SECOND-BEST ROAD PRICING THROUGH HIGHWAY FRANCHISING

Erik T. Verhoef*
Department of Spatial Economics
Free University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Phone: +31-20-4446094
Fax: +31-20-4446004
Email: everhoef@econ.vu.nl

This version: 25/01/06

Key words: Traffic congestion, second-best pricing, highway franchising

JEL codes: R41, R48, D62

Abstract
This paper considers the welfare impacts of a range of franchising regimes for congestible highways. For a single road in isolation, it is shown that a competitive auction with the level of road use as the decision criterion produces the socially optimal road (in terms of capacity and toll level), provided neutral scale economies characterize highway operations. The auction outperforms various alternatives, in which the bidders are asked to minimize the toll level or toll revenues, or to maximize capacity or the bid for the franchise. When second-best network aspects are taken into account, the patronage-maximizing auction is no longer optimal. With unpriced congestion on parallel capacity, the second-best highway would generate losses and the zero-profit condition becomes binding. The auction produces a below-optimal capacity. With unpriced congestion on serial capacity, the auction produces an above-optimal capacity. However, the patronage-maximizing auction does replicate the second-best optimum under a zero-profit constraint in both cases. An inquiry into the degree of generality of this result reveals that the first-order conditions suggest that this similarity would carry over to generalized networks, of undetermined size and shape. But second-order conditions are not fulfilled in general, and also corner solutions may occur. A numerical example is used to illustrate that the patronage-maximizing auction may then achieve the least efficient among the possible zero-profit roads.

*Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam.

1. Introduction

The private supply of highway capacity offers one way to deal with growing traffic congestion in the face of insufficient public funds to finance new capacity, and insufficient support for public road pricing. Private involvement in highway supply is not exceptional. Around one-third of the Western European highway network is currently under concession, with a strong concentration in the more Southern countries of France, Spain, Italy and Portugal. Some of the value pricing projects in the US involve private pay-lanes. And private toll roads are an increasingly common phenomenon in developing countries.

Proclaimed potential advantages of private over public highways include cost-efficiency, innovativeness, and availability of funds. A main disadvantage is the divergence between the private objective of profit maximization and the social objective of welfare maximization (*e.g.* Edelson, 1971; Mills, 1981; Mohring, 1985). An important question is whether there are ways, particularly through the design of auctions for highway concessions, to make the private operator behave more closely in line with welfare maximizing price and capacity setting. Such strategies might preserve the advantages of private involvement, while limiting the potential disadvantages. Moreover, a properly designed auction would provide incentives to minimize the cost of supplying the capacity chosen, and would give the government an objective way to select a road operator among a larger set of candidates. And of course, the use of auctions or comparable allocation mechanisms seems unavoidable in the awarding of concessions for highways, anyway. It is therefore important to have a proper understanding of the potential efficiency impacts of the design of such auctions.

This paper investigates one particular aspect of auctions for highway concessions, namely the extent to which the choice of the criterion used in the selection of the winning bid affects the efficiency of the resulting highway. The choice of criterion becomes relevant whenever the regulator is unsure about the optimal capacity and toll (schedules); possibly because these depend in part on the (efficiency of) the operator's other choices, for example during the construction phase. Under complete certainty, a criterion that awards the concession only to a bidder offering the optimal capacity and toll (schedule) would of course suffice to achieve the socially most desirable outcome, and the auction becomes a formality. We will mainly consider auctions in which private bidders are free to select highway capacity and toll, but will also briefly consider more limited auctions in which capacity is predetermined.

To focus attention, some simplifying assumptions will be made. First, we consider stationary traffic conditions with homogeneous users. Second, we ignore specific distortions that might arise from strategic interactions between bidders in the auction by considering competitive auctions only. There is no *a priori* reason to expect that these interactions would systematically affect the ranking of the various criteria that we will consider, although the welfare gains (or losses) from each criterion would of course be different under non-competitive bidding. Third, we will not formally model demand uncertainty and contract renegotiation. And fourth, we will assume that the government has sufficient power and credibility to enforce fulfilment of the bid, and can punish violations such that the net profitability of winning the auction but not living up to it would be negative and hence below that of not winning. The set-up of the analysis is similar to that of Ubbels and Verhoef (2004), but extends it by considering the impacts of an unpriced complement (in addition to an unpriced substitute); by considering maximum patronage as candidate criteria and, for pre-specified capacities, minimum toll revenues; and by using a different (non-linear, BPR-type) congestion cost function in the numerical model. But most importantly, unlike Ubbels and Verhoef (2004), this paper considers the (second-best) optimality of auctions in generalized networks, of undetermined size and shape – and finds a counter-example where a candidate second-best optimal auction in fact produces a minimum of achievable surplus levels.

The plan of the paper is as follows. Section 2 starts with some theoretical backgrounds. Section 3 considers the performance of a number of auctions for a single road, while Section 4 moves on to introduce network aspects. Section 5 considers the second-best optimality of the most promising auction on generalized networks. Section 6 concludes.

2. Theoretical backgrounds

This section provides some theoretical backgrounds for our analysis. Section 2.1 identifies the conditions for surplus-maximizing and profit-maximizing road capacities and tolls, and thus identifies the direction in and extent to which an auction should ideally affect the private operator's choices, compared to unrestricted freedom in setting the toll and capacity. Three cases are discussed: the benchmark of an isolated road, and two second-best cases allowing for simple network spill-overs, namely where either an unpriced substitute or an unpriced complement is available. Section 2.2 addresses the profitability of surplus-maximizing roads, and therewith identifies the desirability for auctions designed so as to push bidding companies to a zero-profit bid.

2.1. Welfare maximizing and profit maximizing tolls and capacities

Single road

Consider a single road with capacity K, which is used under stationary traffic conditions by homogeneous drivers with an aggregate inverse demand function $D(N)$, where N denotes the equilibrium flow of traffic. The average user cost includes all variable costs incurred by the user, including travel time, and depends, through congestion, on N and K. It is denoted $c(N,K)$. The generalized price faced by road users, $p(N,K)$, is equal to the sum of $c(N,K)$ and a toll τ if levied. The per-unit-of-time capacity cost depends on the road's capacity and is denoted $C^c(K)$. Ignoring external costs other than congestion, the first-best optimal toll and capacity, defined so as to maximize social surplus S, can be determined by solving the following Lagrangian:

$$\Lambda = \int_0^N D(n)\mathrm{d}n - N \cdot c(N,K) - C^c(K) + \lambda \cdot \big(c(N,K) + \tau - D(N)\big) \tag{1}$$

where the first three terms represent the objective, and the fourth term the equilibrium constraint (λ is the Lagrangian multiplier). The set of first-order conditions (w.r.t. N, K, τ and λ) can be solved to yield two familiar conditions:

$$\tau = N \cdot c_N \tag{2a}$$

$$-N \cdot c_K = C_K^c \tag{2b}$$

where subscripts denote partial derivatives. Equation (2a) shows that the optimal toll should be set equal to the marginal external congestion costs, while (2b) shows that the marginal benefits of capacity expansion (the l.h.s.) should be equal to the marginal cost (the r.h.s.).

An unrestricted private road operator would maximize profits by solving:

$$\Lambda = N \cdot \tau - C^c(K) + \lambda \cdot \left(c(N,K) + \tau - D(N)\right) \tag{3}$$

The set of first-order conditions now yield:

$$\tau = N \cdot c_N - N \cdot D_N \tag{4a}$$

$$-N \cdot c_K = C_K^c \tag{4b}$$

The profit-maximizing toll includes the marginal external congestion costs from (2a), but adds to this a standard monopolistic mark-up that increases as demand becomes less elastic. The latter term has the conventional interpretation; the internalization of congestion is motivated by the fact that any reduction in congestion costs can be turned into revenues for the operator by increasing the toll accordingly. Internalizing the congestion externality therefore contributes to the profit.

Interestingly, the profit-maximizing optimality condition for capacity choice (4b) is the same as for the social optimum (2b). The intuition is that the operator can, as it were, turn all savings in average user cost into toll revenues and hence profits on a dollar-by-dollar basis when increasing capacity while keeping the generalized price p and hence total demand N fixed. The profit-maximizing trade-off is therefore identical to the surplus-maximizing trade-off. Of course, the difference between (2a) and (2b) will generally cause the profit-maximizer to evaluate (4b) for a smaller N than a surplus-maximizer would consider, producing a lower optimal capacity. As a corollary, when demand is perfectly elastic so that (4a) becomes equal to (2a), a profit-maximizing operator would set its instruments such that the optimum is achieved, and no further regulation is warranted. Because estimates of demand elasticity for road transport usually indicate (in absolute terms) elasticities well below –1 (*e.g.* Hanly, Dargay and Goodwin, 2002), this observation is of limited use for practical policy making.

Under first-best pricing and capacity choice, equations (2a) and (2b) would hold for every single link in a network. And a private operator would attempt to add to this a monopolistic mark-up, as in (4a), for every origin-destination pair in the network. Indeed, network extensions become analytically more challenging when second-best conditions apply elsewhere on the network. An important type of second-best distortion would be the existence of untolled, congested links, with capacities set arbitrarily. Solving the resulting second-best optimal or revenue-maximizing tolls and capacities for generalized networks, of undetermined size and shape, can yield tedious expressions (see Verhoef, 2002ab) that elude easy interpretation. More insightful expressions can be obtained by considering two particular extensions of the one-link network considered above: one with an unpriced parallel link (a substitute), and one with an unpriced serial link (a complement). Figure 1 shows these simple networks, both serving a single origin-destination pair OD and both assumed to consist of an unpriced link *U* of given capacity, and a link *T* for which both a toll and capacity can be set.

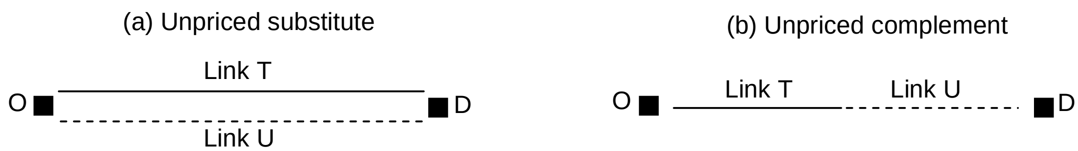

Figure 1. Simple two-link networks with an unpriced substitute (a) and an unpriced complement (b)

Unpriced substitute

The case of second-best congestion pricing with an unpriced substitute, in Figure 1.a, has been considered by various authors, including Lévy-Lambert (1968). The inclusion of capacity as a second policy instrument has been less common. Using superscripts *U* and *T* to denote the untolled and the tolled alternative, respectively, and assuming that K^U is to be treated as given, the surplus-maximizing second-best toll τ^T and capacity K^T can be found by solving the following Lagrangian:

$$\begin{aligned}\Lambda = & \int_0^{N^U+N^T} D(n)\mathrm{d}n - N^U \cdot c^U(N^U,K^U) - N^T \cdot c^T(N^T,K^T) - C^{c,U}(K^U) - C^{c,T}(K^T) \\ & + \lambda^U \cdot \left(c^U(N^U,K^U) - D(N^U+N^T)\right) + \lambda^T \cdot \left(c^T(N^T,K^T) + \tau^T - D(N^U+N^T)\right)\end{aligned} \tag{5}$$

The set of first-order conditions (w.r.t. N^U, N^T, K^T, τ^T, λ^U and λ^T) can be solved to yield:

$$\tau^T = N^T \cdot c^T_{N^T} - N^U \cdot c^U_{N^U} \cdot \left(\frac{-D_N}{c^U_{N^U} - D_N}\right) \tag{6a}$$

$$-N^T \cdot c^T_{K^T} = C^{c,T}_{K^T} \tag{6b}$$

where D_N denotes the slope of the (single) demand function.

The second-best optimal toll (6a) is the same as the one reported by Lévy-Lambert (1968), and is therefore unaffected by the possibility of also setting capacity for route *T*. A more detailed interpretation of this toll expression can be found in Verhoef *et al.* (1996), but note that it is below the marginal external congestion on route *T* in order to optimize the congestion spill-over onto route *U*. The second-best optimal capacity rule (6b) is similar to the first-best rule (2b). Given the equilibrium use level of route *T* (N^T) and the associated generalized equilibrium price p^T=c^T+τ^T, it is optimal to set capacity K^T such that the flow N^T is achieved through a combination of K^T and τ^T that minimizes the social cost of carrying N^T. The optimality condition for capacity is therefore the same as for a road without substitute.

The profit-maximizing toll and capacity follow from:

$$\begin{aligned}\Lambda = {} & N^T \cdot \tau^T - C^{c,T}(K^T) \\ & + \lambda^U \cdot \left(c^U(N^U, K^U) - D(N^U + N^T) \right) + \lambda^T \cdot \left(c^T(N^T, K^T) + \tau^T - D(N^U + N^T) \right)\end{aligned} \tag{7}$$

The set of first-order conditions (w.r.t. N^U, N^T, K^T ,τ^T, λ^U and λ^T) now yields:

$$\tau^T = N^T \cdot c^T_{N^T} - N^T \cdot D_N \cdot \left(\frac{c^U_{N^U}}{c^U_{N^U} - D_N} \right) \tag{8a}$$

$$-N^T \cdot c^T_{K^T} = C^{c,T}_{K^T} \tag{8b}$$

The tax rule (8a) is again not affected by the possibility to set capacity: the same rule was found in Verhoef, Nijkamp and Rietveld (1996) who keep capacity fixed. Note that, in contrast to the second-best toll in (6a), this tax rule adds a positive term to the common first term (that represents the marginal external cost on the tolled route). As for the single link, the optimality conditions for surplus-maximizing and profit-maximizing capacity, (6b) and (8b), are the same – and for the same reason. The equilibrium capacities will differ only because the point of evaluation differs.

Unpriced complements

Prior literature has paid considerably less attention to second-best pricing with an unpriced complement than with an unpriced substitute. Maintaining the assumption of a single origin-destination pair, and considering control over instruments at one of the two links only, the network of Figure 1.b emerges. The second-best optimum can be found by adapting (5) to the new network configuration (note that all travellers use both links U and T):

$$\begin{aligned}\Lambda = {} & \int_0^N D(n)\mathrm{d}n - N \cdot c^U(N, K^U) - N \cdot c^T(N, K^T) - C^{c,U}(K^U) - C^{c,T}(K^T) \\ & + \lambda \cdot \left(c^U(N, K^U) + c^T(N, K^T) + \tau^T - D(N) \right)\end{aligned} \tag{9}$$

The set of first-order conditions (w.r.t. N, K^T ,τ^T, and λ) can be solved to yield:

$$\tau^T = N \cdot \left(c^T_N + c^U_N \right) \tag{10a}$$

$$-N \cdot c^T_{K^T} = C^{c,T}_{K^T} \tag{10b}$$

Intuitively, the second-best optimal toll perfectly internalizes the marginal external congestion costs for both links jointly. The rule that defines optimal capacity has the by now familiar form. Again no welfare effects from link U are present in this rule, which reflects that indirect effects of changes in K^T upon congestion on link U cancel because the toll in (10a) already perfectly internalizes this congestion.

Finally, the profit-maximizing choice of instruments can be found from solving:

$$\Lambda = N \cdot \tau^T - C^{c,T}(K^T) + \lambda \cdot \left(c^U(N,K^U) + c^T(N,K^T) + \tau^T - D(N)\right) \tag{11}$$

The set of first-order conditions (w.r.t. N, K^T, τ^T, and λ) now yields:

$$\tau^T = N \cdot \left(c_N^T + c_N^U - D_N\right) \tag{12a}$$

$$-N \cdot c_{K^T}^T = C_{K^T}^{c,T} \tag{12b}$$

The profit-maximizing toll is a straightforward generalization of (4a), like the surplus-maximizing toll (10a) was from (2a). And also in this final network, the rule dictating profit-maximizing capacity is the same as the one for surplus maximization. The equilibrium capacities will again differ only because the point of evaluation differs.

2.2. Optimality, self-financing and maximized profits

Mohring and Harwitz (1962) showed that an optimally designed road – *i.e.*, with an optimal capacity and an optimal toll – will be exactly self-financing, provided some technical conditions are satisfied. These technical conditions can be summarized as follows: (1) road capacity should be a continuous variable; (2) there should be constant returns to scale in congestion technology (equiproportional changes in use and capacity leave average user cost unaffected); and (3) there should be constant economies of scale in highway construction (the cost per unit of capacity is independent of total capacity).[1] This 'self-financing' theorem has been shown to extend to each road individually in a full network, and therefore also to the network in aggregate, provided each link is optimally priced and all capacities are optimized (Yang and Meng, 2002). The theorem also extends to dynamic models (Arnott, De Palma and Lindsey, 1993); in present-value terms when allowing for adjustment costs and depreciation (Arnott and Kraus, 1998); when maintenance and durability are considered (Newbery, 1988); and when input markets are not competitive (Small, 1999).

Empirical evidence suggests that conditions (2) and (3) may hold at least approximately: empirical estimates of the ratio of long-run average and marginal costs are often relatively close to unity (Small, 1992, Sections 3.4, 3.5).[2] Profits or deficits under optimal design and pricing of a road will then be relatively small. Condition (1) seems unrealistic for a single road because the number of lanes is discrete. But capacity per lane can be varied by widening lanes, by resurfacing, or by re-grading or straightening a stretch of road. And when this is not the case, an optimally designed road might still be self-financing

---

[1] More generally, the original result in fact states that the degree of self-financing, measured as the ratio of toll revenues to capacity cost, is equal to the elasticity of capacity cost with respect to capacity. This implies exact self-financing under neutral scale economies.

[2] More recently, Levinson and Gillen (1998) report a point estimate for the ratio between long-run average and marginal cost of 0.92 for auto, but 1.45 for single trucks and 1.96 for combination trucks, suggesting mild diseconomies for passenger cars but considerable economies for trucks.

over the longer run, when periods of undercapacity and overcapacity alternate as demand grows over time.

We can illustrate the self-financing theorem in our model by replacing the general cost function $c(N,K)$ by the somewhat less general $c(N/K)$ (securing constant returns in congestion technology), and the capacity cost function $C^c(K)$ by $\gamma^c \cdot K$, with γ^c denoting a constant cost per unit of capacity (securing constant economies of scale in highway construction). Observe that:

$$\frac{\partial c(N/K)}{\partial K} \equiv -\frac{N}{K} \cdot \frac{\partial c(N/K)}{\partial N} \tag{13a}$$

All conditions for optimal capacity choice encountered thus far were of the same type, which can be rewritten (using our assumption of a constant cost per unit of capacity) as:

$$-N \cdot c_K = \gamma^c \tag{13b}$$

Multiplying both sides by K yields:

$$-K \cdot N \cdot c_K = C^c \tag{13c}$$

or, using (13a):

$$N \cdot N \cdot c_N = C^c \tag{13d}$$

The l.h.s. of (13d) gives total capacity cost when capacity is set according to (13b), which turns out to be equal to total toll revenues under the first-best pricing rule of (2a). This means that, whenever (13b) is satisfied but the toll rule deviates from (2a), optimal capacity choice will result in an unbalanced budget. There will be a financial surplus if the toll exceeds the marginal external congestion cost (as for all instances of private pricing we have considered, as well as for the case of optimal pricing with an unpriced congested complement). There will be a deficit when the reverse holds (as for optimal pricing with an unpriced congested substitute). For a road in isolation – or, indeed, a road in an otherwise optimally managed network – equation (13d) confirms the self-financing theorem.

These results have implications for the potential of competitive auctions to achieve efficiency. At least when no subsidies are part of the auction, a competitive auction would drive profits to zero. For a road in isolation, and for which the constant-returns-to-scale assumptions are fulfilled, the optimal profit is zero, too. This means that there could be competitive auctions that would have the optimum road as an outcome. For a road with unpriced complements or substitutes, this would not generally be the case.

3. Competitive auctions for an isolated road

A competitive auction can be defined as one in which a sufficiently large number of sufficiently equally efficient non-cooperative bidders are active, so that there is no scope for strategic behaviour, and the bidders do not believe they will stand a chance of winning the auction when submitting a bid with a lower than their best score on the selection criterion used in the auction. The assumption of a competitive auction, while empirically questionable, allows a ‘clean’ analysis of the performance of different selection criteria for an auction, without imperfections in the bidding process complicating the analysis. Such imperfections would in reality of course be of potentially decisive importance for the overall efficiency. However, the questions of to exactly what extent the performance of an auction would be affected, and of whether the ranking of the different selection criteria would be systematically affected, are left for future research.

For a competitive auction thus defined, any selection criterion that can be improved upon as long as profits are positive would cause bidders to be pushed closer towards a zero-profit bid. We will call such criteria ‘profit-exhausting’. All criteria we will consider in the sequel will be profit-exhausting. A somewhat unrealistic example of a criterion that would not be profit-exhaustive would be the criterion of ‘social surplus’ when an unpriced congested complement is available; compare equations (13d) and (10a) above.

Indeed, if information would be so complete that a meaningful auction with ‘social surplus’ as the criterion were possible, it would not be hard to define the optimal criterion for an auction. In practice, more easily observable criteria will have to be used. The set to be considered below is based on practical examples, earlier proposals in the literature, and on an attempt to find a relatively efficient criterion. The criteria to be considered are: a maximum bid for the right to build and operate the road (“*Bid*”), a maximum capacity supplied (“*Cap*”); a minimum toll charged (“*Toll*”); a minimum toll charged for a pre-defined capacity (“*Toll-cap*”); and the maximization of the use level or patronage of the new capacity (“*Pat*”).

If an auction is profit-exhausting, the occurrence of a ‘winner’s curse’ is of course not inconceivable: the winning bid is from the party that holds the most optimistic expectations about market opportunities, and that therefore stands a considerable risk of incurring losses once operation commences. We will not formalize the existence of a dispersion of expectations across potential bidders. But one way of dealing with this problem in reality would be to ask bidders to supply, along with their bid, a detailed account of the predicted use levels, travel times, toll levels, and road design. This would allow verification of the plausibility of the travel times as a function of road characteristics (capacity) and patronage, as well as patronage as a function of travel time and toll level. For the latter test, existing transport network models could be used; and further insights can be obtained by comparing bids with each other. With a deviation above a certain threshold, the bid could be ignored to protect the bidder from a winner’s curse, or clarification or revision could be demanded.

And finally, a credible and effective penalty should of course exist for under-performance compared to the bid, if wining. This penalty should be such that the firm should make a net loss from winning the auction and underperforming afterwards, and might be coupled to the government obtaining the right to set the toll levels when persisting underperforming occurs.

A numerical model: a single road

We will illustrate the discussion of the various possible selection criteria using the results of a small numerical simulation model. The model is highly stylized, but nevertheless calibrated so as to be representative for a highway that is congested during peak times. The average user cost function is modelled according to the well-known BPR-formulation (Small, 1992):

$$c(N/K) = \alpha \cdot t_f \cdot \left(1 + \beta \cdot \left(\frac{N}{K}\right)^{\chi}\right) \tag{14}$$

where α is a parameter reflecting the value of time (set at 7.5 in our model, according to conventional Dutch estimates), t_f is a parameter reflecting the free-flow travel time (set at 0.5, implying 60 km for a 120 km/hr highway), and β and χ are parameters that are set at 0.15 and 4, respectively – conventional values for the BPR-function.

The units of capacity are chosen such that one conventional traffic lane would correspond to K=1500. This implies a doubling of travel times at a use level of around 2400 vehicles per hour. This is roughly in accordance to the flow at which, empirically, travel times double for a single highway lane and the maximum flow on a lane is reached (*e.g.* Small, 1992, Fig. 3.4, p. 66). A maximum flow, however, is not defined for BPR functions.

The price of capacity, γ^{ς}, is set equal to 7. With a unit of time of one hour, this parameter ought to reflect the hourly capital costs. To derive a value from empirical construction cost estimates, an assumption has to be made on whether the model aims to represent stationary traffic conditions throughout a day, or during peak hours only. Our parameterization concerns the latter. The value of 7 was then derived by dividing the estimated average yearly capital cost of one highway lane kilometre in The Netherlands (€ 0.2 million) by 1100 (220 working days times 5 peak hours per working day; assuming two peaks) and next by 1500 (the number of units of capacity corresponding with a standard highway lane), and finally multiplying by 60 (the number of kilometres corresponding with a free-flow travel time of half an hour). Only welfare effects in peak hours are therefore considered in the numerical exercise, and it is assumed that off-peak travel is so modest that both the optimal off-peak toll and the marginal benefits of capacity expansion would be negligible. To maintain consistency, all selection criteria to be considered below, where relevant, would also apply to peak hours only. And finally, no relevant welfare effects arise outside the peak, and therefore no toll revenues are supposed to be raised.

Finally, it is assumed that a linear inverse demand function applies:

$$D(N) = \delta_0 - \delta_1 \cdot N \tag{15}$$

A choice of δ_0 = 31.21 and δ_1 = 0.00462, together with K = 3000, produced a desired benchmark equilibrium where an equilibrium road use of N = 5000 causes equilibrium travel time t to be around two times the free-flow travel time t_f, while equilibrium demand elasticity ε is equal to –0.35. Because there are no toll revenues, profit π is negative in the benchmark equilibrium. (This benchmark equilibrium will not be interpreted as some initial situation in the single-road analysis; that is, capacity will be allowed to become smaller than the benchmark level.)

	Benchmark	Optimum	Bid	Cap	Toll	Pat
t / t_f	2.16	1.37	1.37	1.01	n.r.[b]	1.37
ε	-0.35	-0.52	-2.05	-1.29	-∞	-0.52
K	3000.00	3530.77	1765.39	5807.29	0.00	3530.77
τ	0.00	5.58	15.82	13.78	n.r.[b]	5.58
N	5000.00	4430.50	2215.25	2949.19	0.00	4430.50
c	8.09	5.14	5.14	3.79	n.r.[b]	5.14
p	8.09	10.72	20.96	17.57	n.r.[b]	10.72
D	8.09	10.72	20.96	17.57	31.21	10.72
π	-21000.00	0.00	22686.70[a]	0.00	0.00	0.00
S	36787.70	45373.40	34030.00	20104.80	0.00	45373.40
ω	0	1	-0.32	-1.94	-4.28	1.00

[a] The figure shown in fact gives the bid. After making this bid, profit will become equal to zero.
[b] Not relevant.

Table 1. Numerical results for a single road

The optimum configuration is depicted in the second column of Table 1. As expected, profits are exactly zero in the optimum. Optimal capacity *K* is higher and optimal road use *N* is lower than in the benchmark. As a result, travel times are lower (1.37 times the free-flow travel time, compared to 2.16 in the benchmark).

Let us now turn to the various criteria for auctions. The first of these, *Bid*, forces the private operator to set the profit-maximizing toll and capacity identified in (4a) and (4b) (the net profit, after the sum promised in the bid has been paid, will of course be zero). This leads to a toll that is nearly three times as high as the optimal toll, and a capacity that is exactly half the optimal capacity (as can be expected with a linear demand function and constant long-run marginal cost). The final row in Table 1 shows an efficiency indicator ω, which is for a particular equilibrium calculated as the social surplus in that equilibrium minus that in the benchmark, divided by social surplus in the optimum minus that in the benchmark. It therefore gives the share of first-best surplus gains relative to the benchmark that a particular auction achieves; a negative value denotes a surplus below the benchmark level. This is for example the case for the auction *Bid*. The poor performance of this policy is in accordance with the rather pessimistic predictions of efficiency impacts of profit-maximizing congestion pricing in various earlier studies (*e.g.* Verhoef and Small, 2004).

A second auction, *Cap*, asks bidders to offer a capacity as large as possible. Because the toll is not restricted to be set optimally, the likely result is that capacity would exceed the optimal level: in the current numerical example it is nearly twice as large. The high capital costs are covered by a toll that is nearly as high as the profit maximizing toll, because it maximizes revenues given the capacity chosen. The resulting relatively small level of road use, in combination with the relatively large capacity, cause social surplus to be even lower than under *Bid*. Note that both auctions *Cap* and *Bid* will apply profit-maximizing tolls given the capacity chosen. But whereas in *Bid*, the capacity will be optimized given the inefficiently small use level, as implied by (4b), *Cap* will distort the capacity choice given the use level, by making capacity the bidder's maximand.

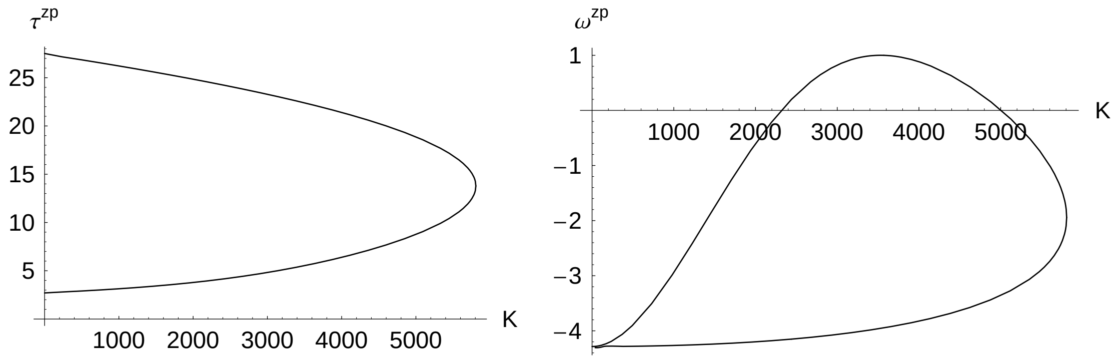

Figure 2. Zero-profit toll (left panel) and relative efficiency (right panel) as a function of capacity

A third auction, '*Toll*', would award the concession to the bidder requiring the lowest toll. Although this criterion may seem reasonable at first sight, it will generally fail to produce an interior solution with a positive capacity. The left panel of Figure 2 illustrates why. It shows the zero-profit contour in the K-τ space. This contour can be denoted $\tau^{zp}(K)$: the correspondence between zero-profit tolls and capacity. First note that for all capacities below the maximum capacity that can be offered without a deficit (*i.e.*, below the solution from auction '*Cap*', around 5807 in the numerical model), there are in fact two toll levels that produce zero profits. For any capacity chosen, it is the lower of these two toll levels that would result with the criterion 'toll'. The area bounded by the contour $\tau^{zp}(K)$ and the vertical axis corresponds with positive profits; the area outside the contour with negative profits. Sufficient conditions for a backward-bending pattern to arise is that the inverse demand function intersects both axes and that the absolute value of the elasticity of demand with respect to toll[3], denoted ε_τ, decreases monotonously in N. All revenue levels below the maximum revenue for a certain capacity (at $\varepsilon_\tau = -1$ for that capacity) can then be realized as a higher-toll – lower-demand combination (the upper segment of the contour $\tau^{zp}(K)$) or as a lower-toll – higher-demand combination (the lower segment of $\tau^{zp}(K)$). Because the profit-maximizing toll for a certain capacity lies between the two zero-profit tolls, and because of our assumption of ε_τ decreasing monotonously in N, profit is increasing in τ for a given capacity on the lower segment, and decreasing in τ on the upper segment.

Near a zero capacity, the minimum zero-profit toll would be a declining function of capacity under rather general conditions. Under the constant-economies-of-scale assumptions, this would be true provided demand is not perfectly elastic. A sketch of a proof is as follows. Consider a certain capacity and the zero-profit toll below the revenue-maximizing toll for that capacity. Then imagine a simultaneous equiproportional reduction of capacity and road use. This would leave average cost c unaltered. Therefore, to support this reduction as an equilibrium while marginal benefits D have increased, τ should rise. This means that a profit will be made: capacity and road use have fallen in the same proportion by construction, so

[3] The elasticity of demand with respect to toll ε_τ differs from the conventional demand elasticity ε because: (a) the toll τ differs from the generalized price p=c+τ, and (b) dN/dp ≠ dN/dτ (because dc/dN ≠ 0).

that maintaining zero-profits would require the toll to remain at its original level, instead of rising. Because revenues are increasing in the toll level in this range, the toll should subsequently fall in order to return to zero profits. Hence, the minimum zero-profit toll is a declining function of capacity. As a result, the auction '*Toll*' will not result in a winning bid with a positive capacity: bidders will be pushed towards bidding a zero capacity.

The right panel of Figure 2 depicts $\omega^{zp}(K)$: the correspondence between capacity and ω under zero-profit tolling. The upper segment of $\omega^{zp}(K)$ corresponds with the lower segment of $\tau^{zp}(K)$, and reversely. It shows that the outcome of the auction '*Toll*' is the least efficient among the possible zero-profit combinations of K and τ. This is a general result under not-perfectly-elastic demand, and is caused by the fact that under zero-profit tolling, marginal benefit is equal to average total cost (user cost and capital cost together). With a falling demand function, total benefit therefore exceeds total cost for any positive use level, and social surplus is always higher than in absence of the road (as resulting from '*Toll*').

The first-best optimum is on the lower segment of the contour $\tau^{zp}(K)$. An auction that would pre-specify the optimal capacity and next use the minimum toll as the criterion therefore would in principle be successful in achieving the first-best optimum. This auction '*Toll-cap*' would be close to the one proposed by Engel, Fisher and Galetovic (1996), who propose an auction with as the criterion the minimization of the net present value of toll revenues (NPR) before the highway is to be transferred to the government. The setting of Engel *et al.* (1996) is rather different from that in this paper. They are primarily concerned with the promotion of cost-effectiveness in construction and the avoidance of renegotiation of contracts under demand uncertainty; but they treat the choice of capacity as exogenous and ignore the effect of toll setting on social welfare (in fact, they assume that the social objective is to minimize the expected value of tolls paid). Demand uncertainty and renegotiation are ignored in the present paper, but the impacts on social welfare are, in contrast, central.[4]

From that perspective, a number of observations can be made concerning Engel *et al.*'s NPR-auction. The first is that Figure 1 implies that over the relevant range of capacities, any target level of toll revenues below maximum revenues could be achieved by two toll levels, with strongly diverging welfare implications. A criterion that is phrased in terms of toll revenues – be it per-unit-of-time or in present value terms – cannot discriminate between these two tolls. As a consequence, there is no guarantee that whichever toll revenue is raised per unit of time, it is raised using the more efficient toll. In the numerical example, even for the optimal capacity would the higher-toll equilibrium produce only very limited benefits compared to the no-road situation (ω is close to –4). A second observation is that the NPR-auction does not direct the operator towards an optimal toll revenue per unit of time: it is only the net present value that matters. If after transferring the road to the government, tolling is discontinued, the auction may cause pricing to be non-optimal both before and after the

---

[4] Cost-effectiveness in construction is not considered explicitly in this paper, and firms are assumed to always operate on the capacity cost function C^c. A competitive profit-exhausting auction would, however, always secure cost-effectiveness in construction. Explicit consideration of cost-effectiveness in construction would therefore not affect the conclusions.

transfer. It is therefore uncertain whether an NPR-auction would indeed produce an optimal outcome, especially when social surplus would be the social objective.

The final auction we consider is '*Pat*': the auction that awards the concession to the operator that offers the highest patronage. The final column in Table 1 shows that the outcome of this auction coincides with the first-best optimum. Again, this result can be expected to carry over to more general settings, as long as the constant-returns-to-scale-assumptions are fulfilled. The intuition is as follows. First, observe that the maximization of demand requires the minimization of generalized price, and hence of $\tau + c$. Next, zero profits imply that τ is equal to capacity cost per user, so that the generalized prize becomes $C^c/N + c$. Minimizing this with respect to K, as the bidder would need to do to maximize N, yields (for any positive N) an expression that is equal to (4b): the optimality condition for optimal capacity as a function of road use N. Because of the constant-returns-to-scale assumptions, the generalized price therefore coincides with the long-run marginal cost (user cost and capacity cost jointly), so that also N is optimized. And the joint occurrence of an optimal N and an optimal $K(N)$ secures achievement of the first-best optimum.

The perhaps counter-intuitive conclusion is therefore that, provided the constant-returns-to-scale conditions are fulfilled, the competitive profit-exhausting auction that maximizes social surplus is the one that maximizes traffic flow.

4. Second-best network effects

An important simplification of the above analysis concerns the neglect of network effects. This is acceptable when studying a road in isolation, or under the hypothetical assumption of first-best pricing throughout the rest of the network. It can also be considered instructive to deliberately ignore network complications, because doing so allows concentration on the primary efficiency impacts of the various auctions, independent of second-best network spill-overs. But network effects are likely to be important in reality, and may, as we shall see below, have significant impacts on the performance of auctions. To maintain focus and keep the exposition transparent, we will first consider two very simple networks in what follows, which would represent the most important types of second-best network issues that could arise. Section 4.1 considers the situation where an unpriced perfect substitute for the new road is available (*i.e.*, a parallel road), while Section 4.2 considers an unpriced complement (*i.e.*, a serial road). Generalized networks will be considered later in Section 5.

4.1. Unpriced substitute

The existence of an unpriced substitute road naturally reduces the potential profitability of the new road. This effect can be substantial, which is illustrated by the fact that when interpreting the rather heavily congested benchmark road from the previous section as pre-existing initial capacity, no profitable capacity-toll combination for additional, priced capacity appears to be possible. Also the second-best optimum, for which the capacity of and toll on the tolled new parallel is optimized under the constraint that initial capacity remains untolled, consequently produces a financial deficit. The results in Table 2 show that only 7% of the capacity cost for the second-best optimal toll road would be covered by the revenues from the second-best toll.

	Benchmark	Optimum	Second-best
K^U	3000.00	3000.00	3000.00
K^T	0.00	530.77	1227.48
K	3000.00	3530.77	4227.48
τ^U	0.00	5.58	0.00
τ^T	0.00	5.58	0.40
N^U	5000.00	3764.47	4010.11
N^T	0.00	666.03	1540.27
N	5000.00	4430.50	5550.38
c^U	8.09	5.14	5.55
c^T	8.09	5.14	5.14
π^T	0.00	0.00	-7974.39
S	36787.70	45373.40	42235.80
ω	0	1	0.63
Cost coverage T	0	1	0.07

Table 2. Numerical results for an unpriced substitute: original parameterization

Under this parameterization, no bids can be expected in an auction if it does not include the possibility of subsidies. Ubbels and Verhoef (2004) explore the possibilities for and properties of auctions with subsidies. In the present paper, we do not consider such auctions, motivated by the observation that if the required subsidy would be so large (93% of the construction costs in the numerical example), a government would most likely prefer to carry 100% of the construction costs and keep the road in public hands altogether. An auction that raises only 7% of the construction cost and hence requiring a 93% subsidy, while meaning loss of (direct) control over the highway operation, does not seem to be a very attractive option when social surplus maximization is the overall objective.

In order to get an idea of the performance of zero-subsidy auctions in the presence of an unpriced substitute, the parameterization has to be adjusted, so as to create the possibility of zero-profit bids with positive capacity. This was achieved in the numerical model by reducing the initial capacity from 3000 to 1500. As a result, the benchmark equilibrium travel time becomes 4.3 times as high as the free-flow travel time. For such a heavily congested road, zero-profit bids for additional priced capacity do become possible, and the results for the different criteria are shown in Table 3.

The first-best optimum (for which pricing on both roads is allowed) is of course the same as that for the road in isolation. Because initial unpriced capacity is relatively small, the second-best equilibrium achieves a relative efficiency of ω=0.91, which is substantial. However, because of the second-best nature of this equilibrium, the toll is set according to equation (6a), producing a toll that is only 16% of the first-best toll. As a result, a substantial deficit will occur on the operation of the tolled road: the cost coverage for the new road in the second-best equilibrium is only 16%, despite the fact that tolled capacity makes up nearly two thirds of total capacity. Again, it would seem more attractive to keep the road in public hands than to design an auction that would, if successful in reproducing the second-best optimum, require a subsidy of 84% of total construction cost.

	Benchmark	Optimum	Second-best	Bid	Cap	Toll	Pat	Second-best zp.
K^U	1500.00	1500.00	1500.00	1500.00	1500.00	1500.00	1500.00	1500.00
K^T	0.00	2030.77	2649.79	633.25	1430.90	0.00	1287.74	1287.74
K	1500.00	3530.77	4149.79	2133.25	2930.90	1500.00	2787.74	2787.74
τ^U	0.00	5.58	0.00	0.00	0.00	0.00	0.00	0.00
τ^T	0.00	5.58	0.87	8.27	6.95	n.r.[b]	5.58	5.58
N^U	3251.72	1882.24	2124.34	3053.85	2868.88	3251.72	2814.60	2814.60
N^T	0.00	2548.26	3325.02	794.61	1441.85	0.00	1615.90	1615.90
N	3251.72	4430.50	5449.36	3848.46	4310.73	3251.72	4430.50	4430.50
c^U	16.17	5.14	6.01	13.41	11.28	16.17	10.72	10.72
c^T	16.17	5.14	5.14	5.14	4.33	n.r.[b]	5.14	5.14
π_T	0.00	0.00	–15661.60	2138.11[a]	0.00	0.00	0.00	0.00
S	13941.20	45373.40	42480.00	25873.10	32453.50	13941.20	34873.40	34873.40
ω	0	1	0.91	0.38	0.59	0.00	0.67	0.67
Cost coverage T	0	1	0.16	1	1	1	1	1

[a] The figure shown in fact gives the bid. After making this bid, profit will become equal to zero.

[b] Not relevant.

Table 3. Numerical results for an unpriced substitute: adjusted parameterization

If insufficient public funds are available to finance this investment, the question rises of how attractive zero-profit roads might be. The final column in Table 3 shows the second-best optimum under an additional zero-profit constraint; so, the best achievable benchmark outcome for zero-profit auctions. The toll is substantially higher and capacity lower than in the second-best optimum. However, ω still reaches a level of 0.67 when the zero-profit constraint is added to the second-best problem. The levels of τ^T and c^T are the same as for the first-best equilibrium, which is caused by the facts that the auction induces the operator to minimize total cost for any given N^T while keeping capacity self-financing. This means selecting the same K/N ratio, and the same c and τ, as for the first-best optimum.

The auction '*Pat*' again achieves the second-best optimum (under the zero-profit constraint). Maximizing N^T under a zero profit constraint requires minimization of average user cost plus capital cost per user. Because of the network equilibrium condition, the minimization of the generalized price on road T implies that average user cost on road U are also minimized. N^U, and $N^T + N^U$, are therefore maximized – and so is therefore total benefit.[5]

[5] The outcome of the auction '*Pat*' appears to be independent of whether it is the patronage of the new capacity (N^T) or of both roads together ($N^T + N^U$) that is used as the criterion. Maximizing N^T through minimizing the generalized price on that road also maximizes N^U, because the generalized prices on both roads will be equalized in equilibrium.

For a given K^U, and given $\tau^U = 0$, these are the same conditions that define the second-best zero-profit equilibrium.

We can be brief on the other criteria. '*Bid*' still does not perform very good (ω=0.38), which is caused by the large discrepancy between the revenue-maximizing and the surplus-maximizing second-best toll; compare (6a) and (8a). '*Cap*' performs relatively good (ω=0.59), because there is now not much scope to expand capacity of the new road beyond the second-best zero-profit level without running into losses. And '*Toll*', as before, will not result in the supply of a positive capacity. Note that all ω's are positive. The reason is that road users as a group can only benefit from the supply of additional capacity that is to be used voluntarily, while profits will be zero. Social surplus, therefore, can only increase.

The absence of subsidization possibilities combined with the absence of pricing on initial capacity causes the maximum achievable welfare gains to be around two thirds of those from first-best pricing and capacity choice. The size of the relative loss, around one third in this example, evidently depends on the assumed initial conditions, and may in some cases become so large that the overall efficiency gain from the auction becomes unacceptably small. Would there, in such cases, be a possibility to enhance the social benefits from the auction by changing its set-up? One possible strategy, based on the observation that the source of the reduced efficiency gains is the existence of initial unpriced capacity, would be to stipulate that the winning bidder will have to buy the existing road against the best estimate of the current construction costs for the same capacity, and to allow the winning bidder to apply a congestion toll on this existing capacity. Provided the implied capital cost per unit of capacity for the initial capacity are the same as a bidder's cost per unit of new capacity, and provided the initial capacity is smaller than the capacity a bidder would choose in an auction, he will then in fact face the same problem as for the road in isolation. The auction '*Pat*' would consequently again achieve the first-best optimum. Therefore, there certainly may be ways to avoid particularly unattractive network spill-overs through auctions, by making the compulsory purchase of the associated links part of the concession.

4.2. Unpriced complement

The logical companion problem to the existence of an unpriced substitute is the existence of an unpriced complement. Table 4 shows the numerical results, for which in order to maintain comparability, the assumption was made that half the road's length would remain unpriced and at the benchmark capacity (3000). This segment thus functions as the unpriced complement *U*, while the other half (*T*) would be subject to the auction. The free-flow travel times t_f and prices of capacity γ^c therefore become 0.25 and 3.5 for both links, respectively.

Equations (10ab) already showed that there will be a financial surplus in the second-best optimum, because the second-best optimal toll also internalizes the congestion externality on the unpriced complement. The third column in Table 4 shows that in the numerical example, the revenues will consequently be more than twice as large as the capacity cost. The second-best optimum with an additional zero-profit constraint defines the best possible outcome for profit-exhausting auctions. The final column in Table 4 shows that ω drops to

0.67 (the similarity with the unpriced-substitute case in Table 3 is a coincidence). Again, the auction '*Pat*' is the only auction that reproduces the second-best zero-profit equilibrium: maximization of N under a zero-profit constraint apparently again implies maximization of the social surplus under that same constraint – given the inability to adjust K^U. The reason is that under zero-profit pricing the generalized price is equal to average total cost (that is: average user cost and per-user capital cost for link T jointly; capital cost for link U are ignored but are fixed anyway). Social surplus is therefore given by consumer surplus, and this is maximized when N is maximized by minimizing the generalized price.

	Benchmark	Optimum	Second-best	Bid	Cap	Toll	Pat	Second-best zp.
K^U	3000.00	3530.77	3000.00	3000.00	3000.00	3000.00	3000.00	3000.00
K^T	3000.00	3530.77	3276.57	2005.84	11440.40	0.00	3816.79	3816.83
τ^U	0.00	2.79	0.00	0.00	0.00	n.r.[b]	0.00	0.00
τ^T	0.00	2.79	6.76	14.98	14.07	n.r.[b]	2.79	2.79
$\tau = \tau^U+\tau^T$	0.00	5.58	6.76	14.98	14.07	n.r.[b]	2.79	2.79
N	5000.00	4430.50	4111.52	2516.98	2845.70	0.00	4789.46	4789.46
c^U	4.05	2.57	2.87	2.01	2.10	n.r.[b]	3.70	3.70
c^T	4.05	2.57	2.57	2.57	1.88	n.r.[b]	2.57	2.57
c	8.09	5.14	5.44	4.59	3.98	n.r.[b]	6.27	6.27
π^T	-10500.00	12357.70	16318.40	30690.60[a]	0.00	0.00	0.00	0.00
S	36787.70	45373.40	44893.50	34834.40	8218.64	-10500.00	42523.50	42523.50
ω	0	1	0.94	-0.23	-3.33	-5.51	0.67	0.67
Cost coverage T	0	1	2.42	1	1	1	1	1

[a] The figure shown in fact gives the bid. After making this bid, profit will become equal to zero.

[b] Not relevant.

Table 4. Numerical results for an unpriced complement

The relative ranking of '*Bid*' and '*Cap*' has reversed compared to the unpriced-substitute case. '*Bid*' leads to the profit maximizing outcome and therefore avoids the potentially substantial overinvestment in link *T*'s capacity that the revenues from implicit congestion pricing of link U allow. Indeed, K^T could be expanded up to more than three times its second-best level without running into losses; compare '*Cap*' and 'Second-best' in Table 4. Because '*Cap*' aggravates this distortion, its efficiency is relatively low. '*Toll*', finally, again suffers from the problem of not producing a positive equilibrium capacity.

Apart from making the compulsory purchase of the unpriced link part of the auction, as for the unpriced-substitute case, a simpler solution to the problem of over-investment appears possible in this case, and that would be to inform the bidders that they will be charged a toll equal to the marginal external congestion cost on link U for every user passing that link. This would take away the 'excess profits' and leave the private bidders facing the same conditions as in Section 3, meaning that '*Pat*' would again reproduce the optimum.

5. Generalized networks

The results so far look promising for the *Pat* auction. Given the restriction to zero-profit configurations, *Pat* was seen to reproduce the associated (zero-profit) second-best outcome in all three networks considered so far. The question is how general this result is. There is reason to doubt whether it carries over to more general networks, because it is then no longer true that the link under consideration serves all relevant origin-destination pairs. The maximization of the patronage of that link may then in fact raise travel costs for OD-pairs not served by it, through induced congestion elsewhere in the network. This could cause a deviation between the *Pat* auction and the zero-profit second-best road.

Analytical results

The generality of the second-best optimality of the *Pat* auction can be assessed by comparing the first-order conditions for two constrained optimization problems, both defined for generalized networks of undetermined size and shape. The first problem considers the second-best optimum of constrained maximization of social surplus when the toll and capacity can be optimized on only one single link. The other considers the *Pat* auction and has the link's flow as the objective, under the same constraints. If the two Lagrangians produce optimality conditions that are possibly mutually inconsistent, similarity of the two equilibria can be rejected for generalized networks.

We extend the notation from the previous sections as follows. There are M markets or OD-pairs, distinguished by index m; there are L links or arcs, distinguished by index l; and there are R routes or paths, distinguished by indices r or ρ (when a second index is required). We use dummies δ_{rm} ($\delta_{\rho m}$) to denote, when equal to 1, that route r (ρ) serves OD-pair m, and δ_{lr} ($\delta_{l\rho}$) to denote (also when equal to 1) that link l is part of route r (ρ). Furthermore, a dummy δ_r^A is used to indicate, when equal to 1, that route r is potentially 'active', meaning that it is in the equilibrium considered among the least cost routes for the associated OD-pair which itself has a positive flow. Finally, the link under consideration, for which the toll and capacity can be set, is denoted by $l{=}l^*$.

The second-best optimal choice for the toll and capacity under a zero-profit constraint can then be found by solving the following Lagrangian (note that OD-flows and link-flows are all expressed in terms of route flows):

$$\begin{aligned}\Lambda = {} & \sum_{m=1}^{M} \int_{0}^{\sum_{r=1}^{R}\delta_{rm}\cdot N^r} D^m(n)\mathrm{d}n - \sum_{l=1}^{L}\sum_{r=1}^{R}\delta_{lr}\cdot N^r\cdot c^l\left(\sum_{\rho=1}^{R}\delta_{l\rho}\cdot N^\rho, K^l\right) - \sum_{l=1}^{L} C^{c,l}(K^l) \\ & + \sum_{r=1}^{R}\delta_r^A\cdot\lambda^r\cdot\left(\sum_{l=1}^{L}\delta_{lr}\cdot\left(c^l\left(\sum_{\rho=1}^{R}\delta_{l\rho}\cdot N^\rho, K^l\right)+\tau^l\right) - \sum_{m=1}^{M}\delta_{rm}\cdot D^m\left(\sum_{\rho=1}^{R}\delta_{\rho m}\cdot N^\rho\right)\right) \\ & + \lambda^{l^*}\cdot\left(\sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\cdot\tau^{l^*} - C^{c,l^*}(K^{l^*})\right)\end{aligned} \qquad (16)$$

The first three main terms define the objective of social surplus. The constraints with multipliers λ^r are Wardropian equilibrium conditions, which will be invoked in the optimality conditions below only for active routes (with δ_r^A=1). The final constraint, with multiplier λ^{l^*}, gives the zero-profit condition for link l^*. Apart from this constraint, the Lagrangian in (16) is similar to those considered in Verhoef (2002ab), who studies second-best tolling on a sub-set of links, but for given capacities. These are the first-order conditions:

$$\begin{aligned}\frac{\partial\Lambda}{\partial N^r} &= \sum_{m=1}^{M}\delta_{rm}\cdot D^m(\cdot) - \sum_{l=1}^{L}\delta_{lr}\cdot c^l(\cdot) - \sum_{l=1}^{L}\sum_{\rho=1}^{R}\delta_{lr}\cdot\delta_{l\rho}\cdot N^\rho\cdot\frac{\partial c^l(\cdot)}{\partial N^r}\\ &\quad + \sum_{\rho=1}^{R}\delta_\rho^A\cdot\lambda^\rho\cdot\left(\sum_{l=1}^{L}\delta_{lr}\cdot\delta_{l\rho}\cdot\frac{\partial c^l(\cdot)}{\partial N^r} - \delta_{rm}\cdot\delta_{\rho m}\cdot\frac{\partial D^m(\cdot)}{\partial N^r}\right) \qquad \forall r:\delta_r^A = 1\\ &\quad + \lambda^{l^*}\cdot\delta_{l^*r}\cdot\tau^{l^*} = 0\end{aligned} \tag{17a}$$

$$\frac{\partial\Lambda}{\partial K^{l^*}} = -\sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\cdot\frac{\partial c^{l^*}(\cdot)}{\partial K^{l^*}} - \frac{\partial C^{c,l^*}(\cdot)}{\partial K^{l^*}} + \sum_{r}^{R}\delta_r^A\cdot\lambda^r\cdot\delta_{l^*r}\cdot\frac{\partial c^{l^*}(\cdot)}{\partial K^{l^*}} - \lambda^{l^*}\cdot\frac{\partial C^{c,l^*}(\cdot)}{\partial K^{l^*}} = 0 \tag{17b}$$

$$\frac{\partial\Lambda}{\partial\tau^{l^*}} = \sum_{r=1}^{R}\delta_r^A\cdot\delta_{l^*r}\cdot\lambda^r + \sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\cdot\lambda^{l^*} = 0 \tag{17c}$$

$$\frac{\partial\Lambda}{\partial\lambda^r} = \sum_{l=1}^{L}\delta_{lr}\cdot\left(c^l(\cdot)+\tau^l\right) - \sum_{m=1}^{M}\delta_{rm}\cdot D^m(\cdot) = 0 \qquad \forall r:\delta_r^A = 1 \tag{17d}$$

$$\frac{\partial\Lambda}{\partial\lambda^{l^*}} = \sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\cdot\tau^{l^*} - C^{c,l^*}(K^{l^*}) = 0 \tag{17e}$$

while:

$$\delta_r^A = 0 \quad \text{iff} \quad \sum_{l=1}^{L}\delta_{lr}\cdot\left(c^l(\cdot)+\tau^l\right) - \sum_{m=1}^{M}\delta_{rm}\cdot D^m(\cdot) > 0 \tag{17f}$$

These first-order conditions are now to be compared to those characterizing the *Pat* auction equilibrium, which can be derived from the Lagrangian that uses N^{l^*} as the objective and otherwise has the same constraints as (16):

$$\begin{aligned}\Lambda &= \sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\\ &\quad + \sum_{r=1}^{R}\delta_r^A\cdot\lambda^r\cdot\left(\sum_{l=1}^{L}\delta_{lr}\cdot\left(c^l\left(\sum_{\rho=1}^{R}\delta_{l\rho}\cdot N^\rho, K^l\right)+\tau^l\right) - \sum_{m=1}^{M}\delta_{rm}\cdot D^m\left(\sum_{\rho=1}^{R}\delta_{\rho m}\cdot N^\rho\right)\right)\\ &\quad + \lambda^{l^*}\cdot\left(\sum_{r=1}^{R}\delta_{l^*r}\cdot N^r\cdot\tau^{l^*} - C^{c,l^*}(K^{l^*})\right)\end{aligned} \tag{18}$$

The first-order conditions are:

$$\frac{\partial \Lambda}{\partial N^r} = \delta_{l^*r} + \sum_{\rho=1}^{R} \delta_{\rho}^{A} \cdot \lambda^{\rho} \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \delta_{l\rho} \cdot \frac{\partial c^l(\cdot)}{\partial N^r} - \delta_{rm} \cdot \delta_{\rho m} \cdot \frac{\partial D^m(\cdot)}{\partial N^r} \right) + \lambda^{l^*} \cdot \delta_{l^*r} \cdot \tau^{l^*} = 0 \qquad \forall r : \delta_r^A = 1 \tag{19a}$$

$$\frac{\partial \Lambda}{\partial K^{l^*}} = \sum_{r}^{R} \delta_r^A \cdot \lambda^r \cdot \delta_{l^*r} \cdot \frac{\partial c^{l^*}(\cdot)}{\partial K^{l^*}} - \lambda^{l^*} \cdot \frac{\partial C^{c,l^*}(\cdot)}{\partial K^{l^*}} = 0 \tag{19b}$$

$$\frac{\partial \Lambda}{\partial \tau^{l^*}} = \sum_{r=1}^{R} \delta_r^A \cdot \delta_{l^*r} \cdot \lambda^r + \sum_{r=1}^{R} \delta_{l^*r} \cdot N^r \cdot \lambda^{l^*} = 0 \tag{19c}$$

$$\frac{\partial \Lambda}{\partial \lambda^r} = \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l(\cdot) + \tau^l\right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) = 0 \qquad \forall r : \delta_r^A = 1 \tag{19d}$$

$$\frac{\partial \Lambda}{\partial \lambda^{l^*}} = \sum_{r=1}^{R} \delta_{l^*r} \cdot N^r \cdot \tau^{l^*} - C^{c,l^*}(K^{l^*}) = 0 \tag{19e}$$

while:

$$\delta_r^A = 0 \quad \text{iff} \quad \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l(\cdot) + \tau^l\right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) > 0 \tag{19f}$$

The two sets of equations (17) and (19) will generally not produce the same solutions for the Lagrangian multipliers (see also the numerical example below), which might suggest, at first glance, a discrepancy between the associated tolls and capacities. However, there is an essential similarity between the sets of equations (17) and (19). Equations (17c) and (19c) both imply the following relation between λ^{l^*} and the multipliers λ^r for the routes passing l^*:

$$\lambda^{l^*} \cdot N^{l^*} = -\sum_{r=1}^{R} \delta_r^A \cdot \delta_{l^*r} \cdot \lambda^r \tag{20}$$

(N^{l^*} is a shorthand for the total link flow on link l^*). Equation (20) means that the second-best investment rule of (17b) can be rewritten as:

$$\left(1 + \lambda^{l^*}\right) \cdot \left(-N^{l^*} \cdot \frac{\partial c^{l^*}(\cdot)}{\partial K^{l^*}} - \frac{\partial C^{c,l^*}(\cdot)}{\partial K^{l^*}} \right) = 0 \tag{17b′}$$

and the *Pat* investment of (19b) rule as:

$$\lambda^{l^*} \cdot \left(-N^{l^*} \cdot \frac{\partial c^{l^*}(\cdot)}{\partial K^{l^*}} - \frac{\partial C^{c,l^*}(\cdot)}{\partial K^{l^*}} \right) = 0 \tag{19b′}$$

For both conditions (17b′) and (19b′), the term between the large brackets repeats the conventional investment rule first encountered in equation (4b). Because λ^{l^*} in (19b′) reflects the marginal effect upon the optimized objective (N^{l^*}) from a relaxation of the zero-profit constraint (*i.e.*, from increasing the infrastructure budget), it will typically be positive, so that the conventional investment rule would be optimal when the objective is to maximize patronage under a zero-profit constraint. The intuition is as before: to maximize the use of the link, the generalized price should be minimized, which under zero-profit conditions means that the sum of capacity and user cost be minimized. Note that no neutral-scale-economies assumption is required for this to be true also in generalized networks.

More surprising, also for the second-best problem it is typically optimal to apply the conventional investment rule. Only when λ^{l^*} would happen to be exactly equal to –1 in the second-best optimum would the capacity choice seem immaterial for the value of the objective that can be achieved. With λ^{l^*} continuous, the probability that this occurs (exactly) in the second-best optimum would seem to be zero (there is no particular reason why $\lambda^{l^*} = -1$ should occur with a greater probability than any other value)[6] – and even if it does occur, the conventional investment rule would still be optimal.

What makes the conventional investment rule appropriate in this case? The explanation starts with a reminder that the first-order conditions (17) only define an extremum, not necessarily a maximum. The result can thus be interpreted as follows: when the conventional investment rule is applied, a marginal change in capacity under zero-profit tolling does not change social surplus. This result, in turn, can be understood after separating the possible effects on social surplus into two components: social surplus for all users passing the link considered (including the link's capacity cost), and social surplus for all other users. When the conventional investment rule is applied, the first component is maximized, for the same reason as given for the simpler networks considered earlier. The rule minimizes the total cost and the generalized price for this first group, and therefore maximizes their benefits. The first surplus component, involving all users of the link, is therefore insensitive with respect to small changes in capacity when the conventional investment rule is applied. But also the second surplus component, involving all other users, is insensitive. These other users can only be affected through congestion effects, which would occur when a small change in capacity of the link would lead to a small change in patronage. But because patronage is maximized at this capacity (because zero-profit tolling applies and implies that minimization of total costs on the link leads to minimization of the generalized price), no such indirect effects on surplus for other users will occur. Therefore, application of the conventional investment rule under zero-profit pricing leads to an extremum in social surplus. Whether it is always a maximum still needs to be determined, but if there is an interior maximum, it requires use of the conventional investment rule unless $\lambda^{l^*} = -1$ would happen to apply.

Such 'quasi first-best' capacity choice seems at odds with second-best pricing rules when unpriced congestion occurs elsewhere on a network, as studied by Verhoef (2002ab). Why would, with fixed capacities, a tax be set differently from the conventional (Pigouvian) first-best rule to account for these indirect effects, but would a capacity rule under zero-profit

[6] Also note that the interpretation of λ^{l^*}=–1 makes it unlikely to be true in a second-best optimum: it would mean that a marginal increase in the capacity budget for the link under consideration, above the zero-profit budget, would lead to an equally large loss in social surplus. This appears unlikely to be the case in the direct vicinity of the second-best optimum; see also the numerical example below.

tolling not be adjusted compared to the conventional first-best rule? In addition to the explanation just given, it is important to realize that the use of the tax instrument as such causes transfers of wealth only, while adaptations of capacity causes direct costs in that the sum of user and capacity cost on the link is no longer satisfied. Taxes are therefore a much more efficient (not socially wasteful) means of confronting users of the link to a certain extent with unpriced welfare effects elsewhere in the network.

And finally, it should be emphasized that the investment rule (17b′) applies only in an interior second-best optimum, where a positive capacity for link l^* is preferred to a zero capacity, and even then a check is in order to verify whether the investment rule produces a locally maximized, not minimized social surplus. In other words, intuitive doubts against this result may in fact correspond to situations where the investment rule does not produce a global second-best optimum for either of these two possible reasons. Before turning to these possibilities in the context of a numerical example, let us first finish the discussion of (17b′) and (19b’) for cases where the former does correspond to a global second-best optimum.

In both cases (17b′) and (19b′), therefore, the road operator will expand capacity on the link under control up to that point where, given that zero-profit tolling applies, the direct marginal benefit of capacity expansion (on link l^* itself) equals the marginal cost. If this equality occurs on one unique point along the link’s zero-profit contour in the K-τ space (such as shown in the right panel of Figure 2), the two equilibria must entail the same combination of K^{l^*} and τ^{l^*}. Whether this would be the case in general is not sure, as this may depend on the specific assumptions on the cost functions for road use and capacity provision. But for the constant returns-to-scale case, uniqueness of satisfaction of (17b′) and (19b′) along the zero-profit contour is easily established under a rather mild additional assumption. First note that when increasing τ along the zero-profit contour in the left panel of Figure 2, the ratio N/K on the link is gradually decreasing (superscripts l^* are dropped for convenience): otherwise the toll could not rise while keeping profits constant at zero.[7] Because the marginal cost for capacity C_K^c is now constant, uniqueness of satisfaction of $-N \cdot c_K = C_K^c$ along the zero-profit contour is guaranteed if the left-hand side is monotonously decreasing along the zero-profit contour when moving in the direction of an increasing τ. For this, in turn, to be true, it is more than sufficient to assume that $c''\geq 0$: the derivative of the user cost function with respect to the ratio N/K is not falling in that ratio. We can then write $-N \cdot c_K = \mathrm{d}c/\mathrm{d}(N/K) \cdot (N^2/K^2)$, which clearly increases in N/K and hence decreases when moving along the zero-profit contour in the direction of increasing τ.

In conclusion, the conventional investment rule applies both for an interior second-best optimum under a zero-profit constraint, and for maximizing patronage under the same constraint. If this rule can be satisfied for only one unique zero-profit capacity (which was not proven to be true in general but was shown to be plausible for the constant-returns case), the sets of first-order conditions (17) and (19) produce the same equilibrium. Provided equations

[7] This would be true for any not perfectly elastic derived demand for the use of the link, regardless of whether it would produce a backward-bending zero-profit contour as shown in Figure 2, a rising one, or a falling one.

(17) correspond to an interior maximum, the second-best optimality under a zero-profit constraint of *Pat* found for simple networks in Section 4 would then indeed carry over to generalized networks. This does not require the equilibrium values of individual Lagrangian multipliers to be equal in the two optimization problems, as can be verified when comparing conditions (17a) and (19a), and as we will also see in the numerical example below.

A numerical example: extending the serial links example

A simple illustration of the above results for a network with multiple OD-pairs, not all using the link under consideration, can be constructed by adding a second OD-pair that only uses link U in the serial-roads network of Figure 2(b), so that the two groups have the same origin O but different destinations D^A and D^B. Figure 3 shows the resulting network, where the original OD-pair is now distinguished with superscripts A and the new pair with B.

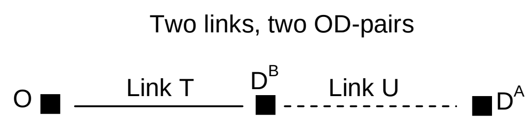

Figure 3. A simple two-link network for studying local versus global maxima

The demand parameters were recalibrated such that, with base capacities of 3000 for both links, both OD-pairs have an equilibrium demand of 5000 and a demand elasticity of –0.35. This was achieved by setting $\delta_0^A = 156.8$, $\delta_1^A = 0.0232$, $\delta_0^B = 141.2$, and $\delta_1^B = 0.0209$.

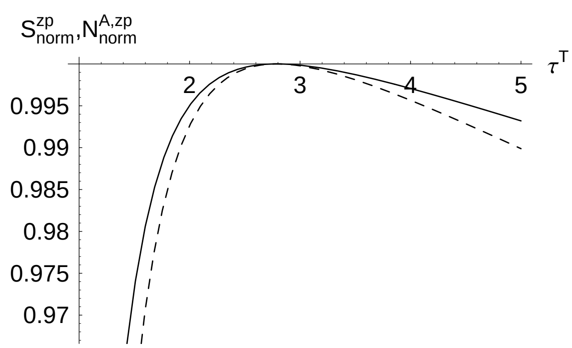

Figure 4. Surplus (solid) and patronage (dashed) for zero-profit equilibria as a function of toll (Both normalized at maximum = 1)

Figure 4 shows for this network the courses of social surplus S and patronage N^A as functions of τ^T when capacity K^T (not shown) is adjusted for each τ^T to maintain zero profits (S and N^A are both normalized at their maximized values, hence the subscript '*norm*' in the Figure). Both curves reach their maximum at the same toll level of τ^T = 2.79, the by now familiar level consistent with operations along the long run cost function (with t_f=0.25 and γ^c=3.5). The associated equilibria are then, of course, also identical in terms of variables N^A (4958.3), N^B

(5016.6), K^T (3951.4), and S. But the Lagrangian multipliers differ in value. In the second-best case, we find λ^A = 2628.2 and λ^B = 2918.7 for the two routes, and λ^T = –0.53006 for the zero-profit constraint. For patronage maximization the values are λ^A = –31.713, λ^B = 12.599, and λ^T = 0.006396. Note that not only the values of the multipliers but also their ratios and even sign patterns differ between the two sets. These differences illustrate that the similarity between the solutions for the two Lagrangians (16) and (18) derives from the fact that (20) applies for both cases, and not from some implicit equivalence between the two objectives.

Note that λ^T, the multiplier associated with the zero-profit constraint, is negative in the second-best equilibrium. This reflects that the second-best toll τ^T would be higher without this zero-profit constraint (as high as τ^T = 59.7 at K^T = 2472.0, with N^A = 3102.0 and N^B = 5675.0 in the second-best optimum). A relaxation of the constraint (setting revenues higher than capacity cost) would therefore reduce social surplus; hence the negative multiplier. This underlines that also under neutral scale economies, the zero-profit constraint will generally be binding when there is unpriced congestion elsewhere in the network.

We can build upon this same example to illustrate that equations (17) need not always define a globally maximized social surplus (given the constraints). One possibility would be that a local maximum defined by equations (17) is dominated by a global maximum at the corner solution where the link under consideration is completely eliminated. A second possibility is that equations (17) define a local (and possibly global) minimum. The extension we make to the previous example to illustrate these possibilities is to allow the value of time for group B, not using the link under consideration, to exceed that of group A. This raises the externalities that group A cause on group B. We consider three cases: vot^B=vot^A as in Figure 4 above (case 1), vot^B=2·vot^A (case 2), and vot^B=3·vot^A (case 3). In the latter cases, the demand parameters for group B are adjusted so as to maintain the same equilibrium use levels and demand elasticities as in case 1.

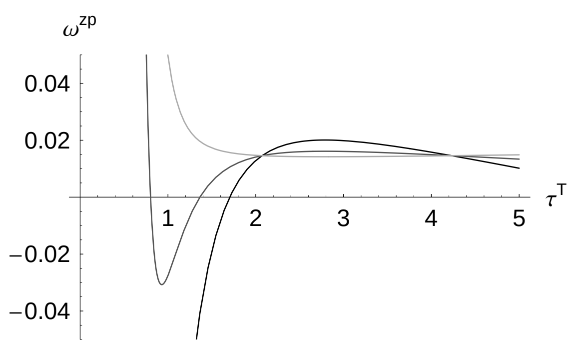

Figure 5. Local vs. global maxima and minima: relative efficiency for zero-profit equilibria as a function of toll for vot^B=vot^A *(black),* vot^B=2·vot^A *(dark grey), and* vot^B=3·vot^A *(light grey)*

Figure 5 shows for each of these cases the relative efficiency by toll level under zero-profit capacity setting. First of all it can be verified visually (and was checked numerically) that each curve yields an extremum at the focal toll level $\tau^T = 2.79$, which maximizes patronage of link *T* for each of these cases.

For case 1, $\tau^T = 2.79$ is a local and a global maximum, so that the *Pat* auction indeed produces the second-best zero-profit outcome.

In case 2, $\tau^T = 2.79$ again entails a local maximum. However, the congestion externality imposed by type-*A* drivers upon type-*B* drivers has become more important. A sufficiently large reduction in τ^T (to a value just below 1) brings us to a point where link *T* is operated so inefficiently (in terms of user cost plus capacity cost) that the social benefit of reducing the patronage of link *T* further, in terms of reduced externalities imposed upon group *B*, outweighs the loss in benefits for the inefficiently served group *A*. This effect is strong enough to make the highest possible surplus occurring when link *T* is effectively shut down (a toll level $\tau^T = 0.733$ corresponds with a zero capacity).

In case 3, $\tau^T = 2.79$ entails a local minimum. It still maximizes patronage of link *T*, but this minimizes social surplus because the congestion externality imposed upon group *B* is now the dominant welfare effect. In other words, conditions (17b′) and (19b′) still characterize the first-order conditions for capacity choice, but now define a local minimum for social surplus versus a local maximum for patronage.

The three cases thus illustrate the limitations of only comparing first-order conditions for Lagrangians: this may lead to a neglect of global (constrained) optima when these occur at a corner of the feasible space, and may also lead to the selection of a local minimum when a local maximum is strived for. Both limitations may cause the success of the *Pat* auction in achieving the second-best zero-profit outcome to break down.[8]

A question that remains open for further research is how likely such cases are to occur in reality. Note in particular that the examples where the *Pat* auction does not lead to a local and global second-best zero-profit auction all involve cases where the link under consideration is relatively unattractive from a social perspective. In this respect, the examples resemble the Braess paradox – and Appendix 1 shows how in a Braess-type network, the *Pat* auction may indeed lead to a local and global minimum of social surplus. Our analysis has shown that, provided the link is selected carefully in the sense that its patronage does not produce excessive external costs elsewhere in the network, the *Pat* auction keeps its attractive properties independent of the shape and size of the network. As long as links to be auctioned are selected with certain care, the potential problems indicated need not become manifest in actual applications.

6. Conclusion

Shortage of funds for road expansion, political unacceptability of public road pricing, and perhaps expectations of higher efficiency from private operations may all be factors that cause

[8] The equilibrium values of the Lagrangian multiplier associated with the zero-profit constraint are -0.53 in case 1, -0.80 in case 2, and -1.06 in case 3. This is consistent with our earlier hypothesis that an equilibrium value equal to –1 is unlikely to occur in a global second-best zero-profit optimum.

the private provision of toll roads to become an attractive option to cope with growing traffic congestion. Concessions for private road operation will typically be auctioned. This paper showed that the selection criterion used in such auctions may have a decisive impact on the efficiency of the resulting winning bid. A maximum possible bid for the right to build and operate the road pushes the bidders towards a profit-maximizing design, which is typically quite different from a surplus-maximizing road. The maximization of capacity typically leads to excessive capacity, in combination with a revenue-maximizing toll given that capacity; both reducing surplus below achievable levels. The minimization of tolls pushes bidders towards a zero capacity, unless capacity is set *a priori*. The minimization of toll revenues does the same, and in addition suffers from the fact that there may be multiple toll levels that, given a capacity, yield the same revenue but differ strongly in welfare impacts. However, an auction that asks to maximize patronage appeared to reproduce the first-best road in absence of network spill-overs and under neutral-scale-economies. It results in the second-best zero-profit configuration when network spill-overs exist. This was shown to be true in a few simple networks, but also to carry over to generalized networks, of undetermined shape and size, provided the external costs caused by the link's users elsewhere on the network are not so small that a complete absence of the link is in fact preferable to any zero-profit combination of toll and capacity.

Many important questions that need further consideration can be identified. A first one is whether a credible and efficient penalty system can be thought of that would guarantee the winning bidder to live up to the bid. A second one is whether a mechanism can be developed to cope with demand uncertainty and avoid renegotiation of contracts. A third one involves extension of the current analysis to larger, ideally 'generalized' networks, and multiple time periods (notably peak – off-peak). A fourth one involves strategic behaviour and interactions during the bidding process. The list could probably be extended easily, and illustrates that there is still sufficient potential for future research.

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 (1998) "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.

Braess, D. (1968) "Über ein Paradoxon aus der Verkehrsplanung" *Unternehmenforschung* **12** 258-268.

Edelson, N.E. (1971) "Congestion tolls under monopoly" *American Economic Review* **61** (5) 872-882.

Engel, E., R. Fisher and A. Galetovic (1996) "Highway franchising: pitfalls and opportunities" *American Economic Review, Papers and Proceedings* **87** (2) 68-72.

Hanly, M., J. Dargay and P. Goodwin (2002) "Review of income and price elasticities in the demand for road traffic". ESRC TSU publication 2002/13, Centre for Transport Studies, University College London.

Lévy-Lambert, H. (1968) "Tarification des services à qualité variable: application aux péages de circulation" *Econometrica* **36** (3-4) 564-574.

Mills, D.E. (1981) "Ownership arrangements and congestion-prone facilities" *American Economic Review, Papers and Proceedings* **71** (3) 493-502.

Mohring, H. (1985) "Profit maximization, cost minimization, and pricing for congestion-prone facilities" *Logistics and Transportation Review* **21** 27-36.

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

Newbery, D.M. (1988) "Road damage externalities and road user charges" Econometrica **56** 295-316.

Sheffi, Y. (1985) Urban *Transportation Networks: Equilibrium Analysis with Mathematical Methods* Prentice-Hall, Englewood Cliffs NJ.

Small, K.A. (1992). *Urban Transportation Economics. Fundamentals of Pure and Applied Economics*. Harwood, Chur.

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

Ubbels, B. and E.T. Verhoef (2004) "Auctioning concessions for private roads" Discussion paper TI 2004-008/3, Tinbergen Institute, Amsterdam-Rotterdam.

Verhoef, E.T. (2002a) "Second-best congestion pricing in general static transportation networks with elastic demands" *Regional Science and Urban Economics* **32** 281-310.

Verhoef, E.T. (2002b) "Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points" *Transportation Research* **36B** 707-729.

Verhoef, E.T., P. Nijkamp and P. Rietveld (1996) "Second-best congestion pricing: the case of an untolled alternative" *Journal of Urban Economics* **40** (3) 279-302.

Verhoef, E.T. and K.A. Small (2004) "Product differentiation on roads: second-best congestion pricing with heterogeneity under public and private ownership" *Journal of Transport Economics and Policy* **38** (1) 127-156.

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.

Appendix 1. Another numerical example: the Braess paradox

The importance of verifying whether the first-order conditions of a Lagrangian such as (16) lead to a maximized surplus can be further illustrated by considering a variant of the celebrated Braess Paradox (Braess, 1968). Figure A1 shows the network, which initially consists of 4 links 1-4 serving one OD-pair, making up two routes: 12 (using links 1 and 2) and 34 (using links 3 and 4). The link under consideration is link 5, that would connect links 1 and 4 and hence open up a third route 154. The Braess paradox might then arise when links 1 and 4 are relatively short but heavily congested, while links 2 and 3 have the opposite characteristics. The opening up of link 5 might then create so much extra congestion that its net welfare effects, even when provided at zero capacity cost, would be negative (see also Sheffi, 1985).

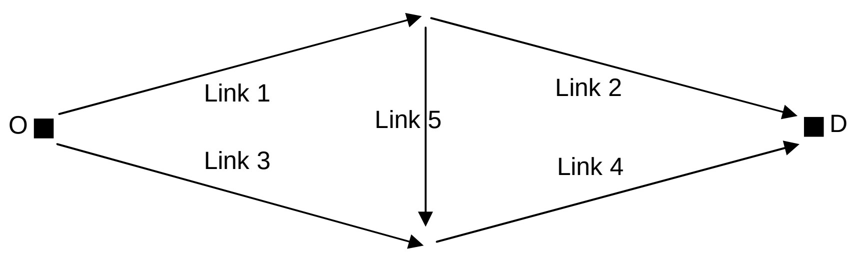

Figure A1. Network for considering the Braess Paradox

This was achieved in the numerical example by making links 2 and 3 uncongestible and assigning them a fixed average cost $c^2 = c^3 = 10$. Links 1 and 4 were assigned BPR cost functions as in (14), with free flow travel times of 0.25, capacities of 3000, and parameters α, β, χ, and γ^f as before. The demand parameters were calibrated so as to obtain an equilibrium flow of 5000 on both initial routes 12 and 34, and again a demand elasticity of –0.35. This required $\delta_0 = 54.17$ and $\delta_1 = 0.00401$. Finally, link 5's cost function was assumed to be identical to that for links 1 and 4, with the exception that a free-flow travel time of 0.1 applies.

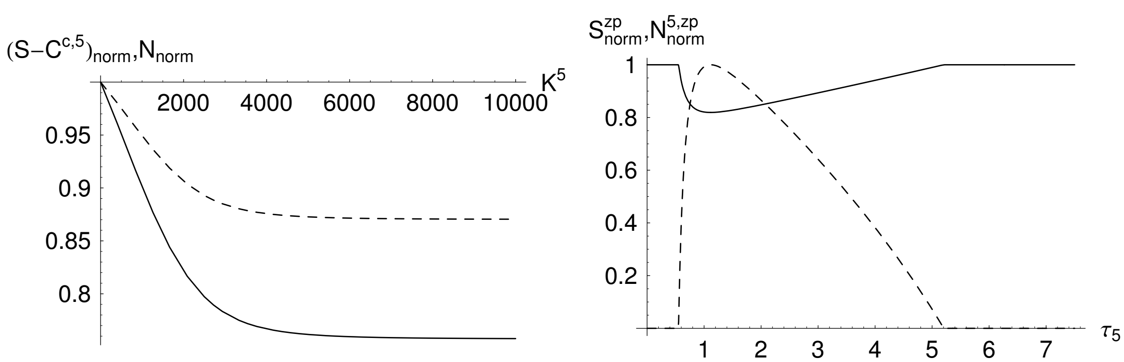

Figure A2. Braess paradox: Social surplus net of capacity cost (solid) and patronage (dashed) as a function of capacity of link 5 at zero tolls (left panel); and social surplus (solid) and link 5's patronage (dashed) for zero-profit equilibria as a function of toll (right panel)

(All normalized at maximum = 1)

The left panel shows how both social surplus, even net of capacity cost, and also patronage fall when expanding the capacity of link 5 in absence of tolling. The Braess Paradox therefore applies on this particular network. Although the Braess Paradox is known not to occur under first-best pricing, it does survive the introduction of second-best zero-profit pricing, at least in this example. The right panel shows how between toll levels of approximately 0.55 and 5.22, there are positive capacities at which link 5 can be supplied under zero-profit conditions. The dashed line shows how the link's patronage is positive, but the solid line shows that social surplus (now including the capacity cost for link 5) is below the level obtained in absence of link 5.

The right panel also shows that the maximization of patronage again minimizes social surplus.[9] The extrema occur at a toll $\tau^T = 1.116$, which implies the same flow-capacity ratio ($N^5/K^5 = 1.255$) as we found before, characterizing operation the long-run cost function (note that link 5 has a free-flow travel time of 0.1). In other words, conditions (17b′) and (19b′) still characterize the first-order conditions for capacity choice, but again define a local minimum for social surplus versus a local maximum for patronage.

[9] The equilibrium has the following properties: the route flows are $N^{12} = N^{34} = 2608.36$ and $N^{154} = 3833.79$ (link-flows and OD-flow can be derived from this); $K^5 = 3055.23$ and $\tau^5 = 1.11569$; $p^{12} = p^{34} = p^{154} = 17.8554$, and $S = 164351$ (capacity costs for links 1 – 4 are ignored).