Serial private infrastructures#

Version of 14 August 2013

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 private supply of two congestible infrastructures that are serial, where the consumer has to use both in order to consume. Four market structures are analysed: a monopoly and 3 duopolies that differ in how firms interact. It is well known that private supply leads too high usage fees, and that a serial duopoly leads to even higher fees than a monopoly, as firms are monopolists on their sections. But, as this paper finds, a duopoly can also lead to a different capacity rule than the first-best one, and this distortion differs from with two parallel facilities. Finally, four auction formats for the right to build and operate facilities are investigated. With a bid auction, the competition is on how much to pay the government. This auction leads to the same outcome as no auction. An auction on the facility's capacity leads to an even lower welfare than no auction, as firms set overly large high capacities. Conversely, an auction on the generalised price or number of users leads to the first-best outcome, which they do when the facilities go to one or two winners and both with serial as with parallel facilities.

Keywords: private supply, congestible facilities, auctions, serial facilities, parallel facilities, imperfect substitutes
JEL codes: D43, L13, L51, R41, R42

#This is a post-print of the article forthcoming in Transportation Research Part B, see http://www.journals.elsevier.com/transportation-research-part-b-methodological/ and http://www.sciencedirect.com/science/journal/01912615

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

1. Introduction

There is a rising interest in the private supply of facilities such as airports, roads, energy networks, public transport, telecommunication 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 makes use of substantial market power.

The literature on private supply tends to focus on a monopoly, or a duopoly on parallel facilities that are perfect substitutes. This paper investigates two serial facilities, where a user has to use both facilities in order to consume. This setting is very 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 driving or using public transport, one often uses different facilities in succession, such as two roads or bus and then rail. For comparison, this paper also looks at parallel facilities, and generalises this setting to imperfect substitutes.[1] This paper investigates four market structures: a monopoly and three duopolies that differ in how firms interact. With the first "open-loop" duopoly, firms take the actions of the others 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 Bertrand-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. Table 1 summarises the market structures for ease of reference.

Table 1: The 4 market structures

Monopoly	One firm controls both facilities
Open-loop duopoly	In this single-stage Nash game, firms set their fees and capacities at the same time.
Closed-loop duopoly	In this two-substages Nash game, firms first set capacities and then fees. The capacity setting takes into account the effect on Nash-equilibrium fees.
Stackelberg duopoly	Capacities are set sequentially and then fees are set in a Bertrand-Nash fashion. The leader's capacity setting considers the follower's capacity choice and the Nash-equilibrium fees.

The 3 serial duopolies have the same rule for fee setting, which leads to even higher fees than with a monopolist: due to the "double marginalisation" there are basically two monopolies after each other.[2] The monopolist and "open-loop" duopolist use the same capacity rule as the first-best

[1] Imperfect substitutes are common: for example, the air or seaports of an area differ in their facilities and accessibility; when choosing between travelling by air, 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).

[2] This is a general result. Economides and Salop (1992) show, without congestion, that prices are always higher with a serial duopoly than a monopoly, as a duopolist ignores that its mark-up lowers the other's profit. Airlines may use code-sharing to limit double marginalisation and lower total fees paid and increase profits. Conversely, with substitute airline cooperative code-sharing raises prices (see, e.g. Bruechner, 2001).

case. With serial facilities, a “closed-loop” duopolist sets a higher capacity in order to lower the competitor’s Nash-equilibrium fee and this in turn raises its profit by attracting more users. The lower Nash-equilibrium fees also increase welfare. The serial Stackelberg leader sets an even higher capacity in order to raise the competitor’s capacity and thereby lower its Nash-equilibrium fee further. Conversely, with parallel facilities, a “closed-loop” firm sets a lower capacity to raise the fee of the other (see also De Borger and Van Dender, 2006), while the Stackelberg leader sets a higher capacity (see also Van den Berg and Verhoef, 2012). The competition with parallel facilities also means that fees are lower than with a monopolist and thus welfare is higher, while with serial facilities the fees are higher with competition. Therefore, the market structures have very different effects in the two network settings.

Private supply leads to too high fees, and in some settings a different congestion level than first-best. To overcome this, the government could regulate: e.g., set a maximum price. But for this the regulator needs information: what are consumers willing to pay, and what are the costs? This can be difficult and costly to ascertain. An alternative—where the regulator needs no information—is regulation via auctions: the competition in the auction will endogenously determine the outcome.

We study four perfectly-competitive auctions on: bid (i.e. how much to pay to the government), capacity, patronage (i.e. number of users) and generalised price (which is usage cost plus fee[3]). The two facilities can be auctioned off to a single firm or to two (where the facilities are auctioned simultaneously). Accordingly, following the auction there are still 4 types of market structures. As the auction is perfectly competitive, a winning firm makes a zero profit.[4] A bid auction leads to the same capacities and fees as no regulation, as this gives the highest profit to pay to the government. The capacity auction leads to very high capacities, and has a much lower welfare than no intervention. The patronage and generalised-price auctions result in the first-best outcome. Hence, these auctions maximise welfare without the need for acquiring data or regulation.

These results are qualitatively the same as with a single link. But the auction types also differ in how robust their effects are to the network structure, if there are one or two winners, and the market structure. The generalised-price and patronage auctions attain the first-best outcome regardless of these issues; the relative effects of the other two auctions depend on them. These auctions have been analysed before for a single facility or with an unpriced alternative by, for example, Ubbels and Verhoef (2008) and Verhoef (2007). But not for when multiple facilities are auctioned, and in reality there are typically multiple facilities that interact.[5] Wu et al. (2011a) analyse perfectly-competitive auctions in a general network with open-loop competition, but not for the more realistic closed-loop and Stackelberg games. There have also been studies on

[3] Here, usage fee is not a cost, since it is transferred from the user to the operator and not lost.
[4] There can be a “normal profit” or market conform return to capital included in the costs.
[5] Verhoef (2008) studies a sequential-entry market structure which is related to the Stackelberg game: the auctioning of facilities is sequential, 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 make a loss when a new entry occurs, as after they entered (but before the next entry) they made a zero profit.

unregulated monopolistic supply or duopolistic supply with perfect substitutes, but not with serial or imperfect-substitute parallel facilities.[6]

Table 2: The 4 auction formats

Bid auction	Bidders compete on how much they pay to the government.
Capacity auction	Bidders compete on the capacity on the infrastructure they will build.
Generalised-price auction	Bidders compete on the generalised price the consumer will face.
Patronage auction	Bidders compete on the patronage (number of users) their facility will have.

The remaining paper is outlined as follows. Section 2 discusses the literature, and the Section 3 this paper's model set-up and assumptions. Section 4 turns to the first-best and monopoly cases, before Sections 5 and 6 analyse the serial and parallel duopolies. The numerical example is presented by Section 7. Section 8 discusses the auctions. Finally, Section 9 discusses some caveats to the research and directions for future research, and Section 10 concludes.

2. Literature review

2.1. Review of the different market structures

Before turning to the modelling, it is important to discuss the different types of market structures and the literature on them. In the first-best case, the 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 for a given number of users. As showed by Mohring and Harwitz (1962), under some general assumptions that hold here, the first-best outcome leads to zero profit and thus the system is self-financing.

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. But firms do add a mark-up to the fee so as to maximise their profit (Zhang and Zhang, 2006; Basso and Zhang, 2007; Xiao et al., 2007; Wu et al., 2011b).

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 than with open-loop game, and they do so to induce higher Nash-equilibrium fees. This also means that usage cost is higher than in the first-best and welfare is further reduced. Conversely, this paper shows that, with serial facilities, firms set higher capacities, as this lowers the competitor's fee and this raises profit by attracting more users. The Stackelberg leader's capacity choice considers the effects on the follower's capacity and the Nash fee setting. We find that with serial facilities, under some common

[6]A related literature studies different local governments supplying serial roads for trough traffic and local traffic (e.g. De Borger et al., 2007). Yuen et al. (2008) study two serial facilities: a gateway (e.g. airport or port) that is used by carriers and an access road that is used by some of the gateway users and local traffic. The gateway facility maximizes the joint profits of the gateway and carriers, the access road maximizes its local welfare.

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. Then, the actions of the intermediate firms also need to be modelled (Zhang and Zhang, 2006; Basso and Zhang, 2007; Mantin, 2012; Matsumura and Matsushima, 2012), 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; Wan and Zhang, 2012), such as transport facilities to an (air/sea)port or register capacity at a supermarket.

2.2. Auctions versus direct regulation or negotiation

This paper uses perfectly competitive auctions. In reality auctions are not perfectly competitive: there tend to be a limited number of bidders (e.g. Lalive and Schmutzler, 2011) and bidders are heterogeneous (e.g. have different (expectations of) demand and cost conditions). An auction with imperfect competition will leave some rent (i.e. expected profit) for the winner (just as a market with imperfect competition leaves profits for the firm); but the outcome approaches the one here with perfect competition as the auction becomes more competitive. An advantage of auctions over direct negotiation or regulation is that they ensure that the most efficient firm wins the franchise (see, e.g., Krishna, 2010). Moreover, the competition in the auction limits the rents and leads to a better outcome for the government than direct negotiation. This holds even if the auction is most-simply designed—for which no information is needed—while the negotiation procedure is optimally designed, for which large amounts of costly information gathering and designing are needed (Bulow and Klemperer, 1996).

An issue with regulation is that if some instruments remain unregulated the choices of these may be distorted. For instance, Averch and Johnson (1962) find (for a facility without congestion) that a rate-of-return (ROR) maximum (i.e. a max profit per unit of capital) distorts a monopolist's choice capital. Sheshinski (1976) shows that a maximum fee alters the choice of service quality (e.g. level of congestion). Spence (1975) studied when an unregulated monopolist sets the socially optimal quality and the effects of ROR regulation. Oum et al. (2004) study price and ROR regulation of a private airport, and their effects on quality.

To deal with distortions due to regulation, Harvey (2011) proposes a scheme where the government sets and receives the tolls while the operator receives a payment that is linear and depends on the realised user cost. The optimal linear contract ensures the first-best outcome. This scheme has an obvious relation with the linear contracts commonly used in procurement, where the payment depends on realised cost (see, e.g., Laffont and Tirole, 1993).

Tan et al. (2010) analyse Pareto optimal Build-Operate-Transfer contracts for private congestible facilities. They show that—under some assumptions that hold here—it is optimal to give the franchise for the entire lifespan of the facility, as extending the franchise allows for making the contract more favourable for welfare while keeping profit the same. This paper will, however, look at a static setting, and hence ignore the lifespan of the facility. They also investigate regulation and conclude that: i) ROR regulation leads to too few users and a too high capacity, ii) fee regulation leads to a too small capacity, iii) patronage and mark-up regulation lead to a Pareto optimum.

3. Assumptions and model set-up

There are two congestible facilities. The generalised 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 the model tractable, it is assumed that inverse demand is linear. With the imperfect-substitute facilities i and j of the **parallel case** (where i=1,2 and $j \neq i$), the inverse demand for i follows:

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

where q_i is i's number of users. The inverse demand measures the willingness to pay in terms of the generalised price. The d_0 is the maximum willingness to pay, d_1 measures the effect of the own number of users, and d_2 the cross effect (i.e. how much the willingness to pay for facility i decreases if there are more users of j).[7] With substitutes d_2 is positive: if it were negative, the facilities would be complements, and the users of i would accept a higher generalised price when there are more users of j. If d_1=d_2, the facilities are perfect substitutes; when d_1>d_2, they are imperfect substitutes; when d_2=0, the two demands are independent. Assuming that there are no income effects, consumer benefit (B) is the line integral of the two inverse demands and is independent of the path used for integration:[8]

$$B = \int_{(0\ ,\ 0)}^{(q_i,\ q_j)} \Big(D_i(x_i, x_j)\mathrm{d}x_i + D_j(x_j, x_i)\mathrm{d}x_j \Big) = q_i \cdot (d_0 - \frac{d_1 \cdot q_i}{2}) + q_j \cdot (d_0 - \frac{d_1 \cdot q_j}{2} - q_i \cdot d_2). \tag{2}$$

Conversely, with **serial facilities**, inverse demand follows:

$$D(q) = d_0 - d_1 \cdot q; \tag{3}$$

[7] These parameters are assumed to be the same for both facilities for notational conveniance, but this assumption is not crusual. It must hold that $d_1 \geq d_2 \geq 0$.

[8] See the textbook Mas-Colell et al. (1997); or, for a transport application, Kraus (2003).

and consumer benefit (B) is the integral the single inverse demand:

$$B = \int_0^q \left(D(x) \mathrm{d}x \right) = q \cdot (d_0 - \frac{d_1 \cdot q}{2}).$$

For user equilibrium with parallel facilities, the generalised price of each facility has to equal its inverse demand; with serial facilities, the inverse demand has to equal the generalised price of using both.

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

$$C_i^{cap} = k \cdot s_i; \tag{4}$$

where k is the cost of a unit of capacity. Both with serial as substitute facilities, consumer surplus is consumer benefit minus total usage cost (i.e. $c_i \cdot q_i + c_j \cdot q_j$) and total fee payments (i.e. $f_i \cdot q_i + f_j \cdot q_j$). Profit for a facility is total fee payment minus its capacity cost:

$$\Pi_i = f_i \cdot q_i - k \cdot s_i. \tag{5}$$

Welfare (W) equals consumer benefit minus total usage and capacity cost (or equivalently consumer surplus plus profits):

$$W = B - q_i \cdot c_i - q_j \cdot c_j - k \cdot s_i - k \cdot s_j; \tag{6}$$

where c_i is the usage cost of facility i. It is thus assumed that taxation is costless.

The facilities are 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. For ease of exposition, the discussion assumes the following functional form:

$$c_i = \chi_i + \delta_i \left(q_i / s_i \right)^n . \tag{a.i}$$

Here, χ_i and δ_i are facility specific positive constants, and n is the same for all. The cost function (a.i) is 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 the equilibrium usage cost with Vickrey (1969) bottleneck congestion. Throughout we will assume that there is an interior solution, and ignore the possibility of less interesting corner solutions.

4. First-best and monopolistic outcomes

This section briefly reviews the first-best and monopoly cases (for an extensive overview see Small and Verhoef (2007)). In all cases, there are two facilities. In the *First-Best* (FB) outcome, which maximises welfare, the fee of a facility equals the marginal external cost (MEC) on it, where the MEC is marginal social cost external to the user's choice (i.e. they equal the derivative of total cost minus usage cost). Such a fee ensures that users consider all marginal social costs in their decision. Moreover, capacity is set to minimise social cost by equating marginal capacity cost, *k*, to the usage cost decrease from the marginal capacity expansion:

$$f_i^{FB} = q_i \cdot \partial c_i / \partial q_i, \tag{7}$$
$$k = -q_i \cdot \partial c_i / \partial s_i. \tag{8}$$

A monopolist also uses capacity rule (8). The intuition is as follows, for given number of users, a capacity increase decreases cost and this allows the fee to be increased by this cost decrease. The firm hence internalises usage costs, as any decrease in total usage cost can be met by an equal increase in toll revenue. The monopolist even has 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 (see also Xiao et al., 2007).[9]

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

$$f_i^{PM} = q_i \cdot \partial c_i / \partial q_i + q_i \cdot d_1 + q_j \cdot d_2. \tag{9}$$

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*. The closer substitutes the facilities are (i.e. d_2 is higher for given d_1), the higher the fees, since this increases the strength of the third effect.

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

$$f_{A+B}^{SM} = q \cdot \partial c_A / \partial q + q \cdot \partial c_B / \partial q + q \cdot d_1. \tag{10}$$

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

5. 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 generalised 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 an open-loop game (where fee and capacity are set at the same

[9] Moreover, since the fee is higher than the Marginal External Cost, given the fee, the welfare-maximising capacity would be higher than what the monopolist sets in order to correct for the fee leading to too few users (Small and Verhoef, 2007, p.172).

time). Conversely, a parallel firm typically sets a lower capacity as it wants to increase the other facility's fee. In the serial Stackelberg game, the leader sets a higher capacity, as it wants a higher follower's capacity as this increases the number of users. Note that in all settings we look at a non-cooperative equilibrium; if the duopolists would collude, they could attain higher profits and with the serial case this even raises welfare as the duopolistic fees are above the monopolistic ones.

5.1. The number of users and changes in capacities and fees

As one would expect, in equilibrium the 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 generalised price of using both facilities: $D=c_i+c_j+f_i+f_j$. By differentiating this user-equilibrium condition w.r.t. the two fees and two capacities, and solving the resulting four equations, one can find how the user-equilibrium q changes with these variables (where superscript u indicates the user equilibrium):

$$\frac{\partial q^u}{\partial f_i} = \frac{\partial q^u}{\partial f_j} = -\frac{1}{Y_1} < 0, \tag{11}$$

$$\frac{\partial q^u}{\partial s_i} = -\frac{\partial c_i / \partial s_i}{Y_1} = \frac{\partial c_i}{\partial s_i}\frac{\partial q^u}{\partial f_i} > 0; \tag{12}$$

and Y_1 follows:

$$Y_1 = \partial c_i / \partial q + \partial c_j / \partial q + d_1. \tag{13}$$

As the third part of (12) 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 generalised price due to the capacity change.

5.2. Fee setting

In setting its fee, firm *i* takes the capacities and fee of *j* as given and maximises its profit w.r.t. its fee:

$$\Pi_i = f_i \cdot q - k \cdot s_i. \tag{14}$$

The first-order condition of the fee is $\partial \Pi_i / \partial f_i = q + f_i \cdot \partial q_i^u / \partial f_i = 0$. Inserting (11) for $\partial q_i/\partial f_i$ and rewriting, gives the profit-maximising fee rule:[10]

$$f_i = f_j = q \cdot \left(\partial c_i / \partial q + \partial c_j / \partial q + d_1 \right) = q \cdot Y_1. \tag{15}$$

[10] The second-order conditions also hold.

A firm not only asks the MEC on its own facility, but also that on the other: any decrease in usage cost be it on *i* or *j*, for a given *q*, can be matched by a fee increase. The term $d_1 \cdot q_i$ in (15) 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. The total fee is thus even higher than with a monopoly due to the "double marginalisation" from each firm ignoring that increasing its price also lowers the profit of the other.

For the closed-loop duopoly and Stackelberg games, one needs to know how (Bertrand-)Nash-equilibrium fees in the second-stage (indicated by Superscript ^{NE}) change with the capacities. For this we need to know the best responses (indicated by superscript ^{R}) to changes in the competitor's fee and the capacities; the Nash-equilibrium fees for given capacities are at the intersection of the best-response functions:

$$\begin{aligned} f_i^{NE}\left(s_i,s_j\right) &= f_i^R\left(s_i,s_j,f_j^{NE}\right), \\ f_j^{NE}\left(s_i,s_j\right) &= f_j^R\left(s_i,s_j,f_i^{NE}\right). \end{aligned} \tag{16}$$

This section summarises the results, for mathematical derivations see Appendix A.1. It is the best response for *i* to decrease its fee with *j*'s fee, and $-½ \geq \partial f_i^R/\partial f_j \geq -1$. This is because a higher f_j lowers the number of users, and this lowers the MEC and mark-up parts of *i*'s fee.

The best response for *i* is typically to decrease its fee if its capacity is higher, as this lowers 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. It can be shown that $\partial f_i^R/\partial s_i \leq 0,$ and only if usage costs are linear in the ratio q/s_i (i.e. n=1), is the best response function flat.[11] Finally, it is a best response for *i* to decrease its fee with *j*'s capacity under the same conditions as for *i*'s capacity.

To find the derivatives of Nash-equilibrium fees to capacities, one differentiates system (16) and solves the result:

$$\begin{aligned} \frac{\partial f_i^{NE}}{\partial s_i} &= \frac{\partial f_i^R/\partial s_i + \partial f_i^R/\partial f_j \cdot \partial f_j^R/\partial s_i}{1-\partial f_i^R/\partial f_j \cdot \partial f_j^R/\partial f_i} \leq 0, \\ \frac{\partial f_j^{NE}}{\partial s_i} &= \frac{\partial f_j^R/\partial s_i + \partial f_j^R/\partial f_i \cdot \partial f_i^R/\partial s_i}{1-\partial f_i^R/\partial f_j \cdot \partial f_j^R/\partial f_i} \leq 0. \end{aligned} \tag{17}$$

Using the above discussion, it can be shown that the Nash-equilibrium fees decrease with the capacities, unless usage costs are linear in q/s_i when the fees are independent of capacity. The reason why NE prices generally decrease with capacity is that a capacity expansion lowers the congestion externality and therewith the part of the fee that ensures that users internalise the

[11] 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.

externality they impose. There is also a counteracting effect that the higher capacity also attracts more users which increases congestion and the mark-up; but only with linear congestion is the second effect as large as the first.

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

With the open-loop game, capacity and fee setting occur simultaneously, and firm *i* maximises its profit of (14) given the actions of the other. The first order condition for capacity setting is:

$$\frac{\partial \Pi_i}{\partial s_i} = \left(\frac{\partial q^u}{\partial s_i} \right) f_i \; - k = 0 \,; \tag{18}$$

which using (12) and (15) can be rewritten to the same capacity rule as with the first-best:

$$k = -q \cdot \partial c_i / \partial s_i . \tag{19}$$

Hence, just as with a monopolist, the firm internalises the usage costs.

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

With the closed-loop game, the Nash-capacity setting precedes the Nash-fee setting. The firm takes into account that if its sets a higher capacity this lowers the competitor's Nash-equilibrium fee that results from the following stage (unless we have linear congestion, when there is no effect), and thus increases profit by attracting more users. This leads firms to set higher capacities than with an open-loop game, and hence user cost is lower.

The firm again maximises (14) by optimising capacity but now the Nash-equilibrium fees are a function of s_i and, the taken as given, s_j. The first order condition for capacity is:

$$\frac{\partial \Pi_i}{\partial s_i} = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q + \frac{dq^u}{ds_i} \, f_i \; - k = \frac{\partial f_i^{NE}}{\partial s_i} \cdot q + \left(\frac{\partial q^u}{\partial s_i} + \frac{\partial q^u}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_i} + \frac{\partial q^u}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} \right) f_i \; - k = 0; \tag{20}$$

where q is the number of users, $\partial q^u/\partial f_i$ is the derivative of the number of users to *i*'s fee, and $\partial f_j^{NE} / \partial s_i$ the derivative of *j*'s Nash-equilibrium fee to *i*'s capacity. This can be simplified to:[12]

$$k = \underbrace{-\frac{\partial c_i}{\partial s_i} q}_{>0} + \underbrace{\overset{<0}{\frac{\partial q^u}{\partial f_j}} \overset{\leq 0}{\frac{\partial f_j^{NE}}{\partial s_i}} \overset{>0}{f_i}}_{\geq 0} \; . \tag{21}$$

[12] Using the f.o.c. for fee setting, one gets that $f_i \cdot \partial q^u/\partial f_i = -q$; and by using (12) and (12) one gets that $\left(\partial q^u / \partial s_i\right) f_i = \left(-\left(\partial c_i / \partial s_i\right)/Y_1\right)\left(q \cdot Y_1\right) = -\partial c / \partial s_i \cdot q > 0$. Inserting these two results into (20) and rewriting results in the simplified rule (21). 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, $(\partial q^u / \partial f_i)(\partial f_i^{NE} / \partial s_i) f_i = -\partial f_i^{NE} / \partial s_i \cdot q$. Hence, the effect of its capacity choice via its own fee on profit drops out. This occurs for any cost function, and, as we will see, it also holds for parallel facilities.

This equation for the closed-loop capacity rule differs from (19) for the open-loop game by the addition of the second term on the right side. This term is zero when costs are linear in q/s_i, as then $\partial f_j^{NE}/\partial s_i = 0$; for any other form of (a.i), the term is positive. Therefore, with linear costs, the open- and closed-loop games have the same outcome; while for other congestion functions, capacity is higher with a closed-loop game than a open-loop.[13] 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.[14] As both firms do this, capacities are higher than with the open-loop game; and thus fees are lower and welfare is higher.

5.5. 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 (21), 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. Its f.o.c. for capacity becomes:

$$\begin{aligned}\frac{\partial \Pi_i}{\partial s_i} &= \frac{df_i^{NE}}{ds_i}\cdot q + \frac{dq^u}{ds_i} f_i - k = 0 \\ &= \left(\frac{\partial f_i^{NE}}{\partial s_i} + \frac{\partial f_i^{NE}}{\partial s_j}\frac{\partial s_j^R}{\partial s_i}\right)\cdot q + \left(\frac{\partial q^u}{\partial s_i} + \frac{\partial q^u}{\partial f_i}\frac{\partial f_i^{NE}}{\partial s_i} + \frac{\partial q^u}{\partial f_j}\frac{\partial f_j^{NE}}{\partial s_i} + \frac{\partial q^u}{\partial s_j}\frac{\partial s_j^R}{\partial s_i}\right) f_i - k\end{aligned}; \tag{22}$$

The leader's f.o.c. for its capacity only differs from the follower's in the addition of the effect of the leader's capacity on the follower's capacity ($\partial s_j^R/\partial s_i$) and therewith the effect on the number of users ($\partial q^u/\partial s_j$) and the leader's fee ($\partial f_i^{NE}/\partial s_j$).

This makes the leader's capacity rule:

$$k = -\frac{\partial c_i}{\partial s_i} q + \frac{\partial q^u}{\partial f_j}\frac{\partial f_j^{NE}}{\partial s_i} f_i + \frac{\partial s_j^R}{\partial s_i}\left(\frac{\partial f_i^{NE}}{\partial s_j} q + \frac{\partial q^u}{\partial s_j} f_i\right); \tag{23}$$

which differs from (21) for the closed-loop setting by the addition of the third term on the right side. This section briefly describes how the capacity setting is affected by being a leader, Appendix A.2 gives the underlying 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 $\partial s_j^R/\partial s_i$ gives the induced change in s_j, and is positive.

The two terms between brackets measure the effects of this induced change in s_j on marginal revenue: the first item gives the profit-lowering effect that a higher s_j lowers *i*'s fee, the second

[13] If the capacity were the same, the right side of (21) would be larger than the left side. By increasing s_i, this equation is balanced again.

[14] Naturally, in the open-loop setting this effect on the competitor's fee of the own capacity also occurs, but as it then takes this fee as given, this does not affect the capacity choice

item gives the profit-increasing effect that more users will be attracted. The sign of the sum of these two terms between brackets is uncertain. However, when the power of the congestion function (n) is not larger than 4, the sum is positive; with n>4, this is likely, but not certain.

Concluding, for a congestion function with $n \leq 4$, the leader sets a higher capacity than with a Nash setting 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 functions, this is likely.

6. 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. Conversely, with serial facilities a firm wanted a lower competitor's fee and higher capacity.

6.1. 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 the number of users of i by increasing the generalised price. A higher fee or lower capacity on j increases the number of i by increasing j's price.

To prove this, we can use that the generalised prices on i and j should equal their respective inverse demands. Then, by differentiating these two user-equilibrium conditions to the two fees and two capacities, and solving the resulting system of 4 equations, one gets how the number of users in equilibrium changes with these instruments:[15]

$$\frac{\partial q_i^u}{\partial f_i} = -\frac{\partial c_j / q_j + d_1}{Z_1} < 0, \tag{24}$$

$$\frac{\partial q_j^u}{\partial f_i} = \frac{\partial q_i^u}{\partial f_j} = \frac{d_2}{Z_1} > 0; \tag{25}$$

where:

$$Z_1 = \left(\partial c_i / \partial q_i + d_1\right)\left(\partial c_j / \partial q_j + d_1\right) - d_2 d_2 > 0. \tag{26}$$

Similarly, the effects on the numbers of users of the capacities follow:

$$\frac{\partial q_i^u}{\partial s_i} = -\frac{\partial c_i}{\partial s_i}\frac{\partial c_j / q_j + d_1}{Z_1} = \frac{\partial c_i}{\partial s_i}\frac{\partial q_i}{\partial f_j} > 0, \tag{27}$$

$$\frac{\partial q_j^u}{\partial s_i} = \frac{\partial c_i}{\partial s_i}\frac{d_2}{Z_1} = \frac{\partial c_i}{\partial s_i}\frac{\partial q_j}{\partial f_i} < 0. \tag{28}$$

[15] 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_i/s_i (i.e. following (a.i) with n=1).

6.2. Fee setting (parallel facilities)

Firms set their fees in a Nash fashion and take capacities as given. The first-order condition of the fee for profit maximisation is $\partial \Pi_i / \partial f_i = q_i + f_i \cdot \partial q_i^u / \partial f_i = 0$, which is the same as with serial facilities except that the q now has an index i to show that it is the number of users of facility i and that $\partial q_i^u / \partial f_i$ follows (24). This results in the fee rule of:

$$f_i = q_i \cdot \partial c_i / \partial q_i + q_i \cdot d_1 - q_i \frac{d_2 \cdot d_2}{\partial c_j / \partial q_j + d_1}. \tag{29}$$

Note that firms compete on generalised prices and not quantity. However, they still make a profit, as facilities are imperfect substitutes and because there is congestion which softens the price competition. The first term in this rule equals the marginal external cost (MEC). The sum of the second and third term gives the mark-up: the second term is the mark-up for a monopolist of a single facility, the third term 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 the fees. With independent demands (i.e. d_2=0), the third term is zero, and the fee the highest; 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.

The Nash-equilibrium fees (as a function of the capacities), are again at the intersection of the best-response functions. The derivatives of the Nash-equilibrium fees follow:

$$\begin{aligned} \frac{\partial f_i^{NE}}{\partial s_i} &= \frac{\partial f_i^R / \partial s_i + \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial s_i}{1 - \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial f_i} < 0, \\ \frac{\partial f_j^{NE}}{\partial s_i} &= \frac{\partial f_j^R / \partial s_i + \partial f_j^R / \partial f_i \cdot \partial f_i^R / \partial s_i}{1 - \partial f_i^R / \partial f_j \cdot \partial f_j^R / \partial f_i} < 0. \end{aligned} \tag{30}$$

NE prices decrease with capacity because a capacity expansion lowers the congestion externality and therewith the part of the fee that ensures that users internalise the externality they impose. The derivatives of the best-response functions w.r.t. to the competitor's fee and the capacities can be found by writing (29) in implicit form, and using the implicit function theorem. Appendix B.1 provides the mathematical details, this section summarises the implications. This corresponds with the findings of De Borger and Van Dender (2006) for linear usage costs.

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

Now, each firm takes the fee and capacity of the other as given in setting its capacity. Maximising profit to capacity again results in basically the same f.o.c. as in the serial case of (31) (although the effects of capacity on the number of users follows (27)-(28) and the fee (29)). The first order conditions are also generally the same as in the serial case. Hence, these conditions are left out. Still, the signs of the effects in the f.o.c.'s and capacity rule may differ, and thus we will discuss the capacity rules. The open-loop capacity rule equals the first-best rule:

$$k = -q_i \cdot \partial c_i / \partial s_i \tag{32}$$

6.4. 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 Nash-equilibrium (NE) fees f_i^{NE} and f_j^{NE}. Appendix B.2 shows that the capacity rule under closed-loop competition is:

$$k = -\overbrace{q_i \cdot \partial c_i / \partial s_i}^{<0} + \overset{>0}{\frac{\partial q_i^u}{\partial f_j}} \overset{<0}{\frac{\partial f_j^{NE}}{\partial s_i}} \overset{>0}{f_i} \ . \tag{33}$$

Just as with serial competition, the only difference between the formula here and (32) for the open-loop game is the 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*. The new term is negative,[16] which implies that *i* sets a lower capacity than without the separate stages (since otherwise the right side of (33) could not equal the constant *k*). This also means that 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.

6.5. 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:

$$k = -\frac{\partial c_i}{\partial s_i} q_i + \frac{\partial q_i^u}{\partial f_j} \frac{\partial f_j^{NE}}{\partial s_i} f_i + \frac{\partial s_j}{\partial s_i} \left(\frac{\partial f_i^{NE}}{\partial s_j} q_i + \frac{\partial q_i^u}{\partial s_j} f_i \right). \tag{34}$$

Analytically, it difficult to say what effect is of the new concern in (34). One would expect that the leader would set a larger capacity to increase its market power, just as without congestion; [17] and this is also the numerical example below and Van den Berg and Verhoef (2012) find.[18]

[16] Following Section 5.2 and Appendix B.1, $\partial f_i / \partial s_j < 0$ is certain under (a.i). The other two items of the new term in (33) are positive for any congestible usage cost.

[17] 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 of (34) are typically negative, and hence the sum of the two is generally negative. The intuition is that a lower competitor's capacity increases your number of users and allows you to ask a higher fee, both of which increase profit.

6.6. Serial vs parallel facilities

As discussed, with serial facilities competition actually raises fees above those with a monopoly and thus lowers profits and welfare. Conversely, with parallel facilities, competition lowers fees and raises welfare relative to with one operator, but still lowers total profit. Comparing when capacity setting precedes fee setting (closed-loop duopoly) and when these are simultaneous decisions (open-loop): with parallel facilities, the closed-loop has a lower capacity than the open-loop and thus lower consumer surplus and welfare; with serial facilities, capacities are higher with the closed-loop than the open-loop. Still both with parallel as well as with serial facilities profits are higher in the closed-loop duopoly than the open-loop. With parallel facilities, to increase its market power a Stackelberg leader tends to sets a higher capacity than without the extra strategic consideration (i.e. in the closed-loop). With a serial leader tends to set a higher capacity than in the closed-loop to reduce the fee of the follower, this raises the profit of the leader, but the leader does make a smaller profit than the follower.

7. Numerical example

This section illustrates the model with a numerical example. The calibration in Table 3 builds on Verhoef (2007). For a base-case calibration without congestion pricing, there are 5000 users and the elasticity w.r.t. the own generalised 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.i) 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,[19] 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 Stackelberg games without auctions.

Table 3: Calibrations of the numerical example

	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
n	4	4

[18] 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.

[19] 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).

7.1. Serial facilities

Table 4 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 as all private games have much lower capacities. The regime has an arbitrarily chosen capacity, and is only there for comparison. In the *first-best* (FB) case, capacity is set following (8) 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 (10) adds a mark-up to the fee; but although the number of users and welfare are much lower, the usage cost is the same as in the first-best. Welfare with a monopolist is much lower than in the first-best case.

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, but the fees are still much higher with competition. 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.

Firms want their serial competitor to set a lower fee (for any level of this fee). To achieve this in the *closed-loop* game, each firm sets a higher capacity than it otherwise would: they set a capacity of 856 with the open-loop and now 1045. Hence, firms have lower usage costs and fees and attract more users than without this strategic consideration. Therefore, welfare is 13% higher in the closed-loop than the open-loop duopoly; although welfare is still far below that with a monopoly. 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 4: 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 duopoly. 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; and thus, 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 5, 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 lower than with open-loop game.

Comparing the *Stackelberg* and closed-loop games, it shows that the Stackelberg lowers welfare game even further, because the leader increases its market power with its capacity setting. The parallel leader makes a larger profit than the follower.

Table 5: 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.

8. Auctions

8.1 Theory

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. The auctions are perfectly competitive as biddings firms are ex ante the same before the auction and there are more bidders than facilities. Accordingly, the auction outcome, given the following competition, is there where firms attain zero profit.[20,21]

The four auction types are bid (highest payment to the government wins), capacity (highest offered capacity wins), generalised-price (lowest price wins), and patronage (highest number of users wins). These auctions have been studied for a single facility, but not for multiple auctioned facilities, whereas in reality there typically multiple facilities that interact. The outcome with a bid auction is the same as without an auction, but the profit is transferred to the government. With a capacity auction, the capacities are much larger than is socially optimal. The patronage and generalised-price auctions lead to the first-best outcome. These results correspond with those for a single facility in Verhoef (2007) and Ubbels and Verhoef (2008).

This section assumes that the government either has to give the facilities to one winner or to two. If it could choose between the two options, then with the price or patronage auction the outcome would be the same regardless of the choice. With the capacity and bid auctions, there is a single best choice. For instance, with the common bid auction and serial facilities it is best to offer the facilities to one winner, so as to prevent the double marginalisation and raise revenue. Conversely, with parallel facilities welfare is higher with two winners, while the revenue for the government is lower. Still, even if the government could optimise the choice of how many winners there will be, it remains much better for welfare to use a patronage or price auction.

8.2. Results for the numerical model

The *bid* auction attains the same outcome as no auction, as this outcome leads to the highest profit to pay to the government. Therefore, for the effects of this auction, see Tables 4 (serial) and 5 (parallel). For the three other auctions, Table 6 gives the results with serial facilities and Table 7

[20] 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 chance of winning.

[21] It is assumed that bidders cannot collude, and renegotiation is not possible. There is no incentive for a bidder (given the actions of the other bidders) to misrepresent its type and bid anything else than the zero profit outcome: bidding, for instance, a higher transfer would ensure winning the auction but firm would make a loss (after the money is transferred to the government), bidding a lower transfer would ensure that the bidder never wins the auction.

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, for example with 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.[22]

With a *bid* auction, a firm is willing to offer a premium in order to be a monopolist instead of a duopolists. Therefore, whenever firms are allowed to win both facilities, the monopoly outcome will be the result. The duopoly only happens when firm are only allowed to win one facility.

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

Conversely, the *patronage* and *generalised-price* auctions attain the first-best outcome in all of the set-ups. They do so when the auction has 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 6: Auctions for serial facilities

	Monopoly on capacity	A duopoly on capacity	A duopoly or Monopoly on patronage	A duopoly or Monopoly on generalised 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 7: Auctions for parallel facilities

	Monopoly on capacity	A duopoly on capacity	A duopoly or Monopoly on patronage	A duopoly or Monopoly on generalised 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

[22] 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.

9. Discussion

The numerical examples indicate how sensitive the results are to changes in the parameterizations of the model. In the numerical example, the oligopolies attain much lower welfares than the first-best or even the base case without congestion pricing. 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 generalised price, the stronger the competition with a parallel duopoly, and the higher welfare. When the model is re-calibrated with only the cross-price sensitivity 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 capacity cost is linear, the first-best outcome has zero profit and thus the generalised-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, as the government may not be able to provide it or have difficulty observing the cost structure of the firm. Alternatively, one could accept that with non-linear costs, the fee will be (slightly) different than first-best.

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 generalised-price auctions: if the cost of public funds is high enough, the revenue from a bid auction becomes so valuable that this auction is preferable.

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. An interesting study in this area is Tan and Yang (2012a), who investigate Pareto optimal BOT-contracts under uncertain demand conditions, where capacity is set while demand is uncertain, while the usage fee is set when the demand function is known.

I ignore preference heterogeneity, which is of course present in reality and can have important effects. Such heterogeneity is especially important with substitutes, since then the two facilities could offer differentiated products and attract different types of users, thereby raising welfare and profits (see, e.g., Luski, 1976; Reitman 1991; Winston and Yan, 2011). Edelson (1971) and Mills (1981) find that with heterogeneity, a monopolist may even ask a fee that is below first-best; and Tan and Yang (2012b) find that a single monopolistic road could have higher, equal, or lower level of congestion than first-best depending on the shape of the value of time distribution.

10. 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 firms interact. All private settings without government intervention result in much lower welfare than the first-best. Moreover, in a duopolistic setting the capacity rule can be different than the first-best rule. In the closed-loop setting, a firm strategically sets a different capacity to affect the competitor's Nash-equilibrium usage fee, which is set in a separate stage following the capacity setting. In the Stackelberg setting, the leader further changes his capacity in order to also affect the follower's capacity rule (while still considering the effect on the Nash-equilibrium fees).

There are marked differences in the effects of the different duopolies with the two network structures. The closed-loop setting leads to higher capacities and welfare than with a open-loop setting with a serial network, while in the parallel case the reverse holds. The Stackelberg leader sets a higher capacity than the follower in both cases; but the leader has a lower profit than the follower in the serial case (i.e. there is a second-mover advantage), and a higher profit in the parallel case.

Welfare is much lower in the private settings than first-best, and thus this paper also investigated the effects of 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 know 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 harms welfare, since each firm is a monopoly on its section. This problem does not occur with the two auctions that seem preferable—on generalised price and number of users (patronage)—since these lead to the first-best outcome regardless of if 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 lead to lower welfare. Hence, auctions need to be carefully designed: a "wrong" auction can actually be worse for welfare than no regulation.

Acknowledgements

I am grateful for the helpful comments of the editor, Hai Yang, and the reviewers. I also thank Erik Verhoef, Paul Koster, Eva Gutiérrez, and Ken Small for their suggestions. Financial support from the ERC (AdG Grant #246969 OPTION) is gratefully acknowledged. The usual disclaimer applies.

Appendix A. Mathematical discussion for the serial facilities

A.1. Best responses of the own fee

The fee rule (15) in text implicitly defines the best-responses. By writing it in implicit form and using the implicit function theorem, one finds the slopes of the best-responses (indicated by superscript R):

$$w_1 = -f_i + q\cdot\left(\partial c_i/\partial q + \partial c_j/\partial q + d_1\right) = -f_i + q\cdot Y_1 = 0; \tag{35}$$

where $Y_1>0$ follows (13), q is the number of users, and f_i is i's fee.

This gives for the reaction to f_j:

$$\begin{aligned}\frac{\partial f_i^R}{\partial f_j} &= -\frac{\partial w_1/\partial f_j}{\partial w_1/\partial f_i} = -\frac{\left(\partial q^u/\partial f_j\right)\left(\partial c_i/\partial q + \partial c_j/\partial q + d_1 + q\left(\partial^2 c_i/\partial q^2 + \partial^2 c_j/\partial q^2\right)\right)}{-1+\partial q^u/\partial f_i\left(\partial c_i/\partial q + \partial c_j/\partial q + d_1 + q\left(\partial^2 c_i/\partial q^2 + \partial^2 c_j/\partial q^2\right)\right)},\\ &= -\frac{\left(-1/Y_1\right)\left(Y_1+Y_2\right)}{-1+\left(-1/Y_1\right)\left(Y_1+Y_2\right)} = -1+\frac{Y_1}{2Y_1+Y_2}<0;\end{aligned} \tag{36}$$

where $Y_2=q(\partial^2 c_i/\partial q^2+\partial^2 c_j/\partial q^2)\geq 0$. This equation not only implies that the fee of i decreases with j's fee, but also that $-½\geq \partial f_i^R/\partial f_j \geq -1$.[23] For linear in q/s_i congestion, the slope is $-1/2$; it would be -1, 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 stronger the response.

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

$$\begin{aligned}\frac{\partial f_i^R}{\partial s_i} &= -\frac{\partial w_1/\partial s_i}{\partial w_1/\partial f_i} = -\frac{q\dfrac{\partial^2 c_i}{\partial q\partial s_i}+\left(\dfrac{\partial q^u}{\partial s_j}\right)\left(\partial c_i/\partial q + \partial c_j/\partial q + d_1 + q\left(+\partial^2 c_j/\partial^2 q\right)\right)}{-1+\partial q^R/\partial f_i\left(\partial c_i/\partial q + \partial c_j/\partial q + d_1 + q\left(\partial^2 c_i/\partial q^2 + \partial^2 c_j/\partial q^2\right)\right)},\\ &= -\frac{q\dfrac{\partial^2 c_i}{\partial q\partial s_i}+\left(-\dfrac{\partial c_i}{\partial s_i}\dfrac{1}{Y_1}\right)\left(Y_1+Y_2\right)}{-1+\left(-1/Y_1\right)\left(Y_1+Y_2\right)} = \frac{q\dfrac{\partial^2 c_i}{\partial q\partial s_i}+\left(-\dfrac{\partial c_i}{\partial s_i}\right)\left(1+Y_2/Y_1\right)}{2+Y_2/Y_1}.\end{aligned} \tag{37}$$

In the last version of (37), the denominator is positive: since Y_1 and Y_2 are positive. In the numerator, 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 linear congestion costs (i.e. n=1 in (a.i)), it is a best-response not to change the fee to changes in the own capacity (i.e. $\partial f_i^R/\partial s_i = 0$); for any other form satisfying (a.i), the best-response function is downward sloping with the own capacity:[24]

[23] Indeed this condition holds for any congestible cost function, which increases convexly in the number of users and decreases strictly convexly with capacity.

[24] For general costs, the slope could be positive

$$\left.\frac{\partial f_i^R}{\partial s_i}\right|(\text{a.i}) = -\frac{(n-1)n\cdot\delta_i\cdot d_1\cdot s_j^{\ n}}{s_i\left(2d_1\cdot q\cdot s_i^{\ n}\cdot s_j^{\ n}+(n-1)n\cdot q^n\left(\delta_i s_j^{\ n}+\delta_j s_i^{\ n}\right)\right)}\leq 0. \tag{38}$$

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

Turning to the response of f_i to s_j, one finds:

$$\frac{\partial f_i^R}{\partial s_j} = -\frac{\partial w_1/\partial s_j}{\partial w_1/\partial f_i} = -\frac{q\dfrac{\partial^2 c_j}{\partial q\partial s_j}+\left(\dfrac{\partial q^u}{\partial s_j}\right)\left(\partial c_i/\partial q+\partial c_j/\partial q+d_1+q\left(\partial^2 c_i/\partial q^2+\partial^2 c_j/\partial q^2\right)\right)}{-1+\left(\partial q^u/\partial f_i\right)\left(\partial c_i/\partial q+\partial c_j/\partial q+d_1+q\left(\partial^2 c_i/\partial q^2+\partial^2 c_j/\partial q^2\right)\right)},$$

$$= -\frac{q\dfrac{\partial^2 c_j}{\partial q\partial s_j}+\left(-\dfrac{\partial c_j}{\partial s_j}\dfrac{1}{Y_1}\right)\left(Y_1+Y_2\right)}{-1+\left(-1/Y_1\right)\left(Y_1+Y_2\right)} = \frac{q\dfrac{\partial^2 c_j}{\partial q\partial s_j}+\left(-\dfrac{\partial c_j}{\partial s_j}\right)\left(1+Y_2/Y_1\right)}{2+Y_2/Y_1}\leq 0. \tag{39}$$

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

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 (23) 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 (23) needs to be positive, and this condition is restated below:

$$\frac{\partial s_j}{\partial s_i}\left(\frac{\partial f_i^{NE}}{\partial s_j}q+\frac{\partial q^R}{\partial s_j}f_i\right)>0. \tag{40}$$

Here, subscript j indicates the follower and i the leader. This term measures the effects of the induced change in the follower's capacity on marginal revenue. 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.i):

$$\left.\left(\left(\frac{\partial f_i^{NE}}{\partial s_j}q\right)+\left(\frac{\partial q^R}{\partial s_j}f_i\right)\right)\right|(\text{a.i}) = \left(-\frac{d_1(n-1)nq^{2+n}s_i^{\ n}\delta_j}{s_j(3d_1qs_i^{\ n}s_j^{\ n}+n(1+2n)q^n(s_i^{\ n}\delta_j+s_i^{\ n}\delta_j))}\right)+\left(n\delta_j\frac{q^{1+n}}{s_j^{\ 1+n}}\right)$$

$$= \frac{nq^{1+n}\delta_i\left[-(n-4)d_1\cdot q\cdot s_i^{\ n}s_j^{\ n}+n(1+2n)q^n(s_j^{\ n}\delta_i+s_i^{\ n}\delta_j))\right]}{s_j^{\ 1+n}\left(3d_1qs_i^{\ n}s_j^{\ n}+n(1+2n)q^n(s_i^{\ n}\delta_j+s_i^{\ n}\delta_j)\right)}. \tag{41}$$

Here, the denominator is positive. Hence, (41) is positive if the term between square brackets in the numerator is positive. Thus, the question is whether the below condition holds:

$$-(n-4)d_1\cdot q\cdot s_i^{\ n}s_j^{\ n}+n(1+2n)q^n(s_j^{\ n}\delta_i+s_i^{\ n}\delta_j))>0. \tag{42}$$

The second term of (42) always is positive; the first term is non-negative when $n \leq 4$, and negative when $n>4$. Hence, (42) is positive under $n \leq 4$ (as is the case for the linear and BPR congestion costs). For $n>4$ it might be negative, but it 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. A positive slope of this reaction function is intuitive: 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.i), $\partial s_j/\partial s_i$ is always positive, since the two first terms dominate (and with linear costs, the third term is even zero).

To investigate this, first write (21)—which gives follower j's capacity rule—in implicit form and then use $f_j \cdot \partial q^u/\partial s_j = -q \cdot \partial c_j/\partial s_j$:

$$w_2 = -k + f_j^{NE} \frac{\partial q^u}{\partial s_j} + \frac{\partial q^u}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_j} f_j^{NE} = 0. \tag{43}$$

Then, by using the implicit function theorem, one gets:

$$\frac{\partial s_j}{\partial s_i} = -\frac{\partial w_2 / \partial s_i}{\partial w_2 / \partial s_j}; \tag{44}$$

where the denominator can be shown to be negative as it is the second-order condition of j's capacity choice. Hence, $\partial s_j/\partial s_i$ is positive, if the numerator, $\partial w_2/\partial s_i$, is positive:

$$\partial w_2 / \partial s_i = \left(\frac{\partial^2 q}{\partial s_j \partial s_i} + \frac{\partial q}{\partial f_i} \frac{\partial^2 f_i^{NE}}{\partial s_j \partial s_i} + \frac{\partial^2 q}{\partial f_i \partial s_i} \frac{\partial f_i^{NE}}{\partial s_j} \right) f_j + \left(\frac{\partial q}{\partial s_j} + \frac{\partial q}{\partial f_i} \frac{\partial f_i^{NE}}{\partial s_j} \right) \frac{\partial f_j^{NE}}{\partial s_i}. \tag{45}$$

Under (a.i), this simplifies to:

$$\left.\frac{\partial w_2}{\partial s_i}\right|(\text{a.i}) = \frac{n^2(1+2n)q^{1+2n}\delta_i\delta_b(d_1{}^2(2+n^2)q^2 s_i{}^{2n} s_j{}^{2n} + d_1 n(1+n)(1+2n)q^{1+n} s_i{}^n s_j{}^n (s_j{}^n \delta_i + s_i{}^n \delta_j) + n^3(1+2n)q^{2n}(s_j{}^n \delta_i + s_i{}^n \delta_{j)}{}^2)}{s_i s_j (d_1 q s_i{}^n s_j{}^n + nq^n(s_j{}^n \delta_i + s_i{}^n \delta_j))(3d_1 q s_i{}^n s_j{}^n + n(1+2n)q^n(s_j{}^n \delta_i + s_i{}^n \delta_j))^2} > 0; \tag{46}$$

and thus 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 (29) in text implicitly determines the response function (superscript R indicates a best-response): $f_i^R[s_i, s_i, f_j]$. To see the effects on f_i^R from changes in s_i, s_j and f_j, one writes (29) in implicit form and use the implicit function theorem:

$$w_3 = -f_i + q_i \cdot \left(\partial c_i / \partial q_i + d_1 - \frac{d_2 \cdot d_2}{\partial c_j / \partial q_j + d_1} \right) = 0. \tag{47}$$

For the best-response of f_i to f_j this gives:

$$\begin{aligned}
\frac{\partial f_i^R}{\partial f_j} &= -\frac{\partial w_3 / \partial f_j}{\partial w_3 / \partial f_i} = -\frac{\frac{\partial q_i^u}{\partial f_j}\left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^{\,2}}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial q_i^2} \right) + \frac{\partial q_j^u}{\partial f_j}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,2}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}{-1 + \frac{\partial q_i^u}{\partial f_i}\left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^{\,2}}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial q_i^2} \right) + \frac{\partial q_j^u}{\partial f_i}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,2}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)} \\
&= \frac{d_2 \left(\frac{f_i}{q_i} + q_i \frac{\partial^2 c_i}{\partial q_i^{\,2}} \right) - \left(\partial c_j / \partial q_j + d_1\right)\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,2}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}{Z_1 + \left(\partial c_i / \partial q_i + d_1\right)\left(\frac{f_i}{q_i} + q_i \frac{\partial^2 c_i}{\partial q_i^{\,2}} \right) - \left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,3}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}.
\end{aligned} \tag{48}$$

where the last version uses eqs. (24) and (25) for the derivatives of q_i and q_j w.r.t. the fees, the definition from (29) for f_i, and Z_1 follows (26). The denominator is positive, since it is Z_1 multiplied by the second-order condition of f_i. An increase in the fee of j has two effects on i's fee: 1) it increases the number of users on i and thus i's fee; and 2) it lowers the number of users on j, which lowers i's fee (as this increases congestion on j, making j a less attractive substitute). Under (a.i), the sum of the effects is positive under (a.i), and it can be show that $0>\partial f_i^R/\partial f_j>1$.[25]

The response of i's fee to its capacity is:

$$\begin{aligned}
\frac{\partial f_i^R}{\partial s_i} &= -\frac{\partial w_3 / \partial s_i}{\partial w_3 / \partial f_i} = -\frac{q_i \frac{\partial^2 c_i}{\partial q_i \partial s_i} + \frac{\partial q_i^R}{\partial s_i}\left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^{\,2}}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial q_i^2} \right) + \frac{\partial q_j^R}{\partial s_i}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,2}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}{-1 + \frac{\partial q_i^R}{\partial f_i}\left(\frac{\partial c_i}{\partial q_i} + d_1 - \frac{d_2^{\,2}}{\partial c_j / \partial q_j + d_1} + q_i \frac{\partial^2 c_i}{\partial q_i^2} \right) + \frac{\partial q_j^R}{\partial f_i}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,2}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)} \\
&= \frac{Z_1 \cdot q_i \frac{\partial^2 c_i}{\partial q_i \partial s_i} - \frac{\partial c_i}{\partial s_i}\left(\left(\partial c_j / q_j + d_1\right)\left(\frac{f_i}{q_i} + q_i \frac{\partial^2 c_i}{\partial q_i^{\,2}} \right) \right) + \frac{\partial c_j}{\partial s_i}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,3}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}{Z_1 + \left(\left(\partial c_j / \partial q_j + d_1\right)\left(\frac{f_i}{q_i} + q_i \frac{\partial^2 c_i}{\partial q_i^{\,2}} \right) \right) - \left(\frac{\partial^2 c_j}{\partial q_j^{\,2}} \frac{q_i d_2^{\,3}}{\left(\partial c_j / \partial q_j + d_1\right)^2} \right)}.
\end{aligned} \tag{49}$$

In the last version of (49), the denominator is again positive. In the numerator, the first term is negative; the second positive; and the last negative or zero; it is non-positive, and it is strictly negative unless n=1 (see De Borger and Van Dender (2006) for such linear costs).[26] A higher capacity of i lowers i's fee. There are 3 effects of a higher s_i: i) it lowers congestion and thus the

[25] Under a general congestion function, it can be shown that if firms are symmetric, the fee also always decreases with the competitor's. If firms are really asymmetric this may not hold.

[26] For general cost functions, $\partial f_i^R/\partial s_i$ is typically negative, unless the firms are very ex-post asymmetric.

Marginal External Cost part of the fee, ii) it lowers the mark-up by lowering the number of users of j and thereby the usage cost of j, iii) due to the induced increase in the number of users of i, which limits the decrease of the MEC and raises the mark-up. The best-response fee is non-increasing in the own capacity, and the curve is only flat when usage cost is linear in the ratio, for other congestion function the best-response fee decreases with the own capacity.

The $\partial f_i^R/\partial s_j$ is negative:

$$\frac{\partial f_i^R}{\partial s_j}=\frac{\partial w_3/\partial s_j}{\partial w_3/\partial f_i}=-\frac{\frac{\partial^2 c_j}{\partial q_j\partial s_j}\left(\frac{q_i d_2^{\,2}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)+\frac{\partial q_i^u}{\partial s_j}\left(\tau_i+q_i\frac{\partial^2 c_i}{\partial q_i^2}\right)+\frac{\partial q_j^u}{\partial s_j}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}}\frac{q_i\cdot d_2^{\,2}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)}{-1+\frac{\partial q_i^u}{\partial f_i}\left(\frac{\partial c_i}{\partial q_i}+d_1-\frac{d_2^{\,2}}{\partial c_j/\partial q_j+d_1}+q_i\frac{\partial^2 c_i}{\partial q_i^2}\right)+\frac{\partial q_i^u}{\partial f_i}\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}}\frac{q_i d_2^{\,2}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)}$$

$$=\frac{\frac{\partial^2 c_j}{\partial q_j\partial s_j}\left(\frac{Z_1 q_i d_2^{\,2}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)+\frac{\partial c_j}{\partial s_j}d_2\left(\frac{f_i}{q_i}+q_i\frac{\partial^2 c_i}{\partial q_i^2}\right)+\frac{\partial c_j}{\partial s_j}\left(\frac{\partial c_i}{\partial s_i}+d_1\right)\frac{\partial^2 c_j}{\partial q_j^{\,2}}\left(\frac{q_i\cdot d_2^{\,2}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)}{Z_1+\left(\left(\partial c_j/\partial q_j+d_1\right)\left(\frac{f_i}{q_i}+q_i\frac{\partial^2 c_i}{\partial q_i^2}\right)\right)-\left(\frac{\partial^2 c_j}{\partial q_j^{\,2}}\frac{q_i d_2^{\,3}}{\left(\partial c_j/\partial q_j+d_1\right)^2}\right)}. \tag{50}$$

Here, the denominator is negative, while in the numerator 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 thus 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 the third term gives 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). Therefore, $0>\partial f_i^R/\partial f_j>1$, $\partial f_i^R/\partial s_i\leq 0$ and $\partial f_i^R/\partial s_j<0$.

All this implies that the Nash-equilibrium fee always decrease with the own and competitor's capacities.

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

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

$$\Pi_i = f_i \cdot q_i - k\cdot s_i; \tag{51}$$

which gives an f.o.c. for capacity setting of:

$$\frac{\partial \Pi_i}{\partial s_i}=0=\frac{\partial f_i^{NE}}{\partial s_i}\cdot q_i+\frac{dq_i^u}{ds_i}f_i-k=\frac{\partial f_i^{NE}}{\partial s_i}\cdot q_i+\left(\frac{\partial q_i^u}{\partial s_i}+\frac{\partial q_i^u}{\partial f_i}\frac{\partial f_i^{NE}}{\partial s_i}+\frac{\partial q_i^u}{\partial f_j}\frac{\partial f_j^{NE}}{\partial s_i}\right)f_i-k. \tag{52}$$

Maximising profit with respect to the fee gives:

$$f_i\cdot \partial q_i^u/\partial f_i+q_i=0; \tag{53}$$

which can be rewritten to fee condition in (29) by using (24). By inserting (27) for $\partial q_i^R/\partial s_i$ and rewriting, one gets that $f_i\cdot\partial q_i^R/\partial s_i=-q_i\cdot\partial c_i^u/\partial s_i$. From (53) one gets that $f_i\cdot\partial q_i^u/\partial f_i=-q_i$. Inserting all this into (52), and rewriting, results in condition (33) 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 $\partial q_i^u / \partial f_i \cdot \partial f_i^{NE} / \partial s \cdot f_i = -\partial f_i^{NE} / \partial s_i \cdot q_i$—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.

References

Acemoglu, D., Bimpikis, K. Ozdaglar, A., 2009. Price and capacity competition. *Games and Economic Behavior* 66(1), 1–26.

Averch, H., Johnson, L.L, 1962. Behavior of the Firm Under Regulatory Constraint. American Economic Review 52(5), 1052–1069.

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

De Borger, B., Dunkerley, F. Proost, S., 2007. Strategic investment and pricing decisions in a congested transport corridor. *Journal of Urban Economics* 62(2), 294-316.

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.

Brueckner, J.K., 2001. The economics of international codesharing: an analysis of airline alliances. *International Journal of Industrial Organization* 19(10), 1475-1498.

Bulow, J., Klemperer, P., 1996. Auctions Versus Negotiations. *American Economic Review* 86(1), 180-194.

Economides, N., Salop, S.C., 1992. Competition and integration among complements, and network market structure. *Journal of Industrial Economics* 40(1), 105–123.

Edelson , N.M., 1971. Congestion tolls under monopoly. *American Economic Review* 61(5), 873–882.

Engel, E., Fisher, R., Galetovic, A., 1997. Highway franchising: pitfalls and opportunities. *American Economic Review* 8(2), 68–72.

Harvey, M., Commercial road supply with incentive regulation. In: Australasian Transport Research Forum 2010 Proceedings, 29 September – 1 October 2010, Canberra, Australia. Retrieved from http://www.atrf.info/papers/2010/2010_Harvey.pdf (on 8 August2013).

Kraus, M., 2003. A new look at the two-mode problem. *Journal of Urban Economics* 54(3), 511–530.

Krishna, V., 2010. *Auction Theory*, 2nd edition. Academic press, Amsterdam:.

Laffont, J.J., Tirole, J., 1993. *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge.

Lalive, R., Schmutzler, A., 2011. Auctions vs negotiations in public procurement: which works better? (Version of 8 August 2011). *CEPR Discussion Papers* 8538.

Luski, I., 1976. On partial equilibrium in a queuing system with two servers. *Review of Economic Studies* 43(3), 519–525.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. *Microeconomic Theory*. Oxford University Press, New York.

Mantin, B., 2012. Airport complementarity: private vs. government ownership and welfare gravitation. *Transportation Research Part B* 46(3), 381-388.

Matsumura, T., Matsushima, N., 2012. Airport privatization and international competition, *Japanese Economic Review* 63(4), 431–450;

Mills, D.E., 1981. Ownership Arrangements and Congestion-prone Facilities. *American Economic Review (Papers and Proceedings)* 71(3), 493–502.

Mohring, H., Harwitz, M., 1962. *Highway benefits: an analytical framework*. Northwestern University Press, Evanston.

Oum, T.H., Zhang, A., Zhang, Y., 2004. Alternative forms of Economic Regulation and their Efficiency Implications for Airports. *Journal of Transport Economics and Policy* 38(2), 217–246.

Reitman, D., 1991. Endogenous quality differentiation in congested markets. *Journal of Industrial Economics* 39(6), 621–647.

Sheshinski, E., 1976. Price, quality and quantity regulation in monopoly situations. *Economica* 43(170), 127–137.

Spence, A.M., 1975. Monopoly, quality, and regulation. *Bell Journal of Economics* 6(2), 417–429.

Small, K.A., Verhoef, E.T., 2007. *The Economics of Urban Transportation*. Routledge, Abingdon.

Tan, Z.J., Yang, H., 2012a. Flexible build-operate-transfer contracts for road franchising under demand uncertainty. *Transportation Research Part B* 46(10), 1419–1439.

Tan, Z.J., Yang, H., 2012b. The impact of user heterogeneity on road franchising. *Transportation Research Part E* 48(5), 958-975.

Tan, Z.J., 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* 46(8), 971–983.

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

Wan, Y., Zhang, A., 2013. Urban Road Congestion and Seaport Competition. *Journal of Transport Economics and Policy* 47(1), 55–70.

Winston, C., Yan, J., 2011. Can privatization of U.S. highways improve motorists' welfare? *Journal of Public Economics* 95(7–8), 993–1005.

Wu, D., Yin, Y., Lawphongpanich, S., 2011a. Optimal selection of build-operate-transfer projects on transportation networks. *Transportation Research Part B* 45 (10), 1699–1709.

Wu, D., Yin, Y., Yang, H., 2011b. The independence of volume–capacity ratio of private toll roads in general networks. *Transportation Research Part B* 41(1), 96–101.

Xiao, F., Yang, H., Han, D., 2007. Competition and efficiency of private toll roads. *Transportation Research Part B* 41(3), 292–308.

Yuen, A., Basso, L.J., Zhang, A., 2008. Effects of Gateway Congestion Pricing on Optimal Road Pricing and Hinterland. *Journal of Transport Economics and* Policy 42(3), 495–526.

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