TI 2008-056/3
Tinbergen Institute Discussion Paper

Pricing, Capacity and Long-run Cost Functions for First-best and Second-best Network Problems

Erik T. Verhoef[1]
Andrew Koh[2]
Simon Shepherd[2]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam;*
[2] *Institute for Transport Studies, University of Leeds.*

PRICING, CAPACITY AND LONG-RUN COST FUNCTIONS FOR FIRST-BEST AND SECOND-BEST NETWORK PROBLEMS

Erik T Verhoef[*]

Department of Spatial Economics

VU University Amsterdam

De Boelelaan 1105

1081 HV Amsterdam

+31-20-5986090

everhoef@feweb.vu.nl

Andrew Koh[**] and Simon Shepherd[**]

Institute for Transport Studies

University of Leeds

Leeds - LS2 9JT

+44-113-3435325

A.Koh@its.leeds.ac.uk

S.P.Shepherd@its.leeds.ac.uk

Abstract

This paper considers the use of 'long-run cost functions' for congested networks in solving second-best network problems, in which capacity and tolls are instruments. We derive analytical results both for general cost and demand functions and for specific functional forms, namely Bureau of Public Roads cost functions and constant-elasticity demand functions. The latter are also used in a numerical simulation model. We consider second-best cases where only a subset of links in a network is subject to tolling and/or capacity choice, and cases with and without a self-financing constraint imposed. We will demonstrate that, under certain assumptions, second-best long-run cost (or actually: generalized price) functions can be derived for most of the cases of interest, which can be used in an applied network model as a substitute for the conventional short-run user cost functions. Doing so reduces the dimensionality of the problem and should therefore be helpful in speeding up procedures for finding second-best optima.

1. Introduction

Road pricing is gaining increasing momentum as a possible instrument in dealing with traffic congestion. The concept is firmly based in micro economic theory: Pigou (1920) was the first to recognize that traffic congestion entails an external cost, and that Pareto efficiency requires a toll equal to the marginal external congestion costs. This result remains the first-best pricing rule when moving from single roads to full networks, as demonstrated by Beckmann, McGuire and Winsten (1956), Dafermos (1973) and Yang and Huang (1998); and also applies in the long-run when link capacities are optimized (Mohring and Harwitz, 1962). For practical implementations of road pricing, however, the assumptions that underlie first-best analysis are often unacceptably strong. These assumptions include in the first place that underpricing of congested traffic is, as it were, the only market failure in the entire economy: on all markets that are directly or indirectly related to the traffic considered, prices should be equal to marginal social cost. Therefore, if congestion occurs for commuting, this assumption means that spatial labour markets function perfectly and that no pre-existing labour taxes exist. Secondly, the assumptions entail that the regulator has perfect (pricing) instruments, meaning that all users on all links can be tolled on all times, with perfect toll differentiation possible. In reality, this is of course usually not the case. An important example, also addressed in this paper, concerns the case where some but not all links in the network are subject to tolling, as is the case for schemes that employ toll cordons, toll areas, specific toll roads, or so-called 'pay-lanes', for which unpriced parallel capacity remains offered. Violations of both types of assumptions brings the analysis in the realm of second-best pricing, which means that tolls are set to maximize a social objective (usually social surplus) under one or more constraints as described above. It is then for example not possible to optimize prices outside the transport network of interest, or it is not possible to charge all network users with the individually optimal toll.

With the growing number of practical applications of road pricing, also the study of second-best has gained increasing interest in the road pricing literature. A recent overview is provided by Small and Verhoef (2007). Specific to the theme of second-best network problems are early studies by Marchand (1968) and Lévy-Lambert (1968), who studied what has now become known as the 'classic two-route problem', where unpriced parallel capacity (a road, or a lane) is offered in addition to priced capacity. Various aspects of this problem have more recently been studied. For example, Verhoef, Nijkamp and Rietveld (1996) considered various ownership regimes, including private monopoly. Braid (1996) considered the problem in the context of Vickrey's (1969) bottleneck model. Liu and McDonald (1998) considered an empirical application, namely the Californian SR-91. De Palma and Lindsey (2000) focused on competition between operators of the two links. Verhoef and Small (2004) looked at heterogeneous drivers, with different values of time. Rouwendal and Verhoef (2004) paid explicit attention to capacity choice in addition to toll setting. De Borger, Proost and Van Dender (2005) also considered capacity choice and tolling, but in the context of a

private duopoly. Some main conclusions from these studies are that the second-best toll is usually below the first-best level, and achieves only a certain part of the possible welfare gains from pricing – the size of which depends on elasticity of demand and heterogeneity of travelers, among other determinants. Not surprisingly, private ownership usually raises profits and lowers welfare, compared to public operation.

The two-route problem has also been extended to larger networks. Verhoef (2002ab) considered the more general second-best network pricing problem, where a certain sub-set of links in a network of undetermined size and shape can be tolled. As may be expected, the resulting toll expressions are tedious, and depend on demand and cost elasticities in the network, besides the marginal external congestion costs on the tolled and untolled links. This reflects that tolls are set to optimize network-wide social surplus and therefore should take into account congestion on untolled links, as well as the extent to which pricing on the tolled links aggravates it. Verhoef (2002b) also studies the related question of which links to toll, if only a sub-set can be tolled. This entails questions like "where should a toll cordon be positioned?" and "should we have a cordon at all?". Sumalee, May and Shepherd (2005) show that a formal approach to these questions may lead to a rather different pattern of toll points over a network, and higher welfare gains, than does an approach based on expert opinions.

While the more recent papers on the classic two-route problem often consider capacity choice in addition to the toll instrument, this issue seems to have received considerably less attention in the context of second-best policies for larger networks. This is an important omission from the perspective of actual policy making which, of course, usually involves tolling on bigger networks. Undoubtedly, the analytical complexity of combined second-best toll and capacity setting explains at least partly this lack of attention. And for numerical network simulation modeling, one may expect that the combination of two instruments on a link may greatly complicate procedures of finding second-best optimality, due to what is sometimes referred to as the 'curse of dimensionality' (Bellman, 1961).

In this paper we aim to show how the use of 'long-run cost functions' for congested networks can help solve second-best network problems in which both capacities and tolls are instruments. We derive analytical results both for general cost functions and for cost functions with a specific functional form, namely the well-known Bureau of Public Roads (BPR) formula. We consider second-best cases where only a subset of links in a network is subject to tolling and a subset is subject to capacity optimization, where these subsets may or may not overlap. We also consider the case where for some – possibly all – links that are subject to both tolling and capacity optimization, a zero-profit constraint is imposed, so that toll revenues collected on the link are constrained to be equal to the link's capacity cost. Our paper has a simple structure: Section 2 will present our analytical results, and Section 3 provides a numerical application. Section 4 will conclude.

We will demonstrate that, under the assumptions to be spelled out below, second-best long-run cost (or actually: generalized price) functions can be derived for most of the cases of interest, which can be used in an applied network model as a substitute for the conventional

short-run user cost functions. Doing so reduces the dimensionality of the problem and should therefore be helpful in speeding up procedures used for finding second-best optima. Of course, two special cases of our second-best problem deserve special attention, because they produce strong benchmark insights. These are the cases where the network collapses to a single link, and the first-best case where on a larger network, all capacities and tolls can be set. We will start the next analytical section with these two special cases.

2.   Pricing, capacity and long-run cost functions: analytical results

This section presents our analytical results on the relation between pricing, capacity and long-run cost functions for congested traffic. We will start simple, by reviewing the basic results for single roads, and next move to a number of second-best network problems.

   A number of assumptions underlie all our derivations. First, we consider stationary-state (or static) traffic conditions. Road users are homogeneous: they have the same value of time and the same (marginal) impact on link travel times. The travel time on a link therefore depends only on its aggregate use. Demand is not perfectly inelastic, and the inverse demand for an OD-pair m (for "market") is denoted $D^m(N^m)$. The user cost $c^l(\cdot)$ on a link (l) depends on the link flow N^l and on the link's capacity K^l. The link capacity cost $C^{c,l}$ depends on K^l only. Our (Marshallian) measure for benefits is the conventional area below the inverse demand, while total costs are the combination of user cost and capacity costs. We will consider public operators only, who seek to maximize social surplus, which is defined as total benefits minus total (user and capacity) costs. A toll τ^l may be levied; its receipts constitute a transfer that in itself represents no social welfare gain (such gains may, of course, arise from toll-induced changes in demand).

   Besides general formulations, in which functional forms are not specified, we will use more specific formulations when these yield additional analytical insights and, of course, in the numerical simulation model. For the inverse demand function, this specific form assumes that it takes on a constant-elasticity form:

$$D^m(N^m) = \delta^m \cdot \left(N^m\right)^{1/\eta^m},$$
(1)

where δ^m is a scale parameter and η^m is the elasticity of demand with respect to the generalized price.

   The specific user cost function employs the Bureau of Public Roads (BPR) travel time function:

$$p^l = c^l + \tau^l = \alpha \cdot t_f^l \cdot \left(1 + \beta \cdot \left(\frac{N^l}{K^l}\right)^\chi\right) + \tau^l,$$
(2)

where p^l is generalized price, α is the value of time, t_f^l the free-flow travel time, and β and χ are parameters. Note that c^l only contains time costs. This is one of many user cost functions exhibiting constant returns to scale in congestion technology (Small and Verhoef, 2007), meaning that it is homogeneous of degree zero in the flow-capacity ratio N^l/K^l.

The specific capacity cost function assumes neutral scale economies in road construction:

$$C^{c,l}(K^l) = \rho \cdot t_f^l \cdot K^l,$$ (3)

where ρ is a parameter that reflects the unit price of capacity, expressed per unit of time (*i.e.*, it represents costs of interest and depreciation) and per unit of distance (represented by the free-flow travel time, which is – for a given free-flow speed – a perfect measure of distance).

2.1.   Single link results

It is instructive to start our exposition with some results for the single-link case; *i.e.*, ignoring network complications. In this case, there is a one-to-one correspondence between "the" link and "the" OD-pair, so that we can suppress indices *m* and *l*.

The first-best benchmark is the case where the regulator can set both the toll and the capacity of the single road. A constraint is that road users will enter the road up to the point where marginal benefits $D(N)$ are equal to the generalized price $p=c(N,K)+\tau$. The associated Lagrangian looks as follows:

$$\Lambda = \int_0^N D(n)\mathrm{d}n - N \cdot c(N,K) - C^c(K) + \lambda \cdot \big(c(N,K)+\tau - D(N)\big).$$ (4)

The first three main terms define the objective of social surplus and the final term gives the constraint just mentioned; λ is the Lagrangian multiplier. The set of equations defined by the first-order conditions with respect to *N*, *K*, τ and λ can be solved to yield the following two well-known policy rules (*e.g.* Small and Verhoef, 2007, p. 164):

$$\tau = N \cdot \frac{\partial c}{\partial N},$$ (5a)

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

while for the Lagrangian multiplier we find:

$$\lambda = 0.$$ (5c)

Equation (5a) stipulates that the optimal toll should be set equal to the marginal external congestion costs, while equation (5b) shows that the marginal benefits from capacity expansion (the left-hand side) should be equal to the marginal cost (the right-hand side). In addition, equation (5c) reflects that the constraint has no impact on the level of welfare that can be achieved (recall that the equilibrium value of a Lagrangian multiplier gives the marginal impact on the optimized objective from a marginal relaxation of the associated constraint).

Mohring and Harwitz (1962) have demonstrated how the joint application of the tax rule (5a) and the investment rule (5b) result in a zero financial surplus for the road operator, provided three technical conditions are fulfilled: (i) capacity is a continuous variable; (ii) there are constant returns to scale in congestion technology, so that $c(N,K)$ can be written in the

form $c(N/K)$; and (iii) there are neutral scale economies in road construction, so that $C^c(K)$ can be written as $\rho \cdot K$. As a matter of fact, units of capacity can always be chosen such that latter condition is satisfied, namely by defining a measure of capacity that is proportional to (minimized) capacity cost; what matters is whether the combined effect of conditions (ii) and (iii) is achieved. The quickest way to prove the Mohring-Harwitz result in the current setting is to multiply both sides of (5b) by K, and to replace the resulting term $-K \cdot \partial c/\partial K$ on the left-hand side by $N \cdot \partial c/\partial N$ (the equality follows from Euler's theorem, as well as from a combined application of the chain and quotient rules of differentiation). The left-hand side of (5b) then gives total toll revenues ($N \cdot N \cdot \partial c/\partial N$), while the right-hand side, $K \cdot \partial C^c/\partial K$, gives total capacity cost if the marginal cost of capacity $\partial C^c/\partial K$ is constant (and therefore equal to a constant unit price ρ).

For our purposes, it is interesting to see whether we can derive a long-run cost function from the solutions (5). In closed form, this is possible only when the functional forms of the cost functions c and C^c are specified. Consider the specific functions of (2) and (3). Equation (5b) then becomes:

$$-N \cdot \alpha \cdot t_f \cdot \beta \cdot \chi \cdot \left(\frac{N}{K} \right)^{\chi-1} \cdot \frac{N}{-K^2} = \rho \cdot t_f \quad \Rightarrow \quad \tilde{K} = \tilde{N} \cdot \left(\frac{\alpha \cdot \beta \cdot \chi}{\rho} \right)^{\frac{1}{\chi+1}}, \tag{6a}$$

where a tilde ($\sim$) denotes the long-run optimal value.[1] Note that long-run optimal capacity $\tilde{K}$ increases linearly with the level of use, N, as might be expected with neutral scale economies. Inserting $\tilde{K}$ from (6a) into (2) and (3) gives long-run average user costs $\tilde{c}$ and long-run capacity cost $\tilde{C}^c$. The latter can be divided by $\tilde{N}$ to obtain long-run average capacity cost $\tilde{c}^c$, while the sum of $\tilde{c}$ and $\tilde{c}^c$ gives the long-run average 'total' cost, $\tilde{c}^t$. And finally, we can determine the long-run optimal toll level $\tilde{\tau}$ by evaluating (5a) for a ratio N/K implied by (6a). All this results in the following expressions:

$$\tilde{c} = t_f \cdot \left(\alpha + (\alpha \cdot \beta)^{\frac{1}{\chi+1}} \cdot \left(\frac{\rho}{\chi} \right)^{\frac{\chi}{\chi+1}} \right), \tag{6b}$$

$$\tilde{c}^c = t_f \cdot (\alpha \cdot \beta \cdot \chi)^{\frac{1}{\chi+1}} \cdot \rho^{\frac{\chi}{\chi+1}}, \tag{6c}$$

$$\tilde{c}^t = \tilde{c} + \tilde{c}^c, \tag{6d}$$

$$\tilde{\tau} = t_f \cdot (\alpha \cdot \beta \cdot \chi)^{\frac{1}{\chi+1}} \cdot \rho^{\frac{\chi}{\chi+1}}, \tag{6e}$$

$$\tilde{p} = \tilde{c} + \tilde{\tau}. \tag{6f}$$

---

[1] Equation (6a) differs from the corresponding equation in Small and Verhoef (2007, p. 108) because in the present case, (1) no "duration" of the peak period is specified (their q is normalized to 1 here); (2) the cost of capacity is defined per unit of distance (so that T_f drops out here); and (3) we define directly ρ as the per-unit-of-time cost of a unit of capacity (they multiply a cost recovery factor (which they denote ρ) by investment costs).

The equality of $\tilde{c}^c$ and $\tilde{\tau}$ shows that the road is exactly self-financing in the long run.

A relevant second-best case for the single-link model concerns the situation where optimal pricing is not in place. We consider the case where there is no toll at all, so τ=0. The Lagrangian of (4) then becomes:

$$\Lambda = \int_0^N D(n)\mathrm{d}n - N \cdot c(N,K) - C^c(K) + \lambda \cdot \big(c(N,K) - D(N)\big). \tag{7}$$

The set of equations defined by the first-order conditions with respect to N, K, and λ can be solved to yield the following investment rule:

$$-(N-\lambda) \cdot \frac{\partial c}{\partial K} = \frac{\partial C^c}{\partial K}, \tag{8a}$$

with:

$$\lambda = \frac{N \cdot \dfrac{\partial c}{\partial N}}{\dfrac{\partial c}{\partial N} - \dfrac{\partial D}{\partial N}}. \tag{8b}$$

These form a special case of the investment rule presented by Small and Verhoef (2007, p. 172), who consider a multi-period model and allow for any arbitrary toll level (not just 0). The interpretation of the investment rule is intuitive: the marginal benefits from capacity expansion, on the left-hand side of (8a), are calculated as if fewer travellers than N are present on the road (note that λ in (8b) is positive). This reflects that additional capacity will induce additional demand, which in itself is however socially undesirable (at the margin) because congestion is underpriced. The correction factor or shadow price λ in (8b) consequently 'deflates' the direct benefits of capacity expansion to account for undesired induced demand, and quite naturally λ increases when congestion is more severe (in the numerator) or when equilibrium demand is more responsive to average cost changes (in the denominator). Both would provide reason for weighting induced demand effects more heavily.

Does the investment rule from (8ab) imply a "second-best long-run cost function"? In principle it does, but we were unable to derive a closed-form expression for it, even for the case with explicit functions. The second-best long-run cost function should follow from the solution of the following second-best variant of (6a):

$$\left(\frac{-\dfrac{\partial D}{\partial N}}{\dfrac{\partial c}{\partial N} - \dfrac{\partial D}{\partial N}} \right) \cdot \alpha \cdot t_f \cdot \beta \cdot \chi \cdot \left(\frac{N}{K} \right)^{\chi+1} = \rho \cdot t_f. \tag{9a}$$

The first term, in large brackets, can be rewritten as:

$$-\frac{\dfrac{\partial D}{\partial N}}{\dfrac{\partial c}{\partial N}-\dfrac{\partial D}{\partial N}}=\frac{-\eta}{\dfrac{\partial c/\partial N}{c/N}-\eta}\quad\text{with}\quad\frac{\partial c/\partial N}{c/N}=\frac{\chi}{\dfrac{1}{\beta}\cdot\left(\dfrac{K}{N}\right)^{\chi}+1},\tag{9b}$$

where we have used that, in equilibrium, $c=D$. The first term in the denominator is not constant for the BPR function, which prevents a closed-form solution from being available. However, (9a) and (9b) together imply that the solution of $\tilde{K}$ as a function of N depends on the ratio K/N only, which suggests that for a model in which (1)-(3) hold, $\tilde{K}$ is proportional with $\tilde{N}$; that is, if for instance the scale factor δ in (1) grows over time and capacity is adjusted in a second-best optimal fashion, we expect the second-best optimal levels of average user cost and average capacity cost to remain constant.

2.2. First-best networks

The single-link results help interpreting the results we obtain for full networks. We will consider a number of regimes for full networks, which differ in terms of the constraints that apply for the network regulator when setting tolls and capacities. The first-best case, where no such constraints apply at all, is the least realistic regime, but also the analytically most transparent one. The first-best tolls and capacities are most easily determined by imagining that the regulator can set route-specific tolls, τ^r, besides the link-specific capacities, K^l. From the resulting route-toll pattern, we can next determine which link-specific tolls τ^l are consistent with the first-best tolls τ^r. We find the optimal tolls and capacities by solving the following Lagrangian:

$$\begin{aligned}\Lambda=\sum_{m=1}^{M}\int_{0}^{\sum_{r=1}^{R}\delta_{rm}\cdot N^r}D^m(n)\mathrm{d}n-\sum_{l=1}^{L}\sum_{r=1}^{R}\delta_{lr}\cdot N^r\cdot c^l\left(\sum_{\rho=1}^{R}\delta_{l\rho}\cdot N^\rho,K^l\right)-\sum_{l=1}^{L}C^{c,l}(K^l)\\
+\sum_{r=1}^{R}\delta_r^A\cdot\lambda^r\cdot\left(\sum_{l=1}^{L}\delta_{lr}\cdot c^l\left(\sum_{\rho=1}^{R}\delta_{l\rho}\cdot N^\rho,K^l\right)+\tau^r-\sum_{m=1}^{M}\delta_{rm}\cdot D^m\left(\sum_{\rho=1}^{R}\delta_{\rho m}\cdot N^\rho\right)\right)\end{aligned}\tag{10}$$

Note that we write OD-flows N^m as the sum of route flows N^r, where δ_{rm} is a dummy that takes on the value of 1 if route r serves OD-pair m, and 0 otherwise. Likewise, we write link flows N^l as the sum of route flows N^r, where δ_{lr} is a dummy that takes on the value of 1 if link l is part of route r, and 0 otherwise. Note that indices r and ρ are both used to denote routes; the latter for summations over routes when keeping r fixed (the index ρ will be used only in expressions for with general functions, so where ρ does not appear as the unit price of capacity). 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 $\delta_r^A=1$), defined as routes that users may choose in equilibrium because the generalized price equals the minimum for that OD-pair. These are the first-order conditions:

$$\frac{\partial \Lambda}{\partial N^r} = \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) - \sum_{l=1}^{L} \delta_{lr} \cdot c^l(\cdot) - \sum_{l=1}^{L} \sum_{\rho=1}^{R} \delta_{lr} \cdot \delta_{l\rho} \cdot N^{\rho} \cdot \frac{\partial c^l(\cdot)}{\partial N^r}$$

$$+ \sum_{\rho=1}^{R} \delta_{\rho}^{A} \cdot \lambda^{\rho} \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \delta_{l\rho} \cdot \frac{\partial c^l(\cdot)}{\partial N^r} - \delta_{rm} \cdot \delta_{\rho m} \cdot \frac{\partial D^m(\cdot)}{\partial N^r} \right) = 0 \qquad \forall r : \delta_r^A = 1 \tag{11a}$$

$$\frac{\partial \Lambda}{\partial K^l} = -\sum_{r=1}^{R} \delta_{l,r} \cdot N^r \cdot \frac{\partial c^l(\cdot)}{\partial K^l} - \frac{\partial C^{c,l}(\cdot)}{\partial K^l} + \sum_{r}^{R} \delta_r^A \cdot \lambda^r \cdot \delta_{lr} \cdot \frac{\partial c^l(\cdot)}{\partial K^l} = 0 \qquad \forall l, \tag{11b}$$

$$\frac{\partial \Lambda}{\partial \tau^r} = \lambda^r = 0 \qquad \forall r : \delta_r^A = 1, \tag{11c}$$

$$\frac{\partial \Lambda}{\partial \lambda^r} = \sum_{l=1}^{L} \delta_{lr} \cdot c^l(\cdot) + \tau^r - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) = 0 \qquad \forall r : \delta_r^A = 1, \tag{11d}$$

while:

$$\delta_r^A = 0 \quad \text{iff} \quad \sum_{l=1}^{L} \delta_{lr} \cdot c^l(\cdot) + \tau^r - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) > 0. \tag{11e}$$

The result that all Lagrangian multipliers λ^r are zero is typical for first-best problems (*e.g.*, Small and Verhoef, 2007, p. 141). It greatly simplifies the solution of (11). In particular, substitution of (11c) and (11d) in (11a), and (11c) in (11b), gives the following policy rules:

$$\tau^r = \sum_{l} \delta_{lr} \cdot N^l \cdot \frac{\partial c^l}{\partial N^l} \qquad \forall r : \delta_r^A = 1, \tag{12a}$$

$$-N^l \cdot \frac{\partial c^l}{\partial K^l} = \frac{\partial C^{c,l}}{\partial K^l} \qquad \forall l, \tag{12b}$$

where we use N^l to denote link-flows (which in (11) were still written in function of the appropriate route flows). The optimal toll rule of (12a) implies that first-best link tolls are not necessarily unique,[2] but that one possible solution in terms of link tolls is the straightforward generalization of the single-link toll of (5a):

$$\tau^l = N^l \cdot \frac{\partial c^l}{\partial N^l} \qquad \forall l. \tag{12a'}$$

From this point onward, we assume that the toll rule (12a) is operationalized as specified in (12a′).

The first-best policy rules for links full network are therefore a straightforward extension of the single-link results. This means that the long-run optimum can be found using the long-run function already identified for the single-link model. For a setting in which the specific functional forms (1)-(3) apply, this means:

---

[2] An example would be two serial links carrying exactly the same travelers. A constant can then be added to the toll on the one link and subtracted from that on the other, without changing route tolls and therefore behaviour.

$$\tilde{K}^l = \tilde{N}^l \cdot \left(\frac{\alpha \cdot \beta \cdot \chi}{\rho} \right)^{\frac{1}{\chi+1}},$$ (13a)

$$\tilde{c}^l = t_f^l \cdot \left(\alpha + (\alpha \cdot \beta)^{\frac{1}{\chi+1}} \cdot \left(\frac{\rho}{\chi} \right)^{\frac{\chi}{\chi+1}} \right),$$ (13b)

$$\tilde{c}^{c,l} = t_f^l \cdot (\alpha \cdot \beta \cdot \chi)^{\frac{1}{\chi+1}} \cdot \rho^{\frac{\chi}{\chi+1}},$$ (13c)

$$\tilde{c}^{t,l} = \tilde{c}^l + \tilde{c}^{c,l},$$ (13d)

$$\tilde{\tau}^l = t_f^l \cdot (\alpha \cdot \beta \cdot \chi)^{\frac{1}{\chi+1}} \cdot \rho^{\frac{\chi}{\chi+1}},$$ (13e)

$$\tilde{p}^l = \tilde{c}^l + \tilde{\tau}^l.$$ (13f)

Because (13f) contains constant parameters only, the network optimum can be determined very easily, by replacing the conventional short-run generalized price functions by their long-run counterparts of (13f). The resulting levels of $\tilde{N}^l$ imply the long-run capacities $\tilde{K}^l$ via (13a). Applying these long-run capacities in the original short-run cost functions should then result in a network that has toll levels of (13e) also as the short-run optimal tolls; that is, satisfying (12a′), which for the BPR function implies:

$$\tau^l = \alpha \cdot t_f^l \cdot \beta \cdot \chi \cdot \left(\frac{N^l}{K^l} \right)^{\chi},$$ (13g)

It is easily verified that (13g), with the ratio $\tilde{N}^l / \tilde{K}^l$ as implied by (13a) substituted for N^l/K^l, indeed produces a toll level equal to (13e).

For the specific functional forms chosen, the long-run optimum can therefore be determined as the equilibrium of a network model with the flat link-specific long-run generalized price functions replacing the conventional short-run generalized price functions.

2.3. Second-best pricing and capacity choice

2.3.1. No self-financing constraint

We next consider the situation where not all links are subject to toll and capacity optimization. The regulator then faces a problem that is characterized by the following Lagrangian:

$$\Lambda = \sum_{m=1}^{M} \int_0^{\sum_{r=1}^{R} \delta_{rm} \cdot N^r} D^m(n)\mathrm{d}n - \sum_{l=1}^{L} \sum_{r=1}^{R} \delta_{lr} \cdot N^r \cdot c^l \left(\sum_{\rho=1}^{R} \delta_{l\rho} \cdot N^\rho, K^l \right) - \sum_{l=1}^{L} C^{c,l}(K^l)$$
$$+ \sum_{r=1}^{R} \delta_r^A \cdot \lambda^r \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l \left(\sum_{\rho=1}^{R} \delta_{l\rho} \cdot N^\rho, K^l \right) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m \left(\sum_{\rho=1}^{R} \delta_{\rho m} \cdot N^\rho \right) \right)$$ (14)

Compared to the first-best problem in (10), we now work with link-based tolls from the outset. Denote links at which a toll can be set as links for which $\delta_l^\tau = 1$, and for which capacity can be set as $\delta_l^K = 1$ (both dummies are zero otherwise). The Lagrangian in (14) 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:

$$\frac{\partial \Lambda}{\partial N^r} = \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) - \sum_{l=1}^{L} \delta_{lr} \cdot c^l(\cdot) - \sum_{l=1}^{L}\sum_{\rho=1}^{R} \delta_{lr} \cdot \delta_{l\rho} \cdot N^\rho \cdot \frac{\partial c^l(\cdot)}{\partial N^r}$$

$$+ \sum_{\rho=1}^{R} \delta_\rho^A \cdot \lambda^\rho \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \delta_{l\rho} \cdot \frac{\partial c^l(\cdot)}{\partial N^r} - \delta_{rm} \cdot \delta_{\rho m} \cdot \frac{\partial D^m(\cdot)}{\partial N^r} \right) = 0 \qquad \forall r : \delta_r^A = 1 \tag{15a}$$

$$\frac{\partial \Lambda}{\partial K^l} = -\sum_{r=1}^{R} \delta_{l,r} \cdot N^r \cdot \frac{\partial c^l(\cdot)}{\partial K^l} - \frac{\partial C^{c,l}(\cdot)}{\partial K^l} + \sum_r \delta_r^A \cdot \lambda^r \cdot \delta_{lr} \cdot \frac{\partial c^l(\cdot)}{\partial K^l} = 0 \qquad \forall l : \delta_l^K = 1, \tag{15b}$$

$$\frac{\partial \Lambda}{\partial \tau^l} = \sum_{r=1}^{R} \delta_r^A \cdot \delta_{lr} \cdot \lambda^r = 0 \qquad \forall l : \delta_l^\tau = 1, \tag{15c}$$

$$\frac{\partial \Lambda}{\partial \lambda^r} = \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l(\cdot) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) = 0 \qquad \forall r : \delta_r^A = 1, \tag{15d}$$

while:

$$\delta_r^A = 0 \quad \text{iff} \quad \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l(\cdot) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) > 0. \tag{15e}$$

In contrast to the first-best case, the Lagrangian multipliers λ^r are now not all individually equal to zero. Recall that the equilibrium value of a Lagrangian multiplier reflects the marginal impact of a relaxation of the associated constraint upon the optimized value of the objective. A non-zero value of λ^r therefore denotes that social surplus could be increased by introducing a marginal route-specific toll (if $\lambda^r > 0$) or subsidy (if $\lambda^r < 0$) in addition to the existing link tolls in the second-best optimum. With imperfect pricing, there will generally be routes for which this is indeed the case.

It turns out that the possibility to control some of the link capacities does not simplify the second-best toll formula that can be deduced from (15a)–(15e), which was presented by Verhoef (2002a). Of course, second-best optimization of capacities will affect the second-best equilibrium toll levels, and the achievable social surplus, but the general toll expression remains unaltered. Because this formula is tedious and has no closed-form solution with the Lagrangian multipliers λ^r substituted out, we will not repeat it here. The intuition behind the unchanged toll formula is, in fact, rather straightforward. The toll was originally determined for arbitrary capacities. Obviously, these capacities could have been set optimally by pure chance, in which case the toll formula should still be correct. The result is also consistent with toll choice being a short-run decision, while capacity choice is a long-run decision.

Because there is no closed-form solution for the multipliers λ^r, it may seem that also the investment rule of (15b) yields little further insight. But this ignores that for links at which both capacity and toll is optimized, substitution of (15c) in (15b) implies:

$$-N^l \cdot \frac{\partial c^l(\cdot)}{\partial K^l} = \frac{\partial C^{c,l}(\cdot)}{\partial K^l} \qquad \forall l : \delta_l^K = \delta_l^\tau = 1 . \tag{15b$'$}$$

Comparing (15b$'$) with (12b), we find that the second-best investment rule for a link on which both capacity and toll can be set, on an otherwise not perfectly priced network, is the same as for first-best policies (nevertheless, because flows N^l generally differ between the first-best and the second-best optimum, the actual equilibrium capacity levels will of course also generally not be equal). The intuition is that, with a link toll τ^l available, the regulator can perfectly control the generalized price p^l on that link, and therefore the flow N^l. Whichever combination $\{p^l, N^l\}$ the regulator chooses, social surplus is always maximized if that combination is achieved against minimized (average) social cost. And this induces the regulator to operate 'along the long-run cost function', which is defined by the minimization of social cost given the flow N^l, as represented by (15b$'$).

For our specific functional forms model, we can derive the same long-run average cost components as we found in the first-best case:

$$\tilde{K}^l = \tilde{N}^l \cdot \left(\frac{\alpha \cdot \beta \cdot \chi}{\rho} \right)^{\frac{1}{\chi+1}} , \tag{16a}$$

$$\tilde{c}^l = t_f^l \cdot \left(\alpha + (\alpha \cdot \beta)^{\frac{1}{\chi+1}} \cdot \left(\frac{\rho}{\chi} \right)^{\frac{\chi}{\chi+1}} \right) , \tag{16b}$$

$$\tilde{c}^{c,l} = t_f^l \cdot (\alpha \cdot \beta \cdot \chi)^{\frac{1}{\chi+1}} \cdot \rho^{\frac{\chi}{\chi+1}} , \tag{16c}$$

$$\tilde{c}^{t,l} = \tilde{c}^l + \tilde{c}^{c,l} . \tag{16d}$$

However, since there is no closed-form solution for the second-best toll $\tilde{\tau}^l$, we also do not have a closed-form solution for the long-run generalized price $\tilde{p}^l$; *i.e.*, no equivalent expressions for (13e) and (13f) can be derived.

To what extent, then, is the solution in (16) helpful in applied modelling? Probably it still is. Knowing the optimal ratio between second-best optimal flow and capacity in advance greatly simplifies numerical procedures to find the second-best optimum, since the dimensions are reduced by one: capacity is implied by flow. This, of course, is true only for capacities on links that are also tolled. For links for which capacity can be adjusted, but no toll can be set, equation (15b) implies that the investment rule depends on equilibrium values of the Lagrangian multipliers λ^r. Because there are no closed-form solutions for these multipliers, there will also be no closed-form solutions for the associated long-run cost functions.

2.3.2. *Self-financing constraint*

As a final case we consider the second-best situation described above with an additional constraint added, namely that some of the links for which the toll and capacity are set should be self-financing. For a first-best network and under the appropriate conditions, this constraint will be satisfied automatically, as already discussed above. But for the second-best case where untolled links exist, the toll formulae change compared to the first-best case of Pigouvian taxes equal to marginal external costs, while the investment rule for a tolled link remains the same – as the comparison of (15b′) with (12b) revealed. We can therefore no longer expect exact self-financing to apply for links on which both the toll and capacity is optimized. Imposing it as a constraint, therefore, may be expected to affect both the link's second-best toll and its capacity.

Denote the links for which the self-financing constraint is imposed with a dummy variable $\delta_l^\pi = 1$ (again, it is zero otherwise). Note that we assume that a self-financing constraint will apply only for links on which both the tolls and the capacity can be set, but that there may be links without a self-financing constraint on which either the toll, or the capacity, or both can be set. The associated Lagrangian follows closely the one specified in Appendix A of Verhoef (2007), who considered the case where only for one link in a generalized network the toll and capacity can be set under a self-financing constraint. It reads as follows:

$$
\begin{aligned}
\Lambda = \sum_{m=1}^{M} \int_0^{\sum_{r=1}^{R} \delta_{rm} \cdot N^r} D^m(n) \mathrm{d}n - \sum_{l=1}^{L} \sum_{r=1}^{R} \delta_{lr} \cdot N^r \cdot c^l \left(\sum_{\rho=1}^{R} \delta_{l\rho} \cdot N^\rho, K^l \right) - \sum_{l=1}^{L} C^{c,l}(K^l) \\
+ \sum_{r=1}^{R} \delta_r^A \cdot \lambda^r \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l \left(\sum_{\rho=1}^{R} \delta_{l\rho} \cdot N^\rho, K^l \right) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m \left(\sum_{\rho=1}^{R} \delta_{\rho m} \cdot N^\rho \right) \right). \\
+ \sum_{l=1}^{L} \delta_l^\pi \cdot \lambda^l \cdot \left(\sum_{r=1}^{R} \delta_{lr} \cdot N^r \cdot \tau^l - C^{c,l}(K^l) \right)
\end{aligned}
\tag{17}
$$

The final constraint is the self-financing constraint, and a non-zero value of λ^l reflects that the constraint is binding. The first-order conditions are:

$$
\begin{aligned}
\frac{\partial \Lambda}{\partial N^r} = \sum_{m=1}^{M} \delta_{rm} \cdot D^m(\cdot) - \sum_{l=1}^{L} \delta_{lr} \cdot c^l(\cdot) - \sum_{l=1}^{L} \sum_{\rho=1}^{R} \delta_{lr} \cdot \delta_{l\rho} \cdot N^\rho \cdot \frac{\partial c^l(\cdot)}{\partial N^r} \\
+ \sum_{\rho=1}^{R} \delta_\rho^A \cdot \lambda^\rho \cdot \left(\sum_{l=1}^{L} \delta_{lr} \cdot \delta_{l\rho} \cdot \frac{\partial c^l(\cdot)}{\partial N^r} - \delta_{rm} \cdot \delta_{\rho m} \cdot \frac{\partial D^m(\cdot)}{\partial N^r} \right) , \\
+ \sum_{l=1}^{L} \delta_{lr} \cdot \delta_l^\pi \cdot \lambda^l \cdot \tau^l = 0 \qquad \forall r : \delta_r^A = 1
\end{aligned}
\tag{18a}
$$

$$
\begin{aligned}
\frac{\partial \Lambda}{\partial K^l} = -\sum_{r=1}^{R} \delta_{l,r} \cdot N^r \cdot \frac{\partial c^l(\cdot)}{\partial K^l} - \frac{\partial C^{c,l}(\cdot)}{\partial K^l} + \sum_r^{R} \delta_r^A \cdot \lambda^r \cdot \delta_{lr} \cdot \frac{\partial c^l(\cdot)}{\partial K^l} \\
- \delta_l^\pi \cdot \lambda^l \cdot \frac{\partial C^{c,l}(\cdot)}{\partial K^l} = 0 \qquad \forall l : \delta_l^K = 1
\end{aligned}
\tag{18b}
$$

$$\frac{\partial \Lambda}{\partial \tau^l} = \sum_{r=1}^{R} \delta_r^A \cdot \delta_{lr} \cdot \lambda^r + \sum_{r=1}^{R} \delta_l^\pi \cdot \delta_{lr} \cdot N^r \cdot \lambda^l = 0 \qquad \forall l : \delta_l^\tau = 1, \tag{18c}$$

$$\frac{\partial \Lambda}{\partial \lambda^r} = \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l (\cdot) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m (\cdot) = 0 \qquad \forall r : \delta_r^A = 1, \tag{18d}$$

$$\frac{\partial \Lambda}{\partial \lambda^l} = \sum_{r=1}^{R} \delta_{lr} \cdot N^r \cdot \tau^l - C^{c,l} (K^l) = 0 \qquad \forall l : \delta_l^\pi = 1, \tag{18e}$$

while:

$$\delta_r^A = 0 \quad \text{iff} \quad \sum_{l=1}^{L} \delta_{lr} \cdot \left(c^l (\cdot) + \tau^l \right) - \sum_{m=1}^{M} \delta_{rm} \cdot D^m (\cdot) > 0 . \tag{18f}$$

Despite the tedious appearance of this set of first-order conditions, they imply a simple and intuitive investment rule, and a corresponding long-run cost function, for those links on which a self-financing constraint applies (and the toll and capacity can therefore both be set). To see this, first observe that (18c) implies:

$$N^l \cdot \lambda^l = -\sum_{r=1}^{R} \delta_r^A \cdot \delta_{lr} \cdot \lambda^r \qquad \forall l : \delta_l^\pi = 1 . \tag{18c$'$}$$

Substitution of (18c$'$) into (18b) implies:

$$\left(1 + \lambda^l \right) \cdot \left(-N^l \cdot \frac{\partial c^l (\cdot)}{\partial K^l} - \frac{\partial C^{c,l} (\cdot)}{\partial K^l} \right) \qquad \forall l : \delta_l^\pi = 1 . \tag{19a}$$

This means that for any value of λ^l, the conventional first-best investment rule of (12b) applies. Multiplying both terms inside the large brackets with K^l, observing that the second term then gives (minus) total capacity cost $K^l \cdot \partial C^{c,l}/\partial K^l$, and using Euler's theorem (implying that $-K^l \cdot \partial c^l/\partial K^l = N^l \cdot \partial c^l/\partial N^l$) as we did before under (5), it is easily demonstrated that exact self-financing for the links with $\delta_l^\pi = 1$ requires:

$$\tau^l = -K^l \cdot \frac{\partial c^l}{\partial K^l} = N^l \cdot \frac{\partial c^l}{\partial N^l} \quad \forall l : \delta_l^\pi = 1 . \tag{19b}$$

Equation (19b) implies that self-financing on these links requires the toll to be set exactly according to the Pigouvian rule, which also applies in the first-best case. In other words, for these links, on which the toll and capacity can be optimized under a self-financing constraint, the conventional first-best rules apply; independent of whether there are other links in the network on which a toll is set, capacity is optimized, or a self-financing constraint applies. For our specific example, this means that equations (13a)-(13g) apply for these links, so that second-best equilibrium results should be relatively easy to find.

3. Numerical application

We now turn to a discussion of numerical results that illustrate our main analytical findings. Besides a small exercise with a single-link model, the same network model is used throughout

this section. It is designed to capture the most important types of link interactions in the simplest possible configuration, and therefore has both serial and parallel links. The network is illustrated in Figure 1.
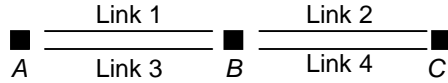

Figure 1. The network for the numerical example

There are two origin-destination pairs, *AC* and *BC*, and four links labelled 1–4. The two inverse demand functions have the constant-elasticity form of equation (1). We use a demand elasticity of $\eta=-0.35$ for both OD-pairs. The scale parameters δ amount to $4.0630 \cdot 10^{12}$ for *AC* and $1.4056 \cdot 10^{12}$ for *BC* (these are in fact rounded values; they were calibrated to produce a base equilibrium with flows of 4000 for *AC* and 3000 for *BC*, respectively).

The user cost functions take on the BPR form of equation (2) for all links. We set $\beta=0.15$ and $\chi=4$, which are conventional values for the BPR function. Furthermore, we assume that each link has a length of 60 kilometres, which, with a free-flow speed of 120 km/hr, corresponds with $t_f=30$ for all links (we measure time in minutes). All links have an initial capacity of $K^l=1500$. This was chosen to approximate one highway lane: $K^l=1500$ implies a doubling of travel times at a flow of around 2400 vehicles per lane 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 and Verhoef, 2007, Figures 3.2 – 3.4). A maximum flow, however, in itself is not defined for BPR functions. With travel times measured in minutes, the value of time α was set at 0.125 (Euros), implying a value of time of 7.5 Euros per hour, which is in line with the "official" Dutch value.

The capacity cost function is as shown in (3). Only the parameter ρ needs to be determined. We follow Verhoef and Mohring (2007), who proposed a unit price of capacity of 7 Euros for a 60 km road, for use with a BPR function.[3] Because we define ρ in this paper as a value that is to be multiplied with t_f, and t_f equals 30 minutes for a 60 km road, we set $\rho=7/30$ (approximately 0.233).

Table 1 shows the equilibrium values of the key variables in the resulting base-equilibrium. Clearly, there is a divergence between average and marginal user costs on all links, reflecting the unpriced congestion externality.

---

[3] They wrote: "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)."

	Flow	Average user cost €	Marginal cost €	Toll €	Capacity	Revenue €	Capacity cost €	Volume capacity ratio
Link 1	2 000	5.53	12.64	0	1500	0	10 500	1.33
Link 2	3 500	20.42	87.12	0	1500	0	10 500	2.33
Link 3	2 000	5.53	12.64	0	1500	0	10 500	1.33
Link 4	3 500	20.42	87.12	0	1500	0	10 500	2.33
OD-pair *AC*	4000	25.95						
OD-pair *BC*	3000	20.42						

Table 1: Numerical model: base equilibrium

3.1. Single-link results

For the single link model, we used the link-based parameters specified above, while the single OD-related demand parameters were the same as those for the OD-pair *AC*. We use the model to verify our expectation just below (9b), stating that because the solution of $\tilde{K}$ as a function of *N* depends on the ratio *K/N* only, $\tilde{K}$ will be proportional to $\tilde{N}$. That is, if for instance the scale factor δ in (1) grows over time and capacity is adjusted in a second-best optimal fashion, we expect the second-best optimal levels of average user cost and average capacity cost to remain constant. We test this hypothesis for three demand elasticities: the base value of η=–0.35, a low elasticity case with η=–0.1, and a high elasticity case with η=–0.7. We present the long-run average user cost $\tilde{c}$ and the long-run average capacity cost $\tilde{c}^c$ for different values of δ for each of these elasticities, and also show, for comparison, $\tilde{c}$ and $\tilde{c}^c$ as they would apply under optimal tolling, from (6b) and (6c).
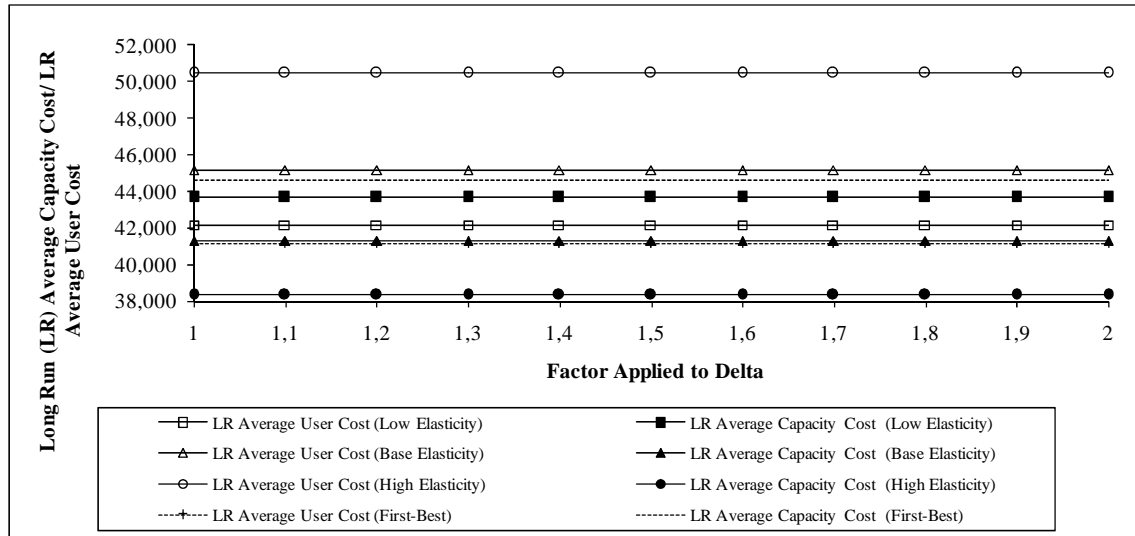

Figure 2. Long-run average costs for the single-link model: second-best (no tolling) under various demand elasticities, and first-best

Figure 2 shows the results. We draw two conclusions. The first is that, as expected, the second-best long-run average cost functions are flat, so also without pricing these are constants, independent of equilibrium demand levels. The second is that the second-best long-run average cost levels approach their first-best counterparts more closely as demand becomes less elastic. This is consistent with the expression we found for the Lagrangian multiplier λ associated with the user equilibrium condition, in equation (8b), which showed that the adjustment made to the conventional first-best investment rule diminishes as demand becomes less elastic. Note that the average user costs $\tilde{c}$ are, without pricing, above the first-best level, and increasingly so when demand becomes more elastic. The average capacity costs $\tilde{c}^c$, in contrast, are below the first-best level, and increasingly so when demand becomes more elastic. Both results reflect the finding of equations (8a) and (8b), namely that a greater correction factor λ is applied to the actual flow N, and therefore less capacity per user is supplied, when demand is more elastic and should therefore be discouraged more strongly to limit socially unwarranted induced demand effects.

3.2.   First-best networks

We now move to first-best pricing and capacity choice for a full network. As noted above, because (13f) contains constant parameters only, the network optimum can be determined very easily, by replacing the conventional short-run generalized price functions by their long-run counterparts of (13f). The resulting levels of $\tilde{N}^l$ imply the long-run capacities $\tilde{K}^l$ via (13a). Applying these long-run capacities in the original short-run cost functions should then result in a network that has toll levels of (13e) also as the short-run optimal tolls; that is, satisfying (13g) for the BPR functions used here.

	Flow	Average user cost €	Marginal cost €	Toll €	Capacity	Revenue €	Capacity cost €	Volume capacity ratio
Link 1	4 240.61	5.14	10.72	5.58	3379.44	23 656.10	23 656.10	1.255
Link 2	7 999.45	5.14	10.72	5.58	6374.96	44 624.70	44 624.70	1.255
Link 3	35.46	5.14	10.72	5.58	28.26	197.81	197.81	1.255
Link 4	35.46	5.14	10.72	5.58	28.26	197.81	197.81	1.255
OD-pair *AC*	4 276.07	10.28						
OD-pair *BC*	3 758.84	5.14						

Table 2: Numerical model: first-best optimum

Table 2 shows the resulting flows, costs, tolls and capacities resulting from the application of this process. All links are self-financing and that the volume capacity (N^l/K^l) ratio is constant for all links. The first-best solution is characterized by an increase in total capacity on both sections of the network. The fact that the capacities of links 3 and 4 have become small, while those of 1 and 2 have increased, is meaningless: the solution in Table 2 is just one of an infinite number of possible solutions. Due to the symmetry of the network there is no unique optimum in terms of individual link capacities and flows. Only the sum of capacities for links 1 and 3, and 2 and 4, matter under first-best pricing.

The total welfare change compared to the base equilibrium is € 92 868.75. We next verify that this is an optimum, at least locally, by perturbing the optimal tolls and capacities by a small positive or negative value.[4] Table 3 shows the welfare values for perturbations of +/- 1%. The results confirm that we have indeed found an optimum welfare solution. The welfare surface is, not surprisingly, rather flat around this optimum, and that for perturbations of the low capacities (3,4) the welfare levels are nearly identical to the original solution.

The example thus illustrates how the long-run cost functions can indeed be used to easily identify the long-run optimum for a congested network, with both tolls and capacities optimized. We have already applied the same method successfully on a bigger network, namely one for Edinburgh, and solved this modified traffic assignment problem using the SATURN (Van Vliet, 1982) model. For reasons of space, we will not present these results here.

Perturbation +1%	Social surplus gain €	Perturbation -1%	Social surplus gain €
Toll Link 1	92 868.56	Toll Link 1	92 868.55
Toll Link 2	92 868.30	Toll Link 2	92 868.29
Toll Link 3	92 868.70	Toll Link 3	92 868.70
Toll Link 4	92 868.70	Toll Link 4	92 868.70
Cap. Link 1	92 863.76	Cap. Link 1	92 863.76
Cap. Link 2	92 859.92	Cap. Link 2	92 859.64
Cap. Link 3	92 868.74	Cap. Link 3	92 868.74
Cap. Link 4	92 868.74	Cap. Link 4	92 868.74

Note: First-best social surplus gain is € 92868.75

Table 3: Numerical model: perturbations around first-best optimum

3.3.   Second-best pricing and capacity choice

For the second-best policies, we consider the case where only link 2 is subject to tolling and capacity choice. The network configuration makes sure that our example allows for interaction with both a parallel link and with serial links.

3.3.1. No self-financing constraint

As pointed out in Section 2.3.1, there is generally no closed-form solution for the second-best toll when no self-financing constraint applies. For the present small network, a closed-form expression might still be available, but we prefer to pretend to be in the situation that will also be faced when working with bigger networks, so that numerical methods have to be applied to find the second-best tolls. In this situation, we can use the so-called 'cutting constraint algorithm' (CCA) of Lawphongpanich and Hearn (2004) to solve for the second-best toll and capacity. For details of how this is implemented, we refer to Koh, Shepherd and Sumalee (2007). This algorithm is mathematically guaranteed to converge to a local equilibrium point of the problem. Table 4 shows the results obtained by using the CCA algorithm.

---

[4] Each variable (toll or capacity) is perturbed separately with all others held at the optimal values from Table 2.

Toll (€)	Capacity	Flow	Social surplus gain (€)	Revenue (€)	Capacity cost (€)	Capacity implied by flow and (16a)
2.67	5219.85	6553.75 (N/K =1.255)	81 679 (87.95% of first-best)	17 467.55	36 538.93	5222.84

Table 4: Second-best pricing and capacity choice for link 2 alone (no self-financing constraint): results for link 2

Table 4 shows that in this case, second-best toll revenues are insufficient to cover the capacity costs. The social surplus gain of € 81 679 comparedto the base equilibrium is approximately 88% of the first-best gain. This gain is therefore quite large, which is a direct consequence of the assumed network configuration and the fact that initial base capacities are relatively small.

The final column gives the capacity that equation (16a) would give for the flow level obtained by the CAA. This is practically equal to the capacity that the algorithm itself finds. This confirms the validity of equation (16a), and suggests that its use directly in the algorithm might make it more rapid as it reduces the dimensions of the optimization problem. We intend to investigate this possibility further in follow-up research.

3.3.2. Self-financing constraint

The final case concerns the above second-best problem with an additional self-financing constraint. The discrepancy between revenues and capacity cost in Table 4 of course indicates that the constraint will be binding. To find the associated toll and capacity, we can use the rules given in (13a) to (13f). As for the first-best case, we thus we solve a standard user equilibrium traffic assignment problem, with the only difference that for link 2 we replace the conventional short-run user cost function by the (constant) generalized price of (13f). Again, once we have found the link-flow from this adapted network problem, the capacity on the link to be optimised (link 2) can be inferred from (13a), and the tolls can be computed from either (13e) or (13g), which should give the same answer.

Toll (€)	Capacity	Flow	Social surplus gain (€)	Revenue (€)	Capacity cost (€)
5.58	4 408.08	5 531.37 (N/K =1.255)	78 618 (84.65% of first-best)	30 856.54	30 856.54

Table 5: Second-best pricing and capacity choice for link 2 alone with self-financing constraint: results for link 2

The results (for link 2) are shown in Table 5. Note that the gain in social surplus is now 84.65% of the first-best case, and therefore lower than that without the binding additional constraint on self-financing.

We have verified the results from Table 5 by solving the same problem with the CCA, which indeed produced the same results. Moreover, to verify that the solution we have found is at least optimal locally, we carried out a perturbation analysis for the toll and capacity

shown in Table 5. When perturbing the toll (capacity) for link 2, we applied a bisection search method to find a corresponding level of capacity (toll) that would preserve the self-financing result. The results are reported in Table 6 below. It is clear that all the other perturbation levels and the accompanying capacities that satisfy the self-financing constraint would not provide a higher level than the welfare that we found. Again, the region around this second-best optimum is rather flat, and the welfare solutions are very close. Note that the capacity perturbation of +1% and the toll perturbation of +1% are nearly identical; the same is true for the two –1% perturbations.

Toll €5.58	Adjusted Capacity	Toll Revenue	Capacity Cost	Welfare
+1%: 5.63	4 451.83	31 162.82	31 162.81	78 607.59
–1%: 5.52	4 363.66	30 545.64	30 545.65	78 607.19
Capacity 4408.08	Adjusted Toll	Toll Revenue	Capacity Cost	Welfare
+1%: 4 452.16	5.63	31 165.10	31 165.10	78 607.43
–1%: 4 364.00	5.52	30 547.95	30 547.97	78 607.39

Note: Second-best social surplus gain is €78 618 (see Table 5)

Table 6: Second-best pricing and capacity choice for link 2 alone with self-financing constraint: perturbations around second-best optimum

4.   Conclusion

This paper considered the use of 'long-run cost functions' for congested networks in solving second-best network problems, in which capacity and tolls are instruments. We considered second-best cases where only a subset of links in a network is subject to tolling and a subset is subject to capacity optimization, where these subsets may or may not overlap. We also considered cases with and without a self-financing constraint imposed.

Our results demonstrate that, under certain assumptions, second-best long-run cost (or actually: generalized price) functions can be derived for most of the cases of interest, which can be used in an applied network model as a substitute for the conventional short-run user cost functions. Doing so reduces the dimensionality of the problem and therefore helps in speeding up procedures used for finding second-best optima.

The most straightforward cases were the first-best problem, in which tolls and capacities can be set on all links, and the second-best problem where a self-financing constraint applies on a link for which both instruments can be set. It turned out that the long-run cost functions for these two cases are in fact the same. To find the optimum in a network of any size or shape, it suffices to replace the conventional short-run user cost functions by the long-run generalized price function. From the resulting network equilibrium, optimal capacities and tolls can be immediately derived.

For the second-best cases where only the toll or the capacity can be set, no such easy answers are available. Nevertheless, for the case of tolling, it is true that the ratio of second-best optimal link flow and link capacity can be determined beforehand, which is valuable

information that of course should be exploited in algorithms designed to find the second-best optimum. But there is no closed-form expression for the long-run second-best optimal toll. The same is true for second-best capacity for links without tolling. The latter result was shown in the context of a single-link model, but there is no reason to expect that the conclusion would change in a bigger network. Even though things are more complicated in these cases, we can observe that the combination of our findings is that for the most general second-best network problem – where only tolls can be set on some links, only capacities on some other links, and both on yet a third subset of links – there is at most one 'instrumental dimension' for each link: either toll, or capacity, or toll with capacity implied immediately by the link flow. This reduction in dimensionality should have positive consequences for the speed of solution algorithms. The extent to which this is indeed the case is an issue that we intend to address in a follow-up paper.

References

Beckmann, M.J., C.B. McGuire and C.B. Winsten (1956) *Studies in the Economics of Transportation* Yale University Press, New Haven, CT.

Bellman R. (1961) *Adaptive Control Processes: A Guided Tour* Princeton University Press, Princeton, NJ.

Braid, R.M. (1996) "Peak-load pricing of a transportation route with an unpriced substitute" *Journal of Urban Economics* **40** 179-197.

Dafermos, S. (1973) "Toll patterns for multiclass-user transportation networks" *Transportation Science* **7** 211-223.

De Borger, B., S. Proost and K. van Dender (2005) "Congestion and tax competition in a parallel network" *European Economic Review* **49** 2013-2040.

De Palma, A. and R. Lindsey (2000) "Private roads: Competition under various ownership regimes" *Annals of Regional Science* **34** 13-35.

Koh A., S. Shepherd and A. Sumalee A (2008) "Second-best toll and capacity optimisation in networks" Proceedings of the Hong Kong Society for Transportation Studies.

Lawphongpanich, S. And D.W. Hearn (2004) "An MPEC approach to second-best toll pricing" *Mathematical Programming* **101B** (1) 33-55.

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

Liu, L.N. and J.F. McDonald (1998) "Efficient congestion tolls in the presence of unpriced congestion: A peak and off-peak simulation model" *Journal of Urban Economics* **44** 352-366.

Marchand, M. (1968) "A note on optimal tolls in an imperfect environment" *Econometrica* **36** 575-581.

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* London: Routledge.

Sumalee, A., A. May and S. Shepherd (2005) "Comparison of judgmental and optimal road pricing cordons" *Transport Policy* **12** 384-390.

Van Vliet, D. (1982) "SATURN - a modern assignment model" *Traffic Engineering and Control* **23** 578-581.

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. (2007) "Second-best road pricing through highway franchising" *Journal of Urban Economics* **62** 337-361.

Verhoef, E.T. and H. Mohring (2007) "Self-financing roads" Discussion paper TI 2007-068/3, Tinbergen Institute, Amsterdam-Rotterdam.

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

Verhoef, E.T. and J. Rouwendal (2004) "Pricing, capacity choice and financing in transportation networks" *Journal of Regional Science* **44** 405-435.

Verhoef, E.T. and K.A. Small (2004) "Product differentiation on roads: constrained congestion pricing with heterogeneous users" *Journal of Transport Economics and Policy* **38** 127-156.

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

Wardrop, J.G. (1952) "Some theoretical aspects of road traffic research" *Proceedings of the Institute of Civil Engineers (Part II)* **1** (2) 325-378.

Yang, H. and H.-J. Huang (1998) "Principle of marginal-cost pricing: how does it work in a general road network?" *Transportation Research* **32A** (1), 45-54.