Cost Recovery of Congested Infrastructure under Market Power

Erik T. Verhoef

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

COST RECOVERY OF CONGESTED INFRASTRUCTURE UNDER MARKET POWER[*]

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

Abstract

The famous Mohring-Harwitz theorem states that, under certain technical conditions, the degree of self-financing of congested infrastructure is equal to the elasticity of the capacity cost function in the optimum, so that under neutral scale economies exact self-financing applies. Although the theorem has been proven to remain valid for various extensions of the basic set-up for which it was originally derived, it breaks down when the infrastructure is used by operators with market power when competing in Cournot fashion, the case in point often being oligopolistic airlines at a congested airport. This paper proposes a regulatory scheme, not involving lump-sum payments or budget constraints in the optimal pricing problem, that restores self-financing for congested infrastructure for this market form. What is more, under the proposed scheme, exact self-financing applies independent of the elasticity of the capacity cost function. The result remains true both for the case where operators treat the tolls parametrically, and for "manipulable" tolls, designed to account for the fact that operators with market power can be expected to be aware of, and exploit, the fact that toll are not truly parametric, but instead depend on their own behaviour.

1. Introduction

The celebrated theorem on self-financing of infrastructure, due to Mohring and Harwitz (1962), states that under certain technical conditions, the degree of self-financing of congested infrastructure is equal to the elasticity of the capacity cost function in the optimum. Hence, under neutral scale economies, where this elasticity is unity, the revenues from optimal pricing – following Pigou's (1920) prescription that the congestion toll be equal to the marginal external cost – will be exactly equal to the capital cost associated with the supply of the optimal size of capacity.

Subsequent research has shown that this theorem, originally derived in the rather basic context of a static model for a single road with homogeneous users, survives various extensions that make the setting more realistic (for reviews of this literature, see for example De Palma and Lindsey, 2007; and Small and Verhoef, 2007). In particular, the theorem remains true for full networks (Yang and Meng, 2002); with heterogeneous users (Arnott and Kraus, 1998a); with dynamic congestion technologies such as the bottleneck model of Vickrey (1969) (Arnott, de Palma and Lindsey, 1993) but also more generally (Arnott and Kraus, 1998a); when including maintenance and wear and tear (Newbery, 1989); when allowing for variable prices for inputs such as land (Small, 1999); and in present value terms when considering the long run (Arnott and Kraus, 1998b). But the theorem breaks down in other circumstances. One is where capacity is lumpy, not continuous. Another one, under consideration in this paper, is when infrastructure users possess market power.

A burgeoning literature, most of which concerns aviation, has discussed how optimal congestion tolls for actors with market power may be far below the conventional Pigouvian prescription. The reason is that under Nash-Cournot behaviour, operators with market power internalize congestion imposed upon their own services (Daniel, 1995; Brueckner, 2002), so that the optimal congestion toll should include only marginal congestion effects on other firms' services. Quite intuitively, Brueckner (2002) finds that for uniform values of time and marginal congestion effects, this implies firm-specific tolls that are proportional to a fraction $(1 - s_i)$, with s_i being the firm's market share at the airport. All else equal, larger firms should then face lower tolls. Because the optimal investment rule for capacity remains unaltered compared to the conventional model of congestion with atomistic users, while the optimal pricing rule is adapted, the Mohring-Harwitz rule breaks down. As a consequence, the prospects for fully self-financing infrastructure are reduced for such markets, compared to the case of atomistic users. This is reinforced when the tolls take into account the fact that profit-maximizing Cournot competitors apply a demand-related mark-up, which increases when demand becomes less elastic and vanishes only in the limiting case of perfectly elastic demand.[1] This calls for tolls that are adjusted downward further compared to the tolls proposed by Brueckner (2002), and that even may become negative when the (negative) demand-related market power component in the toll outweighs the (positive) congestion

---

[1] Both Oum and Zhang (1993) and Brander and Zhang (1990) conclude that actual airline pricing behaviour appears to be closer to Cournot than to Bertrand behaviour.

component (Pels and Verhoef, 2004). This is consistent with Buchanan's (1969) rule for optimal environmental taxation of a monopolist.

Later contributions have identified market configurations that would have better prospects for self-financing under market power as optimal tolls move closer to atomistic levels. For example, this was found for Stackelberg competition as opposed to Nash competition (Brueckner and Van Dender, 2008); for differentiated Bertrand competition as opposed to Cournot competition (Silva and Verhoef, 2011); and for dynamic bottleneck congestion as opposed to static congestion (Silva, Verhoef and Van den Berg, 2012). Obviously, these findings are of less help, the more accurately the market form is best described by the conventional Cournot-Nash model with static flow congestion.

Basso and Zhang (2006) and Zhang and Zhang (2006) included the role of airports in modelling congestion between airlines at airports. Zhang and Zhang (2006) study how a budget constraint would then affect the behaviour of a welfare-maximizing airport – surprisingly finding that there will be overinvestment in capacity at the margin.

This paper addresses the problem of financing congested infrastructure when operators have market power from a different perspective. A regulatory scheme is proposed, not involving lump-sum payments or budget constraints in the optimal pricing problem, that restores self-financing for congested infrastructure for this market form. In fact, under the proposed scheme, exact self-financing applies independent of the elasticity of the capacity cost function. What is more, a balanced net budget not only applies at the aggregate level (*i.e.*, for the airport operator), but also for each firm individually. The result remains true both for the case where operators treat the tolls parametrically, and for "manipulable" tolls as proposed by Brueckner and Verhoef (2010). These are tolls that are designed to allow for the fact that operators with market power can be expected to be aware of, and exploit, the fact that tolls are not truly parametric, but instead depend on their own behaviour.

The finding that exact self-financing holds independent of the elasticity of the capacity cost function makes application of the scheme probably more attractive in reality. That is, substantial net surpluses or deficits from airport operations seem more likely to cause political and social opposition than cases where an airport is close to breaking even. The same can be said of balanced budgets for each firm individually. The by now conventional Cournot result that tolls should be inversely proportional to the operator's size is easily interpreted as "unfair", making its application less attractive from the political perspective. A scheme that leads to a balanced budget in each operator's contributions to infrastructure finances seems, seen from that perspective, more attractive.

The basic idea behind the scheme is simple. It applies the notion that non-atomistic operators have an incentive to contribute voluntarily to the supply of capacity. The balanced-budget result stems from the fact that the willingness to make such contributions naturally increases with the firm's market share. The lower marginal tolls are therefore just compensated for through higher contributions to capacity provision. The balanced budgets, in turn, result from the fact that the toll revenues are just sufficient to cover the subsidies on capacity provision that are needed to make the firms collectively supply the first-best aggregate capacity.

The budgets will be balanced exactly only if a particular equi-proportionality condition is fulfilled: each firm's share in the total capacity cost should be equal to its share in the total output. It will be shown that this condition, although intuitively plausible and natural, will not be satisfied spontaneously in the decentralized optimum. For exact self-financing to apply, the regulator would therefore have to make sure the condition is satisfied.

The paper will focus on the derivation of the theoretical result, which will be complemented by a simple numerical exercise for illustrative purposes only. The conceptual point that the scheme restores self-financing of congested infrastructure in the first-best optimum, now even independent of the elasticity of the capacity cost function, is the paper's main message. The natural follow-up question, of whether and how this result can be put into practice, will be briefly considered in the paper's concluding section. Arguably, the most pressing challenge in doing this is to cope with the non-stationarity of market equilibria in reality, while capacity investments are irreversible. This would make firms unwilling to contribute to the cost of capacity insofar as it concerns sunk costs from earlier investments. A way to reconcile this might be to tie the supply of capacity to a short-run complementary service, that can be credibly reduced in size if smaller contributions are made. Airport staff would be a good example.

The plan of the paper is as follows. Section 2 presents the optimality conditions that characterize the first-best optimum, and briefly shows how the self-financing result would apply with atomistic users but breaks down with market power of service operators. Section 3 presents the scheme for parametric congestion taxes and capacity subsidies, while Section 4 does the same for their manipulable counterparts. Section 5 concludes.

2.   The degree of self-financing with first-best pricing and central capacity provision

It is instructive to start our exposition with a brief derivation of the classic self-financing result for atomistic users of a congestible facility, even though it follows standard expositions as given in many textbooks such as Small and Verhoef (2007). Let us consider a single market, where users are identical in all respects except for their willingness to pay to use the facility. Demand is characterized by a single inverse demand function $D(N)$, where N is the number of users. Average user cost c increases in N and decreases in capacity K, and to be able to establish a link between toll revenues and capacity cost for general functions, we need to assume that c is a function of the ratio of the two: $c(N/K)$. Besides the total user cost $N \cdot c(\cdot)$, there are costs associated with the supply of capacity, $C_K(K)$. We thus ignore that the latter could be dependent on N, too; Newbery (1989) showed that the inclusion of use-dependent maintenance cost does not upset the classic self-financing result. We also ignore, for now, any other user costs than variable time losses from congestion; including these does not provide additional useful insights.

If the road operator can levy a toll τ, the equilibrium condition for users is that the marginal benefit $D(N)$ be equal to the so-called generalized price $c(\cdot) + \tau$, and the problem of maximizing social surplus S can be written as:

$$\text{Max}_{N,K} \, S = \int_0^N D(x)\mathrm{d}x - N \cdot c(\frac{N}{K}) - C_K(K)$$

$$\text{s.t.:} \, D(N) - c(\frac{N}{K}) - \tau = 0 \tag{1}$$

The first-order conditions are:

$$\frac{\partial S}{\partial N} = D(N) - c(\cdot) - N \cdot \frac{\partial c(\cdot)}{\partial N} = 0 \quad \Leftrightarrow \quad \tau = N \cdot \frac{\partial c(\cdot)}{\partial N} \tag{2a}$$

$$\frac{\partial S}{\partial K} = -N \cdot \frac{\partial c(\cdot)}{\partial K} - C_K'(K) = 0 \tag{2b}$$

where a prime denotes a derivative for a function with one argument. Equation (2a) gives the standard Pigouvian toll; equation (2b) states that the marginal benefit of capacity expansion should be equal to the marginal cost. A decisive determinant for the degree of self-financing is the elasticity of the capacity cost function:

$$\kappa \equiv C_K' \cdot \frac{K}{C_K} \tag{3}$$

The other equation that is needed to derive the self-financing result is Euler's rule, which implies that:

$$N \cdot \frac{\partial c(\cdot)}{\partial N} = -K \cdot \frac{\partial c(\cdot)}{\partial K} \tag{4}$$

For future reference, it is useful to point out that, by using the shorthand variable $F \equiv N/K$ for the volume-capacity ratio, equation (4) can be verified as follows:

$$N \cdot \frac{\partial c(\cdot)}{\partial N} = N \cdot \frac{\partial c(\cdot)}{\partial F} \cdot \frac{\partial F}{\partial N} = \frac{\partial c(\cdot)}{\partial F} \cdot \frac{N}{K}$$

$$-K \cdot \frac{\partial c(\cdot)}{\partial K} = -K \cdot \frac{\partial c(\cdot)}{\partial F} \cdot \frac{\partial F}{\partial K} = \frac{\partial c(\cdot)}{\partial F} \cdot \frac{N}{K} \tag{5}$$

Multiplying (2b) by K, and substituting (4) into the resulting first term, and (3) in the second, gives:

$$R \equiv N \cdot N \cdot \frac{\partial c(\cdot)}{\partial N} = \kappa \cdot C_K \quad \Rightarrow \quad \frac{R}{C_K} = \kappa \tag{6}$$

where R denotes total toll revenues. Equation (6) shows the classic self-financing result: in the first-best optimum, the degree of self-financing, R/C_K, is equal to the elasticity of the capacity cost function, κ.

Next, let us consider the case of Cournot competitors providing services using the same congestible facility. We allow for demand functions and cost functions to be firm-specific, but assume for convenience that firms' outputs enter their congestion cost functions symmetrically: the additional delay due to an extra user is independent of the identity of the firm that serves this user. We denote individual firms with index i. The way in which congestion effects enter the firms' optimization problems does not depend on whether these costs are initially incurred by the users (*i.e.*, involves the users' valuation of delays), by the firms themselves (involving the firms' valuations), or constitute a combination of both. In any

of these cases we may treat $D_i(\cdot) - c_i(\cdot)$ as the firm's net revenue per user served.[2] Firm i's short-run optimization problem can therefore be written as:

$$\underset{N_i}{\text{Max }} \Pi_i = N_i \cdot D_i(N_i) - N_i \cdot c_i\left(\frac{N_i + N_{-i}}{K}\right) - N_i \cdot \tau_i \tag{7}$$

The variable N_{-i} represents the sum of all other firms' outputs, so that total use $N = N_i + N_{-i}$. The toll τ_i is now allowed to be firm-specific, and is modelled as a per-user toll that the firm pays. In the context of aviation, this analytical set-up implies that passenger loads per aircraft are treated as given. The first-order condition with respect to N_i is:

$$\frac{\partial \Pi_i}{\partial N_i} = D_i(N_i) + N_i \cdot D_i'(N_i) - c_i(\cdot) - N_i \cdot \frac{\partial c_i(\cdot)}{\partial N_i} - \tau_i = 0 \tag{8}$$

The corresponding social surplus maximization problem looks as follows:

$$\underset{\forall N_i, K}{\text{Max }} S = \sum_i \int_0^{N_i} D_i(x)\mathrm{d}x - \sum_i N_i \cdot c_i\left(\frac{N_i + N_{-i}}{K}\right) - C_K(K) \tag{9}$$

The first-order conditions are:

$$\frac{\partial S}{\partial N_i} = D_i(N_i) - c_i(\cdot) - N_i \cdot \frac{\partial c_i(\cdot)}{\partial N_i} - \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial N_i} = 0 \quad \forall i \tag{10a}$$

$$\frac{\partial S}{\partial K} = \sum_i -N_i \cdot \frac{\partial c_i(\cdot)}{\partial K} - C_K'(K) = 0 \tag{10b}$$

Subtracting (10a) from (8) gives the optimal toll for operator i:

$$\tau_i = \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial N_i} + N_i \cdot D_i'(N_i) \quad \forall i \tag{11}$$

The toll consists of two terms. The first is a congestion component, which subtracts firm-internal congestion costs from the conventional Pigouvian congestion toll in (2a) (Brueckner, 2002). The second is a term that corrects the demand-related mark-up that the Cournot operator would apply (Pels and Verhoef, 2004). Note that, with a downward-sloping inverse demand function, the second term is negative, reflecting that a subsidy is needed to stimulate a Cournot profit-maximizer to set output beyond the point where marginal revenue – which is below marginal benefit – is equal to marginal cost.

The combination of pricing according to (11) and investments according to (10b) will lead to a smaller surplus, or a larger deficit, than policies according to (2a) and (2b). To see why, first observe that given the vector of use levels N_i, the highest possible revenues occur when all demands are perfectly elastic, so that the second term in (11) vanishes. So let us consider that case. Total toll revenues R^C (the superscript C is added to remind us we are considering the congestion component in (11) only) are then equal to:

$$R^C = \sum_i (N - N_i) \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial N} \tag{12a}$$

---

[2] Note that $D_i(\cdot)$ gives the inverse demand in terms of the generalized price; *i.e.*, the sum of fare and other costs incurred by passengers such as travel time. To the extent that $c(\cdot)$ represents user-incurred costs, it thus reduces the willingness to pay for fares on a dollar-by-dollar basis, and therefore enters the firm's profit function in the same way as would congestion costs incurred directly be the firm itself.

Substituting, from (3), $C'_K = \kappa \cdot C_K / K$ into (10b) gives:

$$C_K = \frac{1}{\kappa} \cdot \sum_i K \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial K} \tag{12b}$$

Because c_i is a function again of $F \equiv N / K$, the following degree of self-financing follows from (12ab):

$$\frac{R^C}{C_K} = \frac{\sum_i (N - N_i) \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial F} \cdot \frac{1}{K}}{\frac{1}{\kappa} \cdot \sum_i K \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial F} \cdot \frac{N}{K^2}} = \kappa \cdot \frac{\sum_i \frac{(N - N_i)}{K} \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial F}}{\sum_i \frac{N}{K} \cdot N_i \cdot \frac{\partial c_i(\cdot)}{\partial F}} \tag{13}$$

This is below the degree implied by (6), approaching it only when each individual N_i approaches zero; *i.e.*, as the atomistic case is approached asymptotically. Obviously, with demand not perfectly elastic, the degree of self-financing will be further below the expression on the right-hand side of (13), because of the further downward adjustments on tolls.

3.   The degree of self-financing with in the first-best optimum with optimally subsidized voluntary capacity contributions: parametric tolls and subsidies

3.1.   Analytics

Self-financing of infrastructure may bring a number of advantages, especially if it is near exact self-financing. Not only would it mean that there is no need for levying distortionary taxes elsewhere in the economy to raise the funds needed to supply capacity, there is also a potentially substantial advantage in the transparency and perceived logic and fairness of financial flows: it is the users of the infrastructure who pay for its provision – but they are paying no more than that. This may be an appealing message from the political perspective. It therefore seems worthwhile to explore ways of restoring self-financing when it breaks down; in this case, due to the existence of market power of operators. Moreover, there is an intellectual challenge in designing such a scheme, as it provides deeper insight into the underlying reason why the theorem breaks down under Cournot competition. The simple answer is of course that the tolls are adjusted downwards in (11) compared to (2a), while the investment rule does not change between (2b) and (10b). A somewhat deeper answer is that, behind the combination of optimal policy rules (10b) and (11) for the case of operators with market power, there is a fundamental asymmetry in that the toll rule is adjusted to properly reflect market power of Cournot oligopolists, while the investment rule is not. More in particular, the above analysis ignores that exactly because the operators have market power, they also have an incentive to contribute to the provision of capacity.

The schemes to be considered in this paper exploit this incentive for operators to contribute voluntarily to capacity. The schemes thus bear resemblance to mechanisms designed to optimize the voluntary provision of a public good (*e.g.*, Falkinger, 1996; see also Bergstrom, Blume and Varian, 1986). A quite predictable problem in such mechanisms is that non-rivalness and non-excludability usually causes incentives to free-ride, so that under-provision typically results if the aggregate provision is left exclusively dependent on the voluntary contributions only. That is not any different in the current problem. Given that the

self-financing result in its original form applies to the first-best outcome, we will only consider schemes that effectively address this problem, and that are designed to achieve the first-best outcome in terms of both use levels and aggregate capacity.

Let us consider the case where the firm not only has to choose its output level N_i, but also the amount of capacity it supplies, K_i. Aggregate capacity enters the cost functions $c_i(\cdot)$ in the same way as above, but is now composed of firm-specific contributions: $K = \sum_i K_i$. Anticipating that free-rider behaviour would result in under-provision of capacity, we equip the regulator with a second instrument, namely firm-specific subsidies σ_i on capacity provision. Consider the case of simultaneous decisions on capacity and output. If conventional capacity is irreversible in practice, as is true for most transport infrastructures, this would be consistent with a practical set-up as mentioned above, where there is a critical variable factor that is perfectly complementary to capacity, and that has to be put into use in order to make capacity effective and productive. Operating staff would be a natural example. The consideration of simultaneous – not sequential – capacity and output decisions has the overriding analytical advantage that the single-stage structure of the problem as assumed in the conventional analysis is maintained here, thus introducing no further differences between the cases that would complicate the comparison. Furthermore, to prevent demand-related market power terms in optimal tolls from clouding the pure results on self-financing, we assume that all demands are perfectly elastic, so that firm-specific marginal benefits become constant terms D_i. Under these assumptions, firm i's profit maximization problem becomes:

$$\underset{N_i,K_i}{\text{Max}}\, \Pi_i = N_i \cdot D_i - N_i \cdot c_i(\frac{N_i + N_{-i}}{K_i + K_{-i}}) - N_i \cdot \tau_i - C_{K,i}(K_i; K_{-i}) + \sigma_i \cdot K_i \qquad (14)$$

The arguments in the capacity cost function $C_{K,i}$ are written in this specific way to make sure that also for $\kappa \neq 1$, capacity cost can satisfy:

$$\sum_i C_{K,i}(K_i; K_{-i}) = C_K(K) \quad \text{if} \quad \sum_i K_i = K \qquad (15a)$$

In other words, the aggregate (or social) capacity cost function is not altered due to the decomposition. Moreover, the formulation allows the marginal cost of capacity expansion to be treated as equal across firms, and equal to the marginal social cost of capacity, by assuming:

$$\frac{\partial C_{K,i}(K_i; K_{-i})}{\partial K_i} = C_K'(K) \quad \forall i \qquad (15b)$$

Note that at the margin, a firm faces the full marginal social cost of capacity, reflecting that other firms will not make matching additional contributions in response to a marginal increase in this firm's contribution. For that reason, the partial derivative in (15b), relevant to characterize the firm's behaviour, will generally be different from the partial derivative of a function – if it would be continuous and smooth – that could be used to describe how (15a) is achieved over infra-marginal units of capacity. An individual firm does not assume that such a function would remain valid if it would change its own contribution to capacity, but instead the firm takes other firms' contributions as given.

The first-order condition of (14) for N_i is similar to the one in (8), except for the vanished market power term, while also a first-order condition for K_i now applies:

$$\frac{\partial \Pi_t}{\partial N_i} = D_i - c_i(\cdot) - N_i \cdot \frac{\partial c_i(\cdot)}{\partial N_i} - \tau_i = 0 \tag{16a}$$

$$\frac{\partial \Pi_t}{\partial K_i} = -N_i \cdot \frac{\partial c_i(\cdot)}{\partial K} - C_K'(K) + \sigma_i = 0 \tag{16b}$$

Note that we assume here that tolls τ_i and subsidies σ_i are treated as parametric by firms; manipulable tolls and subsidies are considered in Section 4 below. To derive the optimal rules for τ_i and σ_i, we contrast (16ab) with the social optimality conditions. These are given by (10ab), with $D_i(N_i)$ in (10a) replaced with the constant D_i. As a result, the firm-specific toll now only contains marginal external congestion costs imposed on other firms:

$$\tau_i = \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial N} \quad \forall i \tag{17a}$$

Next, the investment rule (16b) implies that, for all firms to face an incentive to contribute to capacity in equilibrium, the subsidies σ_i should be allowed to vary over firms such that it can exactly compensate for differences in the first term, which represents the firm's own marginal benefit of capacity expansions. This in fact occurs for the optimal choice of subsidies σ_i, namely those that set each firm's subsidy equal to the marginal benefits that a capacity expansion brings to all other firms. It is easily verified that this turns, for each firm individually, the problem of optimizing capacity according to (16b), into the social problem with the outcome as given in (10b):

$$\sigma_i = \sum_{j \neq i} -N_j \cdot \frac{\partial c_j(\cdot)}{\partial K} \quad \forall i \tag{17b}$$

The question of interest is now whether a simultaneous implementation of (17ab) would lead to a re-establishment of the self-financing result. To investigate this, define the total toll revenues received from firm i as R_i, and the expenses on subsidies to firm i as E_i. Straightforward computations then show:

$$R_i = N_i \cdot \tau_i = N_i \cdot \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial N} = \frac{N_i}{K} \cdot \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial F} = \frac{N_i}{N} \cdot F \cdot \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial F} \quad \forall i \tag{18a}$$

$$E_i = K_i \cdot \sigma_i = K_i \cdot \sum_{j \neq i} -N_j \cdot \frac{\partial c_j(\cdot)}{\partial K} = \frac{K_i \cdot N}{K^2} \cdot \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial F} = \frac{K_i}{K} \cdot F \cdot \sum_{j \neq i} N_j \cdot \frac{\partial c_j(\cdot)}{\partial F} \quad \forall i \tag{18b}$$

The two final expressions in (18a) and (18b) are equal if the firm's share in aggregate use, N_i/N, is equal to the firm's share in aggregate capacity, K_i/K. And, naturally, if $R_i = E_i$ is true for each firm, it is also true in aggregate, so that $R = E$. This leads to:

Proposition 1: Under the assumptions made, the degree of self-financing of a congestible facility with optimal parametric congestion tolls and optimal parametric subsidies for voluntary provision of capacity by Cournot-Nash competitors is equal to unity, independent of the elasticity of the capacity cost function, if each competitor's share in capacity provision is

the same as its share in total use. Moreover, for each competitor individually, the sum of tolls paid will then be equal to the sum of subsidies received.

Proof: (18a) and (18b) imply the firm-specific result; summing over firms gives the aggregate result.□

Proposition 1 presents a remarkable result. Not only does the application of optimally subsidized voluntary contributions to capacity restore self-financing of congested infrastructure under Cournot competition; exact self-financing now applies independent of the elasticity (κ) of the capacity cost function. This is an attractive property, as it reinforces the potential political advantages of self-financing mentioned earlier, including the absence of a need to use distortionary taxes to raise funds, and the perceived fairness of the scheme. Moreover, this perceives fairness may be reinforced because there is not only a balanced budget in the aggregate, but also in the financial flows between the regulator and each firm individually. Whereas the Cournot tolls of (17a) alone would have the politically possibly somewhat inconvenient property that the largest operators should face the lowest tolls, and *vice versa*, there appears to be what many may perceive as a strong sense of fairness in a scheme where in the aggregate, for each firm the sum of tolls and subsidies exactly cancel.

There is one important proviso: for exact self-financing to apply, it is indeed needed that for each firm the share in capacity provision, K_i/K, is the same as the share in total use, N_i/N. Because the firm's optimality condition (16b) involves marginal cost and benefits of capacity only, and is independent of the size of the firm's own infra-marginal capacity contributions, there is no equilibrium mechanism that would secure this equi-proportionality to be spontaneously satisfied in equilibrium. If it is not satisfied, the aggregate result in Proposition does not necessarily break down: undoubtedly, one could construct examples where non-zero net budgets per firm still result in a zero aggregate net budget. But it is certainly no longer guaranteed to hold true. This implies that in the implementation of the scheme, there is a need to design additional rules of the game such that the equality of ratio's will be achieved in equilibrium; at least, if the situation with zero firm-specific and aggregate net budgets is considered desirable.

A consequence of the regulator's budget being zero also when the elasticity κ is unequal to unity, is that the incidence of the surplus or deficit that would otherwise occur for the regulator will now be shifted onto the firms. To see why, it is sufficient to observe that the optimum is independent of how it is achieved. System-wide use levels, capacity, fare revenues, user cost, consumer surplus, and capacity cost, will therefore be the same, so that the deficit or surplus, if no longer occurring with the regulator, must be passed on to the operators.

3.2. *Numerical example*

To illustrate the results contained in Proposition 1, let us consider a small numerical example. There are three operators, assumed to be asymmetric to make sure that the results do not hinge on any symmetry between firms. Each firm faces its own perfectly elastic demand, with

marginal willingness to pay D_i. Variable cost functions are also firm-specific, and take on the well-known BPR-form:

$$c_i(F) = \alpha_i \cdot \left(1 + \beta \cdot \left(\frac{\sum_j N_j}{K} \right)^{\chi} \right) + \delta_i \cdot N_i \quad i, j = \{1, 2, 3\} \tag{19a}$$

The final term represent rising firm-internal marginal cost, which is needed in order to have unique interior equilibria with perfectly elastic demands.

We use a constant-elasticity capacity cost function, for which a composite constant multiplicative term is constructed that secures that the same optimum capacity (K_0) is found when varying κ in a sensitivity analysis:

$$C_K(K) = \frac{\gamma}{\kappa \cdot K_0^{\kappa-1}} \cdot K^{\kappa} \tag{19b}$$

Furthermore, we define a number of short-hand variables: mec_i gives the marginal external cost that firm i imposes on the other firms' services; C_V is the total variable cost or the sum of $N_i \cdot c_i$ over firms; c_i will be subdivided into the BPR component representing congested time, $c_{t,i}$, and the firm-internal non-time cost, $c_{nt,i} = \delta_i \cdot N_i$; mpc_i is the marginal private cost for firm i or $\partial N_i \cdot c_i / \partial N_i$; p_i is the perceived generalized price for the firm or $mpc_i + \tau_i$; $cong$ is the indicator for the level of congestion and gives the ratio of equilibrium travel time over free-flow travel time or $1 + \beta \cdot (N/K)^{\chi}$; B is total benefits or the sum of $N_i \cdot D_i$ over firms; msb_K is the marginal social benefit of capacity or the sum of $-N_i \cdot \partial c_i(\cdot)/\partial K$ over firms; $mb_{K,i}$ is the marginal benefit of capacity for firm i or $-N_i \cdot \partial c_i(\cdot)/\partial K$; and I_i is the net transfer from the regulator to the firm, or $I_i = E_i - R_i$.

Parameters				Equilibrium			
Firm 1	*Firm 2*	*Firm 3*	*Other*	*Firm 1*	*Firm 2*	*Firm 3*	*Other*
$D_1=110$	$D_2=105$	$D_3=100$	$\beta=0.15$	$N_1=3719$	$N_2=2261$	$N_3=1225$	$N=7205$
$\alpha_1=9$	$\alpha_2=10$	$\alpha_3=11$	$\chi=4$	$K_1=2676$	$K_2=1627$	$K_3=882$	$K=5185$
$\delta_1=0.01$	$\delta_2=0.015$	$\delta_3=0.025$	$K_0=5185$	$c_{t,1}=14.03$	$c_{t,2}=15.59$	$c_{t,3}=17.15$	$cong=1.56$
			$\kappa=1$	$c_{nt,1}=37.19$	$c_{nt,2}=33.91$	$c_{nt,3}=30.63$	$B=768\ 955$
			$\gamma=30$	$mpc_1=98.80$	$mpc_2=90.43$	$mpc_3=82.59$	$C_V=360\ 921$
				$mec_1=11.20$	$mec_2=14.57$	$mec_3=17.41$	$C_K=155\ 554$
				$p_1=110$	$p_2=105$	$p_3=100$	$S=252\ 480$
				$mb_{K,1}=14.44$	$mb_{K,2}=9.75$	$mb_{K,3}=5.81$	$msb_K=30$
				$\tau_1=11.20$	$\tau_2=14.57$	$\tau_3=17.41$	$R=95\ 926$
				$\sigma_1=15.56$	$\sigma_2=20.25$	$\sigma_3=24.19$	$E=95\ 926$
				$R_1=41\ 655$	$R_2=32\ 945$	$R_3=21\ 327$	$\Pi=252\ 480$
				$E_1=41\ 655$	$E_2=32\ 945$	$E_3=21\ 327$	$\Pi^{\#}=0$
				$\Pi_1=138\ 299$	$\Pi_2=76\ 655$	$\Pi_3=37\ 527$	
				$\Pi_1^{\#}=0$	$\Pi_2^{\#}=0$	$\Pi_3^{\#}=0$	
				$I_1=0$	$I_2=0$	$I_3=0$	

Note: rounding, if applicable, is either to nearest integer or (when decimal point is included) to the second digit

Table 1. First-best optimum for the numerical example: parametric tolls and subsidies

For the assessment of firm profits, it is helpful to distinguish between overall profits Π_i that do take into account that part of profits that is due to diseconomies of scale in the non-time

costs, and a measure $\Pi_i^{\#}$ that excludes this part of profits (by subtracting the term $\delta_i \cdot N_i^2$). The measure $\Pi_i^{\#}$, which will be referred to as "adjusted profits", is more relevant from the conceptual viewpoint, as it reflects the profits insofar as related to capacity costs, congestion costs, and regulatory policies.

The base calibration creates an optimum in which firms differ substantially in their relative sizes, and as a consequence also the tolls differ; as expected, in reversed ranking compared to the outputs. The base calibration assumes that K_i/K equals N_i/N for each firm, and as claimed in Proposition 1, the net transfers to individual firms are zero, and so is therefore the regulator's total budget. Also the adjusted profits $\Pi_i^{\#}$ are all equal to zero. This is in hindsight perhaps not surprising, given that demands are perfectly elastic, so that the profit-maximizing prices are equal to marginal costs; and given that neutral scale economies apply in capacity costs. The regular profits Π_i are greater than zero, due to the diseconomies of scale in the non-time costs. It is easily verified that social surplus S consists of these regular profits only: perfectly elastic demands imply that consumer surplus will be zero.

The numerical example allows us to illustrate how the abandoning of $K_i/K = N_i/N$ causes exact self-financing to break down. To that end, Figure 1 shows the adjusted profit measures $\Pi_i^{\#}$ if the optimal policy rules (17a) and (17b) remain in use to decentralize the social optimum, but compensating changes are made in K_1 and K_3 such that K remains constant (so does K_2). The diagram also shows the firm-specific net transfers, $I_i = E_i - R_i$, and the sum of these, which gives the net deficit for the regulator.
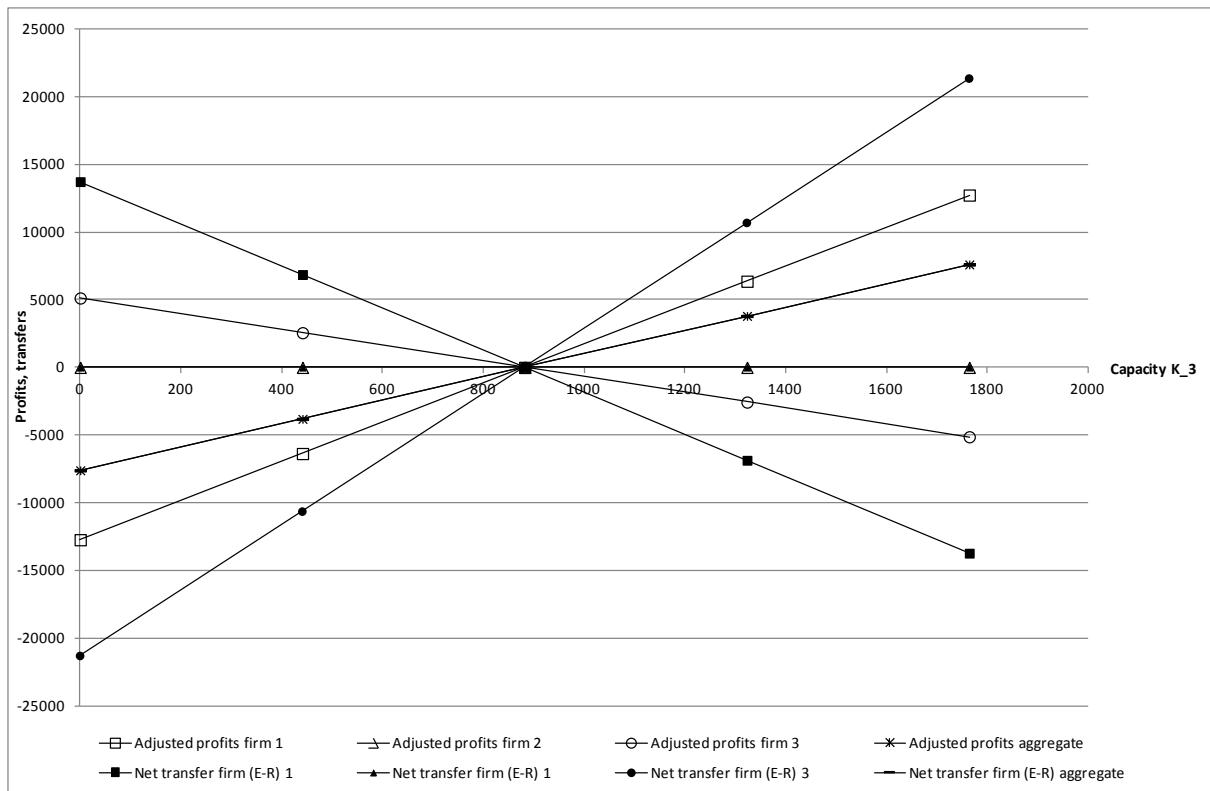

Figure 1. Sensitivity analysis: simultaneous compensating changes in K_1 and K_3 with parametric tolls and subsidies

The base calibration is obviously in the centre of the diagram. Moving to the left, the capacity share of the small firm 3, requiring a high σ_i, decreases. As a result, Firm 3's profits become positive, as the net own contribution to capacity declines. This is true even though the net budget that firm 3 receives from the regulator becomes negative. To the right of the centre, the opposite holds for firm 3. And because the capacity share of the large firm 1 moves in the opposite direction, the results have the opposite sign in both domains. Because Firm 2's capacity share remains constant to its output share, both its adjusted profits and its net budget remain zero.

   To the left of the centre capacity switches from being subsidized according to the needs of the small firm (with a high σ_i) to the needs of the large firm (with a low σ_i). As a result, the regulator's deficit (subsidies minus taxes) becomes negative, implying a surplus. Because prices and costs are not affected, this surplus is exactly equal to the industry-wide losses. The two series perfectly overlap in the graph.

   Finally, this sensitivity analysis was used to verify numerically that also capacity shares unequal to output shares may constitute an equilibrium, under the same set of taxes and subsidies. In other words, this confirms the earlier notion that there is no reason to expect a spontaneous emergence of equi-proportionality of capacities and outputs over firms.
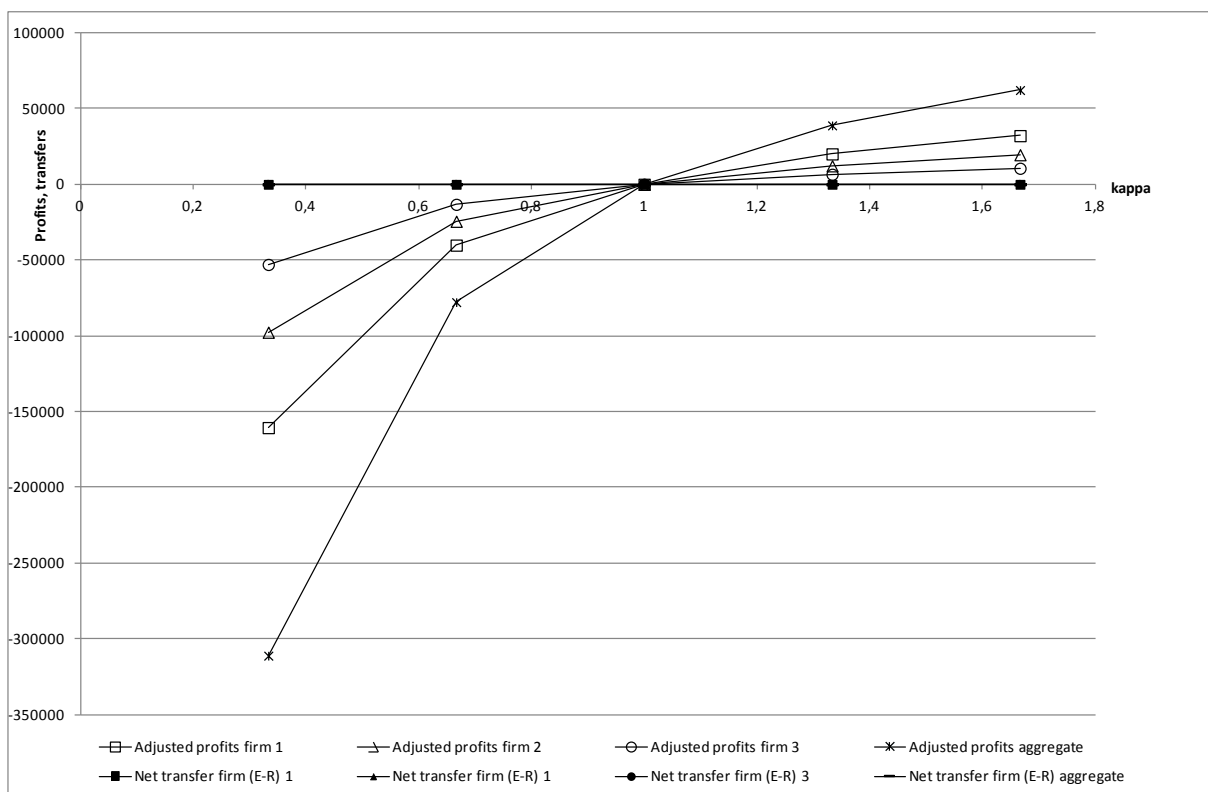
Figure 2. Sensitivity analysis: varying κ with parametric tolls and subsidies

The second sensitivity analysis concerns the elasticity of the capacity cost function, κ. Because of the specific form of (19b), the simultaneous changes in the multiplicative constant and the elasticity, in the exponent, are such that the optimum occurs at exactly the same capacity, and therefore the same prices and outputs apply for each value of κ. Assuming that

$K_i/K = N_i/N$ again applies for each firm, Proposition 1 implies the net transfers should not vary with κ. At the same time, because a change in the elasticity of the capacity cost function means that the total capacity cost will vary with κ, we expect firm-specific and industry-wide profits to vary with κ. Figure 2 confirms this.

The four series representing net transfers perfectly overlap at a constant level of zero, confirming Proposition 1: exact self-financing applies independent of the elasticity κ, with zero net transfers for each firm individually. The profits that arise if there are diseconomies of scale ($\kappa > 1$) rise with κ, and otherwise – quite naturally – increase with the size of the firm. Reversely, losses become bigger the stronger the economies of scale, so for lower values of κ when $\kappa < 1$.

4. Manipulable tolls and subsidies

4.1. Analytics

Brueckner and Verhoef (2010) observed that there is an inconvenient inconsistency in the combined set of assumptions that, on the one hand, firms in a congested Cournot oligopoly would take into account the fact that their own output level affects the aggregate level of congestion at the facility; but, on the other hand, would not be aware of the implication that the congestion tolls are not truly parametric, but instead depend on their own output choices just like the average time delay does. They demonstrate that if firms would take this into account, while the regulator nevertheless sets tolls erroneously assuming firms treat these tolls as parametric, the first-best optimum will not be achieved – quite intuitively, because the regulator's assumption on the behavioural responses to the tolls are not correct. They also propose a solution to this problem: a "manipulable toll", defined as a rule that specifies what the total toll payment will be as a function of the firm's output, which thus allows the per-unit-of-output toll to vary with the output level. This manipulable toll varies such that at every output level, the "marginal" toll ($\tau_i + N_i \cdot \partial \tau_i / \partial N_i$) satisfies (11). For a given number of firms, there is a constant of integration to be chosen for such a toll rule. The most natural choice for this constant is such that – for a duopoly – a carrier i's total toll liability is equal to the "the increase in the other carrier's congestion cost due to carrier i's operations minus the addition to consumer surplus from these operations" (Brueckner and Verhoef, 2010, eq. (22), p. 320).

Given that the behavioural assumption underlying the manipulable toll seems a more natural assumption for Cournot operators than the (usually implicit) assumption underlying the parametric toll, this section investigates to what extent a counterpart of Proposition 1 applies also with manipulable tolling. Because it seems inconsistent to treat tolls as manipulable and subsidies as parametric, or *vice versa*, both will be treated as manipulable.

With two instruments, there are two ways in which the policy could be operationalized. The first is to assume that the subsidy on contributions to capacity and the tax on use of that capacity are defined independently. This would mean that the total congestion toll $R_i(N_i)$ for firm i would be determined as the decrease in all other firms' costs if firm i's output N_i were removed, given all other firms' output choices, and given all capacity choices, including that of firm i:

$$R_i(N_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(\frac{N_i + \sum_{k \neq i} N_k}{K} \right) - c_j \left(\frac{\sum_{k \neq i} N_k}{K} \right) \right) \tag{20a}$$

Note that the derivative of (20a) with respect to N_i is the same as (17a) for any N_i, confirming that (20a) is indeed the optimal manipulable toll. The complementary capacity subsidy $E_i(K_i)$ for firm i would then be determined as the increase in all other firms' costs if firm i's capacity contribution K_i were removed, given all other capacity choices, and given all output levels, including that of firm i:

$$E_i(K_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(\frac{N}{\sum_{k \neq i} K_k} \right) - c_j \left(\frac{N}{K_i + \sum_{k \neq i} K_k} \right) \right) \tag{20b}$$

It is also easily verified that the derivative of (20b) with respect to K_i is equal to (17b) for any K_i, again confirming it is the optimal manipulable rule.

Under the equi-proportionality assumption of Proposition 1, simultaneous application of (20a) and (20b) will not generally lead to exact self-financing, neither in the sense of net transfers at the firm level, nor at the aggregate level. To see why, define the firms' share in the optimal output as ρ_i. Under equi-proportionality, we may then rewrite (20ab) as:

$$R_i(N_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(F^{FB} \right) - c_j \left((1 - \rho_i) \cdot F^{FB} \right) \right) \tag{20c}$$

$$E_i(K_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(\frac{1}{(1 - \rho_i)} \cdot F^{FB} \right) - c_j \left(F^{FB} \right) \right) \tag{20d}$$

Equations (20c) and (20d) will not generally be equal, and certainly not for the convex functional forms that are conventionally used for delay functions. In fact, they will be equal only if $c(\cdot)$ takes on a logarithmic specification such as $c_j(F) = \gamma_j \cdot Log(F)$, because for that case we find:

$$\begin{cases} R_i(N_i) = \sum_{j \neq i} N_j \cdot \gamma_j \cdot \left(Log \left(F^{FB} \right) - Log \left((1 - \rho_i) \cdot F^{FB} \right) \right) = \sum_{j \neq i} N_j \cdot \gamma_j \cdot \left(Log \left(\frac{1}{(1 - \rho_i)} \right) \right) \\ E_i(K_i) = \sum_{j \neq i} N_j \cdot \gamma_j \cdot \left(Log \left(\frac{1}{(1 - \rho_i)} \cdot F^{FB} \right) - Log \left(F^{FB} \right) \right) = \sum_{j \neq i} N_j \cdot \gamma_j \cdot \left(Log \left(\frac{1}{(1 - \rho_i)} \right) \right) \end{cases} \tag{20e}$$

For the second possibility of operationalizing the policy, however, exact self-financing is restored, again independent of the functional form of $c(\cdot)$, and again independent of κ. And that is through a single transfer $T_i(N_i, K_i)$, from the regulator to the firm if positive, which is determined as the increase in all other firms' costs if firm i's capacity contribution K_i and output N_i were simultaneously removed, given all other firms' capacity choices and output levels:

$$T_i(K_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(\frac{\left(\sum_{k \neq i} N_k \right)}{\sum_{k \neq i} K_k} \right) - c_j \left(\frac{\left(N_i + \sum_{k \neq i} N_k \right)}{K_i + \sum_{k \neq i} K_k} \right) \right) \qquad (21a)$$

The partial derivative of (21a) with respect to N_i gives minus (17a), which is as required because (21a) is defined as a subsidy when positive, while that with respect to K_i again gives (17b). In other words, the firm again faces the socially optimal incentives in its choice of N_i and K_i.

It is trivial to show that under this scheme, exact self-financing again applies under equi-proportionality of use- and capacity shares; both in the sense of net transfers at the firm level, and therefore also at the aggregate level. Under equi-proportionality, we may rewrite (21a) as:

$$T_i(K_i) = \sum_{j \neq i} N_j \cdot \left(c_j \left(\frac{(1-\rho_i)}{(1-\rho_i)} \cdot F^{FB} \right) - c_j \left(F^{FB} \right) \right) \qquad (21b)$$

which is evidently equal to zero for all firms. These results are summarized in:

Proposition 2: Under the assumptions made, the degree of self-financing of a congestible facility with an optimal manipulable transfer that simultaneously addresses congestion and the voluntary provision of capacity, used by Cournot-Nash competitors, is equal to unity, independent of the elasticity of the capacity cost function, if each competitor's share in capacity provision is the same as its share in total use. Moreover, for each competitor individually, the size of the transfer will be zero in the optimum. With a separate manipulable congestion toll and manipulable capacity subsidy, exact self-financing applies only for a specific (logarithmic in volume-capacity ratio) congestion cost function.

Proof: (21a) and (21b) imply the firm-specific result on the single transfer; summing over firms produces the aggregate result. Equations (20a-e) imply the result on separate manipulable congestion tolls and capacity subsidies.□

Because the possibility of Cournot operators attempting to manipulate tolls and subsidies seems far from far-fetched, it is reassuring that the self-financing result of Proposition 1 survives in this specific context; at least, for the single transfer policy. And because it again applies independent of the capacity κ, and only under equi-proportionality, the advantages and proviso's of this variant of the policy are similar to those for parametric tolls and subsidies.

4.2. Numerical example

The same numerical example as in Section 3 will be used to illustrate the analytical findings just presented. First, Table 2 gives the results of the policies for the base parameterization of the model. The results above the empty line repeat those given in Table 1, and confirm that the policies achieve the same first-best optimum. The first two firm-specific results give the marginal taxes (on output) and subsidies (on capacity provision) in the optimum. As required,

these are equal to the tolls and subsidies in Table 1. Next come the total transfers between regulator and firms. With separate instruments, in the left half of the table, and given that the congestion function is assumed to be convex in F, the revenues from the manipulable toll are smaller than for the parametric toll in Table 1. This is due to the fact that a firm's marginal external congestion cost decreases as its output is reduced further. In contrast, the expenditures on manipulable subsidies are larger than for their parametric counterparts, because the marginal increase in other firms' costs from marginal reductions in a firm's capacity contributions becomes larger as its capacity is reduced further. As a result, the net transfers I_i and adjusted profits $\Pi_i^{\#}$ are both positive now. Because both were zero in the optimum with parametric tolling and subsidies, and costs and consumer prices have not changed, they are also equal in the present case.

Separate manipulable tolls and subsidies				Single manipulable transfer			
Firm 1	*Firm 2*	*Firm 3*	*Other*	*Firm 1*	*Firm 2*	*Firm 3*	*Other*
N_1=3719	N_2=2261	N_3=1225	N=7205	N_1=3719	N_2=2261	N_3=1225	N=7205
K_1=2676	K_2=1627	K_3=882	K=5185	K_1=2676	K_2=1627	K_3=882	K=5185
$c_{t,1}$=14.03	$c_{t,2}$=15.59	$c_{t,3}$=17.15	$cong$=1.56	$c_{t,1}$=14.03	$c_{t,2}$=15.59	$c_{t,3}$=17.15	$cong$=1.56
$c_{nt,1}$=37.19	$c_{nt,2}$=33.91	$c_{nt,3}$=30.63	B=768 955	$c_{nt,1}$=37.19	$c_{nt,2}$=33.91	$c_{nt,3}$=30.63	B=768 955
mpc_1=98.80	mpc_2=90.43	mpc_3=82.59	C_V=360 921	mpc_1=98.80	mpc_2=90.43	mpc_3=82.59	C_V=360 921
mec_1=11.20	mec_2=14.57	mec_3=17.41	C_K=155 554	mec_1=11.20	mec_2=14.57	mec_3=17.41	C_K=155 554
p_1=110	p_2=105	p_3=100	S=252 480	p_1=110	p_2=105	p_3=100	S=252 480
$mb_{K,1}$=14.44	$mb_{K,2}$=9.75	$mb_{K,3}$=5.81	msb_K=30	$mb_{K,1}$=14.44	$mb_{K,2}$=9.75	$mb_{K,3}$=5.81	msb_K=30
$\partial R_1/\partial N_1$= 11.20	$\partial R_2/\partial N_2$= 14.57	$\partial R_3/\partial N_3$= 17.41	R=55 975 E=474 850	$\partial T_1/\partial N_1$= 11.20	$\partial T_2/\partial N_2$= 14.57	$\partial T_3/\partial N_3$= 17.41	T=0 Π=252 480
$\partial E_1/\partial K_1$= 15.56	$\partial E_2/\partial K_2$= 20.25	$\partial E_3/\partial K_3$= 24.19	Π=671 356 $\Pi^{\#}$=418875	$\partial T_1/\partial K_1$= 15.56	$\partial T_2/\partial K_2$= 20.25	$\partial T_3/\partial K_2$= 24.19	$\Pi^{\#}$=0
R_1=19 069	R_2=20 428	R_3=16 477		T_1=0	T_2=0	T_3=0	
E_1=348 003	E_2=92 119	E_3=34 729		Π_1=138 299	Π_2=76 654	Π_3=37 527	
Π_1=467 232	Π_2=148 345	Π_3=55 778		$\Pi_1^{\#}$=0	$\Pi_2^{\#}$=0	$\Pi_3^{\#}$=0	
$\Pi_1^{\#}$=328933	$\Pi_2^{\#}$=71 691	$\Pi_3^{\#}$=18 251		I_1=0	I_2=0	I_3=0	
I_1=328 933	I_2=71 691	I_3=18 251					

Note: rounding, if applicable, is either to nearest integer or (when decimal point is included) to the second digit

Table 2. First-best optimum for the numerical example

With a single transfer, these transfers become equal to zero again, confirming Proposition 2, and so do the adjusted profits, which is not overly surprising given that costs and prices are the same as in the parametric case, and a zero net transfer applies in both policies.

The profits with the separate manipulable tolls and subsidies, which are equal to net transfers received in that policy, depend strongly on the convexity of the congestion function. Figure 3 demonstrates this by showing these profits and transfers for varying levels of χ, the exponent of the congestion function. The variation in profits is substantial, being more than 16 times as high for χ=6 than for χ=2. This variation is much stronger than that in N (a ratio of 0.89 between the values for χ=6 and χ=2), K (a ratio of 1.62 between the values for χ=6 and to χ=2), or N/K (a ratio of 0.55 between the values for χ=6 and χ=2). This confirms the intuitive expectation that in particular with a more convex congestion function, the single manipulable transfer policy becomes relatively more attractive than the separate manipulable

toll and subsidy, at least when exact self-financing is desirable. Recall that this diverging pattern, for increasing levels of convexity, is consistent with the separate toll and subsidy leading to a zero profit and net transfer for the concave congestion function that results from the logarithmic specification, as shown in (20e).
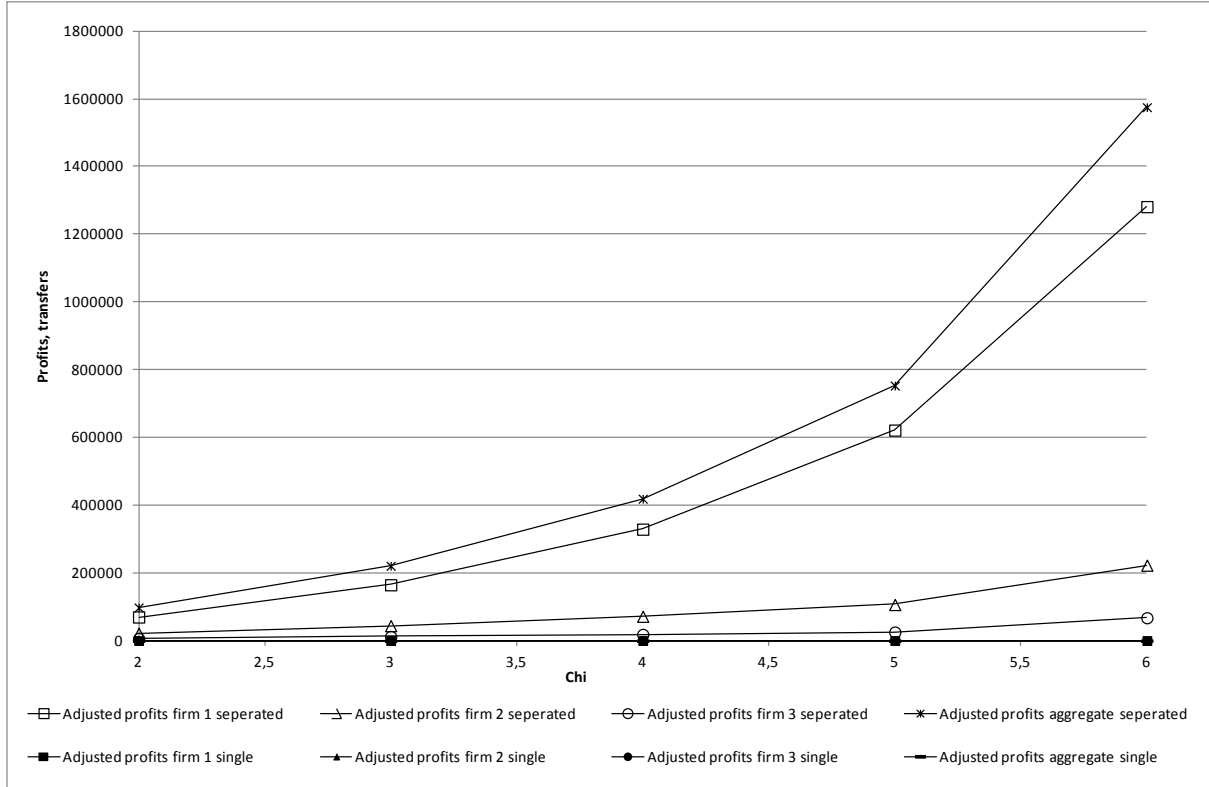

Figure 3. Sensitivity analysis: varying χ with manipulable tolls and subsidies

5. Conclusion

Common wisdom is that the Mohring-Harwitz theorem of self-financing infrastructure breaks down if the infrastructure is used by Cournot-Nash operators. However, the conventional analysis leading to this conclusion is asymmetric in the sense that it assumes that the toll rule is adjusted to properly reflect market power of Cournot oligopolists, while the investment rule is not. That is, the investment rule ignores that exactly because operators have market power, they also have an incentive to contribute to the provision of capacity. This paper proposed a regulatory scheme that utilizes this incentive. It turns out that this scheme restores self-financing, and even stronger than before: under the proposed scheme, exact self-financing now applies independent of the elasticity of the capacity cost function. Moreover, a balanced net budget not only applies at the aggregate level (*i.e.*, for the infrastructure operator), but also for each competitor individually. The result remains true both for the case where operators treat the tolls parametrically, and for manipulable tolls and subsidies as proposed by Brueckner and Verhoef (2010), at least if a single manipulable transfer is used that gives simultaneous incentives to optimize output and capacity. What is needed for the result to apply exactly, though, is that a particular form of equi-proportionality applies: each firm's share in the subsidized contributions to capacity should be equal to its share in total output at

the facility. Irrespective of how natural and intuitive this condition may seem, there is no reason why it should be satisfied spontaneously. Numerical results illustrated the rather intuitive point that a further deviation from equi-proporionality leads to larger discrepancies from exact self-financing.

The scheme seems attractive from different viewpoints. It reinforces the absence of a need to use distortionary taxes to raise funds to finance infrastructure, as a balanced budget now applies independent of the elasticity of the capacity cost function. Furthermore, the scheme may be perceived as relatively "fair", because there is not only a balanced budget in the aggregate, but also in the financial flows between the regulator and each firm individually. This may contrast with the perceived fairness of regular Cournot congestion tolls, which have the property that the largest operators should face the lowest tolls, and *vice versa*.

The paper emphasized the conceptual point that the scheme restores self-financing of congested infrastructure in the first-best optimum, now even independent of the elasticity of the capacity cost function. Practical implementation would raise some challenges that need addressing. Arguably, the most pressing of these is to cope with the non-stationarity of market equilibria in reality, while capacity investments are irreversible. This would make firms unwilling to contribute to the cost of capacity that concerns sunk costs from earlier investments. A way to reconcile this might be to tie the supply of effective capacity to a short-run complementary service, that can be credibly reduced in size if smaller contributions are made. Airport staff would be a good example. A second challenge is how to secure that the equi-proportionality of output and capacity shares is guaranteed, as required for exact self-financing. This requires a careful design of the process through which airport capacity is rationed across firms, and contributions to expansion can be made. An initial allocation of capacity contributions based on expectations of output volumes, leaving the option to make adjustments to these as operations commence, may make the equilibrium allocation sufficiently close to the theoretical case of exact self-financing to be useful in practice. But care should be taken not to introduce perverse incentives in such arrangements. The theoretical results in this paper obtained suggest that it is definitely worthwhile to think through such possibilities carefully. This would be a first priority for further study. Other natural follow-up questions would be to study the schemes for sequential capacity and output decisions (in a two-stage game set-up), and to consider the performance of the scheme under Stackelberg competition, Bertrand competition, and with dynamic congestion.

References

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

Arnott, R. and M. Kraus (1998a) "When are anonymous congestion charges consistent with marginal cost pricing?" *Journal of Public Economics* **67** (1) 45-64.

Arnott, R. and M. Kraus (1998b) "Self-financing of Congestible Facilities in a Growing Economy". In: D. Pines, E. Sadka and I. Zilcha (eds.) (1998) *Topics in Public Economics: Theoretical and Applied Analysis* Cambridge University Press, Cambridge UK, pp. 161-184.

Basso, L.J. and A. Zhang (2006) "Congestible facility rivalry in vertical structures" *Journal of Urban Economics* **61** 218-237.

Bergstrom, T., L. Blume and H. Varian (1986) "On the private provision of public goods" *Journal of Public Economics* **29** (1) 25-49.

Brander, J.A. and A. Zhang (1990) "Market conduct in the airline industry: an empirical investigation" *Rand Journal of Economics* **21** (4), 567-583.

Brueckner, J.K. (2002) "Airport congestion pricing when carriers have market power" *American Economic Review* **92** 1357-1375.

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

Brueckner, J.K. and E.T. Verhoef (2010) "Manipulable congestion tolls" *Journal of Urban Economics* **67** (3) 315-321.

Buchanan, J.M. (1969) "External diseconomies, external taxes, and market structure" *American Economic Review* **59** 174-177.

Daniel, J.I. (1995) "Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues" *Econometrica* **63** 103-130.

De Palma, A. and C.R. Lindsey (2007) "Transport user charges and cost recovery" *Research in Transportation Economics* **19** (1) 29-57.

Falkinger, J. (1996) "Efficient private provision of public goods by rewarding deviations from average" *Journal of Public Economics* **62** (3) 413-422.

Mohring, H. and M. Harwitz (1962) *Highway Benefits: An Analytical Framework* Northwestern University Press, Evanston, Illinois.

Newbery, D.M. (1989) "Cost recovery from optimally designed roads" *Economica* **56** 165-185.

Oum, T.H. and Y. Zhang (1990) "Airport pricing, congestion tolls, lumpy investment and cost recovery" *Journal of Public Economics* **43** 353-374.

Pels, E. and E.T. Verhoef (2004) "The economics of airport congestion pricing" *Journal of Urban Economics* **55** 257-277.

Pigou, A.C. (1920) *The Economics of Welfare* (First Edition). Macmillan, London.

Silva, H. E. and E.T. Verhoef (2011) "Optimal pricing of flights and passengers at congested airports. The effiency of atomistic charges" TI Discussion paper 11-179/3, Tinbergen Institute, Amsterdam-Rotterdam.

Silva, H.E., E.T. Verhoef and V.A.C. van de Berg (2012) "Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion" TI Discussion paper 12-056/3, Tinbergen Institute, Amsterdam-Rotterdam.

Small, K.A. (1999) "Economies of scale and self-financing rules with noncompetitive factor markets" *Journal of Public Economics* **74** 431-450.

Small, K.A. and E.T. Verhoef (2007) *The Economics of Urban Transportation* Routledge, London.

Vickrey, W.S. (1969) "Congestion theory and transport investment" *American Economic Review (Papers and Proceedings)* **59** 251-260.

Yang, H. and Q. Meng (2002) "A note on 'Highway pricing and capacity choice in a road network under a build-operate-transfer scheme'" *Transportation Research* **36A** 659-663.

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