (Y_1 + Y_2)}{-1 + (-1/Y_1)(Y_1 + Y_2)} = \frac{q \frac{\partial^2 c_i}{\partial q \partial s_i} + \left(-\frac{\partial c_i}{\partial s_i} \right) (1 + Y_2 / Y_1)}{2 + Y_2 / Y_1}. \end{aligned} \quad (27)$$

In the last version of (27), the denominator is positive: since Y_1 and Y_2 are positive (moreover, the denominator equals one minus the second-order condition of i ’s capacity, and this condition is

negative since the chosen capacity maximises capacity). In the nominator, the first term is negative and the second positive. The first term measures that a capacity increase lowers the MEC part of the fee, the second measures the effect of the increased number of users. For general costs, the slope in (27) could be positive. But if we impose (a.iii), we can be sure that the slope is non-positive and, unless costs are linear in q/s_i (i.e. $n=1$), f_i decreases with s_i :

$$\left. \frac{\partial f_i^R}{\partial s_i} \right|_{(a.iii)} = - \frac{(n-1)n \cdot \delta_i \cdot d_1 \cdot s_j^n}{s_i \left(2d_1 \cdot q \cdot s_i^n \cdot s_j^n + (n-1)n \cdot q^n (\delta_i s_j^n + \delta_j s_i^n) \right)} \leq 0. \quad (28')$$

Here, δ_i and n are parameters of the cost function, and d_1 of the inverse demand.

Turning to the response of f_i to s_j , one finds

$$\begin{aligned} \frac{\partial f_i^R}{\partial s_j} &= - \frac{\partial w_1 / \partial s_j}{\partial w_1 / \partial f_i} = - \frac{q \frac{\partial^2 c_j}{\partial q \partial s_j} + \left(\frac{\partial q^R}{\partial s_j} \right) \left(\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q \left(\partial^2 c_j / \partial^2 q \right) \right)}{-1 + \left(\partial q^R / \partial f_i \right) \left(\partial c_i / \partial q + \partial c_j / \partial q + d_1 + q \left(\partial^2 c_i / \partial^2 q + \partial^2 c_j / \partial^2 q \right) \right)}, \\ &= - \frac{q \frac{\partial^2 c_j}{\partial q \partial s_j} + \left(-\frac{\partial c_j}{\partial s_j} \frac{1}{Y_1} \right) (Y_1 + Y_2)}{-1 + (-1/Y_1)(Y_1 + Y_2)} = \frac{q \frac{\partial^2 c_j}{\partial q \partial s_j} + \left(-\frac{\partial c_j}{\partial s_j} \right) (1 + Y_2 / Y_1)}{2 + Y_2 / Y_1}. \end{aligned} \quad (28)$$

This derivative is negative under the same conditions for $\partial f_i / \partial s_i$. The effect of s_j on f_i and f_j is the same as the two fees are equal: $\partial f_i / \partial s_j = \partial f_j / \partial s_j$.

The Nash-equilibrium fees for given capacities $f_i^{NE}(s_i, s_j)$ and $f_j^{NE}(s_i, s_j)$ are found at the intersection of the response functions to the competitor's fee:

$$\begin{aligned} f_i^{NE}(s_i, s_j) &= f_i^R(s_i, s_j, f_j^{NE}), \\ f_j^{NE}(s_i, s_j) &= f_j^R(s_i, s_j, f_i^{NE}). \end{aligned}$$

To find the derivatives of Nash-equilibrium fees to capacities, one differentiates this system and solves the result:

$$\begin{aligned} \frac{\partial f_i^{NE}}{\partial s_i} &= \frac{\partial f_i^R / \partial s_i + \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial s_i}{1 - \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial f_i}, \\ \frac{\partial f_j^{NE}}{\partial s_i} &= \frac{\partial f_j^R / \partial s_i + \partial f_j^R / \partial f_i \cdot \partial f_i^R / \partial s_i}{1 - \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial f_i}. \end{aligned} \quad (29)$$

The equilibrium fees generally decrease with the capacities. Using (a.iii), it can be shown that the derivatives are certainly non-positive and they are only zero when congestion costs are linear in the ratio q/s_i , as then fees are independent of capacity).

A.2. *When does the Stackelberg leader set a higher capacity than with the closed-loop game?*

This section discusses when the leader's capacity rule in (16) implies that it sets a higher capacity in order to induce the follower to set a higher capacity and thereby increase profit. For this to hold the last term in (16) needs to be positive:

$$\frac{\partial s_j}{\partial s_i} \left(\frac{\partial f_i^{NE}}{\partial s_j} q + \frac{\partial q^R}{\partial s_j} f_i \right) > 0. \quad (30)$$

This term measures the effects of the induced change in the follower's capacity on the marginal revenue from capacity expansion. If the sum of the two terms between brackets is positive, this implies that a higher s_j increases i 's revenue. However, this sign of the sum is uncertain even with assumption (a.iii):

$$\begin{aligned} \left(\left(\frac{\partial f_i^{NE}}{\partial s_j} q \right) + \left(\frac{\partial q^R}{\partial s_j} f_i \right) \right) \Big|_{\text{(a.iii)}} &= \left(-\frac{d_1(n-1)nq^{2+n}s_i^n\delta_j}{s_j(3d_1qs_i^n s_j^n + n(1+2n)q^n(s_i^n\delta_j + s_i^n\delta_j))} \right) + \left(n\delta_j \frac{q^{1+n}}{s_j^{1+n}} \right) \\ &= \frac{nq^{1+n}\delta_i \left[-(n-4)d_1 \cdot q \cdot s_i^n s_j^n + n(1+2n)q^n(s_j^n\delta_i + s_i^n\delta_j) \right]}{s_j^{1+n} (3d_1qs_i^n s_j^n + n(1+2n)q^n(s_i^n\delta_j + s_i^n\delta_j))}. \end{aligned} \quad (31)$$

Here, the denominator is positive. Hence, (31) is positive if the term between square brackets in the nominator is positive. Thus, the question is whether the below condition holds:

$$-(n-4)d_1 \cdot q \cdot s_i^n s_j^n + n(1+2n)q^n(s_j^n\delta_i + s_i^n\delta_j) > 0. \quad (32)$$

Here, the second term always is positive; the first term is non-negative when $n \leq 4$, as then $-(n-4) \geq 0$, and negative when $n > 4$. Hence, (32) is positive under $n \leq 4$ (as is the case for the popular linear and BPR forms); for $n > 4$ it might be negative. Still, even with $n > 4$, (32) is often positive, as the first (negative) term depends on n while the second (positive) term on n^2 . However, if d_1 is much larger than δ_i and δ_j and/or the ratio q_i/s_i or q_j/s_j is very low, it might be negative.

For the leader to set a higher capacity, j 's capacity should also increase with i 's. This is a logical outcome, as a higher s_i attracts more users, which in turn increases congestion on j , making expansion more attractive for j . There is, however, also the counteracting force that i 's capacity increase lowers j 's fee, making expansion less rewarding.

To investigate this I first write (15)—which gives follower j 's capacity rule—in implicit form and then insert $f_j \cdot \partial q / \partial s_j$ for $-q \cdot \partial c_j / \partial s_j$:

$$w_2 = -k + f_j^{NE} \frac{\partial q}{\partial s_j} + \frac{\partial q^R}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_j} f_j^{NE} = 0.$$

Then, by using the implicit function theorem, one gets

$$\frac{\partial s_j}{\partial s_i} = -\frac{\partial w_2 / \partial s_i}{\partial w_2 / \partial s_j};$$

where the dominator is the second-order condition of j 's capacity choice. Hence, $\partial s_j / \partial s_i$ is positive if the nominator, $\partial w_2 / \partial s_i$, is positive:

$$\partial w_2 / \partial s_i = \left(\underbrace{\frac{\partial^2 q}{\partial s_j \partial s_i}}_{>0} + \underbrace{\frac{\partial q}{\partial f_i}}_{<0} \underbrace{\frac{\partial^2 f_i^{NE}}{\partial s_j \partial s_i}}_{?} + \underbrace{\frac{\partial^2 q}{\partial f_i \partial s_i}}_{<0} \underbrace{\frac{\partial f_i^{NE}}{\partial s_j}}_{?} \right) f_j + \left(\underbrace{\frac{\partial q}{\partial s_j}}_{>0} + \underbrace{\frac{\partial q}{\partial f_i}}_{<0} \underbrace{\frac{\partial f_i^{NE}}{\partial s_j}}_{?} \right) \underbrace{\frac{\partial f_j^{NE}}{\partial s_i}}_{?}.$$

For general functions, the sign of this is unknown. But under assumption (a.iii) it simplifies to

$$\frac{\partial w_2}{\partial s_i} \Big|_{\text{(a.iii)}} = \frac{n^2(1+2n)q^{1+2n}\delta_i\delta_b(d_1^2(2+n^2)q^2s_i^{2n}s_j^{2n} + d_1n(1+n)(1+2n)q^{1+n}s_i^n s_j^n (s_j^n \delta_i + s_i^n \delta_j) + n^3(1+2n)q^{2n}(s_j^n \delta_i + s_i^n \delta_j)^2)}{s_i s_j (d_1 q s_i^n s_j^n + n q^n (s_j^n \delta_i + s_i^n \delta_j)) (3 d_1 q s_i^n s_j^n + n(1+2n)q^n (s_j^n \delta_i + s_i^n \delta_j))^2} > 0;$$

and thus then s_j increases with s_i .

Appendix B: Mathematical discussion for parallel facilities

B.1. Best response functions of a parallel facility's fee

The fee equation (22) in text implicitly determines the response function (superscript R indicates a response): $f_i^R[s_i, s_j, f_j]$. To see the effects on f_i from changes in s_i , s_j and f_j , I write (22) in implicit form and use the implicit function theorem:

$$w_3 = -f_i + q_i \cdot \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2 \cdot d_2}{\partial c_j / \partial q_j + d_1} \right) = 0. \quad (33)$$

For the response of f_i to f_j this gives

$$\begin{aligned} \frac{\partial f_i^R}{\partial f_j} &= -\frac{\partial w_3 / \partial f_j}{\partial w_3 / \partial f_i} = -\frac{\frac{\partial q_i^R}{\partial f_j} \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^2}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j^R}{\partial f_j} \left(q_i d_2^2 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2} \right)}{-1 + \frac{\partial q_i^R}{\partial f_i} \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^2}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j^R}{\partial f_i} \left(q_i d_2^2 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2} \right)} \\ &= \frac{d_2 \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) - q_i d_2^2 \frac{\partial^2 c_j}{\partial^2 q_j} \frac{\partial c_i / \partial q_i + d_1}{(\partial c_j / \partial q_j + d_1)^2}}{Z_1 + \left(\frac{\partial c_j}{\partial q_j} + d_1 \right) \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + q_i d_2^3 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2}}; \quad (34) \end{aligned}$$

where the last version uses eqs. (17) and (18) for the derivatives of q_i and q_j , the definition from (22) for τ_i , and Z_l follows (19). The denominator is positive, since it is Z_l minus the second-order condition of f_i divided by Z_l . In the nominator, the first term is positive, as it gives the fee-increasing effect that q_i is higher when f_j is higher; the second term is negative, as it gives the fee-lowering effect of the lowered q_j . The $\partial f_i/\partial f_j$ is typically positive, as the first effect is typically stronger than the second; only when the firms are very ex-post asymmetric, it may be negative.

If the outcome is symmetric (i.e. firms have the same cost function, usage levels and capacities), $\partial f_i/\partial f_j$ is between 0 and 1. Moreover, if costs follow assumption (a.iii) then $0 > \partial f_i/\partial f_j > -1$ holds even for asymmetric firms and cost functions.

The response of i 's fee to its capacity is

$$\begin{aligned} \frac{\partial f_i^R}{\partial s_i} &= \frac{\partial w_3 / \partial s_i}{\partial w_3 / \partial f_i} = - \frac{q_i \frac{\partial^2 c_i}{\partial q_i \partial s_i} + \frac{\partial q_i^R}{\partial s_i} \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^2}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j^R}{\partial s_i} \left(q_i d_2^2 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2} \right)}{-1 + \frac{\partial q_i^R}{\partial f_i} \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^2}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j^R}{\partial f_i} q_i \left(d_2^2 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2} \right)} \\ &= \frac{Z_1 \cdot \frac{\partial^2 c_i}{\partial q_i \partial s_i} q_i + \left(-\frac{\partial c_i}{\partial s_i} \right) \frac{\partial c_j}{\partial q_j} \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial c_j}{\partial s_j} \left(\frac{\partial^2 c_j}{\partial^2 q_j} \frac{q_i d_2^3}{(\partial c_j / \partial q_j + d_1)^2} \right)}{Z_1 + \left(\frac{\partial c_j}{\partial q_j} + d_1 \right) \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + q_i d_2^3 \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2}}. \end{aligned} \quad (35)$$

In the last version of (35), the denominator is again positive. In the nominator, the first term is negative; the second positive; and the last negative or zero. The nominator is typically negative: when costs follow (a.iii), the nominator is non-positive, and it is strictly negative unless $n=1$ (see De Borger and Van Dender (2006) for such linear costs). For general cost functions, $\partial f_i/\partial s_i$ is typically negative, unless the firms are very ex-post asymmetric.

The $\partial f_i/\partial s_i$ is typically negative:

$$\begin{aligned} \frac{\partial f_i^R}{\partial s_j} &= \frac{\partial w_3 / \partial s_j}{\partial w_3 / \partial f_i} = - \frac{\frac{\partial^2 c_j}{\partial q_j \partial s_j} \left(\frac{q_i d_2^2}{(\partial c_j / \partial q_j + d_1)^2} \right) + \frac{\partial q_i^R}{\partial s_j} \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j^R}{\partial s_j} \left(\frac{\partial^2 c_j / \partial^2 q_j \cdot q_i \cdot d_2^2}{(\partial c_j / \partial q_j + d_1)^2} \right)}{-1 + \frac{\partial q_i^R}{\partial f_i} \left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^2}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + q_i d_2^2 \frac{\partial q_i^R}{\partial f_i} \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2}} \\ &= \frac{\frac{\partial^2 c_j}{\partial q_j \partial s_j} \left(\frac{Z_1 q_i d_2^2}{(\partial c_j / \partial q_j + d_1)^2} \right) + \frac{\partial c_j}{\partial s_j} d_2 \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial c_j}{\partial s_j} \left(\frac{\partial c_i}{\partial s_i} + d_1 \right) \left(\frac{\partial^2 c_j / \partial^2 q_j \cdot q_i \cdot d_2^2}{(\partial c_j / \partial q_j + d_1)^2} \right)}{Z_1 + \left(\frac{\partial c_j}{\partial q_j} + d_1 \right) \left(\tau_i + q_i \frac{\partial^2 c_i}{\partial^2 q_i} \right) + \frac{\partial q_j}{\partial f_i} q_i \cdot d_2^2 \cdot \frac{\partial^2 c_j / \partial^2 q_j}{(\partial c_j / \partial q_j + d_1)^2}}. \end{aligned} \quad (36)$$

Here, the denominator is negative, while in the nominator the first and second terms are negative, and the third non-positive. The first term gives that the higher s_j makes j more competitive and thereby decreases i 's mark-up; the second gives that the induced decrease of q_i lowers the MEC and mark-up parts of the fee, and finally the third term gives the effect that induced increase of q_j increases the user cost on j which tends to enable i to ask a higher fee (only for linear in q_j/s_j congestion is this third term zero). If we again add assumption (a.iii), then we can be certain that $\partial f_i / \partial s_j$ is negative.

The Nash-equilibrium (NE) fees for given capacities, $f_A^{NE}(s_A, s_B)$ and $f_B^{NE}(s_A, s_B)$, are at the intersection of the response functions. The general functions are the same as for the serial case in (29). The fees generally decrease with the capacities, and are certain to do so under (a.iii).

B.2. Capacity rule with the closed-loop game for parallel firms

The capacity rule is found by directly maximising profit to s_i :

$$\Pi_i = f_i \cdot q_i - k \cdot s_i;$$

which gives and f.o.c. for capacity of

$$\frac{\partial \Pi_i}{\partial s_i} = 0 = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q_i + \frac{dq_i^R}{ds_i} f_i - k = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q_i + \left(\frac{\partial q_i^R}{\partial s_i} + \frac{\partial q_i^R}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_i} + \frac{\partial q_i^R}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} \right) f_i - k. \quad (37)$$

When directly maximising profit, the f.o.c. for the fee in the first stage is

$$f_i \cdot \partial q_i^R / \partial f_i + q_i = 0; \quad (38)$$

which can be rewritten to fee condition in (22) by using (17). By inserting (20) for $\partial q_i^R / \partial s_i$ and rewriting, one gets that $f_i \cdot \partial q_i^R / \partial s_i = -q_i \cdot \partial c_i^R / \partial s_i$. From (38) one gets that $f_i \cdot \partial q_i^R / \partial f_i = -q_i$. Inserting all this into (37), and rewriting, results in condition (24) in text.

The effect $\partial f_i^{NE} / \partial s_i \cdot q_i$ —which gives change in revenue due to the change in fee—is cancelled out by the indirect effect $\partial q_i^R / \partial f_i \cdot \partial f_i^{NE} / \partial s_i \cdot f_i = -\partial f_i^{NE} / \partial s_i \cdot q_i$ —which gives the effect of the change in number of users due to this marginal fee change. This is the same as with serial facilities.