PRODUCT DIFFERENTIATION ON ROADS

Constrained Congestion Pricing with Heterogeneous Users

Erik T. Verhoef[1*] and Kenneth A. Small[2+]

[1]Department of Spatial Economics
Free University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Phone: +31-20-4446094
Fax: +31-20-4446004
E-mail: everhoef@econ.vu.nl

[2]Department of Economics
University of Irvine at California
Social Science Plaza
Irvine CA 92697-5100
USA
Phone: +1-949-824-5658
Fax: +1-949-824-2182
E-mail: ksmall@uci.edu

This version: 19 June, 2002

Key words: congestion, road pricing, networks, second-best, value of time
JEL codes: R41, R48, D62

Abstract

We explore the properties of various types of public and private pricing on a congested road network, with heterogeneous users and allowing for elastic demand. Heterogeneity is represented by a continuum of values of time. The network allows us to model certain features of real-world significance: pricing restrictions on either complementary or substitute links, as well as interactions between different user groups on shared links (e.g. in city centers). We find that revenue-maximizing pricing is much less efficient than welfare-maximizing pricing, whether restricted or unrestricted; but this difference is mitigated by the product differentiation made possible with heterogeneous users. Product differentiation also produces some unexpected distributional effects: those hurt most by pricing may be people with moderate rather than low values of time, and first-best pricing can cause congestion levels to *increase* for some users compared to no pricing. Ignoring heterogeneity causes the welfare benefits of a policy close to one currently being used, namely second-best pricing of one of two parallel links, to be dramatically underestimated. Unlike first-best policies, second-best policies are in danger of losing much of their potential effectiveness if heterogeneity is ignored when setting toll levels.

[*]Erik Verhoef is affiliated as a research fellow to the Tinbergen Institute. The research of Erik Verhoef has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

[+]Kenneth Small is grateful to the University of California Transportation Center for financial support, and to Resources for the Future for being host to a year's leave during which some of this work was completed.

1 Introduction

Public-finance economists have long advocated Pigovian taxes and related ‘market-like’ policies to attain better pricing of goods supplied by the public sector. Most such policies are enacted on a piecemeal and limited basis, if at all. Cases in point are the marketable permits established by the US Clean Air Act of 1990 and several heavily restricted pollution trading schemes reviewed by Hahn (1989).

One of the best-studied applications of Pigovian taxes is road pricing. The economic fundamentals were well laid out by Pigou (1920), Knight (1924), Walters (1961), and Vickrey (1963, 1969). The concept is favoured by many transportation policy makers, but mainly in the form of experiments or demonstrations rather than full-scale applications (Small and Gómez-Ibáñez, 1998). Examples include toll rings around city centres in Norway, peak-period toll surcharges on certain French expressways, special tolled express lanes on two freeway segments in southern California, and a single congestion-priced expressway near Toronto.

This history suggests an increasing importance of partial rather than first-best congestion-pricing schemes. Such schemes include privately or publicly operated toll roads parallel to unpriced highways. Depending on the particular scheme, pricing may be prohibited on routes that are either substitutes for or complements of the one that is priced, and may involve either social or private objectives. Thus the analysis requires a model permitting a variety of objectives and pricing constraints. Because much of the purpose of these schemes is to test and shape public opinion, distributional issues are often paramount. Focusing on these turns out to be quite interesting because some of these demonstrations offer highly differentiated products.

In this paper, we directly and simultaneously address issues of second-best policy, public or private objectives, product differentiation, and distribution as they arise from constrained road pricing. The numerical version of our model uses (for its base case) an empirically obtained distribution of values of time for morning peak road use, based on a questionnaire among morning peak road users in the Dutch Randstad area (Verhoef, Nijkamp and Rietveld, 1997). We analyse both substitute and complementary services to the one being priced by using a simple network with both parallel and serial links. Such a set-up can represent, for example, parallel priced and unpriced arterials entering a city center where their users interact on congested streets.

A preview of especially interesting results: We find that ignoring heterogeneity in values of time may cause the welfare benefits of second-best policies to be drastically underestimated, by a factor of nine in our base case. Private pricing is almost always worse than no pricing, except when a private route has significant free-flow speed advantages over the free parallel route. Heterogeneity makes first-best pricing strongly anti-egalitarian, in that it may actually worsen the travel times faced by low-value-of-time users – a paradox explained by its effect of channelling these users onto just a portion of the total capacity but then applying a low price to them. Second-best pricing is much more egalitarian; however, welfare is greatly enhanced if instead of pricing just a small portion of the network, most capacity is priced with only a small portion reserved as a free option. Finally, offering a differentiated product can produce the intriguing possibility that a second-best pricing policy may provide benefits to those who care least and to those who care most about service quality, while hurting those in the middle –hardly an ideal set-up for political success.

Such results pose challenges for the demonstration-project approach to pricing policy. There is a real danger that most of the hoped-for welfare benefits from pricing will be lost, or even turned into disbenefits; or that specific groups will incur perverse results such as higher price and worse service at the same time. On the other hand, dispersion in preferences does offer the potential to reap substantial benefits through product differentiation, which lends itself to an experimental approach. Our model provides a flexible and realistic tool to study these advantages and disadvantages. Moreover, although we will frame the discussion in terms of roads, the modelling framework and insights obtained may be applicable to other types of congested network markets, such as internet providers and telephone service, where there is dispersion in willingness to pay for quality.

2 The analytical model

2.1 Prior literature

Most of the second-best literature addresses two parallel routes where one of the two routes is untolled. Lévy-Lambert (1968), Marchand (1968), and Verhoef, Nijkamp and Rietveld (1996) use the static model of Walters (1961) and Vickrey (1963), while Braid (1996) uses the dynamic bottleneck model of Vickrey (1969). The main conclusions are that the second-best toll trades off route split effects against overall demand effects; that this toll is usually considerably smaller than the first-best toll; and that second-best pricing often leads to much smaller welfare gains than first-best pricing. Liu and McDonald (1998) confirm these results for parameters designed to match one of the California pricing demonstration projects (SR-91 in Orange County). Yang and Huang (1999) endogenize vehicle occupancy and allow for free carpool access to the tolled route.

Revenue-maximizing congestion tolls for a single highway are derived by Edelson (1971) and Mills (1981). When just one of two parallel roads can be priced, Verhoef *et al.* (1996) and Liu and McDonald (1998, 1999) find that the revenue-maximizing price is typically much higher than the second-best price and will achieve very much lower, usually negative, welfare gains. McDonald et al. (1999, pp. 122-124) derive the second-best toll on a link that has both an unpriced substitute and an unpriced complement; but they are unable to say whether the complementary link makes the toll higher or lower. De Palma and Lindsey (2000) consider a variety of ownership regimes, including private and mixed duopolies, both with and without constraints on pricing one of two parallel roads; they focus especially on the effects of time-varying demand patterns and corresponding time-varying tolls. Viton (1995) considers the prospects for a private operator to cover the cost of road construction, reaching optimistic conclusions due to the high toll that can be charged even when in close competition with a free public road.

Very few studies of the two-route problem incorporate heterogeneity. Arnott, De Palma and Lindsey (1992) consider two user groups and two routes within the bottleneck model, but do not consider the case when only one route can be priced. Small and Yan (2001) do consider such a case, also with just two discrete user groups. Mohring (1979) considers a continuous distribution of values of travel time in analyzing bus fares with a parallel automobile route; he does not, however, analyze dispersion separately from mean value of time. Less closely related are the analyses by Train, McFadden and Goett (1987) and by Train, Ben-Akiva and Atherton (1989) of electricity and telephone users, respectively, facing a voluntary choice among alternate rate schedules with different time-of-day characteristics.

Models that treat two discrete user groups, besides providing only a crude approximation to real heterogeneity, result in analytical difficulties because several distinct types of pooled or separated equilibria. In the present paper, we consider instead a continuum of groups. Only two types of equilibria can then occur: pooled (when tolls are absent or exactly equal on the two parallel routes), or fully separated (in all other cases). Moreover, using a continuum of values of time allows intermediate groups to be considered explicitly.

2.2 Basic set-up and equilibrium conditions

In order to focus on the role of heterogeneity and product differentiation, we specify preferences in considerable detail, and we use a network that is simple yet permits varying degrees of differentiation of trip conditions. We omit from our model a number of practical considerations which would affect policy conclusions for any specific facility. We do not include the costs of toll collection or the many possible sources, besides travel delays due to congestion, of differences between price and marginal cost – for example taxes, maintenance costs, accident costs, or air pollution. We treat user preferences for travel as exogenous rather than derived, and capacities as given. Finally, we do not examine the political economy or industrial organization of public and private operation of highways; rather, we use “public” and “private” as shorthand for second-best optimization and revenue maximization, respectively. This means, of course that “public” operation wins any showdown by definition; but the interesting questions we explore are by how much, and depending on what factors?

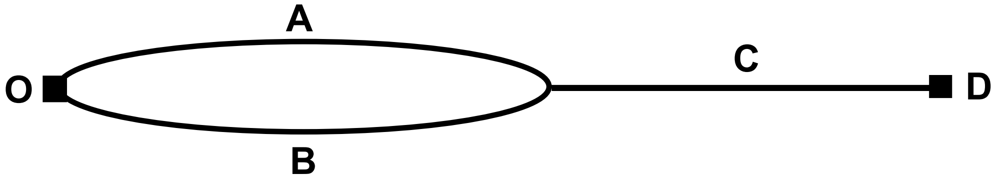


Figure 1. The network considered

The network is shown in Figure 1. There is just one origin-destination pair, OD, connected by two routes: AC (consisting of links A and C) and BC (consisting of links B and C). The travel time on link L is a non-decreasing function of the level of use: $T_L(N_L)$ with $T_L'\geq 0$ (primes are used to denote derivatives). Link L may have a toll, τ_L. Because there are three links but only two routes, there is one redundant toll: a constant can be subtracted from τ_A and τ_B, and added to τ_C, without affecting the price of either route. For convenience, we normalize τ_C to zero except when we wish to require the prices of the two routes to be equal, in which case we normalize $\tau_A=\tau_B=0$ and allow τ_C to represent the single uniform price.

We consider a continuum of (exogenous) values of time, α. For a traveller with value α, the travel cost on a link is $\alpha \cdot T_L$. The density function of value of time across users is denoted N_α; that is, there are $d\alpha \cdot N_\alpha$ users within an infinitesimally small range $[\alpha,\alpha+d\alpha]$. However, this density is endogenous due to the fact that for each value of time, there is a downward-sloping inverse demand function $D_\alpha(N_\alpha)$. Figure 2 shows an example of the resulting demand surface, depicting the one used in the numerical model in Sections 3 and 4.

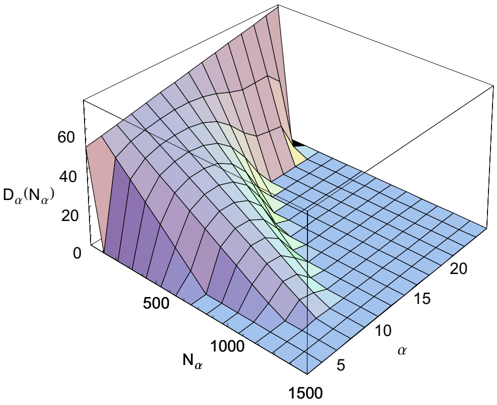


Figure 2. An inverse demand surface

It is worth noting that the correlation between value of time and income is actually far from perfect. Two of the most striking findings from field experience with parallel-route pricing are that many people who use the priced lanes do so intermittently, and that the ranges of incomes using the priced and the free lanes overlap considerably (Sullivan, 1998; Parkany, 1999). Thus, we caution against the temptation to think of the value-of-time distribution as representing the income distribution.

We now consider user equilibrium. Let $N_{\alpha L}$ and $N_{\alpha R}$ be the density functions of users of links L and routes R. The equilibrium conditions are the complementary slackness conditions of Wardrop (1952):

$$N_{\alpha R} \cdot (P_{\alpha R} - D_\alpha) = 0 \tag{1a}$$

$$N_{\alpha R} \geq 0 \tag{1b}$$

$$P_{\alpha R} - D_\alpha \geq 0 \tag{1c}$$

for every α and for both routes R, where $P_{\alpha R}$ is the 'full price' of using route R, defined as:

$$P_{\alpha R} \equiv \alpha \cdot (T_L + T_C) + \tau_L + \tau_C, \quad \{L,R\} = \{A,AC\},\{B,BC\} \tag{1d}$$

These equations state that type-α users will use only the route(s) that have least cost to them, or will not travel at all if all routes have costs exceeding their willingness to pay. The two routes are thus assumed to be perfect substitutes, which makes the model not directly applicable to the case of two competing travel modes. This, instead, would either require an adjustment to the model either on the demand side – *e.g.*, treating the two routes as imperfect substitutes – or on the cost side – *e.g.*, adding a generalized cost penalty to one of the two alternatives, the value of which may differ over individuals.

Formally, we must proceed differently in solving equations (1) depending on whether or not $\tau_A=\tau_B$. If $\tau_A=\tau_B$, positive use can occur on both roads only if travel times are equal, since otherwise all users would choose the road with the lower travel time. In that case, we need an additional condition to obtain a unique equilibrium. The one we choose is that $N_{\alpha A}/N_{\alpha B}=N_A/N_B$ for every α, which may be viewed as a mixed strategy Nash equilibrium.[1] This is a perfectly pooled equilibrium, which we can analyse by merging links A and B into a single link, D, whose travel time is simply a function $T_D(N)$ of total traffic N.[2]

When $\tau_A\neq\tau_B$, both routes can be used only when $T_A\neq T_B$, and there is a separated equilibrium. More precisely, for both routes to be used, $sign\{\tau_B-\tau_A\}=sign\{T_A-T_B\}$ is required. The difference in full price for user α can be written as $(\tau_A+\alpha\cdot T_A)-(\tau_B+\alpha\cdot T_B)$; therefore the critical value of time α^* for which drivers are indifferent between both routes is:

$$\alpha^* = \frac{\tau_B-\tau_A}{T_A-T_B} \tag{2}$$

It is easily checked that, when $\tau_A<\tau_B$, link A is more attractive for all drivers with $\alpha<\alpha^*$, and link B for all drivers with $\alpha>\alpha^*$. Drivers with a relatively low value of time thus only use the link with the lower toll, and similarly for a relatively high value of time and the higher toll.

To complete the model, the following identities are added:

$$N_{\alpha C} = N_{\alpha A} + N_{\alpha B} \tag{3}$$

$$N_L = \int_{\alpha_{\min}}^{\alpha_{\max}} N_{\alpha L}\,\mathrm{d}\alpha \tag{4}$$

where $\alpha_{\min}$ ($\alpha_{\max}$) gives the minimum (maximum) value of time in the population. In the case where $\tau_A<\tau_B$, (4) implies:

[1] The reason that an extra condition is needed is that, although we know that all drivers with all relevant values of time *could* be present on both routes, they do not have to be. For example, with identical routes A and B, one of the possible equilibria satisfying (1) would be if all drivers with a value of time smaller than or equal to the median α were on the one route, and the rest on the other route. By contrast, in the symmetric Nash equilibrium every user plays the same strategy, which is the best strategy given that all other users play that same strategy. That strategy is to choose link A with a probability $p_A=N_A/(N_A+N_B)$. N_A and N_B are then determined from the condition that $T_A(N_A)=T_B(N_B)$. An alternative interpretation, which we owe to an anonymous referee, would be to assume that there are idiosyncratic preferences for route that are distributed independently ofα, in which case our equilibrium condition characterizes the limit as they become negligible.

[2] This function is chosen to be consistent with an allocation $N=N_A+N_B$ such that $T_A(N_A)=T_B(N_B)=T_D(N)$ is satisfied. It has the property that:
$$\frac{1}{T'_D}=\frac{1}{T'_A}+\frac{1}{T'_B}\,.$$

$$N_A = \int_{\alpha_{\min}}^{\alpha^*} N_{\alpha A} \, \mathrm{d}\alpha \tag{4a}$$

$$N_B = \int_{\alpha^*}^{\alpha_{\max}} N_{\alpha B} \, \mathrm{d}\alpha \tag{4b}$$

2.3 Tolling regimes

Despite its simplicity, the network considered allows us to consider a wide variety of toll regimes. Ignoring the possibility of private or mixed duopolies, as considered in De Palma and Lindsey (2000), we still have six possibilities in addition to no tolls. These six consist of either private or public tolls on the entire network, on one of the parallel links (labelled B without loss of generality), or on the serial link. Table 1 summarizes these tolling regimes.

Abbreviation	**Description**	**Tolls on:**
NT	No Tolls	–
Public tolling:		
FB	First-Best tolls on the full network	A and B
SBPL	Second-Best toll on one of the Parallel Links	B
SBSL	Second-Best toll on the Serial Link	C
Private tolling		
PF	Private tolls on the Full network	A and B
PPL	Private toll on one of the Parallel Links	B
PSL	Private toll on the Serial Link	C

Table 1. Tolling regimes

We assume that for the public operator, the objective is to maximize social welfare, W. This is defined as total benefits minus total costs, where benefits are based on the Marshallian measure, which is in this case equal to the volume below the inverse demand surface of Figure 2.[3] When $\tau_{A<}\tau_B$, welfare can be written as:

$$\begin{aligned} W = &\int_{\alpha_{\min}}^{\alpha_{\max}} \int_0^{N_\alpha} D_\alpha(n) \, \mathrm{d}n \, \mathrm{d}\alpha - \int_{\alpha_{\min}}^{\alpha^*} N_{\alpha A} \cdot \alpha \cdot T_A\left(\int_{\alpha_{\min}}^{\alpha^*} N_{aA} \, \mathrm{d}a \right) \mathrm{d}\alpha - \int_{\alpha^*}^{\alpha_{\max}} N_{\alpha B} \cdot \alpha \cdot T_B\left(\int_{\alpha^*}^{\alpha_{\max}} N_{aB} \, \mathrm{d}a \right) \mathrm{d}\alpha \\ &- \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha \cdot \alpha \cdot T_C\left(\int_{\alpha_{\min}}^{\alpha_{\max}} N_a \, \mathrm{d}a \right) \mathrm{d}\alpha \end{aligned} \tag{5a}$$

where α^* is given by (2), and where we have used that $N_{\alpha B}=0$ for $\alpha<\alpha^*$ and $N_{\alpha A}=0$ for $\alpha>\alpha^*$.

When $\tau_A=\tau_B=0$, so when only τ_C is used and the merged link D is considered, the objective is:

[3] It would be possible, in the current framework, to define a social welfare function reflecting distributional concerns. The current specification, however, seems a natural and useful benchmark, one which identifies the distributional effects of a policy that in itself has no redistributional objectives (*i.e.*, in absence of congestion, hypothetical individual-specific tolls would all be equal to zero).

$$W=\int\limits_{\alpha_{\min}}^{\alpha_{\max}}\int\limits_{0}^{N_\alpha} D_\alpha(n)\,\mathrm{d}n\,\mathrm{d}\alpha-\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha\cdot\alpha\cdot T_C\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)\mathrm{d}\alpha-\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha\cdot\alpha\cdot T_D\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)\mathrm{d}\alpha \tag{5b}$$

With private tolling, the objective is assumed to be the maximization of total toll revenues, R. Using dummies δ_L to denote whether a toll is in operation on link L, this objective function can be written as:

$$R=\delta_A\cdot\tau_A\cdot\int\limits_{\alpha_{\min}}^{\alpha^*} N_{\alpha A}\,\mathrm{d}\alpha+\delta_B\cdot\tau_B\cdot\int\limits_{\alpha^*}^{\alpha_{\max}} N_{\alpha B}\,\mathrm{d}\alpha+\delta_C\cdot\tau_C\cdot\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha\,\mathrm{d}\alpha \tag{6}$$

These objective functions can be turned into the relevant Lagrangians by adding terms representing the constraints, of which there is a continuum for all values of α, each representing the requirement of free user choice as embodied in equation (1). These constraints are identical for the public and the private operator. When $\tau_A\neq\tau_B$, the constraints are represented by the following Lagrangian terms:

$$\begin{aligned}&+\int\limits_{\alpha_{\min}}^{\alpha^*}\lambda_{\alpha A}\cdot\left(\alpha\cdot T_A\left(\int\limits_{\alpha_{\min}}^{\alpha^*} N_{aA}\,\mathrm{d}a\right)+\alpha\cdot T_C\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)+\delta_A\cdot\tau_A-D_\alpha(N_\alpha)\right)\mathrm{d}\alpha\\&+\int\limits_{\alpha^*}^{\alpha_{\max}}\lambda_{\alpha B}\cdot\left(\alpha\cdot T_B\left(\int\limits_{\alpha^*}^{\alpha_{\max}} N_{aB}\,\mathrm{d}a\right)+\alpha\cdot T_C\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)+\delta_A\cdot\tau_B-D_\alpha(N_\alpha)\right)\mathrm{d}\alpha\end{aligned} \tag{7a}$$

where $\lambda_{\alpha L}$ is the Lagrangian multiplier for the constraint (1a) for those values of α having positive $N_{\alpha L}$. When $\tau_A=\tau_B=0$, the Lagrangian term is:

$$+\int\limits_{\alpha_{\min}}^{\alpha_{\max}}\lambda_\alpha\cdot\left(\alpha\cdot T_C\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)+\alpha\cdot T_D\left(\int\limits_{\alpha_{\min}}^{\alpha_{\max}} N_a\,\mathrm{d}a\right)+\tau_C-D_\alpha(N_\alpha)\right)\mathrm{d}\alpha \tag{7b}$$

For each of the schemes considered, the optimal toll can then be found by maximizing the appropriate Lagrangian, which is computed by adding the relevant objective function (5a), (5b) or (6) to either (7a) or (7b). Unfortunately, we failed to find a closed form analytical solution for three of the more interesting cases, namely SBPL, PF and PPL, due to a discontinuity at α^*. For the cases where we did find closed form analytical solutions (FB, SBSL, PSL), the tax rules are rather straightforward generalizations of those applying with only a single value of time, as given, for instance, in Verhoef *et al.* (1996). We therefore relegate the derivation and discussion of the first-order conditions, and tolls for the cases where they could be solved, to a separate appendix that is available from the authors upon request. For the other cases, we devised a numerical algorithm which directly finds the toll maximizing the objective function.

3 A numerical model: the base case

In this section we present a numerical model to assess and illustrate the economic properties of these tolling regimes.

3.1 The cost side

The cost side of the model consists of link travel-time functions, describing travel times T_L as a function of usage N_L. The functional form used is

$$T_L = T_{FL} + T_{FL} \cdot b \cdot \left(\frac{N_L}{K_L} \right)^k \tag{8}$$

where b and k are parameters, T_{FL} is the free-flow travel time on link L, and K_L is conventionally called the 'capacity' of link L. (Because there is no maximum flow for this type of congestion function, 'relative capacity' would actually be a better term.) With b=0.15 and k=4, as assumed throughout the simulations, this function is the well-known Bureau of Public Roads formula. For the base-case of the model, intended to represent realistic morning peak traffic situations, it is assumed that link B has 25%, and link A 75%, of their joint capacity (8 000 vehicles per hour), and that they have equal free-flow travel times. This could correspond to a four-lane highway with one lane subject to tolling.[4] Furthermore, it is assumed that link C has the same capacity as A and B combined, and that the free-flow travel time on C is 25% of the total free-flow travel time. The latter is set at 30 minutes, which is a reasonable average for (one-way) commuting trips using main highways in The Randstad and many other urban areas. Table 2 summarizes these base-case parameters.

	Link A	**Link B**	**Link C**
b	0.15	0.15	0.15
k	4	4	4
$\mathbf{T_{FL}}$ (hr)	0.375	0.375	0.125
$\mathbf{K_L}$ (veh/hr)	6000	2000	8000

Table 2. The base-case parameters for the cost functions

3.2 The demand side

The base-case inverse demand surface is depicted in Figure 2 above. It is assumed that for every value of time, the demand function is linear over the relevant range (between the lowest and highest use levels considered), and can thus be written as:

$$D_\alpha = m_\alpha - d_\alpha \cdot N_\alpha \tag{9}$$

Functions m_α and d_α are calibrated to achieve three objectives: (1) a weighted demand elasticity (over all α) of –0.4 in the NT-equilibrium;[5] (2) travel times in the base-case no-toll regime approximately double the free-flow travel time of 0.5 hours implied by Table 2; and (3) a distribution of values of time in the NT-equilibrium similar to that found in an earlier

[4] A common approximation for freeway capacities is 2000 vehicles per hour per lane. For more detailed discussions of capacity, see Small (1992, pp. 61-68) or Transportation Research Board (1998).

[5] We calculated this elasticity assuming that the variable monetary cost of the trip is DFl 12 (6 litres of fuel times DFl 2) in the NT-equilibrium; *i.e.* both the demand plane and the cost level were shifted upwards by DFl 12 to calculate this generalized price elasticity. These variable monetary costs, however, are assumed to be constant over the various tolling regimes considered, and so are ignored in the simulations.

stated preference study for the Dutch Randstad area (Verhoef *et al.*, 1997). [6] The following functions achieves these objectives:

$$m_\alpha = 50 + \alpha \tag{10a}$$

$$d_\alpha = \frac{-0.0434783}{-0.713714 + 0.705429\cdot\alpha - 0.0950357\cdot\alpha^2 + 0.00468093\cdot\alpha^3 - 0.000079\cdot\alpha^4} \tag{10b}$$

The values of time α considered in the simulations range between a minimum of DFl 1.2 and a maximum of DFl 23.8 per hour, with a weighted average value of DFl 9.08 in the base-case described below.[7]

3.3 General results: base case

Table 3 presents results for the various tolling regimes using these base-case parameters.

The first-best (FB) policy produces substantial service differentiation, with travel time 0.204 hours less on link A than on link B. But this policy also produces some surprises. First, welfare is maximized when the facility with the larger capacity (link A) gets the premium service, in contrast to what one might expect from the analogy of first-class service on airplanes and trains. Second, although overall demand is reduced (by 14 percent) compared to the no-toll (NT) regime, congestion on the lower-priced link is actually worse than with no tolls. In the base case, this paradox disappears when the portion of the trip on link C is taken into account – all users then receive faster service in the first-best policy than in the no-toll policy. However, Section 4.2 presents an example where even the *total* travel time for the lower-priced link actually *increases* with optimal tolling. Apparently product differentiation is quite a strong motivation here, calling for a rather low optimal service quality for the segment of the population with lower values of time.

A third surprise is how small the toll differentiation is: the tolls on links A and B differ from each other by only 15 percent. There are two reasons for this. First, although the average value of time of link-B users is smaller, there are more of them (per unit of capacity), and these two effects work in opposite directions on the externality cost of a trip. Second, link-B users interact with higher-value-of-time users on the shared link C, which further increases the marginal cost they impose.

Given the limited degree of optimal toll differentiation, it is not too surprising that the uniform toll policy, SBSL, performs nearly as well in terms of efficiency. It achieves 92 percent of the maximum possible welfare gains, at a uniform toll quite close to the higher of the differentiated FB tolls. Although not shown in the table, one can readily see that most or all low-value-of-time users are worse off with a uniform toll policy than with FB because the

[6] The dashed line in the left panel in Figure 4 below shows this distribution (approximately). It was derived using 961 (93%) of the 1027 respondents for whom a value of time could be calculated: the 7% with the highest values of time were discarded so as to keep a compact distribution. A simple fourth-order polynomial was fitted on the histogram of values of time, split in 12 categories of size DFl 2 (R^2=0.975). Because of the selection, the average value of time used here is DFl 9.08, as opposed to DFl 10.92 for the full set of respondents.

[7] The exchange rate of the Dutch guilder in late 1999 was approximately DFl 2.2≈€1≈\$1.

uniform policy forces them to accept a higher service quality and higher price than they prefer. (We discuss the distributional effects at greater length in the next subsection.)

By contrast, when only one of the parallel links can be priced (SBPL), namely the one with 25% of total capacity, less than one-fourth of the possible welfare gains are achieved. Consistent with the studies reviewed earlier, the second-best toll is much lower than first-best, only DFl 3.31. The reason is that now, welfare gains on link B from raising its price have to be traded off against welfare losses of spill-over traffic on link A, as described in Verhoef *et al.* (1996). Nevertheless there is a surprise for second-best policy as well: as we shall see in Section 4.1, more than twice as great a welfare gain could be achieved with second-best parallel pricing by pricing the high-capacity section of the road instead of the low-capacity section. This result recalls the fact that with first-best pricing, it was the higher-capacity road that received the higher price.

	NT	FB	SBPL	SBSL	PF	PPL	PSL	Free-flow
Rel. use A[a]	1	0.812	1.046	0.854	0.498	1.117	0.527	
Rel. use B[a]	1	1.003	0.831	0.854	0.616	0.533	0.527	
Rel. use C[a]	1	0.860	0.992	0.854	0.527	0.971	0.527	
α* (DFl/hr)	-	5.919	12.996	-	6.138	15.265	-	
Travel time A (hr)	0.729	0.529	0.798	0.563	0.397	0.926	0.402	0.375
Travel time B (hr)	0.729	0.733	0.544	0.563	0.426	0.404	0.402	0.375
Travel time C (hr)	0.243	0.189	0.239	0.188	0.134	0.230	0.134	0.125
Toll A (DFl)	0	9.50	0	0	27.83	0	0	
Toll B (DFl)	0	8.29	3.31	0	27.65	7.98	0	
Toll C (DFl)	0	0	0	9.38	0	0	27.80	
Toll revenues (DFl)	0	99606	8703	101 484	185 603	13 468	185 487	
ω[b]	0/0	1	0.229	0.920	-2.599	-0.272	-2.623	

[a] Use relative to that in NT scenario. The latter is: 9501 on link A, 3167 on link B, 12669 on link C. As discussed in the text, the fact that these exceed link 'capacity' is entirely consistent with the power-law model of equation (8). The NT-use levels are probably best thought of as covering a peak period of about 1.5 hours.

[b] Index of relative efficiency: increase in social welfare (compared to NT) as a share of the increase in social welfare (compared to NT) obtained in the first-best optimum. The latter increase is DFl 16743, or DFl 1.32 per user in the NT equilibrium.

Table 3. Performance of the various toll regimes for the base-case parameters

We now turn to revenue-maximizing tolling by a private operator. Unrestricted tolling extracts a high social cost: welfare is substantially reduced compared to no tolls. The absolute loss is 27 percent of the maximum achievable gain when only link B can be priced (PPL), and 260 percent of that when both links can be priced (PF). The tolls are much higher than the corresponding second-best or first-best optimal tolls: more than twice as high for PPL as for SBPL, and around three times as high for PF as for FB. This is consistent with earlier results, although it is not necessarily the case that revenue-maximizing tolling would always lead to a decrease in social welfare (see Verhoef *et al.*, 1996, and De Palma and Lindsey, 2000). From the point of view of the private operator, being restricted to pricing only the smaller part of the roadway is very costly: it takes away 93 percent of the potential revenues.

There is a surprise in private tolling, as well: the toll differentiation in unrestricted private pricing (policy PF) is negligible. The reason is that the monopoly toll level has reduced total traffic by so much (47 percent) that nearly all congestion is eliminated, making significant service differentiation impossible.

3.4 Distributional results: base case

The numerical simulations allow us to calculate the distribution of welfare effects of the various tolling regimes across people with different values of time. In this sub-section, we present such results for just two public tolling regimes: FB and SBPL. (We ignore SBSL because of its close similarity to FB, and private tolling regimes for reasons of space.)

Figure 3 shows the changes in total and average consumer surplus by value of time, compared to the NT regime. For each value of time, the total change in consumer surplus is given by the change in generalized cost (including toll) for those users who remain on the road, plus the change in surplus for those who leave the road due to tolling.[8] The average change is the total change divided by the level of use in the NT-regime. The figure shows that under first-best tolling (solid line), the average loss in surplus is smaller for people with higher value of time. This result arises, of course, because the price increase is offset by a travel-time decrease, which is valued more by such people. The kink at α^* is due to the fact that the ratio of toll paid to travel time gained differs across the two parallel links.

Figure 4 shows the levels of road use by value of time for the two policies.[9] Since the usage under SBPL is very close to the NT use levels, the dashed line in the left panel also gives an accurate impression of the original distribution of values of time used. Relative use, in the right panel, is defined in the same way as in Table 3. The right panel in Figure 4 follows the same general pattern as that in Figure 3, simply because the change in consumer surplus is closely related to the change in full price, which in turn determines the change in usage.

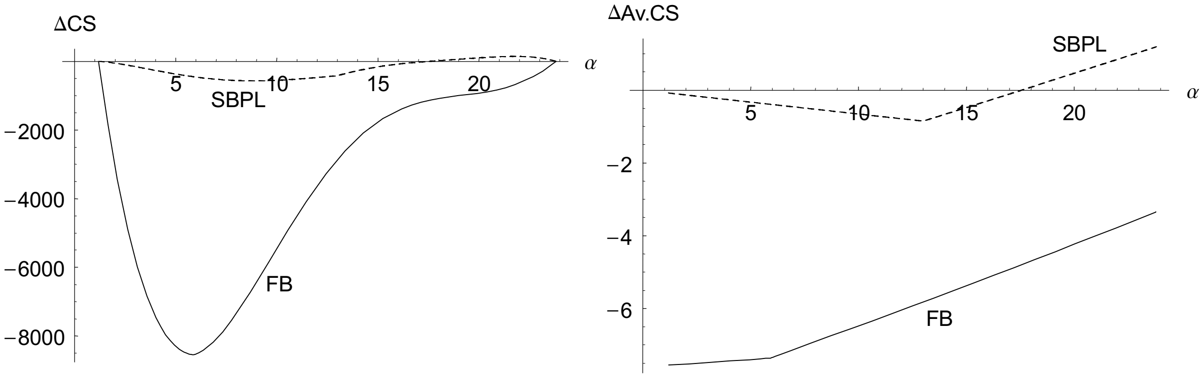


Figure 3. Total (left panel) and average (right panel) change in consumers' surplus, compared to NT, before tax recycling

[8] Units are total consumer surplus in DFl per unit interval of value of time (the latter in DFl/hr).

[9] Units are numbers of users per unit interval of value of time (the latter in DFl/hr).

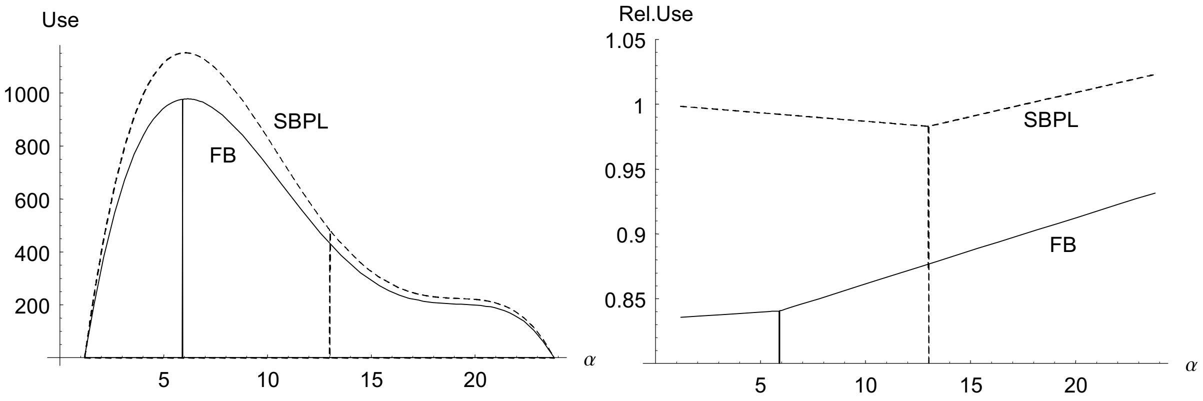


Figure 4. Total (left panel) and relative (right panel) use; the vertical lines indicate α^*

For the SBPL regime (dashed lines in Figures 3 and 4), it is users near the critical value of time who suffer the largest average losses (Figure 3, right panel). The reason is that imposing second-best tolling on link B improves travel time for people taking that route, while worsening it for those taking the other route. It is users near the critical value of time who benefit least from the travel-time reduction on the link with the toll or who suffer most from the travel-time increase on the unpriced link. One could say that the policy caters to the more extreme users, leaving those in the middle disadvantaged. However, none of the consumer-surplus changes are very large, the biggest loss amounting to just DFl 0.85 (US$ 0.40) per trip. These changes are much smaller than under FB, and users with the highest values of time even benefit directly from SBPL. This helps explain why parallel-route pricing appears to be more politically acceptable than first-best tolling.

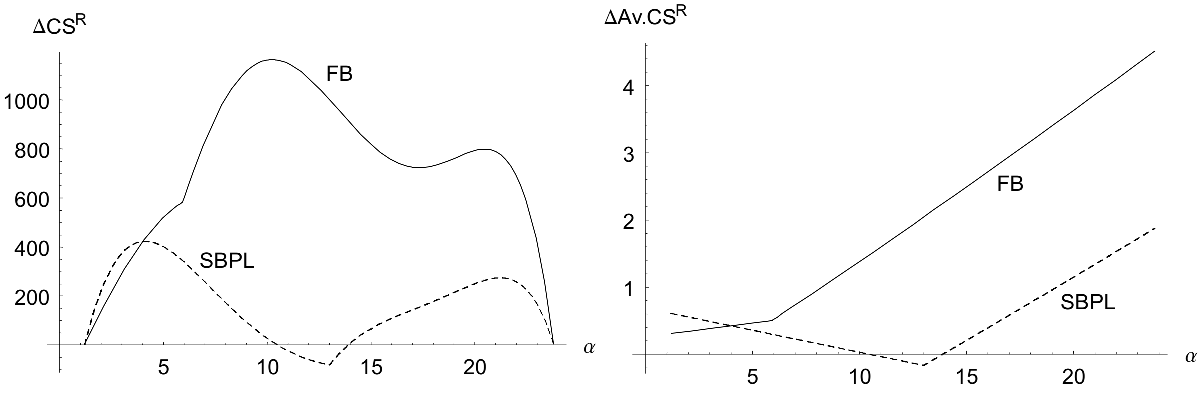


Figure 5. Total (left panel) and average (right panel) change in consumers' surplus, compared to NT, after non-differentiated tax recycling

As expected, the relative attractiveness of the FB and SB regimes may be reversed by redistributing the toll revenues. Figure 5 shows the changes in total and average consumer surplus after applying the simplest possible tax-recycling scheme: an equal redistribution to all initial road users. This simply means an upward shift of each of the curves shown in the right panel of Figure 3. Because revenue is much larger under first- than second-best pricing, the solid curve is shifted up by much more than the dashed curve, so that first-best pricing is now better for all but the very lowest-value-of-time users. In fact, with this toll redistribution, first-best pricing is welfare-enhancing for every user compared to no tolls.[10] When these average surplus changes are multiplied by the level of usage shown in the left panel of Figure 3, the result is the curious double-peaked distribution of change in total consumer surplus shown on the left panel of Figure 5.

It can be expected that the distribution of changes in average consumer surplus under private tolling will show patterns comparable to those shown in the right panel of Figure 3, for the same reasons as outlined above. Of course, the absolute welfare losses will be larger; and since all the private tolling regimes generate net welfare losses, no redistribution could make everyone better off. In practice, private tolls are likely to be restricted by additional regulations, such as rate-of-return caps or direct price regulation; our results provide support for some such restriction.

4 The impact of some key parameters on the performance of the tolling regimes

In this section, we assess the impact of key parameters upon the relative performance of the different tolling regimes. We start by varying parameters related to the cost side of the model, namely the capacities and lengths of the links. We next consider the impact of two characteristics of the demand side: the (weighted) demand elasticity, and the type of distribution of values of time.

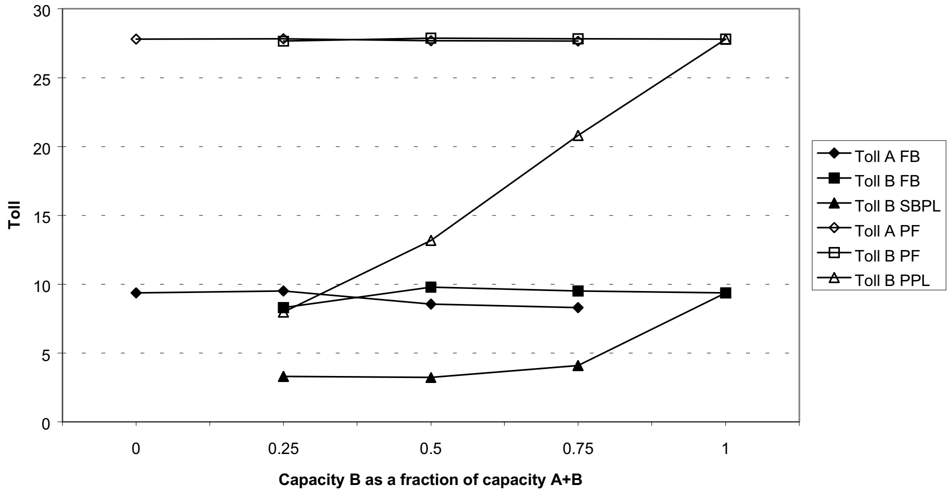
4.1 Varying the relative capacities of the two parallel links

We first consider the impacts of increasing the fraction of the highway subject to tolling, keeping total joint capacity of links A and B fixed. Figures 6a and 6b show the optimal tolls and the relative efficiency ω, respectively, if the capacity of B is increased, in 25% steps, from 0% to 100% of the joint capacity (recall that the base-case is at 25%).[11] Unsurprisingly, the greatest impacts of capacity allocation occur for those policies constraining a parallel link, namely SBPL and PPL. For public tolling, greater capacity of B makes the second-best policy (SBPL) relatively more efficient, because the importance of the unpriced substitute is diminished; at 75% capacity, nearly half the possible welfare gains are realized. These results suggest that from an efficiency viewpoint, and taking into account heterogeneity of users, one public ‘free-lane’ on a four-lane highway is preferable to one public ‘pay-lane’. In other words, it would be better to think of a priced system with a ‘life-line’ type of unpriced service available to those who most need it, rather than an unpriced system with special premium service for the elite.

[10] A similar result in a mode choice model was observed by Small (1983).

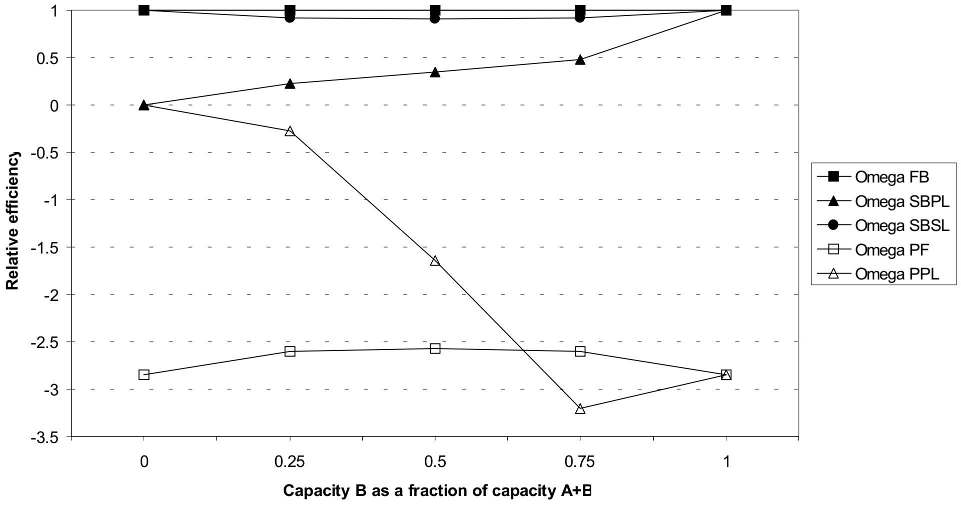
[11] On the left-hand side of these figures, therefore, SBPL and PPL are identical to NT, because no capacity is tolled; whereas on the right-hand side, they are identical to SBSL and PSL, respectively, because all the capacity is tolled. At both extremes, toll differentiation is impossible, so FB is identical to SBSL and PF to PSL.

The opposite holds for private tolling. The private operator, ignoring the efficiency aspects of spill-overs, increases the toll on the parallel route rapidly as its relative capacity increases. This substantially increases the relative welfare losses from PPL, at least up to 75% of capacity. Oddly, once at least 75 % of capacity if allocated to a private operator it is better that all capacity be so allocated; this counterintuitive result, also found by Verhoef *et al.* (1996), occurs because full control of the network avoids inefficient route splits. Finally, the finding of relatively limited price differentiation under FB pricing remains intact.[12]



Note: For graphical clarity, tolls for SBSL and PSL, being close to those for FB and PF, are surpressed.

Figure 6a. Varying the relative capacities of the two parallel links: tolls



Note: For graphical clarity, relative efficiency for PSL, being close to that for PF, is surpressed.

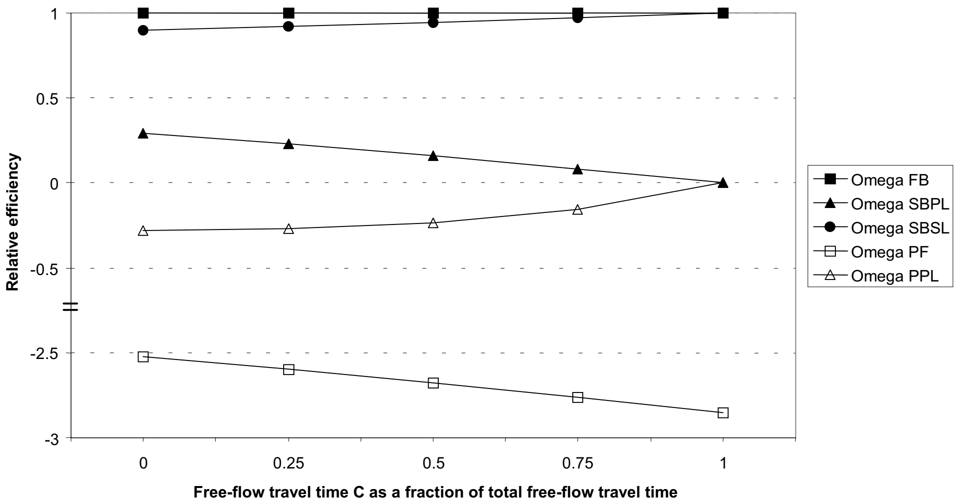
Figure 6b. Varying the relative capacities of the two parallel links: relative efficiency

[12] The FB scheme will have differentiated tolls throughout. The intersection of the lines representing τ_A and τ_B in Figure 6a results from graphical interpolation only, and is near the point where it becomes more efficient to charge a higher toll on link B than on link A, instead of the other way around. A similar argument holds for PF, and also for FB in Figure 8a below.

Together, these results contradict the idea that efficiency always increases monotonically with the degree of privatization. If one insists on a system with both unpriced and priced alternatives, it is more efficient to allow a public operator to price most of the capacity, but a private operator only a small portion of it instead.

4.2 Varying the relative length of the serial link

Most studies ignore the fact that users of two parallel routes will usually not be completely isolated but will share some links, upstream or downstream of the split road section. A similar interaction is likely for other congestible facilities such as telephone trunk lines. Figure 7 shows how this feature affects the relative efficiency of the various tolling regimes considered. Along the horizontal axis, the relative length of the serial link C – represented by its free-flow travel time – is increased in 25% steps, keeping the total free-flow travel time constant. Note that when the relative length of C has become 1, the parallel links effectively disappear so FB becomes identical to SBSL and PF to PSL.



Note: For graphical clarity, relative efficiency for PSL, being close to that for PF, is surpressed.

Figure 7. Varying the relative length of the serial link: relative efficiency

As the relative length of the serial link increases, second-best toll differentiation becomes less viable, so both the public and private tolls on the parallel link fall (even per kilometer) and approach zero. As a result, the relative efficiencies of these regimes approach zero as well. From a societal point of view, this is bad news in the case of the public toll and good news in the case of the private toll. This finding suggests that the relative efficiency gains or losses from parallel route pricing are likely to be overstated in studies ignoring the existence of serial, common used links. For instance, ω_{SBPL} is equal to 0.29 when link C has zero length, but falls to 0.16 when C is equally long as A and B and to 0.08 when C is 3 times as long. Similarly ω_{PPL} changes from –0.28 to –0.16 over the same interval. A similar pattern would be found if instead of increasing the relative length of the serial link, its relative capacity were decreased.

The base-case result that FB tolling actually increases congestion (not shown in diagram) on link B, compared to no toll, remains true when link C has zero length. Therefore, product differentiation alone can cause optimal pricing to increase the travel times of lower-value-of-time users, compared to no pricing. Of course, since FB pricing leads to a potential Pareto improvement, it remains true that these users could be made better off by some lump-sum redistribution of revenues. In practice, this result raises a strong political barrier to optimal pricing – qualified, however, by a reminder that low-value-of-time users are not necessarily the same people from one day to the next.

4.3 Varying the relative length of the parallel links

It is of course possible that the two parallel links are not lanes of the same highway, but are separate roads instead. In that case, the parallel links need not have equal free-flow times. An example is a toll road that parallels an arterial with at-grade intersections.

Figures 8a and 8b show how the tolls and the relative efficiency change if the free-flow travel times on links A and B are changed in opposite directions. The base case is now in the centre of the diagram. As the tolled link B becomes shorter when moving to the left, it requires a relatively higher marginal external cost or a higher toll in order to equalize marginal private costs on the two links. The tolls for link B therefore have the tendency to increase when moving to the left, and to decrease – even becoming negative – when moving to the right.[13]

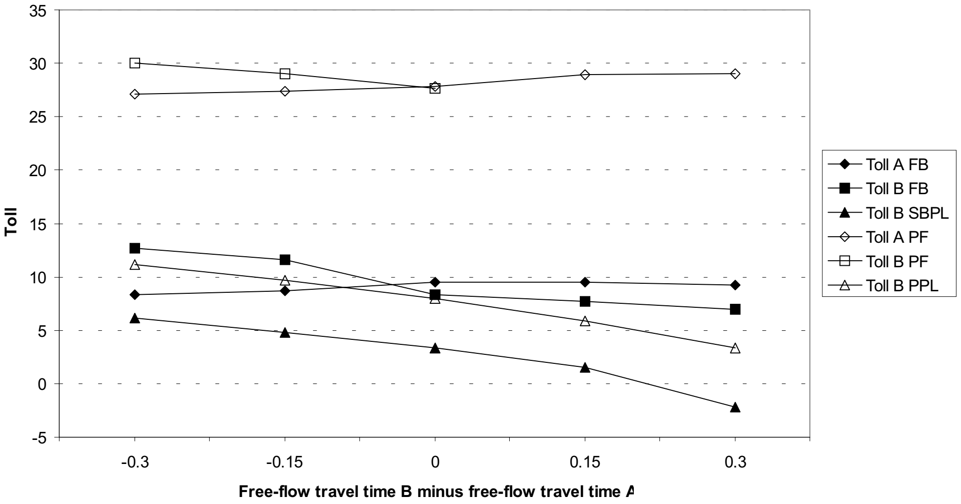
Toll differentiation naturally becomes more important when the two links are of different lengths: that is, when products vary in more dimensions that just amount of congestion.[14] Consequently, the potential welfare gain from fully optimal pricing (FB) increases the more unequal are free-flow travel times. Furthermore, when link B is shorter (left side of Figure 8), there is less disadvantage to being unable to price link A, so the relative welfare gain from SBPL also rises – to just over 50% at a 0.3 hours free-flow travel time difference. A similar result is also found by Verhoef *et al.* (1996, Figures 2 and 5) and Liu and McDonald (1999, Table 1 and p. 187).

Another consequence is that equal prices on the parallel links become increasingly unsatisfactory as the links become more unequal. As a result, ω_{SBSL} decreases rapidly when

[13] For SBPL, there will be a specific combination of parameters for which the second-best optimal toll is actually zero (this combination is not among the plotted points). This requires link B to be longer than link A. The two forces governing the second-best optimal level of the toll – reducing overall traffic, and diverting traffic from link A, where marginal external costs are higher, to link B – then exactly off-set each other. In this case, ω_{SBPL} is zero. Beyond that point, as Figure 8a shows, a subsidy is welfare improving.

[14] This is illustrated by a curious result which appears when free-flow travel time is 0.3 hours less on A than on B. This case produces substantial price differentiation under FB pricing, as seen at the far right of Figure 8a. But the second-best serial pricing for this case (SBSL, not shown in the diagram) produces a toll that is lower than either FB toll – in contrast to all other simulations, where the serial toll lies between the two FB tolls. The reason appears to be that SBSL pricing provides such an inferior option for high-value-of-time users, relative to FB, that it substantially reduces their proportion in the overall composition of traffic. This lowers the marginal cost imposed by any driver sufficiently to result in a second-best toll lower even than the lowest of the two first-best tolls.

moving to the edges of the diagram. The shorter link tends to get the higher price, and carries the higher value of time travellers, both for FB and for PF.[15]



Note: For graphical clarity, tolls for SBSL and PSL, being close to those for FB and PF, are surpressed.

Figure 8a. Varying the relative lengths of the parallel links: tolls

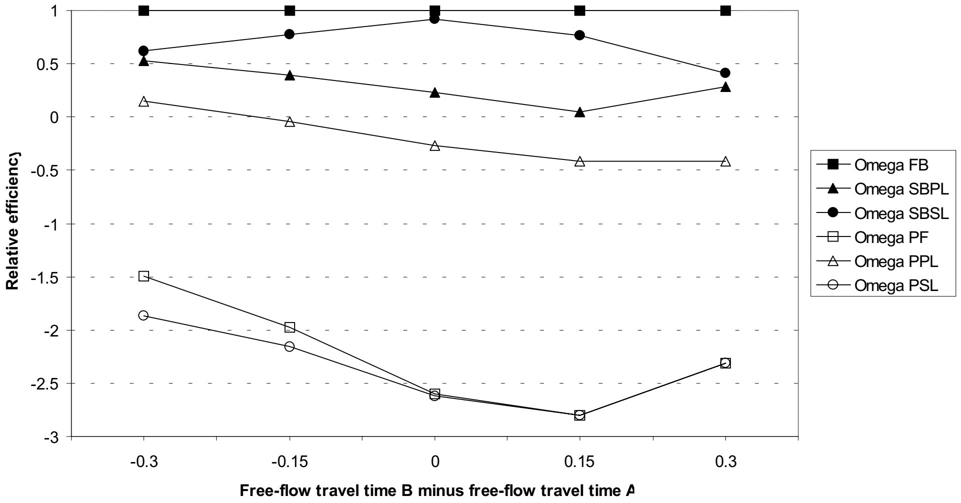


Figure 8b. Varying the relative lengths of the parallel links: relative efficiency

The relative efficiency of PPL declines somewhat more strongly than that of SBPL when moving to the right. In the range where a subsidy would be welfare enhancing when only link B can be tolled, ω for PPL remains low. It does not decrease any further though, since link B has become relatively so unattractive that the monopolist is quite ‘harmless’. On the far left-

[15] It should be noted that the ω’s are in a sense ‘deflated’ when moving to either side of Figure 8b, since the welfare gain with FB increases, due to growing efficiency gains of toll differentiation. Therefore, the same *absolute* welfare change with any given policy would show as a smaller *relative* welfare change.

hand side, in contrast, we witness an instance of private tolling on link B leading to an efficiency gain.

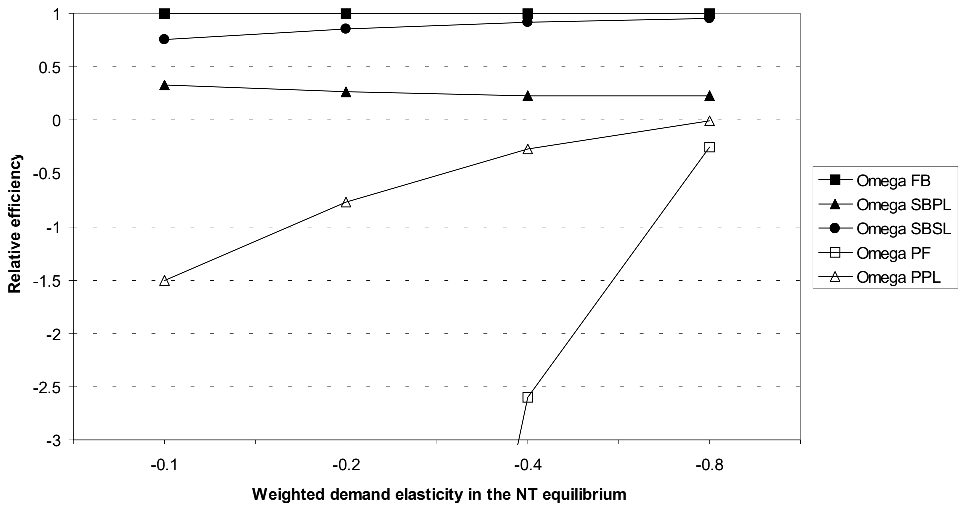
With the other private tolling policies (PF and PSL), the private operator actually has closed down link B at both observations to the right of the base-case by setting the tolls so that link B is not used.

4.4 Varying the overall capacity of the network

Next, we consider the effect of a simultaneous proportional increase of the three links' capacities. We examined the tolls for four values of total capacity: 6000, 8000 (the base case), 10,000, and 100,000 vehicles per hour, all for the same demand surface. The results (not depicted graphically) show that the degree of toll differentiation (in FB and PF) increases with the equilibrium level of congestion. All public tolls, as well as the PPL toll, approach zero as the capacity of the network approaches infinity and congestion vanishes. With PF and PSL, however, the private operator can still extract monopoly profits by tolling, leading to tolls and welfare losses which do not approach zero.[16]

4.5 Varying the total (weighted) demand elasticity

In the next round of simulations, m_α and d_α in equations (10) were changed simultaneously so as to generate different weighted demand elasticities in the NT equilibrium, keeping the total level of road use approximately fixed. (The calculation of demand elasticity is explained in the first footnote to Section 3.2.) Values of approximately –0.1, –0.2, –0.4 (the base case), and –0.8 were produced. Figure 9 shows the effect on relative efficiency.



Note: For graphical clarity, relative efficiency for PSL, being close to that for PF, is surpressed.

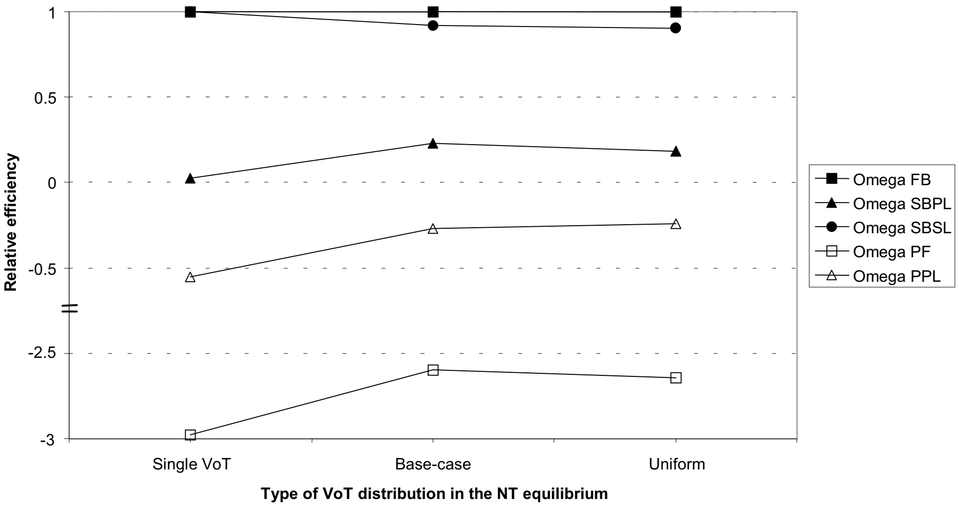
Figure 9. Varying the weighted demand elasticity: relative efficiency

[16] We also used this variation to double-check the logic of our private tolls by confirming that, as expected when congestion is negligible, the monopolist operates at the point where the total demand elasticity (with respect to toll, not full price) is –1.

At a more inelastic demand, the welfare effects of monopolistic pricing become increasingly negative, as is well known from earlier studies (Verhoef *et al.*, 1996). Therefore, for PF and PSL, and to a lesser extent for PPL, ω falls rapidly and at an increasing rate when moving leftwards. A new result is, however, that as demand becomes more inelastic, separation of traffic with different values of time becomes relatively more important for overall efficiency. Therefore, ω_{SBPL} increases and ω_{SBSL} decreases when moving to the left.

4.6 Varying the type of distribution of values of time

Finally, we consider the extent to which the results presented depend on the distribution of values of time. To that end, we redo the base case with two alternative types of distribution: a uniform distribution (which has greater variance of values of time than the base case distribution) and a degenerate distribution with a single value of time. We calibrate on the distribution in the NT equilibrium, since the exact distribution varies between equilibria (see Figure 4). We keep the same weighted average value of time of DFl 9.08 per hour, again in the NT equilibrium; for the uniform distribution, we accomplish this using an interval [1.20,16.96]. The height and price-slope of the demand surface are calibrated to keep total road use and weighted demand elasticity in the NT equilibrium the same as in the base case.



Note: For graphical clarity, relative efficiency for PSL, being close to that for PF, is surpressed.

Figure 10. Varying the type of distribution of values of time: relative efficiency

Figure 10 shows the impacts on relative efficiency. Of course, the significance of toll differentiation disappears with a single value of time;[17] as a result, policies restricted to pricing just one parallel link perform considerably worse than in the base case. Thus ignoring heterogeneity may lead to serious underestimation of the efficiency of parallel link pricing, as suggested also by Small and Yan (2001). Of particular interest, ignoring heterogeneity would lead one to underestimate the relative efficiency of the SBPL policy by a factor of nine (0.025

[17] This is true also of PSL, not shown in the figure, and of PF which, as noted earlier, produces very little toll differentiation even when there is dispersion in values of time.

compared to 0.229 in the base case). This establishes that product differentiation by congestion level is indeed critical to the evaluation of pricing policies that leave parallel roads unpriced. At the other extreme, moving from the base-case to the uniform distribution produces slightly more toll differentiation in the FB case, and thus the second-best policies are slightly worse relatively. These latter differences are small, however, so we conclude that the results of this paper are not sensitive to the exact shape of the value-of-time distribution.

What if an erroneous assumption of a single value of time is carried through to the toll-setting stage? The second-best toll for parallel-route pricing (SBPL, single VoT) is only DFl 0.88, about 27% of the true second-best toll of DFl 3.31 (shown in Table 3). The actual use of this smaller toll when true heterogeneity exists, as in our base case, would lead to a relative welfare gain of ω=0.103. This is 45% of the welfare gain from the correctly calculated toll, which is ω=0.229 (again as shown in Table 3). Therefore, a regulator knowing the average value of time but ignoring its dispersion when setting the toll would lose more than half of the already limited efficiency gains possible from parallel route pricing.

For first-best pricing, in contrast, the predicted optimal toll when ignoring heterogeneity is DFl 9.19, not very different from the truly optimal differentiated tolls of DFl 8.29 and 9.50. The relative welfare gain, applying the former toll, is ω=0.9199; that is, the inefficiency from ignoring heterogeneity is only eight percent. Furthermore, the best one can do with a single toll is ω=0.9203 (the value for SBPL from Table 3). Therefore so long as both parallel links are being priced, the inefficiency from ignoring heterogeneity is almost entirely from adopting uniform pricing, which may actually be optimal once collection costs are accounted for; the further inefficiency from calculating the wrong uniform toll is negligible.

This reconfirms an insight from earlier studies: second-best taxes are not only by definition less efficient than first-best taxes, but in addition are harder to implement optimally because they require more information. First-best tax rules require knowing only the level of marginal external costs in the final equilibrium. The second-best tax rule for parallel-route pricing, as derived for example by Verhoef *et al.* (1996), requires that the regulator also know the demand and cost elasticities. Our results show that in addition it is important to know the distribution of values of time. When such information is lacking or ignored, the resulting inefficiency from non-optimal toll levels is much greater than for first-best taxes.

5 Conclusion

This paper has reconsidered the road-pricing problem in a significantly broader context. We treat partial network pricing in a flexible way by considering two parallel routes followed by a shared link. We account for heterogeneity of users by assuming a continuous distribution of values of time. These innovations capture aspects of real applications of pricing, and they turn out to have significant effects.

Several new results stand out. First, when heterogeneity of road users is considered, travel times in the first-best optimum might actually be higher on one of the routes than in the no-toll equilibrium. This is caused by the use of differentiated tolls to provide a higher-quality service on link A by crowding link B even more.

Second, the most common approach to analyzing the benefits of parallel-route pricing creates two opposing biases. On the one hand, using two parallel routes but ignoring the interaction of users on other parts of the network (link C in our model) causes benefits of second-best pricing to be overstated, because users of the free lanes cause additional external congestion costs elsewhere. On the other hand, ignoring user heterogeneity causes benefits of second-best policies to be understated, by a factor of nine in our base case, because significant efficiency gains due to separation of traffic are omitted. Interestingly, it does not matter much to our results exactly what form the heterogeneity takes.

A third result concerns the distribution of benefits and losses. Under first-best pricing, users with the lowest values of time suffer the greatest average welfare losses or enjoy the smallest average gains. Many discussions of the politics of road pricing have focused on this point. However, the pattern changes when close substitutes of the priced good remain free: then, the users with intermediate values of time suffer most or gain least. It is as though we were to offer airline travellers only propeller planes or supersonic jets; this would cater to the extremes, but a lot of people would want something in between. To the extent that democratic processes cater to median preferences, this may help explain why pricing policies for congestible public facilities have made less political headway than other market-oriented reforms.

Fourth, the degree of toll differentiation that maximizes either welfare or revenue in an unconstrainted setting is smaller than expected. The importance of toll differentiation increases when demand becomes less elastic, and when the parallel links have different free-flow travel times.

Finally, the results confirm a more general insight from studies in second-best pricing: the amount of information required to apply a policy instrument to best advantage increases with the ‘imperfectness’ of this instrument. For the case considered here, this information includes the distribution of values of time and the demand elasticities of users having different values of time. Thus, second-best policies require considerable sophistication in order to achieve their theoretical benefits.

References

Arnott, R., A. de Palma and R. Lindsey (1992) "Route choice with heterogeneous drivers and group-specific congestion costs" *Regional Science and Urban Economics* **22** 71-102.

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

De Palma, A. and R. Lindsey (2000) "Private roads: competition under various ownership regimes" *Annals of Regional Science* **34** (1) 13-35.

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

Hahn, R. (1989) "Economic prescriptions for environmental problems: how the patient followed the doctor's orders" *Journal of Economic Perspectives* **3** (2) 95-114.

Knight, F. (1924) "Some fallacies in the interpretation of social costs" *Quarterly Journal of Economics* **38** 582-606.

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

Liu, N.L. 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.

Liu, N.L. and J.F. McDonald (1999) "Economic efficiency of second-best congestion pricing schemes in urban highway systems" *Transportation Research* **33B** 157-188.

Marchand, M. (1968) "A note on optimal tolls in an imperfect environment" *Econometrica* **36** (3-4) 575-581.

McDonald, John F., Edmond L. d'Ouville, and Louie Nan Liu (1999) *Economics of Urban Highway Congestion and Pricing*. Kluwer, Boston.

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

Mohring, Herbert (1979) "The benefits of reserved bus lanes, mass transit subsidies, and marginal cost pricing in alleviating traffic congestion" *Current Issues in Urban Economics*, ed. by Peter Mieszkowski and Mahlon Straszheim. Johns Hopkins, Baltimore, pp. 165-195.

Parkany, A.E. (1999) *Traveler Responses to New Choices: Toll vs. Free Alternatives in a Congested Corridor*, Ph.D. dissertation, Transportation Science, University of California at Irvine.

Pigou, A.C. (1920) *Wealth and Welfare*. Macmillan, London.

Small, Kenneth A. (1983) "The incidence of congestion tolls on urban highways," *Journal of Urban Economics* **13** (1) 90-111.

Small, Kenneth A. (1992) *Urban Transportation Economics*. Harwood Acaemic Publishers, Chur, Switzerland.

Small, K.A. and J.A. Gómez-Ibáñez (1998) "Road pricing for congestion management: the transition from theory to policy," in: *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*, ed. by K.J. Button and E.T. Verhoef. Cheltenham, UK: Edward Elgar, pp. 213-246.

Small, K.A. and J. Yan (2001) "The value of 'value pricing' of roads: second-best pricing and product differentiation" *Journal of Urban Economics* **49** (2) 310-336.

Sullivan, E. (1998) *Evaluating the Impacts of the SR 91 Variable-toll Express Lane Facility: Final Report*, report to California Department of Transportation. Dept. of Civil and Environmental Engineering, Cal Poly State University, San Luis Obispo, California.

Train, K.E., D.L. McFadden and A.A. Goett (1987) "Consumer attitudes and voluntary rate schedules for public utilities" *Review of Economics and Statistics* **69** (3) 383-391.

Train, K.E., M. Ben-Akiva and T. Atherton (1989) "Consumption patterns and self-selecting tariffs" *Review of Economics and Statistics* **71** (1) 62-73.

Transportation Research Board (1998) *Highway Capacity Manual: Special Report 209* (3rd edition, 1997 update). National Research Council, Washington, D.C.

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., P. Nijkamp and P. Rietveld (1997) "The social feasibility of road pricing: a case study for the Randstad area" *Journal of Transport Economics and Policy* **31** (3) 255-276.

Vickrey, W.S. (1963) "Pricing in urban and suburban transport" *American Economic Review, Papers and Proceedings* **53** (2) 452-465.

Vickrey, W.S. (1969) "Congestion theory and transport investment," *American Economic Review, Papers and Proceedings* **59** (2) 251-260.

Viton, P.A. (1995) "Private roads" *Journal of Urban Economics* **37** (3) 260-289.

Walters, A.A. (1961) "The theory and measurement of private and social cost of highway congestion" *Econometrica* **29** 676-699.

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

Yang, H. and H.-J. Huang (1999) "Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicles" *Transportation Research* **33A** 139-155.

Appendix to:
“Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users” by Erik T. Verhoef and Ken A. Small

Analytical derivation of optimal tolls in the various regimes

In this appendix, we consider the analytical derivation of the optimal tax rules for the various pricing regimes considered in the main text. These results provide insights into the solution and, in three cases (FB, SBSL and PSL), were used to calculate the numerical solutions.

A.1 FB and SBPL: public differentiated tolling

The Lagrangian Λ for schemes FB and SBPL results from adding the objective (5a) and the constraints (7a) (with $N_\alpha=N_{\alpha A}+N_{\alpha B}$). For FB we set the ‘toll-dummies’ $\delta_A=\delta_B=1$ and $\delta_C=0$, while for SBPL we set $\delta_B=1$ and $\delta_A=\delta_C=0$. The first-order conditions can be found by setting the partial derivatives of Λ with respect to each of the following variables equal to zero: $N_{\alpha L}$ (for all α present on L); $\lambda_{\alpha L}$ (for all α present on L); τ_A; and τ_B. When taking these derivatives, equation (2) is substituted for α^*, which therefore depends on τ_A, τ_B, and all N_α (since every N_α appears in the argument of either T_A or T_B). We again assume without loss of generality that $\tau_B>\tau_A$ and we define dummy variable δ_{α^*}, which takes on the value of 1 only when $\alpha=\alpha^*$. The first-order conditions then imply (simplified by using the constraints):

$$\begin{aligned}\frac{\partial\,\Lambda}{\partial\,N_{\alpha A}}=0\Rightarrow\delta_A\cdot\tau_A-\int_{\alpha_{\min}}^{\alpha^*}N_{aA}\cdot a\cdot T_A'\,\mathrm{d}a-\int_{\alpha_{\min}}^{\alpha_{\max}}N_a\cdot a\cdot T_C'\,\mathrm{d}a\\+\int_{\alpha_{\min}}^{\alpha^*}\lambda_{aA}\cdot a\cdot\left(T_A'+T_C'\right)\mathrm{d}a+\int_{\alpha^*}^{\alpha_{\max}}\lambda_{aB}\cdot a\cdot T_C'\,\mathrm{d}a-\lambda_{\alpha A}\cdot D_\alpha'-\delta_{\alpha^*}\cdot\lambda_{\alpha^*B}\cdot D_{\alpha^*}'\\+\frac{T_A'\cdot\left(\delta_B\cdot\tau_B-\delta_A\cdot\tau_A\right)}{\left(T_A-T_B\right)^2}\cdot N_{\alpha^*}\cdot X^*=0\quad\forall\,\alpha\le\alpha^*\end{aligned}\tag{A1a}$$

$$\begin{aligned}\frac{\partial\,\Lambda}{\partial\,N_{\alpha B}}=0\Rightarrow\delta_B\cdot\tau_B-\int_{\alpha^*}^{\alpha_{\max}}N_{aB}\cdot a\cdot T_B'\,\mathrm{d}a-\int_{\alpha_{\min}}^{\alpha_{\max}}N_a\cdot a\cdot T_C'\,\mathrm{d}a\\+\int_{\alpha_{\min}}^{\alpha^*}\lambda_{aA}\cdot a\cdot T_C'\,\mathrm{d}a+\int_{\alpha^*}^{\alpha_{\max}}\lambda_{aB}\cdot a\cdot\left(T_B'+T_C'\right)\mathrm{d}a-\lambda_{\alpha B}\cdot D_\alpha'-\delta_{\alpha^*}\cdot\lambda_{\alpha^*A}\cdot D_{\alpha^*}'\\-\frac{T_B'\cdot\left(\delta_B\cdot\tau_B-\delta_A\cdot\tau_A\right)}{\left(T_A-T_B\right)^2}\cdot N_{\alpha^*}\cdot X^*=0\quad\forall\,\alpha\ge\alpha^*\end{aligned}\tag{A1b}$$

$$\frac{\partial\,\Lambda}{\partial\,\tau_A}=\int_{\alpha_{\min}}^{\alpha^*}\lambda_{\alpha A}\,\mathrm{d}\alpha+\frac{1}{\left(T_A-T_B\right)}\cdot N_{\alpha^*}\cdot X^*=0\quad iff\quad\delta_A=1\tag{A2a}$$

$$\frac{\partial\,\Lambda}{\partial\,\tau_B}=\int_{\alpha^*}^{\alpha_{\max}}\lambda_{\alpha B}\,\mathrm{d}\alpha-\frac{1}{\left(T_A-T_B\right)}\cdot N_{\alpha^*}\cdot X^*=0\quad iff\quad\delta_B=1\tag{A2b}$$

with:

$$\begin{aligned} X^{*} = \alpha^{*}\cdot\left(T_A - T_B\right) + \int_{\alpha_{\min}}^{\alpha^{*}} N_{\alpha A}\cdot\alpha\cdot T_A'\,\mathrm{d}\alpha - \int_{\alpha^{*}}^{\alpha_{\max}} N_{\alpha B}\cdot\alpha\cdot T_B'\,\mathrm{d}\alpha \\ - \int_{\alpha_{\min}}^{\alpha^{*}} \lambda_{\alpha A}\cdot\alpha\cdot T_A'\,\mathrm{d}\alpha + \int_{\alpha^{*}}^{\alpha_{\max}} \lambda_{\alpha B}\cdot\alpha\cdot T_B'\,\mathrm{d}\alpha \end{aligned} \tag{A3}$$

The first two conditions (A1a) and (A1b) involve trading off the direct benefits of road use on the one route against the direct costs on that same route, as well as the indirect costs on the other. The direct costs are represented by the first two terms, which are familiar expressions reflecting the marginal external congestion costs imposed by a vehicle on all others using the same road. Note that the marginal benefits D_α and private travel costs $\alpha\cdot T$ do not appear in (A1a) and (A1b) because they were eliminated by substituting the constraints (7a) into the first-order conditions, causing the tolls τ to appear instead.

Next come four terms involving the Lagrangian multipliers λ_α, each of which gives the shadow price of a constraint which in simplified form is just $\alpha\cdot(T_L+T_C)+\tau_L=D_\alpha$ for which ever link L applies. If we think of D_α as containing an exogenous parameter shifting the inverse demand curve for α-type users downward, we see that λ_α represents the marginal impact on social welfare of such a demand shift. In the first-best optimum, FB, it will turn out that everyone is priced at marginal cost so a demand shift has no welfare impact at the margin and $\lambda_\alpha=0$ for every α. In the second-best optimum SBPL, however, even users of the priced link are paying less than their marginal cost so there is positive social welfare from shifting their demand downward, hence $\lambda_\alpha>0$ for all α. These three terms in equations (A1), then, show that in evaluating the marginal cost of a user with value of time α, one should also consider the indirect effects of this change upon road use by all other users, the latter being caused by the change in travel time (hence full prices) on the two alternative routes, plus an adjustment for the own elasticity of demand (relevant for both routes when α^* is considered). Note that these demand-related terms are the only ones that differ when comparing (A1a) or (A1b) for different values of α present on either link A or B. Therefore, the shadow prices $\lambda_{\alpha L}$ are inversely proportional to the steepness of the demand D_α: when α-users are less sensitive to price differentials, the shadow price $\lambda_{\alpha L}$ decreases in proportion.

The terms related to X^*, defined in (A3), reflect the welfare impact of induced marginal changes in α^*, again via induced changes in travel times. Equation (A3) shows that this impact includes the change in travel time for α^*-drivers transferred from link B to link A, the direct external congestion cost changes of such a transfer on both routes, and indirect welfare effects, like those just discussed.

Equations (A2a) and (A2b) show that when a toll can be charged on a given link, the shadow price for users of that link would average to zero except for the effect of induced shifts to and from the other link (by users with value of time α^*). When both links are tolled, adding (A2a) and (A2b) show that overall, the shadow prices average to zero. In fact, we already noted that they are identically zero in that case.

These equations exhibit a highly inconvenient discontinuity at α^* – which is why the dummy δ_{α^*} was needed. This discontinuity arises from the fact that a marginal increase of use by α^*-users on either route will affect marginal benefits on both routes. As a result, unless all λ's are equal to zero, a closed-form analytical solution to (A1a)-(A2b) cannot be found.[18] To see why, observe that we can solve all λ's for $\lambda_{\alpha^* A} + \lambda_{\alpha^* B}$ from (A1a) and (A1b):

$$\lambda_a = \left(\lambda_{\alpha^* A} + \lambda_{\alpha^* B}\right) \cdot \frac{-D'_{\alpha^*}}{-D'_a} \quad \forall\, a \neq \alpha^* \tag{A4}$$

Substituting (A4) into equations like (A1a) and (A1b) lead to problems of discontinuity at α^*. In the first-best case, because it can be shown that all λ's are zero, the following intuitive tax-rules apply:

$$\tau_A = \int_{\alpha_{\min}}^{\alpha^*} N_{\alpha A} \cdot \alpha \cdot T'_A \,\mathrm{d}a + \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha \cdot \alpha \cdot T'_C \,\mathrm{d}\alpha \tag{A5a}$$

$$\tau_B = \int_{\alpha^*}^{\alpha_{\max}} N_{\alpha B} \cdot \alpha \cdot T'_B \,\mathrm{d}\alpha + \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha \cdot \alpha \cdot T'_C \,\mathrm{d}\alpha \tag{A5b}$$

These tax rules simply state that each toll should be equal to the marginal external cost for that route. With optimal pricing on one route, the optimal price on the other can be determined independently, a normal consequence of the envelope theorem. We can also see that, with λ_α=0, (A2) require X^*=0, which, from (A3), requires that for α^*-users the valued time difference between the two routes be exactly balanced by the difference in externality costs. With first-best tolls applying on both routes, this is indeed the case.

For SBPL, a closed-form analytical solution can be found only if it happens that $\mathrm{N}_{\alpha^*} = 0$, so that no one is indifferent and hence there are no direct spill-over effects between links A and B. We then end up with an independent first-best optimization problem for the priced link. (Similarly, for FB we would end up with two independent first-best optimization problems.) Such a case can only arise if the distribution of values of time is bimodal. It is for this reason that assuming two groups, each with a distinct value of time, permits an analytical solution as in Small and Yan (2001).

A.2 SBSL: Public undifferentiated tolling

[18] *If* one would ignore the terms with δ_{α^*} in (A1a) and (A1b), a closed-form solution can be found, but using the simulation model, it was found to produce second-best taxes considerably different from the optimal second-best taxes. Comparable erroneous simplifications were tested and refuted for other cases where no closed-form solution can be found (PF and PPL).

The second-best public toll on the serial link can be found by solving the Lagrangian consisting of objective (5b) and constraints (7b). The optimal non-differentiating toll on link C can be shown to be equal to:

$$\tau_C = \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha \cdot \alpha \cdot T_C' \,\mathrm{d}\alpha + \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha \cdot \alpha \cdot T_D' \,\mathrm{d}\alpha \tag{A6}$$

We expect this solution to provide typically lower welfare than that computed for the first-best problem, but in fact we need to check because the latter was derived on the assumption that the tolls were unequal. We accomplish this by showing that in SBSL, the same traffic flow can be accommodated at lower total cost by setting τ_A marginally lower and τ_B marginally higher than τ_C as defined by (A6). Doing so would lead to a separation of traffic at α^*, and would induce a marginal shift of users from link B to link A. For simplicity, suppose the two links are identical, so that $T_A=T_B$, $T_A'=T_B'$ and $N_A=N_B$ at the solution to SBSL. Denote the size of the shifted traffic as Δ^*. Because travel times are equal on both links, the change in total travel costs resulting from this marginal tax change can be written as:

$$\Delta^* \cdot \left(\int_{\alpha_{\min}}^{\alpha^*} \alpha \cdot N_\alpha \cdot T_A' \,\mathrm{d}\alpha - \int_{\alpha^*}^{\alpha_{\max}} \alpha \cdot N_\alpha \cdot T_B' \,\mathrm{d}\alpha \right) \tag{A7}$$

The change in travel costs is thus equal to Δ^* times the difference in marginal external congestion costs. With $T_A'=T_B'$ and $N_A=N_B$ this change in cost is negative, because α will be higher on route B. With different routes, the same type of proof can be given by setting the marginally higher toll on the link that carries more traffic in SBSL. It could be the case, however, that counter-examples can be constructed where differences in T_A' and T_B' happen to exactly off-set the differences in $\int \alpha \cdot N_\alpha$ in the SBSL equilibrium.

A.3 PF and PPL: Private differentiated tolling

For PF and PPL, the Lagrangian consists of equations (6) plus (7a). Proceeding as in Section A.1, the first-order conditions imply:

$$\begin{aligned}\frac{\partial \Lambda}{\partial N_{\alpha A}} = 0 \Rightarrow \delta_A \cdot \tau_A + \int_{\alpha_{\min}}^{\alpha^*} \lambda_{aA} \cdot a \cdot (T_A' + T_C')\,\mathrm{d}a + \int_{\alpha^*}^{\alpha_{\max}} \lambda_{aB} \cdot a \cdot T_C' \,\mathrm{d}a - \lambda_{\alpha A} \cdot D_\alpha' - \delta_{\alpha^*} \cdot \lambda_{\alpha^* B} \cdot D_{\alpha^*}' \\ + \frac{T_A' \cdot (\delta_B \cdot \tau_B - \delta_A \cdot \tau_A)}{(T_A - T_B)^2} \cdot N_{\alpha^*} \cdot \left((\tau_B - \tau_A) - \int_{\alpha_{\min}}^{\alpha^*} \lambda_{aA} \cdot a \cdot T_A' \,\mathrm{d}a + \int_{\alpha^*}^{\alpha_{\max}} \lambda_{aB} \cdot a \cdot T_B' \,\mathrm{d}a \right) = 0 \\ \forall\, \alpha \le \alpha^*\end{aligned} \tag{A8a}$$

$$\begin{aligned}\frac{\partial \Lambda}{\partial N_{\alpha B}} = 0 \Rightarrow \delta_B \cdot \tau_B + \int_{\alpha_{\min}}^{\alpha^*} \lambda_{aA} \cdot a \cdot T_C' \,\mathrm{d}a + \int_{\alpha^*}^{\alpha_{\max}} \lambda_{aB} \cdot a \cdot (T_B' + T_C')\,\mathrm{d}a - \lambda_{\alpha B} \cdot D_\alpha' - \delta_{\alpha^*} \cdot \lambda_{\alpha^* A} \cdot D_{\alpha^*}' \\ - \frac{T_B' \cdot (\delta_B \cdot \tau_B - \delta_A \cdot \tau_A)}{(T_A - T_B)^2} \cdot N_{\alpha^*} \cdot \left((\tau_B - \tau_A) - \int_{\alpha_{\min}}^{\alpha^*} \lambda_{aA} \cdot a \cdot T_A' \,\mathrm{d}a + \int_{\alpha^*}^{\alpha_{\max}} \lambda_{aB} \cdot a \cdot T_B' \,\mathrm{d}a \right) \\ = 0 \quad \forall\, \alpha \ge \alpha^*\end{aligned} \tag{A8b}$$

$$\frac{\partial \Lambda}{\partial \tau_A} = \int_{\alpha_{\min}}^{\alpha^*} N_{\alpha A}\,\mathrm{d}\alpha + \int_{\alpha_{\min}}^{\alpha^*} \lambda_{\alpha A}\,\mathrm{d}\alpha$$

$$+\frac{1}{(T_A - T_B)}\cdot N_{\alpha^*}\cdot\left((\tau_B-\tau_A) - \int_{\alpha_{\min}}^{\alpha^*}\lambda_{\alpha A}\cdot\alpha\cdot T'_A\,\mathrm{d}\alpha + \int_{\alpha^*}^{\alpha_{\max}}\lambda_{\alpha B}\cdot\alpha\cdot T'_B\,\mathrm{d}\alpha\right) = 0 \qquad \text{(A9a)}$$

$$\textit{iff} \quad \delta_A = 1$$

$$\frac{\partial \Lambda}{\partial \tau_B} = \int_{\alpha^*}^{\alpha_{\max}} N_{\alpha B}\,\mathrm{d}\alpha + \int_{\alpha^*}^{\alpha_{\max}} \lambda_{\alpha B}\,\mathrm{d}\alpha$$

$$-\frac{1}{(T_A - T_B)}\cdot N_{\alpha^*}\cdot\left((\tau_B-\tau_A) - \int_{\alpha_{\min}}^{\alpha^*}\lambda_{\alpha A}\cdot\alpha\cdot T'_A\,\mathrm{d}\alpha + \int_{\alpha^*}^{\alpha_{\max}}\lambda_{\alpha B}\cdot\alpha\cdot T'_B\,\mathrm{d}\alpha\right) = 0 \qquad \text{(A9b)}$$

$$\textit{iff} \quad \delta_B = 1$$

Again, the first-order conditions are hard to interpret, and we refer to Verhoef *et al.* (1996) for an interpretation of simpler versions. Roughly speaking, the first two conditions consider the direct and indirect effects of marginal changes of road use upon the objective of maximizing revenue, whereas the latter two help to define the Lagrangian multipliers in the (private) optimum considered. Neither PF nor PPL has a closed-form analytical solution.

A.6 PPS: Private undifferentiated tolling

The problem of a private toll on the serial link has a Lagrangian which combines equation (6) and (7b). The first-order conditions are:

$$\frac{\partial \Lambda}{\partial N_\alpha} = \tau_C + \int_{\alpha_{\min}}^{\alpha_{\max}} \lambda_a \cdot a \cdot (T'_C + T'_D)\,\mathrm{d}a - \lambda_\alpha \cdot D'_\alpha = 0 \quad \forall\alpha \qquad \text{(A10a)}$$

$$\frac{\partial \Lambda}{\partial \tau_C} = \int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha\,\mathrm{d}\alpha + \int_{\alpha_{\min}}^{\alpha_{\max}} \lambda_\alpha\,\mathrm{d}\alpha = 0 \qquad \text{(A10b)}$$

Equation (A10a) can be solved for $\int \lambda_\alpha$ by using that $\lambda_\alpha \cdot - D_\alpha'$ is constant for all α. The following pricing rule can then be found:

$$\tau_C = \frac{\int_{\alpha_{\min}}^{\alpha_{\max}} N_\alpha\,\mathrm{d}\alpha}{\int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{-D'_\alpha}\,\mathrm{d}\alpha}\cdot\left(1 + (T'_C + T'_D)\cdot\int_{\alpha_{\min}}^{\alpha_{\max}}\frac{\alpha}{-D'_\alpha}\,\mathrm{d}\alpha\right) \qquad \text{(A11)}$$

This rule is a somewhat complicated variant of the standard revenue-maximizing toll on a congested road. It shares with earlier results (*e.g.* Edelson, 1971; Verhoef *et al.*, 1996) the feature that the toll decreases with the elasticity of demand (the monopolistic mark-up), and increases with the marginal external congestion costs.