

Journal of Transport Economics and Policy, Volume 42, Part 3, September 2008, pp. 463–493

Private Roads: Auctions and Competition in Networks

Erik T. Verhoef

Address for correspondence: Erik T. Verhoef, Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands (everhoef@feweb.vu.nl). The author is affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam.

Abstract

This paper studies the efficiency impacts of private toll roads in initially untolled networks. The analysis allows for capacity and toll choice by private operators, and endogenises entry and therewith the degree of competition, distinguishing and allowing for both parallel and serial competition. Two institutional arrangements are considered, namely one in which entry is free and one in which it is allowed only after winning an auction in which patronage is to be maximised. Both regimes have the second-best zero-profit equilibrium as the end-state of the equilibrium sequence of investments; but the auctions regime approaches this end-state more rapidly: tolls are set equal to their second-best zero-profit levels immediately, and capacity additions for the earlier investments are bigger. When discreteness of capacity is relevant and limits the number of investments that can be accommodated practically, the auctions regime may therefore still result in a more efficient end-state, with a higher social surplus, although the theoretical end-state is the same as under free entry.

Date of receipt of final manuscript: July 2008

1.0 Introduction

Over recent decades, there has been an increasing interest in private involvement in road infrastructure supply. One important reason is that declining government budgets motivate the search for alternative funds for financing desired road-capacity expansions. In addition, there is a rather widespread belief that the private sector would be inherently more efficient and innovative than their public counterparts, so that private roads may be built and operated at lower costs than public ones. Another consideration could be that the public at large may accept the imposition of tolls, generally believed to be important in curbing traffic congestion, more easily from private than from public operators.

There are, however, also potential economic hazards in the private supply of road capacity. Particularly, private toll-road operators would typically be interested in maximising profit rather than social surplus, and socially optimal first-best pricing cannot be expected from them — particularly, because the control of a road (section) will usually imply a certain degree of market power. The impacts on the private operator's price setting has been studied, for instance, by Edelson (1971), Mills (1981), Mohring (1985), Verhoef *et al.* (1996), Verhoef and Small (2004), and De Palma and Lindsey (2000). One recurring and probably not so surprising conclusion from such studies is that profit-maximising private road operators typically set congestion tolls above the optimal level: the profit-maximising toll not only internalises marginal external congestion costs, as the efficient toll does, but adds to this a monopolistic demand-related mark-up that rises as demand becomes less elastic. In addition, even though a profit maximiser has an incentive to offer the socially optimal amount of capacity given the prevailing level of demand, overpricing reduces demand, and hence the private supply of capacity is generally below the optimal level (for some further discussion, see, for example, Small and Verhoef, 2007).

Most studies of private road supply take the number of private suppliers as given. Usually only one operator is considered, sometimes a duopoly (for example, as in De Palma and Lindsey, 2000), but only seldom more. This may lead to a somewhat pessimistic picture of the efficiency of private toll roads: DeVany and Saving (1980) and Engel *et al.* (2004) show how profit-maximising tolls fall as the number of parallel competitors increases, approaching the optimal value as firms become infinitesimally small and competition becomes perfect. The limited attention for this theoretical benchmark result can probably be explained by the fact that perfect competition, with many parallel competitors, seems a rather theoretical option, due to the lumpiness of road infrastructure in practice.

It is not only just the number of competing private road suppliers that determines overall efficiency; it is also their distribution over the network. Small and Verhoef (2007, Ch. 6) illustrate this in a simple example, by studying how tolls and social surplus will vary if a road of a given capacity and length is split up and divided over an increasing number of symmetric private competitors in two contrasting cases: when they compete in parallel as substitutes versus when they compete in series as complements. In accordance with the two studies just mentioned, they find that the tolls approach the optimal level, which just internalises marginal external congestion costs, when the number of parallel competitors approaches infinity. Efficiency thus rises with the degree of competition. In contrast, when the number of serial competitors increases, so does a road user's aggregate toll (for using all serial road segments), and efficiency then falls with the number of competitors. These findings are in accordance with insights that Economides and Salop (1992) provide into the efficiency effects of mergers between serial and parallel firms in network markets. When looking at competition in network markets, it is important to consider explicitly, therefore, the configuration of the network and the distribution of competitors over that network — and in particular to distinguish between serial and parallel competition; that is, competition between substitutes and complements.

Besides competition, auctions for the right to operate a toll road can also be designed so as to improve the overall economic efficiency from private toll roads. Engel *et al.* (1996), for example, argue how a Net Present Value auction may be used to circumvent problems of renegotiation under demand uncertainty. Verhoef (2007) studies how the criterion used for selecting the winning bid in an auction can affect the efficiency of the resulting equilibrium. The classic criterion of the maximum bid pushes bidders towards the monopolistic profit-maximising toll and capacity, with the associated negative impacts on efficiency, and therefore does not seem to be very attractive from the social viewpoint. Perhaps surprisingly, when the winning bid is defined as the one that maximises the use or 'patronage' of the new road, the result will correspond with the second-best zero-profit combination of toll and capacity for the new link. That is the most efficient outcome for which one could reasonably hope when there is unpriced congestion elsewhere on the network (which is why it is second-best), and no subsidies are granted to private road operators bidding competitively (which is why a zero-profit constraint applies).

Verhoef (2007) derives these results for a first tolled link at an exogenous location in an otherwise untolled network. A natural follow-up question, addressed in this paper, is whether this 'patronage auction' retains its attractive properties in a more generalised setting. A first generalisation

is that also the location of the link to be supplied will now be part of the auction, because the franchise will be granted to the bidder that can attract the largest number of users to a new link, irrespective of its location in the network. A second generalisation is that we will now consider a sequence of auctions, each of which can be won by incumbents or entrants, so that entry into the network is introduced endogenously when new firms make the best bids. There are two natural benchmarks against which we can judge the performance of such a sequence of auctions. A first is a free-entry sequence, for which we assume that at each stage, a new link is added by the firm who realises the highest profits from doing so, and who sets the profit-maximising capacity and toll.¹ A second one is the sequence where at each stage, the socially most desirable link is added, with the second-best optimal capacity and toll. Both benchmarks will be considered in this paper. It brings us to the realm of sequential modelling of network evolution, a topic that has recently been addressed also by Levinson and Yerra (2006) and Zhang and Levinson (2006).

This paper thus studies the efficiency impacts of private roads in initially unpriced (hence public) networks. We allow for capacity and toll choice by private operators, and endogenise entry and therewith the degree of competition, allowing for both parallel and serial competition. Two institutional arrangements are considered, namely one in which entry is free and one in which it is allowed only after winning an auction. A number of simplifying assumptions are made for the dual purposes of keeping the analysis manageable and keeping the model transparent, so that an economic interpretation of the results is more easily given. The main assumptions are the following. The congestion externality forms the only relevant market failure. We consider identical road users, and firms that are equipped with identical cost functions for providing road capacity. There are neutral economies of scale in road construction and the congestion technology exhibits constant returns to scale (that is, the travel time functions are homogeneous of degree zero in traffic volume and capacity). Capacity is a continuous variable, but we will address qualitatively the question of how results might change when capacity would become discrete, as it is usually thought to be in reality. Auctions are perfectly competitive: there are no strategic interactions between bidders during the bidding phase and the

¹An anonymous reviewer remarked that this free-entry sequence could therefore also be interpreted as a sequence of 'classic' bid-maximising auctions at each stage. It is true that at each stage, the resulting capacity addition and toll would be the same for both sequences. A main difference would be that under competitive bid-based auctions, the payment of the bid immediately exhausts the profits, so that a firm would typically run into losses after additional parallel capacity is auctioned off at a later moment (this will in fact be illustrated later in Figures 6 and 7). For that reason, we will maintain the terminology and interpretation of a free-entry sequence versus a sequence of (patronage) auctions.

winner will realise a zero profit from carrying out the bid. Evidently, each of these assumptions is debatable empirically, and may thus offer worthwhile extensions for further research. The present paper deliberately focuses on this simplified environment, in the hope of deriving transparent results that are indicative of the main economic forces in this type of problem, which will remain relevant also in a more complex setting that allows for some of the complications just mentioned.

The paper is organised as follows. Section 2 introduces the model and the main assumptions underlying it, and discusses some analytical backgrounds. Section 3 describes the numerical version to be used in this paper. Section 4 contains the simulation results, and Section 5 concludes.

2.0 Model Set-up

2.1 Network configuration

We will consider what is probably the simplest possible network configuration that allows us to incorporate interactions between both serial and parallel roads in a network. This configuration is portrayed in Figure 1. There is a single market for trips between one origin (A) and one destination (B). The road ‘corridor’ between these locations consists of two serial segments a and b , which are connected through an uncongested crossing X .

Initially, there is only untolled capacity on both segments; these are the ‘initial links’ that will be denoted as links a_0 and b_0 . We will study how new links are added to these on both segments, under different institutional ‘regimes’. Note that the initial capacities can be set to zero without problem, so that absence of initial untolled capacity is just a special case of the proposed model.

Each new link covers either segment a or segment b , and is connected to the same crossing X . Because road users consider parallel links to be perfect substitutes, ‘Wardropian’ equilibrium conditions apply on both segments individually. This means that the generalised price faced by users, to be

Figure 1

The Initial Network with Untolled Links on Segment a and Segment b (Upper Diagram), and a Possible Later Configuration after some Links have been Added (Lower Diagram)



defined below, must be equalised on all links on a segment that can carry traffic, and cannot be lower on unused links for that segment (Wardrop, 1952). The lower diagram shows a possible network configuration after three links have been added; two on segment *a* and one on *b*. The dashing in the drawing aims to reflect that one firm has become active on both segments, and a second firm only on segment *a*. The exclusion, by assumption, of possible direct roads between *A* and *B* serves to maintain the original network structure with substitute and complement roads; allowing for more structural changes in the network configuration is an interesting generalisation for future study.

We consider stationary-state congestion and assume that road users are homogeneous. Their inverse demand for travelling between *A* and *B* is given by the inverse demand function $D(N)$, in which N denotes the number of trips per unit of time (or traffic flow).

The average user cost on a certain link l includes all variable costs incurred by the users, including travel time, and depends, through congestion, on the link flow N^l and the link capacity K^l . It is denoted $c^l(N^l, K^l)$. The generalised price faced by users of a link l , $p^l(N^l, K^l)$, is equal to the sum of $c^l(N^l, K^l)$ and a toll τ^l (if levied). Every possible route r uses two links, one on each segment, so that the number of possible routes is equal to the product of the numbers of links on segments *a* and *b*. Equilibrium is characterised by the following Wardropian conditions:

$$\forall r : \begin{cases} p^{a,r}(\mathbf{N}, K^{a,r}) + p^{b,r}(\mathbf{N}, K^{b,r}) - D(N) \geq 0 \\ N^r \geq 0 \\ N^r \cdot (p^{a,r}(\mathbf{N}, K^{a,r}) + p^{b,r}(\mathbf{N}, K^{b,r}) - D(N)) = 0 \end{cases} \quad (1)$$

$$N = \sum_r N^r,$$

where \mathbf{N} is the vector of route flows. Note that a certain link l may carry users from multiple routes, so that the generalised price for a link used by route r may depend on more route flows than just N^r (but, of course, not necessarily on all route flows). The composite superscript s,r denotes the specific link l that route r has on segment s . With condition (1) satisfied, the generalised prices for all used routes are equalised in equilibrium, and are equal to marginal benefits $D(N)$. Because users can freely choose combinations of links from segments *a* and *b*, the equilibrium conditions (1) imply that on both segments the generalised prices for all used links must be equalised.

Assuming that the social objective is to maximise social surplus, defined as user benefits minus user cost minus capacity cost, we can next find the

socially optimal or ‘first-best’ values of K^l and τ^l by maximising, subject to condition (1):

$$S = \int_0^N D(n) \, dn - \sum_l N^l \cdot c^l(N^l, K^l) - \sum_l C^{c,l}(K^l), \quad (2)$$

where $C^{c,l}$ is the capacity cost for link l . Because a link flow N^l is the sum of all route flows N^r for routes that use that link, and aggregate flow N is the sum of all route flows together, objective (2) can be maximised with respect to all route flows to find the short-run optimality conditions (these conditions also apply in the long-run optimum, in which capacity is also optimised). This produces first-order Kuhn–Tucker conditions that will not be written out, because they look very similar to conditions (1). The only difference is that $p^{s,r}$ is replaced by $mc^{s,r}$: the (short-run) marginal user cost on link s,r . Observe that mc^l , in turn, is the sum of the generalised average cost c^l and the marginal external cost $mec^l = N^l \cdot \partial c^l / \partial N^l$:

$$mc^l = \frac{\partial N^l \cdot c^l(N^l)}{\partial N^l} = c^l + N^l \cdot \frac{\partial c^l}{\partial N^l}. \quad (3)$$

Because $p^l = c^l + \tau^l$, the following tolls will consequently achieve short-run optimality:²

$$\tau^l = N^l \cdot \frac{\partial c^l}{\partial N^l}. \quad (4)$$

These are conventional ‘Pigouvian’ toll expressions, equal to mec , as can be found in nearly every transport economics textbook (for example, Small and Verhoef, 2007).

Optimal investment rules are found by optimising objective (2) with respect to link capacities K^l , which gives:

$$-N^l \cdot \frac{\partial c^l}{\partial K^l} = \frac{\partial C^{c,l}}{\partial K^l}. \quad (5)$$

The economic interpretation of equation (5) is straightforward: the marginal benefits of capacity expansion (the left-hand side), consisting of reduced aggregate user cost, should be equal to the marginal cost (the right-hand side). A few relatively straightforward manipulations are

²Optimal toll vectors need not be unique; in the current network, a constant could, for example, be added to all tolls on segment a and subtracted from all tolls on segment b without changing the equilibrium, so that equation (4) does not hold for any link but the optimum is nevertheless achieved. When demand is not perfectly inelastic, all optimal toll schedules produce the same aggregate route tolls, so that the total toll paid (over the full trip) by any individual is the same, irrespective of which among the possible optimal toll patterns is applied. The toll rule of equation (4) is, of course, the most natural and intuitive among these toll patterns.

sufficient to confirm the well-known ‘self-financing’ result of Mohring and Harwitz (1962).³ This result implies that when: (i) capacity is continuous, (ii) there are neutral economies of scale in road construction (that is, the marginal cost on the right-hand side of equation (5) is constant), and (iii) the congestion technology exhibits constant returns to scale (that is, the travel time functions are homogeneous of degree zero in traffic flow and capacity), the total toll revenues equal the total cost of capacity when equations (4) and (5) are both satisfied for all links. The optimal road network is then exactly self-financing: the profits π^l on each link l are all zero. This result is especially significant in the context of the present paper, because it means that if free entry of firms, and competition between them, eventually leads to a zero-profit outcome, this need not be inherently inconsistent with a first-best equilibrium. The same holds for a sequence of competitive auctions that drives down profits to zero. Nevertheless, because there are many possible combinations of tolls and capacities that produce a zero profit, zero profits are of course not a *sufficient* condition for optimality.

Throughout this paper, it will be assumed that the above conditions (i) to (iii), underlying the exact self-financing result, are satisfied. However, because we will allow for the continuing existence of unpriced and congestible initial capacity, the first-best outcome will generally be unattainable. The existence of unpriced congestion will for a parallel tolled link cause a downward adjustment on second-best tolls compared to Pigouvian tolls, so as to reduce congestion spill-overs (for example, Lévy-Lambert, 1968). In contrast, it typically creates an upward bias on the toll on a serial link, which is adjusted in an attempt also (partially) to internalise downstream or upstream congestion. As shown in Verhoef (2007), the existence of unpriced congestion on parallel or serial links does, however, *not* affect the second-best investment rules for newly added priced capacity. Consequently, the investment rule (5) remains valid for a tolled link with unpriced parallel or serial congestion, while the toll rule of (4) does not.

The consequence is that second-best investments and pricing then do not generally result in exact self-financing of newly added tolled capacity, simply because this would require (4) and (5) to be both satisfied.

³The first step is to multiply both sides of equation (5) by K^l . Because, by Euler’s theorem, $-K^l \cdot \partial c^l / \partial K^l = N^l \cdot \partial c^l / \partial N^l$ when c is homogeneous of degree zero as we assume, the left-hand side of equation (5) is then equal to total toll revenues. After the said multiplication by K^l , the right-hand side of equation (5) gives total capacity cost when the marginal cost of capacity $\partial C^{c,l} / \partial K^l$ is constant. Exact self-financing thus applies. The Mohring–Harwitz result is in fact more general than this, and states that the degree of self-financing (the ratio of total revenues and total capacity cost) is equal to the elasticity of the capital cost function with respect to capacity (see also Small and Verhoef, 2007).

Equivalently, when free entry or auctions cause long-run profits on tolled roads to become zero, while unpriced initial capacity remains available, the resulting equilibrium cannot be second-best (which has a non-zero profit or loss). At best, it would correspond to the ‘second-best zero-profit’ configuration (‘second-best’ because there is unpriced congestion elsewhere in the network; ‘zero-profit’ because the new capacity is restricted to produce a zero surplus). In our analysis below, we will therefore use both the ‘first-best’ and the ‘second-best zero-profit’ configuration as benchmarks for assessing the performance of the free-entry regime and the auctions regime.

2.2 Game-theoretic set-up in the ‘free-entry regime’

Let us next turn to the assumed game-theoretic set-up for the ‘free-entry regime’, for which it is assumed that operators are free to add capacity to the network, and are free in setting tolls. Before discussing various aspects of this regime in greater detail, it is useful to sketch the more general structure. A sequence of two-stage games is considered, where each two-stage game defines a ‘round’ in the sequence (the initial equilibrium is ‘round 0’). The second stage in such a game involves Bertrand toll competition between road operators. The first stage involves capacity choice for a single added link: it is assumed that in each round, only one link can be added to the network. Of course, there are multiple candidate road operators and multiple candidate locations (that is, segment a or b) for such an added link. For each candidate operator–link combination, the described two-stage game will be solved, and it is assumed that the operator–link combination that implies the highest profit gain for the associated operator is the one that will materialise. We then move to the next round; that is, the next two-stage game. Note that there is thus full rationality within each two-stage game, while we assume myopia between two-stage games. Let us now turn to the more detailed assumptions and, where needed, their justification.

First, we assume that all firms have access to the same technology, and face the same user cost functions $c^l(N^l, K^l)$ and capacity cost functions $C^{c,l}$.

Next, to account for the sequential character of network development in reality, we choose to consider a sequential game, with only one capacity addition in each round, instead of a game where all potential firms simultaneously decide how much capacity to add on which segment. The moments at which investments are made are exogenously determined in our model. We thus ignore the optimal timing of investments; we do this for simplicity and acknowledge that it offers an important possible extension of the present model. Exogenous timing could be relevant in reality

when the government would not allow multiple road construction projects to be carried out simultaneously.

Between capacity additions, the network configuration is given and the firms then present play a Nash–Bertrand game when setting their tolls. This means that they set their tolls τ so as to maximise their profit π , treating as given any other operators' tolls, as well as all link capacities. Note that this means that a firm operating more than one link sets all his tolls simultaneously, so as to maximise his aggregate profit, summed over all his links together.

Bertrand toll-setting behaviour of road operators, as assumed here, seems intuitively more plausible than a Cournot model, where players would assume that the flows on the competitors' links are fixed. Bertrand competition is therefore common in network models of competing operators (for example, De Palma and Lindsey, 2000; De Borger *et al.*, 2005). Furthermore, Nash behaviour seems a more neutral starting point than having a Stackelberg leader on the network, but this is another issue that may warrant further study (for example, Ubbels and Verhoef, 2008, compare Nash versus Stackelberg behaviour in a two-stage game-theoretic model of two competing governments supplying tolled infrastructure, and find that in their model the difference between the two types of competition is much more pronounced in the toll stage than in the capacity stage).

We now turn to the firm's behaviour when considering whether or not to invest and add capacity to the network. In fact, it is not so straightforward to choose an appropriate specification. A strict adherence to Nash behaviour might lead to a model in which it is assumed that a firm would not expect other firms to change their tolls in response to its own investment — even though the addition will make a non-marginal change to the network. However, this seems a rather naive assumption, especially if it is commonly known from earlier investments that firms do adjust their tolls when the system moves from the one Nash equilibrium to the other. This is why we use the two-stage set-up in each round, which implies that a firm realises that after it will have invested, a new Nash equilibrium in tolls will result.

Each firm, incumbent, or entrant, is assumed to calculate, for both segments of the network, for which level of investment the increase in its profits between the current and the new Nash equilibrium is maximised. *If* the firm invests in a certain round, it will then choose that segment and that capacity level that produces the highest profit gain. However, only one of these candidate investments will be made in each round and we assume that it is the one by the firm that has the highest profit gain from investing in that particular round. The motivation for this assumption

could be that in the absence of entry barriers, the firm expecting the largest profit would be that most likely to invest when only one addition can be made. When deciding on capacity additions, firms are therefore ‘nearly-myopic’: when investing, they optimise by looking no further than the immediate post-investment Nash equilibrium — but they do predict this equilibrium correctly.

It is important to acknowledge that there is some inconsistency in assuming, on the one hand, that the firm realises that, in the second stage, other forms will change their tolls after it has made an investment, and, on the other, assuming that the firm will nevertheless not set its toll and capacity on the new link as a Stackelberg leader. There are two reasons for accepting this inconsistency. One is that we prefer to leave the consideration of Stackelberg behaviour in investing and toll setting for later study, having Nash behaviour as the natural benchmark. The second is that it seems equally (or even more) inconsistent to assume that a firm behaves as a toll-leader when planning an investment, but next voluntarily moves to the role of follower when a next investment is made, by another firm.

Finally, note that the assumption of nearly-myopic behaviour, rather than forward-looking behaviour, is again consistent with Nash behaviour, in the sense that it prevents firms in our model from setting capacities strategically — that is, so as also to affect capacity choice by future entrants in the network; but it does leave open a question of ‘regret’. In particular, an undesirable property of a predicted equilibrium sequence of entries would be that at some moment along the sequence, one of the firms would regret earlier decisions, because it starts running into losses. We shall see that this does not occur in our model: profits will never fall below zero, which is due to our neutral-scale-economies assumptions. Therefore, although profits will fall over time, there is never a reason to regret having entered the market. Moreover, we will see that in the long-run end-point equilibrium, all profits will have fallen to zero, so that all firms will have become indifferent with respect to the capacity they chose when making their investments. The assumption we make on the sequential process is therefore not too harmful, in the sense that it does not lead to persisting different profitability levels for individual firms.

2.3 Auctions

The second regime of interest is the ‘auctions regime’. For this regime, we assume that there is a sequence of auctions in which the right to build a single tolled link on either segment a or b is granted to the firm that makes the best offer. As in the ‘free-entry regime’, we thus have a sequence

of equilibria, in successive ‘rounds’, at the beginning of which the network configuration is changed because a single link is added on either segment *a* or *b*. Following Verhoef (2007), we consider ‘patronage-maximising’ auctions, which in his model reproduce the second-best zero-profit outcome. With this auction, the right to build and operate a new addition to the network is granted to the firm that offers to carry the highest traffic flow on that new piece of capacity. We assume that bidding firms commit to a particular combination of capacity and toll, and that these imply a level of patronage for the new capacity which the regulator can calculate correctly, and use to determine the winning bid. We also assume that neither the toll nor the capacity can be changed when further capacity additions are made to the network later on. There is, therefore, no direct toll competition between firms.

The patronage-maximising auction is ‘profit-exhausting’, at least for firms who are not yet active in the network: under competitive auctioning, as we will assume to apply, it pushes newly entering firms to make an offer that produces a zero profit on the new capacity (Verhoef, 2007). It is important to realise that every traveller on the new link carries, as a ‘generalised’ price, the sum of average user cost and, through zero-profit tolling, average capacity cost. When a bid successfully maximises the use of the link, it must have minimised this sum of average user cost and average capacity cost. The auction therefore induces such newly entering operators to enter ‘according to the long-run cost function’ as we will call it. This means that the post-investment flow on the new link is served against minimised social cost. In other words, the first-best investment rule of equation (5) will apply. Because neutral scale economies apply, the accompanying toll that produces zero profits is also the first-best toll, given in equation (4), and its value is independent of the scale of operations. Therefore, the toll level is the same as it would be in the long-run first-best optimum, and the capacity is the one that minimises social cost for the post-investment flow level on the link.

Because only one right is granted in each ‘round’, the set-up induces firms again to behave nearly myopically, in the sense that they are asked only to maximise the immediate post-investment patronage of the new capacity. When further capacity additions follow later on, they should keep the toll to which they have committed unchanged, but the resulting changes in patronage are not considered to be a violation of their earlier bids.

We again face the question of whether firms may run into losses at a later stage because of the capacities they choose and the tolls to which they have committed. Because new entrants will always enter according to the long-run cost function, and will do so only when demand is large

enough to prevent losses from doing so, and because Wardrop equilibrium conditions imply that equal tolls on parallel links will lead to equal travel times and hence equal flow/capacity ratios on these parallel links, earlier investors need not fear losses as long as they have committed to the long-run first-best toll level.

As explained, new entrants, with no capacity from earlier investments, will indeed choose that toll level; but what about a firm that already has capacity on the serial segment? Because the firm is committed to the toll set earlier on its serial — say, downstream — link, its profit on that downstream link increases when the use of that link increases after an investment on the upstream segment raises the equilibrium flow N . As a result, the firm's aggregate profit, over both segments together, may be maximised at a patronage level for the new upstream link that exceeds the level consistent with entry according to the long-run cost function. This bid would then win from bids according to the long-run cost function, as new entrants would make; but it may involve a toll level below the long-run first-best level, and will then lead to negative profits after further additions on both segments, by other firms, would drive average cost and generalised prices to their long-run first-best levels on the links competing with those of the firm under consideration. The firm would then regret its earlier toll bid. We assume that firms will not make such bids, and apply a lower bound on the tolls they bid that prevent the investment from becoming loss-generating in the future.⁴ Effectively, this means that all capacity additions will be according to the long-run cost function.

2.4 Second-best zero-profit entries

Finally, we briefly characterise a sequence of link additions that we call the 'second-best zero-profit entries' sequence. This sequence is a relevant benchmark for judging the auctions regime. It involves a sequence of capacity additions that are chosen such that each addition has the maximum possible contribution to social surplus, under the constraint that the toll

⁴In the numerical analysis below, this assumption is 'binding' but has only limited consequences. The lower limit becomes binding in round 4, but only if links $a1$ and $a2$ would be operated by the same firm. That firm would then offer a higher patronage for link $b2$ than implied by entry according to the long-run cost function, namely 132 versus 109, at a slightly lower toll level: 2.761 versus 2.789. In round 5, also an operator with only a single link, namely $b1$, would offer a higher patronage for $a3$ than implied by entry according to the long-run cost function: 77 versus 12, again at a slightly lower toll level: 2.712 versus, again, 2.789. These differences in tolls are conceptually significant, as they indicate that the auctions regime and the second-best zero-profit regime are not formally identical in terms of their results if we do not add the assumption that firms will not bid tolls that will imply losses in a later phase. However, the size of the differences is negligible in the present model, especially because the auctions sequence is already very close to the end-state in round 4.

and capacity produce a zero profit on the added capacity — at least before any further capacity addition is made to the network. This sequence allows us to verify whether Verhoef's (2007) finding that the patronage-maximising auction produces the second-best zero-profit road for a first auctioned road, in an exogenously specified network, remains valid for a sequence of auctions in a network, which develops in an endogenous manner through these auctions.

3.0 A Numerical Model⁵

We will illustrate the relative performance of the three 'regimes' of interest by using the results from a numerical simulation model. The model is very similar to that used in Verhoef (2007), and the discussion in this section closely follows his exposition. The model is rather stylised, but still it is calibrated so as to be representative for peak-hour highway congestion. The average user cost function c^l takes on the well-known BPR (Bureau of Public Roads) form (see, for example, Small and Verhoef, 2007):

$$c(N^l/K^l) = \alpha \cdot t_f \cdot \left(1 + \beta \cdot \left(\frac{N^l}{K^l} \right)^\chi \right). \quad (6)$$

The parameter α represents the value of time, and is set at 7.5 in our model, close to the value (in Euros) currently used for official Dutch policy evaluations. The parameter t_f represents the free-flow travel time, and is set at 0.25 for both segments a and b , implying a total trip length of 60 km for a 120 kilometres per hour highway. Finally, β and χ are parameters that are set at 0.15 and 4, respectively; these are conventional values for the BPR-function.

The units of capacity are chosen such that a conventional traffic lane corresponds to $K^l = 1,500$. This implies a doubling of travel times at a use level of around 2,400 vehicles per lane per hour. This is roughly in accordance with the flow at which, empirically, travel times are twice their free-flow values for a single highway lane, and the maximum flow on a lane is reached (for example, Small and Verhoef, 2007, page 74, Figure 3.3). A maximum flow as such, however, is not defined for BPR functions. Note that the BPR function exhibits constant returns to scale in congestion technology: the underlying travel time function is homogeneous of degree zero in N^l and K^l .

⁵The exposition in this section draws heavily from Verhoef (2007).

Next, capacity cost is assumed to be proportional with capacity, so as to secure neutral scale economies in road construction:

$$C^{c,l}(K^l) = \gamma \cdot K^l. \quad (7)$$

The unit price of capacity, γ , is set equal to 3.5 for both segments. Because our unit of time is one hour, this parameter reflects the hourly capital costs. To derive a value from empirical highway construction cost estimates, we have to make an assumption on whether the model aims to represent stationary traffic conditions throughout a day, or during peak hours only. Our parameterisation concerns the latter. The value of 3.5 was derived by dividing the estimated average yearly capital cost of one highway lane kilometre in The Netherlands (€0.2 million)⁶ by 1,100 (220 working days times five peak hours per working day; assuming two peaks), and next by 1,500 (the number of units of capacity corresponding with a standard highway lane), and finally multiplying by 30 (the number of kilometres corresponding with a free-flow travel time of 15 minutes). Only welfare effects in peak hours are therefore considered in our model, 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, no relevant welfare effects are assumed to arise outside the peak, and therefore also no toll revenues are supposed to be raised.

Because firms are assumed to have access to the same technology, the cost functions of equations (6) and (7) apply, with equal parameters, to all firms — incumbents and entrants.

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

$$D(N) = \delta_0 - \delta_1 \cdot N. \quad (8)$$

We choose $\delta_0 = 61.27$ and $\delta_1 = 0.01167$, together with initial capacities $K^{a0} = K^{b0} = 1,500$, to obtain a sufficiently congested benchmark equilibrium, that allows a reasonable number of further capacity additions in a sequence of investments. The initial equilibrium road use of $N = 3,500$ causes equilibrium travel time t to be around 5.4 times the free-flow travel time t_f , which is high but empirically not unreasonable (it corresponds to a speed of around 22 kilometres per hour for a 120 kilometres per hour road). The equilibrium demand elasticity ε is equal to -0.5 in the initial

⁶With an infinitely-lived highway, without maintenance and an interest rate of 4 per cent, this implies investment costs of €5 million per lane-km, or €8 million per lane-mile. This order of magnitude is reasonably in accordance with figures that Litman (2006) presents for the USA. He quotes diverging estimates that suggest that the median investment cost per lane mile would be in the range of \$5 to 10 million; more than a third exceeds \$10 million.

Table 1
Key Characteristics of some Benchmark Equilibria

	<i>Base equilibrium</i>	<i>First-best</i>	<i>Second-best</i>	<i>Second-best zero-profits</i>	<i>Maximum added capacity (zero-profits)</i>
S	60,983	109,466	108,204	98,966	92,133
ω	0	1	0.974	0.783	0.643
π^{a0}, π^{b0}	-5,250	0	-5,250	-5,250	-5,250
π^{a1}, π^{b1}	-	-	-7,382	0	0
π^a, π^b	-5,250	0	-12,632	-5,250	-5,250
K^{a0}, K^{b0}	1,500	3,452	1,500	1,500	1,500
K^{a1}, K^{b1}	-	-	2,234	1,209	1,456
K^a, K^b	1,500	3,452	3,734	2,709	2,956
τ^{a0}, τ^{b0}	0	2.789	0	0	0
τ^{a1}, τ^{b1}	-	-	0.156	2.789	4.145
c^{a0}, c^{b0}	10.212	2.572	2.728	5.362	6.163
c^{a1}, c^{b1}	-	-	2.572	2.572	2.018
$D = p$	20.424	10.723	5.456	10.723	12.326
N^{a0}, N^{b0}	3,500	4,331	1,979	2,815	2,964
N^{a1}, N^{b1}	-	-	2,803	1,517	1,230
N^a, N^b, N	3,500	4,331	4,783	4,331	4,194

Note: Due to the assumed symmetry, equilibrium values for segments a and b are equal for all relevant variables, and are therefore shown in a single row.

equilibrium, while it will be equal to -0.21 in the second-best zero-profit outcome. Averaging over the extremes of the range of use levels considered in our analysis, we therefore find a reasonable -0.35 .

Table 1 provides the values of some of the model's key variables in a few benchmark equilibria. First, as a matter of notation, note that in Table 1 we use a slightly different double index to distinguish links than before, the first character (a or b) still indicates the segment, but now the second character identifies the individual link on that segment. A single index a or b refers to aggregates for a segment, summed over all its links.

The 'base equilibrium', in the first column, is as described above. Because no tolls are charged, the operator's profits π on both segments are negative, reflecting the capacity cost of the initial capacity. The generalised price in the 'first-best' optimum is nearly 50 per cent lower, despite the imposition of a toll. This is a consequence of the capacity expansion, which is in relative terms substantially bigger than the increase in traffic flow. As anticipated, both segments have a zero profit in the first-best optimum.

Next, the 'second-best' equilibrium, in which the initial capacity remains unpriced but the new capacity is tolled, has a remarkably high social surplus. The relative efficiency indicator ω , defined as the increase in social surplus compared to the base equilibrium, relative to the increase obtained through first-best pricing and capacity, amounts to 0.97. This is due to the

fact that the initial capacity is very low, so that the expansion of capacity brings substantial net benefits. Because the second-best toll is below the marginal external cost on the new capacity, a deficit now results.

Imposing a zero-profit condition, as in the ‘second-best zero-profits’ policy,⁷ avoids such a deficit, but at the expense of a lower relative efficiency ($\omega = 0.783$), and by setting a higher toll. In fact, the numerical value of the toll τ , as well as the flow/capacity ratio and therewith the average user cost c , are equal to their first-best counterparts. The intuition is that, under the constraint that the new capacity be self-financing, the best thing to do is to set capacity at a level that implies the minimisation of average social cost (user cost and capacity cost combined). Equilibrium route choice behaviour implies that this also minimises the average user cost on the initial capacity, given that new capacity should be self-financing. Therefore, the sum of the average costs that can be affected (all social cost components except capital cost of the initial capacity) is minimised, and because the generalised price faced by travellers is equal to the resulting average cost level, the social surplus is maximised. In this equilibrium, we are therefore adding new capacity according to the long-run cost function — using a social cost-minimising ratio of traffic flow and capacity, and a toll and a generalised price that would also apply in the long-run first-best optimum.

The final benchmark involves ‘maximum added capacity (zero-profits)’, where again the initial capacity is assumed to remain untolled. This equilibrium is included because it identifies the maximum level of new capacity that one could expect when a zero-profit constraint applies, either because profit-exhausting auctions are used or because free entry of road operators continues until profits are exhausted. In this equilibrium, relative efficiency ω is, not surprisingly, below the level under ‘second-best zero-profits’: 0.643. The toll τ on the tolled capacity, as well as the generalised price, exceeds the first-best level because we are no longer operating along the long-run cost function.

The values presented in Table 1 are the relevant benchmarks against which to assess the values of key variables at various stages during the three regimes of interest, ‘free entry’, ‘auctions’ and ‘second-best zero-profit entries’. This is what we will do in the next section.

⁷It is perhaps important to note the difference between the ‘second-best zero-profits entries’ regime introduced in the previous section, and the ‘second-best zero-profits’ benchmark discussed here. The former imposes a zero-profit constraint on newly added capacity on one of the two segments a or b , keeping capacity at the other segment fixed. This leads to a sequence of capacity additions — alternately on segments a and b , as we will see shortly. The ‘second-best zero-profits’ benchmark equilibrium in Table 1, in contrast, allows for a simultaneous optimisation of newly added (tolled) capacities on both segments. After optimisation, there is of course no scope for further zero-profit capacity additions, so this benchmark involves a single static equilibrium, not a sequence.

4.0 Simulation Results

4.1 Patterns of entry and network growth

A first property of interest of the three regimes concerns the pattern of entries, which is characterised not only by the specific order of additions to segments a and b , respectively, but also by the identity of the firm that makes the investment. For the free-entry regime in our numerical model, we find a very regular pattern of entries, where in every odd 'round' a new firm enters on segment a , while in the even round that follows, an addition to segment b is made by that same firm (given the assumed symmetry in the network we can, without loss of generality, assign the label a to the segment to which the first firm makes the first addition).

Although this pattern is not the only possible equilibrium sequence under free entry,⁸ there is a good economic intuition for why it should be a plausible pattern. After the first investment on segment a , by a firm that we will refer to as firm I (firms will be numbered consecutively by roman numbers in order of entry), it is plausible that segment b is more attractive to enter for a new entrant (firm II) than segment a , because there will be less competition and a smaller aggregate capacity on segment b than on a . It is also immediately clear that segment b must be more attractive than segment a for firm I: we do not expect a possible profitable investment on segment a for firm I if it already optimised the toll and capacity of its added capacity on segment a in round 1. Finally, the incumbent firm I will enjoy a higher profit increase from adding capacity to segment b than a new entrant does, because firm I can maximise the joint profits on both segments, while a new entrant II will end up in a situation of serial competition with the incumbent I. Therefore, in round 2, it is plausible that firm I should add capacity to segment b .

Next, when firm I optimises its link on segment b in round 2, the capacity on segment a is larger than was the capacity on segment b when that same firm I optimised its first investment, on segment a . As a result, it is likely that it chooses a larger capacity for its link on segment b in round 2, than for its link on segment a in round 1. A potential new entrant will, in round 3, therefore find segment a more attractive to enter than segment b ; but also the incumbent firm I would prefer investing on segment a over segment b in round 3, as it has just optimised its link on segment b . The question therefore is this: will a new entrant II foresee greater profits from investing in segment a than the profit increase expected by the

⁸For a different parameterisation of the numerical model, for example, we found three successive entries by the first firm entering, on segments a , b , and then again a .

incumbent firm I? This cannot be said with certainty. The incumbent firm has the advantage that it can avoid competition between links on segment a , so it is likely to end up with higher tolls. But the incumbent firm has the disadvantage that new capacity will reduce demand for its earlier capacity on segment a . It imposes, as it were, a pecuniary externality upon the profitability of its own earlier capacity. The incumbent firm will take into account the implied fall in profits on its earlier capacity, a loss that a new entrant will not face. Depending on which of these two effects dominate, it may be the incumbent or a new entrant who invests on segment a in round 3. In our numerical model, it is the new entrant, whom we will refer to as firm II.

Finally, in round 4, there are six possible entries to consider: the incumbent firms I and II and a new entrant may each add capacity to segments a or b . Because the aggregate capacity is now larger on segment a while the tolls are lower, each of these firms would prefer an investment on segment b . The comparison between the profit gains for firms I and II involves the same trade-off as just described for round 3, and also for round 4 it results in a net advantage for firm II in our numerical model. The comparison between firm II and a new entrant, firm III, involves the same trade-off as described above for round 2, and again it results in a net advantage for firm II. Firm II therefore invests in segment b in round 4.

This pattern of new firms entering on segment a in an odd round and, after that, on segment b in the succeeding even round, is maintained in our numerical model as far as we have tested it (four firms; eight rounds). As stated, this pattern is not the only possible equilibrium sequence, but it is a likely pattern because of the considerations and trade-offs sketched above. Because it results in a configuration with parallel competition on both segments, it suggests that the inefficient pattern of serial monopolists studied by Small and Verhoef (2007), with a single monopolist on each serial segment, will not easily arise spontaneously, in its pure form, as the outcome of free entry of road providers. At the same time, because road users can switch between road providers at the intersection halfway between the two segments in our model, an operator that sets its toll on the one segment is likely to consider the tolls on most of the links on the other segment as given, so it remains to be seen whether the resulting tolls tend towards the competitive level, as suggested by the degree of parallel competition on both segments individually, or to a higher level, as suggested by the persisting pattern of serial competition between segments. We will, of course, turn to this question shortly.

First, we will discuss the pattern of network formation under a sequence of auctions. This pattern appears to be identical to the one just described (although the capacities and tolls will be different): there are alternate

additions to segments a and b (again, we can freely label the segment of the first addition as segment a). As discussed in Section 2.3, the patronage-maximising auction forces bidders to operate with zero profits along the long-run cost function. After a winning bid has been implemented on, for example, segment a , it is therefore impossible to make a further expansion on that segment without running into losses. Hence, if a next bid is made, it must be on segment b ; and because the expansion just made on segment a raises demand, it will generally be possible to make such a bid, involving a positive patronage and capacity. The sequence will thus have alternating additions to segments a and b .

An important difference with the free-entry regime is that, under the auctions regime, the identity of the firms entering is immaterial. The reason is that the bidders set tolls according to the long-run cost function (at a level of 2.789 in our numerical model; see Table 1), and are restricted to remain committed to these toll levels also after further additions are made to the network.

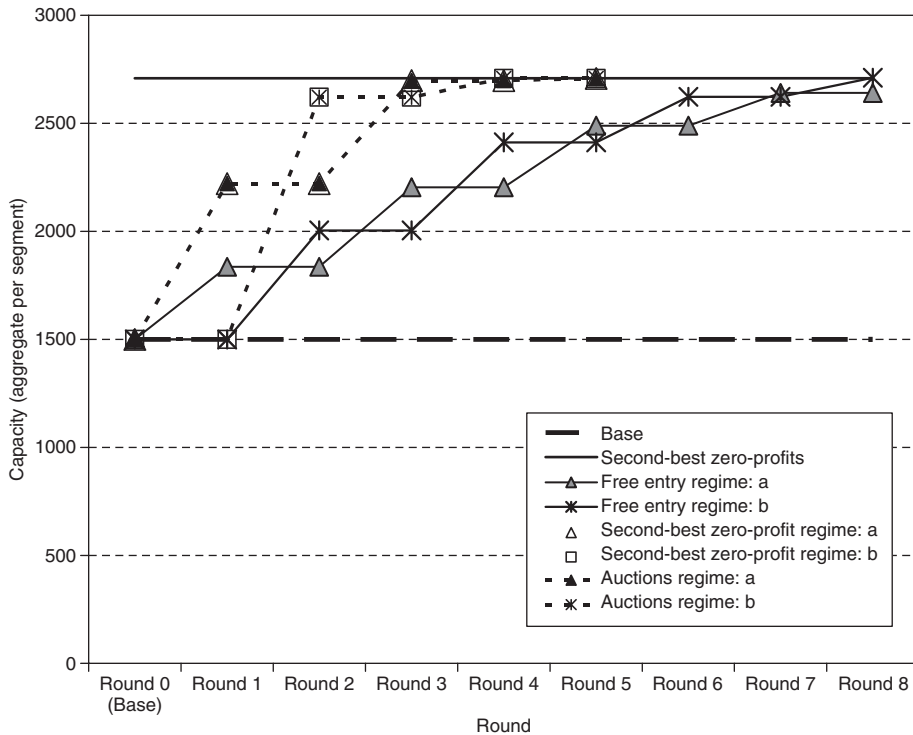
Finally, the same pattern of entries and network formation will arise under the 'second-best zero-profit entries' regime. The intuition is now even simpler. After having optimised the capacity and toll for a new addition on segment a , it is by definition not possible to have a further improvement in social surpluses by revising the capacity on segment a once more. If there is scope for improvement, it must be on segment b , and exactly because the capacities and tolls for both segments are not optimised simultaneously, a sequence will be produced in which there is scope for improvement on the one segment after an increase in capacity on the other segment has induced an increase in demand over the full corridor.

4.2 Development of capacities and tolls

Although the pattern of network development is identical for the three regimes considered, the capacities and tolls involved may, of course, be different. This is illustrated in Figures 2 and 3, which show, for the various regimes and for the two segments a and b separately, the development over rounds of aggregate capacity (summed over a segment) and average toll (that is, averaged for tolled users only, so ignoring users on the untolled initial capacity). The diagrams use two benchmarks: the base equilibrium and the second-best zero-profit equilibrium, both described in Table 1 — where the latter is, of course, to be distinguished clearly from the sequence shown as the second-best zero-profit regime.

Figures 2 and 3 reveal a number of interesting properties of the different regimes. First, it can be noted that the auctions regime and the second-best zero-profits regime produce identical results. This is not too surprising once

Figure 2
The Development of Aggregate Capacity per Segment over Time under Various Regimes

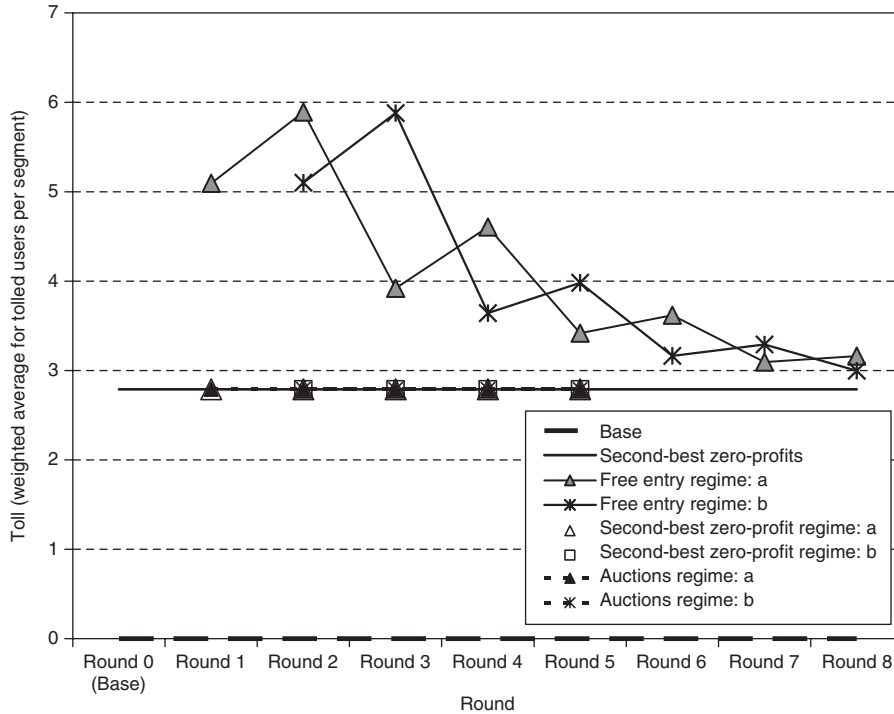


Note: Data points for the second-best zero-profits regime and the auctions regime overlap perfectly.

it is recognised that both induce entries with tolls set according to the long-run cost function, and the capacity maximised under a zero-profit constraint. We can thus confirm that Verhoef's (2007) finding that the patronage-maximising auction produces the second-best zero-profit road for a first auctioned road, on an exogenously determined location in an exogenously specified network, remains valid for a sequence of auctions in a network, which develops in an endogenous manner through these auctions. (However, it should be recalled that this would change, albeit in a numerically modest way, if we would allow firms to make toll bids that imply structural losses later on.)

Second, both the auctions regime and the free-entry regime appear to approach the second-best zero-profits equilibrium more closely as the network develops further. For the auctions regime, this is less surprising as it induces a sequence of additions that are all according to the long-run cost function — which the second-best zero-profits equilibrium does in one shot for both segments jointly. For free entry, the beneficial impacts of increased

Figure 3
The Development of the Weighted Average Toll (for Tolled Users) per Segment over Time under Various Regimes



Note: Data points for the second-best zero-profits regime and the auctions regime overlap perfectly.

parallel competition apparently outweigh the potential caveat that serial competition remains in existence. The reason is that new firms will remain entering a segment as long as the generalised price is above the long-run optimal level. A combination of capacity and toll that brings the flow capacity ratio on the new link equal to the optimal level, and the generalised price to the level provided by the parallel competitors, would then produce a profit — and would therefore induce entry — until the long-run optimal generalised price level is achieved.

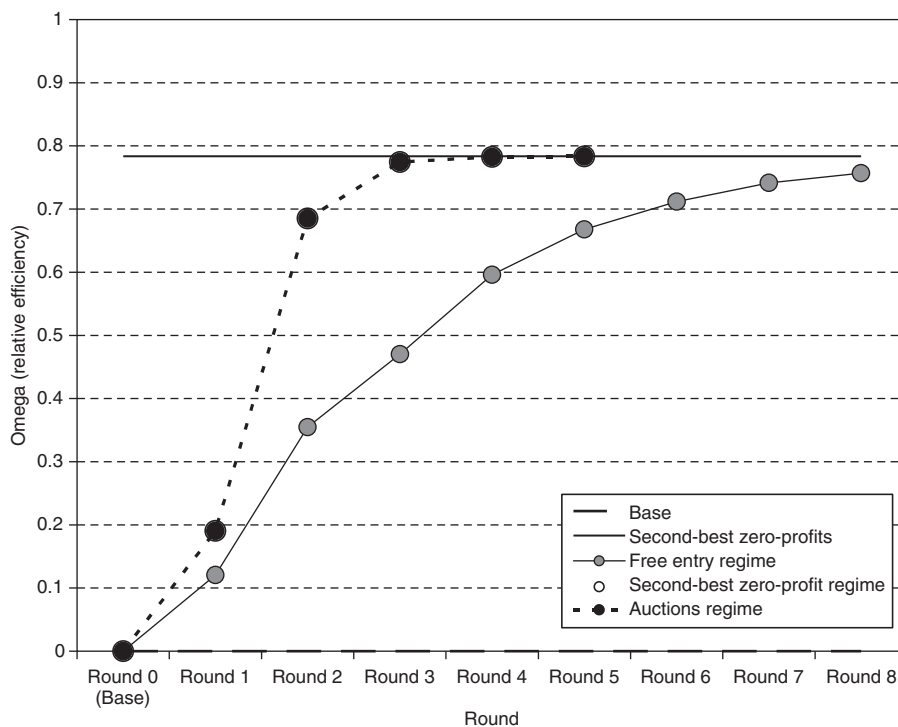
Third, Figure 2 shows that although both sequences have investments that for both segments become smaller over successive rounds, the initial steps are bigger for the auctions regime. That regime also approaches the second-best zero-profit toll level much more rapidly; that is, already from the first capacity addition onwards. For the free-entry regime, the average toll approaches that level only gradually. Note that this decline in tolls is not monotonous: the average toll drops on a segment in a round in which a new firm enters that segment, due to increased competition and

to increased aggregate capacity, but it rises when an investment is made on the other segment, due to the induced increase in demand.

4.3 Development of social surplus

Given that tolls take on the second-best zero-profit levels immediately, and capacities approach those levels more rapidly, it should be no surprise that social surplus rises more rapidly under the auctions regime than with free-entry. Figure 4 shows this by comparing the relative efficiency for both regimes. One possible conclusion from the diagram is that there is actually no need to interfere in entries into the market through the auctioning of concessions, because a process of free entry leads to the same final end-state, namely the second-best zero-profit state. Quite a different conclusion would be that such auctioning is in fact desirable, since the free-entry regime gets sufficiently close to this second-best zero-profit equilibrium only after a sufficient number of competing firms have entered the market, each with relatively small capacities, which might be unrealistic in reality if discreteness of capacity is an issue. It is illustrative that the

Figure 4
The Development of the Relative Efficiency ω under Various Regimes

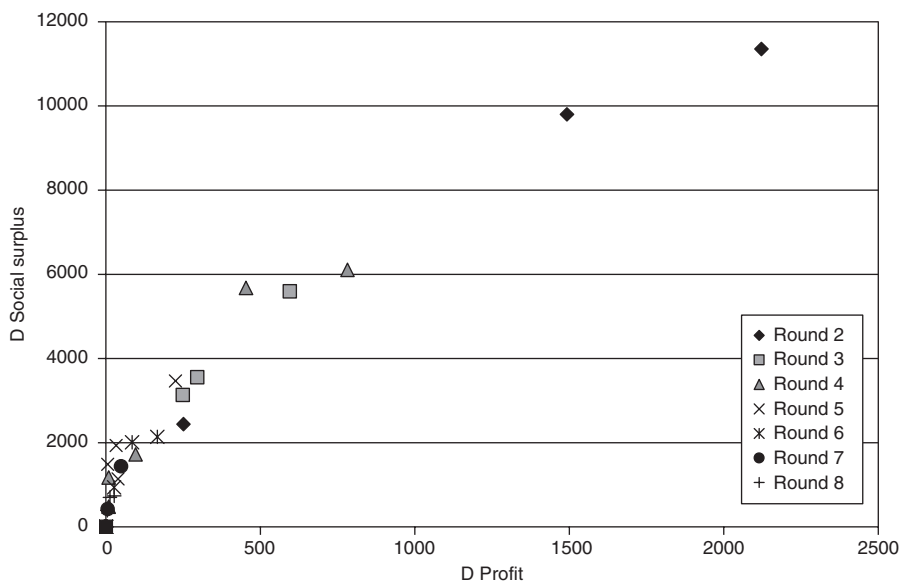


Note: Data points for the second-best zero-profits regime and the auctions regime overlap perfectly.

surplus level achieved already in round 2 with auctions, is under free entry not reached until three firms have entered on both segments in six rounds. Similarly, the welfare gains achieved after two rounds with free entry are only around half of those realised through auctions. These results suggest that the equivalence in the theoretical end-states should not be over-emphasised, and that there may still be a convincing case for preferring auctions over free entry.

One might wonder whether it is possible to speed up the increase in welfare in the free-entry regime by regulating entry, by assigning the right to enter in a round to a specific firm on a specific segment, leaving the capacity and toll at the discretion of the firm. This might offer a second-best instrument that could raise welfare during this regime. It transpires that in the present network, there is no scope for raising welfare this way. In every round studied, it appeared that the specific investment that implies the highest possible profit gain for the investor (that is, the highest gain among those from the $2 \cdot F + 2$ possible investments when F firms are present in the network) is also the investment that leads to the highest gain in social surplus — given that the potential investors themselves choose their tolls and capacities. This is illustrated in Figure 5, which shows for rounds 2 to 8 the $2 \cdot F + 2$ combinations of the change in profit for the potential investor (along the horizontal axis), and the implied change in

Figure 5
Profit Gain by Social Surplus Gain in Rounds 2–8



social surplus. In each round shown, the combination with the highest profit gain also has the highest social surplus gain, and the rank correlation is nearly perfect.⁹

Again, there is a good economic intuition for this result. Investments are especially profitable on a segment that has a relatively limited number of competitors, and therefore relatively high tolls, and a relatively low aggregate capacity, and therefore a relatively high potential for further growth in demand. Both factors would cause another investment to be desirable also from the social perspective, because increased competition will drive down tolls towards socially more desirable levels, and because the investment implies extra capacity, which is socially desirable as long as profits are positive. Furthermore, each firm's attempt to reduce serial competition by having capacity on both segments also contributes to social welfare in the sense that it avoids 'pure' serial competition as considered in the model of Small and Verhoef (2007), where each segment is controlled by a different single operator.

4.4 Development of profitability

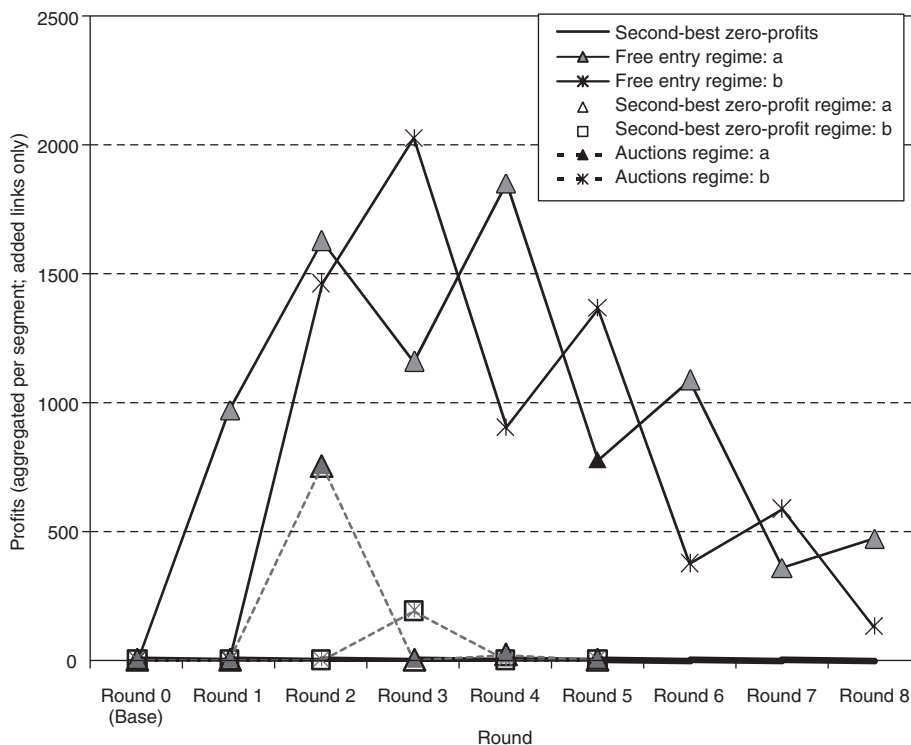
Finally, we discuss how profitability of road operations evolves over time. It is instructive to distinguish between profitability at the level of segments, and at the level of firms.

Figure 6 shows the development of aggregate profitability for both segments under the various regimes (on added, tolled capacity only, so ignoring the cost of initial capacity). For the free-entry regime, the patterns show how, after the first firm *I* has invested, every following entry reduces aggregate profits for the segment on which the investment is made due to increased competition, while it raises profits on the other segment due to increased equilibrium demand. The trend towards zero profits on tolled capacity is clearly visible, although, in accordance with the toll levels depicted in Figure 3 above, profits are still positive with four firms present.

Perhaps surprisingly, also under the auctions regime — which, of course, again coincides with the second-best zero-profit regime in Figure 6 — profits are temporarily made, although each investment yields zero profits in the round it is implemented. These profits result from the interactions between the two segments: a new investment in round *i* that produces zero profits upon completion will become profitable after the capacity on

⁹It is striking how, in Figure 5, the relation between the gain in profits for the investor and the gain in social surplus appears to be rather stable across rounds. It is not hard to imagine the course of a well-fitted square root type of function for the pooled observations; it is less easy to come up with an explanation for this closeness.

Figure 6
The Development of Aggregate Profits per Segment under Various Regimes

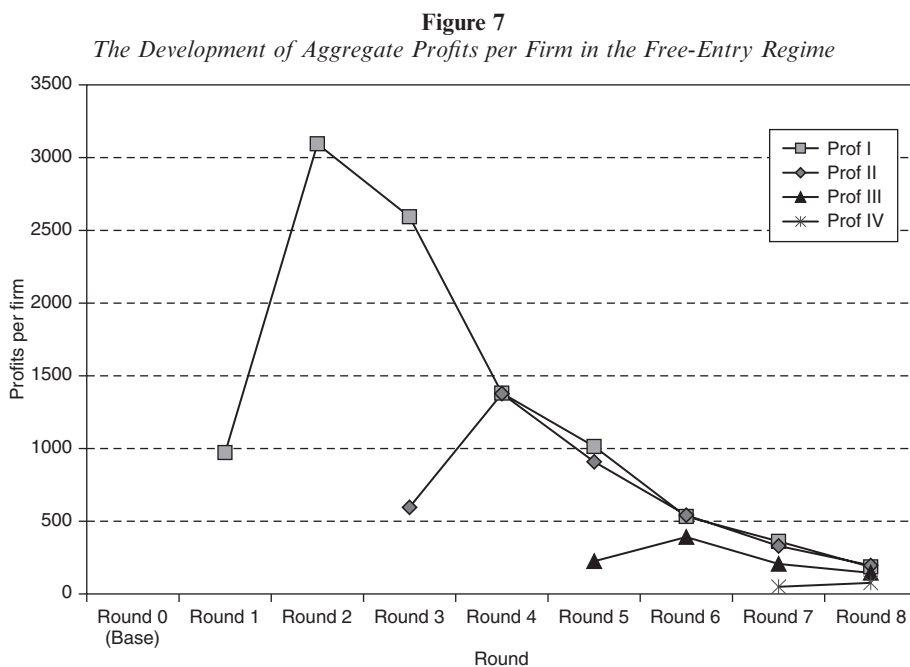


Note: Data points for the second-best zero-profits regime and the auctions regime overlap perfectly.

the other segment, and hence overall equilibrium demand, is increased in round $i + 1$.¹⁰

Figure 7 shows profits by firm in the free-entry regime. For each firm, aggregate profits increase with an own — voluntary — investment, and decrease as other firms add capacity. Not surprisingly, the earlier the firm enters, the higher will its temporary profits be. As the network expands, profits evaporate. The negative impact of later entries upon a firm’s profits creates another problem for more traditional bid-based auctions, where the concession is given to the firm that makes the highest bid. A myopic firm would bid the net present value of profits ignoring later entries, and would therefore suffer losses as soon as further additions to the network

¹⁰It is exactly this mechanism that might induce a firm with capacity on the one segment to bid a toll below the long-run first-best level for capacity on the other segment: this may maximise the firm’s temporary joint profits over both segments together, at the expense of future losses. We discussed this possibility in Section 2.3, and explained that it is numerically insignificant in the present network.



are made. However, even if the firm is more forward-looking, the bid it can make will depend crucially on the assumptions it makes on the time lags between future auctions, something that may be hard to predict also for the regulator. This adds to the more fundamental problem with bid-based auctions already identified in Verhoef (2007), namely that it urges a firm to choose the profit-maximising combination of capacity and toll, rather than the welfare-maximising levels.

5.0 Conclusions

This paper studied the efficiency impacts of private toll roads in initially untolled networks. The analysis allowed for capacity and toll choice by private operators, and endogenised entry and therewith the degree of competition, distinguishing and allowing for both parallel and serial competition. Two institutional arrangements were considered, namely one in which entry is free and one in which it is allowed only after winning an auction. Investments were assumed to be made sequentially. With free entry, the firm expecting the highest profits enters, and with auctions a

concession is granted to the firm that promises to carry the highest traffic flow. The following results stand out.

First, the existence of serial competition does not alter the conclusion obtained by DeVany and Saving (1980) and Engel *et al.* (2004) in the context of parallel competition, namely that entry of more firms drives tolls closer towards socially optimal levels. This is true despite the potentially negative effects that increased serial competition might have on the efficiency of pricing (Small and Verhoef, 2007). The reason is that, with endogenous entries, firms will be ordered over the network such that they occupy capacity on different (serial) segments, and this minimises the excessive-pricing problem that may otherwise characterise network markets with serial competition, identified by Economides and Salop (1992). During the process of entries, neither the size of investments nor the tolls are chosen optimally from the social perspective, but capacity ‘deficits’ will be filled up when later investments are made by other firms, and tolls will be driven down under increased competition.

Second, both sequences considered — the free-entry regime and the auctions regime — have the second-best zero-profit equilibrium as the end-state of the equilibrium sequence of investments. However, the auctions regime approaches this end-state more rapidly: tolls are set equal to their second-best zero-profit levels immediately, and capacity additions in the earlier rounds are bigger. When discreteness of capacity is relevant and limits the number of investments that can practically be accommodated, the auctions regime may therefore result in a more efficient end-state, with a higher social surplus, although the theoretical end-state is the same as under free entry. Consistent with findings for the single patronage-maximising auction in Verhoef (2007), in each round of the auctions sequence, firms are pushed towards bids that imply investment and tolling according to the long-run cost function.

Obviously, the model is still rather abstract, and various important extensions can be envisaged. We name a few, and will also hypothesise whether relaxation of the associated assumption is likely to change this paper’s main conclusions.

First, we considered a rather simple network structure, and also did not allow firms to make fundamental changes to this structure, for example, by adding a direct link between the origin and destination. It would be interesting to consider more general networks in future work. Will this make any fundamental change to the results? The most fundamental change considered in this paper compared to earlier studies with free entry, namely the inclusion of serial competition besides parallel competition, did not undermine the efficiency of free entry — contrary to what one might have expected. Also, the efficiency of auctions was not affected by having

parallel and serial links auctioned. Of course, as also demonstrated in Verhoef (2007), when a bigger network allows us to develop a Braess paradox, things may change drastically. Besides such cases, however, it is not clear why a bigger network would affect the conclusions fundamentally — although this question, of course, needs to be investigated formally. The main arguments underlying the efficiency of free entry and of auctions, as discussed above, seem to remain relevant also in bigger networks.

Second, there is the issue of the timing of investments. Especially in the free-entry regime, this may lead to a complicated dynamic game, where firms not only decide on where to invest and by how much, but also on when to do so. Developing the analytical framework to describe this properly seems a big challenge, even when demand functions are assumed to be stable over time; but will it change the main conclusions? Perhaps not. The prospect of potential profits for new firms, that exists as long as equilibrium tolls exceed the second-best zero-profit level, will induce entry whether or not that entry is optimally timed. In other words, although the outcomes along the free-entry sequence are likely to be different with endogenous timing of investments, the endpoint is likely to be the same. As much can be expected for the auctions regime.

Third, nearly-myopic behaviour when deciding on the size of investments could be replaced by a less naive formation of expectations. Again it is likely that the outcomes along the free-entry sequence will change, but questionable whether the endpoint will be different. A configuration with positive profits is unlikely to be a stable endpoint, so also, then, can one expect entry to make the system converge to the second-best zero-profit equilibrium.

Fourth, we ignored that in reality, the government may be under pressure to add capacity itself at some point in a sequence of additions, particularly if either tolls seem excessive or congestion is severe. The paper did make clear, however, that a sufficiently patient government has a good reason not to do so, since these are also the conditions under which private investments are more likely.

Fifth, the links in our model were identical in terms of length. For an asymmetric network, patronage may have to be weighted with link length in the auction. For free-entry, a relatively short link with a relatively large flow may be particularly attractive when capacity expansion costs are relatively low. Asymmetries therefore pose interesting questions for further development of the ideas presented above.

Of course, it is not difficult to mention some further extensions that may affect the main conclusions. These include the consideration of heterogeneous users, making product differentiation between parallel operators a likely outcome; the consideration of demand uncertainty; the existence

of strategic interactions and market power during auctions; and the replacement of Nash behaviour by Stackelberg leadership in the toll and/or capacity stages of the free-entry game. Future research should inform us of how strongly such changes would affect the model's main conclusions.

References

- De Borger, B., S. Proost, and K. van Dender (2005): 'Congestion and Tax Competition in a Parallel Network', *European Economic Review*, 49, 2013–40.
- De Palma, A. and R. Lindsey (2000): 'Private Roads: Competition under Various Ownership Regimes', *Annals of Regional Science*, 34, 13–35.
- DeVany, A. S. and T. R. Saving (1980): 'Competition and Highway Pricing for Stochastic Traffic', *Journal of Business*, 53, 45–60.
- Economides, N. and S. C. Salop (1992): 'Competition and Integration among Complements, and Network Market Structure', *The Journal of Industrial Economics*, 40, 105–23.
- Edelson, N. E. (1971): 'Congestion Tolls under Monopoly', *American Economic Review*, 61(5), 872–82.
- Engel, E., R. Fisher, and A. Galetovic (1996): 'Highway Franchising: Pitfalls and Opportunities', *American Economic Review, Papers and Proceedings*, 87(2), 68–72.
- Engel, E., R. Fisher, and A. Galetovic (2004): 'Toll Competition among Congested Roads', *Topics in Economic Analysis & Policy*, 4, Article 4.
- Levinson, D. and B. Yerra (2006): 'Self Organization of Surface Transportation Networks', *Transportation Science*, 40(2), 179–88.
- Lévy-Lambert, H. (1968): 'Tarification des Services à Qualité Variable: Application aux Péages de Circulation', *Econometrica*, 36(3–4), 564–74.
- Litman, T. (2006): 'Smart Transportation Investments: Reevaluating the Role of Highway Expansion for Improving Urban Transportation', Unpublished paper, VTPI, Victoria.
- 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.
- Small, K. A. and E. T. Verhoef (2007): *The Economics of Urban Transportation*, Routledge, London.
- Ubbels, B. and E. T. Verhoef (2008): 'Governmental Competition in Road Charging and Capacity Choice', *Regional Science and Urban Economics*, 38(2), 174–90.
- Verhoef, E. T. (2007): 'Second-best Road Pricing through Highway Franchising', *Journal of Urban Economics*, 62(2), 337–61.
- 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), 127–56.

- Wardrop, J. G. (1952): 'Some Theoretical Aspects of Road Traffic Research', *Proceedings of the Institute of Civil Engineers*, 1(2), 325–78.
- Zhang, L. and D. Levinson (2006): 'Economics of Road Network Ownership', Paper presented at the First International Conference on Transport Infrastructure Funding, August 2006, Banff, Canada.