Second-best Regulation of Road Transport Externalities

By Erik Verhoef, Peter Nijkamp and Piet Rietveld*

1. Introduction

In the 1920s, economists like Pigou (1920) and Knight (1924) recognised that "road pricing" offers the first-best solution for optimising congested road-traffic flows. Since then, the severity of traffic congestion has dramatically increased, turning it from a matter of academic interest into one of the most serious problems affecting urbanised areas and transport arteries. In addition, along with the growing levels of road transport, other external costs of road transport (notably environmental effects, noise annoyance and accidents) have become matters of increasing relevance. After seventy years, economists' answers to such market failures in road transport typically still rely heavily on the concept of Pigouvian taxes.

In principle, various types of road charging do exist. From a welfare economic point of view, a system of Electronic Road Pricing (ERP) is generally seen as the first-best technical mechanism for charging regulatory road charges. The reason for its superiority is that the regulatory tax can be differentiated according to the various dimensions affecting the actual marginal external costs of each trip, such as the length of the trip, the time of driving, the route followed and the vehicle used. Hence, individual drivers can be confronted with first-best incentives, inducing optimal (efficient) behavioural responses (assuming full information). However, although the well-known Hong Kong experiment has demonstrated that it is technically possible to operate an ERP scheme successfully nowadays (see Dawson and Catling, 1986; and Hau, 1992), various social and political impediments appear to prevent ERP from being widely introduced and accepted, certainly in the short run. Consequently, recent publications on road pricing have focused on the question of its political feasibility (Borins, 1988; Evans, 1992; Giuliano, 1992; Lave, 1994; and Verhoef, 1994); on methods of introducing road pricing schemes (Goodwin, 1989; Jones, 1991; May, 1992; Poole, 1992; Small, 1992a); and on various alternatives to ERP in traffic demand regulation,[1] such as fuel taxes (Mohring, 1989) and parking policies (Arnott, de Palma and Lindsey, 1991; Glazer and Niskanen, 1992; and Verhoef, Nijkamp and Rietveld, 1995).

* Free University, Amsterdam. Erik Verhoef is also affiliated to the Tinbergen Institute and participates in the VSB-fonds sponsored research project "Transport and Environment". The authors thank an anonymous referee for very useful comments.

[1] See Button (1992) for an overview.

The present paper belongs to this last category. It is concerned with the welfare economic characteristics of second-best alternatives to ERP in managing road transport demand. In particular, attention is focused on the consequences of the regulator's associated incapability of optimal tax differentiation between different types of road users. Throughout the paper, the analyses are restricted to *economic instruments* only (being second-best regulatory user *fees*). The additional shortcomings of physical restrictions in traffic regulations are discussed in Verhoef, Nijkamp and Rietveld (1995). Discussions of related second-best topics in transport can be found in Wilson (1983), and d'Ouville and McDonald (1990) on optimal road capacity supply with suboptimal congestion pricing; Braid (1989) and Arnott, de Palma and Lindsey (1990) on uniform or step-wise pricing of a bottleneck; and Arnott (1979), Sullivan (1983) and Fujita (1989, chapter 7.4) on congestion policies through urban land use policies. Two classic examples on second-best regulation in transport are Lévy-Lambert (1968) and Marchand (1968), studying optimal congestion pricing with an untolled alternative. Finally, it may be noteworthy that the issues considered in this paper bear close resemblance to those discussed in the literature on optimal taxation (see, for instance, Diamond, 1973; Sandmo, 1976; and Atkinson and Stiglitz, 1980).

The structure of the paper is as follows. In Section 2, the optimal undifferentiated fee is derived, and in Section 3 its welfare effects are compared with first-best regulation. Section 4 discusses some aspects related to demand interdependencies, and Section 5 considers a simple form of cost interdependencies, where the regulator is not capable of taxing all road users. Section 6 offers some concluding remarks.

2. The Optimal Common Regulatory Fee

The basic issue considered in this and the next section concerns a road network which is used by different groups of road users, each having their own group-specific level of external costs. We may, for instance, consider *vehicle-specific* externalities. Many types of external costs of road transport depend on the technical characteristics of the vehicle used (for instance, emissions of pollutants, or noise). Efficiency requires regulatory fees to be dependent on such characteristics (that is, on the actual emissions). However, it will often be practically impossible to operate such a delicate tax structure when applying second-best regulatory instruments. For instance, regulatory parking fees (meant to regulate traffic flows rather than parking as such) are in general not suitable for adaptation according to the type of vehicle parked (except for a rough distinction into private cars, vans, buses and trucks); nor are policies such as peak-hour permits. Likewise, many of the suggested alternatives to electronic road pricing are often not capable of differentiation according to individual *trip lengths*. This certainly holds for regulatory parking policies, which are implemented at the end of each trip, and therefore do not enable the regulator to differentiate according to the distance driven. It also holds for instruments such as peak-hour permits. Many (if not all) external costs of road transport, however, are dependent on trip length. Comparable problems may arise with externalities which are *route-* or *time-specific.*

In this section, the optimal common regulatory fee (the optimal second-best undifferentiated or coarse fee, as opposed to the optimal first-best differentiated or fine fee) for such cases is derived. First, we consider an *environmental externality*; then an *intra-sectoral externality* (congestion) is discussed.

2.1 The optimal common regulatory fee for an environmental externality

Suppose we have M different groups of car drivers (denoted $m = 1,2,\ldots,M$), jointly using a certain road network. Each group has its own specific marginal private cost of driving c_m and its own specific marginal external (environmental) cost $\partial E(N_1,N_2,\ldots,N_M)/\partial N_m$ (conveniently denoted $\varepsilon_m(\bullet)$ hereafter); N_m gives the number of trips made by group m. The groups are ordered according to increasing marginal external costs: $\varepsilon_j > \varepsilon_i$ if $j>i$. Now, suppose that the regulatory body wishes to reduce emissions by means of some second-best undifferentiated regulatory tax policy. More ambitiously, it wishes to find the *optimal common regulatory fee.* This fee maximises social welfare under the inherent limitation of the policy, being the impossibility of first-best tax differentiation. Such an optimal common fee can be found by solving the following Lagrangian:

$$\mathcal{L} = \sum_{m=1}^{M} \int_0^{N_m} D_m(n_m)\, dn_m - \sum_{m=1}^{M} c_m \bullet N_m - E(N_1,N_2,\ldots,N_M) + \sum_{m=1}^{M} \lambda_m \bullet [D_m(N_m) - c_m - f] \qquad (1)$$

The first term defines total benefits in the system as the relevant areas under the inverse demand curves D_m. The next term gives the total private cost of driving, for all groups defined as the average cost of a trip c_m, multiplied by the number of trips. Note that we assume an uncongested road network; average private costs are constant, and are therefore equal to marginal private costs. The third term indicates the total environmental external cost E. Finally, the restrictions λ_m indicate that in any equilibrium, for all groups, the inverse demand equals the sum of the marginal private cost c_m and the common regulatory fee (f) (that is, for all groups, the marginal driver has a zero net surplus). The first-order conditions are as follows (with an apostrophe denoting the first derivative):

$$\frac{\partial \mathcal{L}}{\partial N_m} = D_m(N_m) - c_m - \varepsilon_m(\bullet) + \lambda_m \bullet D'_m(N_m) = 0 \qquad m = 1,2,\ldots,M \qquad (1a)$$

$$\frac{\partial \mathcal{L}}{\partial f} = -\sum_{m=1}^{M} \lambda_m = 0 \qquad (1b)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_m} = D_m(N_m) - c_m - f = 0 \qquad m = 1,2,\ldots,M \qquad (1c)$$

The parameters λ_m cause these first-order conditions to differ from the first-best optimal ones. In particular, when in the restrictions in the Lagrangian (1) the common fee f is replaced by differentiated fees r_m, the first-order conditions include $\lambda_m = 0$ for all m. Hence, the first-best optimum is given by $r_m = \varepsilon_m$, which are the standard Pigouvian fees equal to marginal external costs. In the second-best case under consideration here, however, equations (1a) imply the following expressions for λ_m:

$$\lambda_m = -\frac{D_m(N_m) - c_m - \varepsilon_m(\bullet)}{D'_m(N_m)}$$

Therefore, using (1b) and (1c):

$$\sum_{m=1}^{M} \frac{f - \varepsilon_m(\bullet)}{D'_m(N_m)} = 0$$

which gives the following expression for the optimal common fee:

$$f = \sum_{m=1}^{M} \frac{[\varepsilon_m(\bullet)]/[D'_m(N_m)]}{\sum_{i=1}^{M} 1/[D'_i(N_i)]} \tag{2}$$

Equation (2) shows that the optimal common fee is a weighted average of the marginal environmental costs for all groups. Hence, with constant marginal external costs, the optimal common fee is a weighted average of the individual optimal Pigouvian fees (r_m) that would be charged under first-best regulation. This is an appealing result, which in fact bears close resemblance to other results obtained in the literature on optimal taxation (see, for instance, Sandmo, 1976; and Diamond, 1973). The weight attached to the individual optimal fee for a certain group is inversely related to the slope of the demand curve of that group *in the optimum*. This makes sense: the flatter the demand curve, the more distortive deviations from the individual optimal fee are, and hence the larger the weight should be. The slopes of the demand curves in fact capture both the *relative importance* of the groups and their *demand elasticities* (the larger the group and the more elastic the demand, the flatter the demand curve). In order to disentangle these two effects, we use the definition of demand elasticity η (dropping the argument N_m):

$$\eta_m = \frac{dN_m}{dD_m} \bullet \frac{D_m}{N_m}$$

where dN_m/dD_m is simply defined as $1/(dD_m/dN_m)$, and find that the weight w_m attached to group m amounts to:

$$w_m = \frac{1/D'_m}{\sum_{i=1}^{M}(1/D'_i)} = \frac{(\eta_m \bullet N_m)/D_m}{\sum_{i=1}^{M}[(\eta_i \bullet N_i)/D_i]} \tag{3}$$

It may be illustrative to note that, when all demand functions have the following form:

$$D_m(N_m) = \alpha \bullet \ln N_m + \beta_m$$

(where α has the same negative value for all groups), $dD_m/dN_m = \alpha/N_m$ and therefore $\eta_m/D_m = 1/\alpha$ for all values of N_m. This results in weights which are proportional to the group sizes in the (second-best) optimum.

Finally, the fact that the private costs of driving may differ among the groups does not affect the outcome. Such relative cost differences are internal, and therefore do not enter the expression for the optimal common fee.

2.2 The optimal common congestion fee

The foregoing analysis gives the general results for second-best regulation concerning an *environmental externality* (that is, an external cost posed upon actors outside the population of car drivers). We now turn to a special *intra-sectoral externality* associated with road transport: *congestion*. As a matter of fact, this is often seen as one of the most important negative externalities associated with road transport, and traffic regulation is often mainly concerned with the reduction (or optimisation) of congestion. In its most basic form, congestion can be introduced by making the private cost of driving within each group dependent on the number of drivers in that group. Hence, for S groups, we may formulate the following Lagrangian:

$$\mathcal{L} = \sum_{s=1}^{S} \int_0^{N_s} D_s(n_s) \mathrm{d}n_s - \sum_{s=1}^{S} N_s \bullet c_s(N_s) + \sum_{s=1}^{S} \lambda_s \bullet [D_s(N_s) - c_s(N_s) - f] \tag{4}$$

The particular form of the cost functions (the second term) can be interpreted as follows: total costs within a group s (say, C_s) is defined as the group size (N_s) times the average cost (c_s), which in turn depends on the group size. Furthermore, the restrictions show that the individual road user bases his behaviour on the average cost of driving — average social costs are perceived as marginal private costs. Note that the cost functions are assumed to be independent (that is, groups are defined so as to rule out cost interdependencies). Groups may be thought of as users of different radial access roads to a city centre, served by the same parking space where regulatory parking levies are charged. Alternatively, groups may be different cohorts, sufficiently separated in time so as to avoid interdependencies. Clearly, this is a rather restrictive assumption. However, when allowing for inter-group congestion effects, the expression for the optimal common fee can become quite complicated and difficult to interpret (see the Appendix). A relatively simple form of cost-interdependencies is discussed in Section 5 below.

It is easy to show that first-best regulation involves fees r_s, which are equal to the marginal external congestion costs (being the difference between the average and marginal cost of driving):

$$r_s = N_s \bullet c'_s(N_s)$$

However, the regulatory body has to find the optimal common congestion fee by solving (2). The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial N_s} = D_s(N_s) - c_s(N_s) - N_s \bullet c'_s(N_s) + \lambda_s \bullet [D'_s(N_s) - c'_s(N_s)] = 0 \qquad s = 1,2,\ldots,S$$

$$\frac{\partial \mathcal{L}}{\partial f} = -\sum_{s=1}^{S} \lambda_s = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_s} = D_s(N_s) - c_s(N_s) - f = 0 \qquad s = 1,2,\ldots,S$$

which may be solved to yield the following optimal common fee:

$$f = \sum_{s=1}^{S} \frac{[N_s \bullet c'_s(N_s)]/[D'_s(N_s) - c'_s(N_s)]}{\sum_{i=1}^{S} 1/[D'_i(N_i) - c'_i(N_i)]} \tag{5}$$

Like (2), (5) gives a weighted average of the marginal external congestion costs in the second-best optimum. However, the weights are somewhat different, having turned into:

$$w_s = \frac{[1/(D'_s - c'_s)]}{\sum_{i=1}^{S} [1/(D'_i - c'_i)]} = \frac{[1/(d'_s + c'_s)}{\sum_{i=1}^{S} [1/(d'_i + c'_i)]} \tag{6}$$

where d' gives the absolute value of the slope of the demand curve ($d' = -D'$). A group now receives a larger weight, not only the flatter the demand curve (the smaller d_s'), but also the flatter the average cost function (the smaller c_s') in the second-best optimum. The reason is that the weights should be such that the (private) welfare losses due to deviations of the common fee from the optimal individual fees are minimised. With congestion, such welfare losses do not only depend on the slopes of the demand curves, but on those of the average cost functions as well. The steeper both curves, the less responsive the group size to deviations of the optimal common fee from the optimal individual fee, and therefore the smaller the group's weight (relative to the other weights).

3. Evaluating Second-best Policies

3.1 The relative efficiency of second-best policies

A crucial question is, of course, how second-best policies as described by equations (2) and (5) relate to the first-best solution of charging all groups taxes equal to their marginal external costs. In order to shed some light on this issue, we will use the following index of relative welfare improvement, ω:

$$\omega = \frac{W_P - W_0}{W_R - W_0}$$

where W_P is social welfare (social benefits minus social costs) under second-best policies of charging the optimal common fee, W_R is welfare under the first-best policy of differentiated regulation, and W_0 is welfare under non-intervention.[2]

By definition, $W_0 \leq W_P \leq W_R$. Welfare under second-best regulation is at least as high as under non-intervention, which can always be realised by setting $f=0$. Likewise, welfare under first-best policies is at least as high as under (optimal) second-best policies, since this latter can always be realised by setting equal fees for all groups. However, $W_0 < W_P$ when $f>0$; and $W_P < W_R$ if for any $i \neq j$, $r_i \neq r_j$ (where r_m gives the first-best optimal fee for group m) and $D'_i \neq -\infty$ and $D'_j \neq -\infty$. Finally, in order to avoid $\omega = 0/0$, we have to assume that $W_0 < W_R$ (that is, we assume an externality to exist and not all demands to be perfectly inelastic). Under these conditions ($W_0 \leq W_P \leq W_R$ and $W_0 < W_R$), ω may range between zero and one. When $\omega=0$, it is impossible to increase welfare by using second-best policies. On the other hand, $\omega=1$ means that second-best regulation yields the same welfare improvement as does first-best regulation, implying that the instruments are perfect substitutes.

In order to develop an expression for ω, we first consider a simple two-group model with group-specific marginal environmental external costs $\varepsilon_i(N_i)$ ($i=1,2$). Consider the non-trivial case where $0<\varepsilon_1<\varepsilon_2$ and neither demand curve is perfectly inelastic. Hence, $0<r_1<f<r_2$ (f is a weighted average of r_1 and r_2). Let $N_i(c_i)$ give the number of trips made by group i under non-intervention. $N_i(c_i+f)$ and $N_i(c_i+r_i)$ give the number of trips under second-best and first-best policies, respectively (all these values satisfy the inverse demand relations $D_i(N_i)$). The following two inequalities hold: $N_1(c_1+f)<N_1(c_1+r_1)<N_1(c_1)$; and $N_2(c_2+r_2)<N_2(c_2+f)<N_2(c_2)$. Both first-best and second-best policies lead to reduc-

[2] Note that ω is concerned with efficiency considerations only. The redistributive effects of the respective policies are not considered. Furthermore, differences in implementation and intervention costs are ignored.

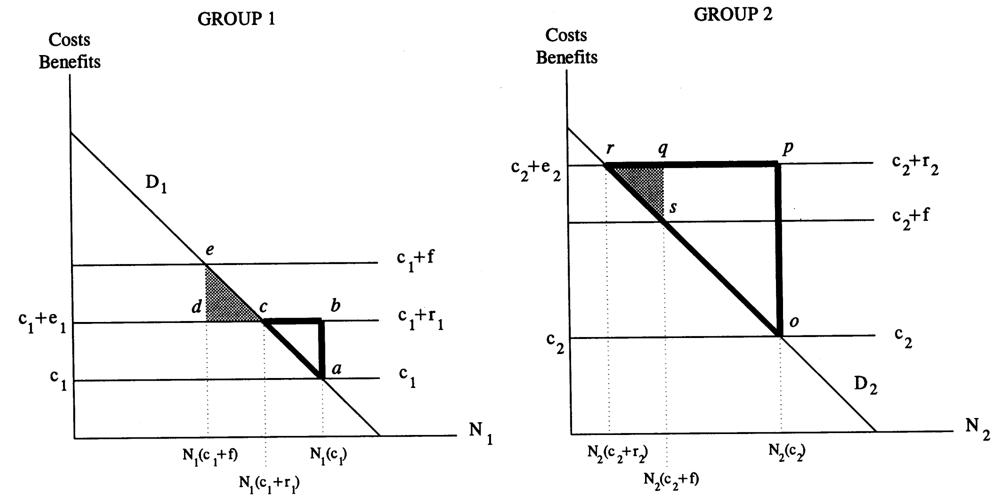

Figure 1
The Welfare Effects of Optimal First-best Regulation (bold) versus Optimal Second-best Regulation (bold minus shaded) for two Groups

tions in participation by both groups, albeit that group 1 is more restricted under second-best policies (since $r_1<f$), while group 2 is more restricted under first-best policies (since $f<r_2$). In this case, ω has the following general form:

$$\omega = \frac{I_1[N_1(c_1+r_1),N_1(c_1)] + I_1[N_1(c_1+f),N_1(c_1+r_1)] + I_2[N_2(c_2+f),N_2(c_2)]}{I_1[N_1(c_1+r_1),N_1(c_1)] + I_2[N_2(c_2+f),N_2(c_2)] + I_2[N_2(c_2+r_2),N_2(c_2+f)]} \tag{7}$$

where $I_i(N_i)$ is defined as the integral of the function resulting from summing the marginal private cost plus the marginal external cost, minus the marginal benefit. The first terms in the numerator and denominator are identical, giving the increase in social welfare that can be realised by levying group 1 their optimal fee r_1. The second term in the numerator, however, gives the welfare loss due to charging the higher second-best fee f, rather than r_1. The third term in the numerator, identical to the second term in the denominator, gives the welfare gain resulting from charging group 2 the fee f. The third term in the denominator, however, gives the additional welfare gain that may be realised by charging group 2 the higher fee r_2, rather than f. All terms are positive, except the second term in the numerator.

Figure 1 provides an illustration for the case where both groups have linear demand curves (D_i), constant marginal private cost (c_i) and constant marginal external cost (e_i). The maximum possible increase in social welfare in comparison with non-intervention is given by the sum of the surfaces of the two bold triangles *abc* and *opr*. This gain can be realised by charging the first-best fees $r_1=e_1$ and $r_2=e_2$. Under second-best regulation, equation (2) may yield an optimal common fee $f(r_1<f<r_2)$. In comparison with the first-

best solution, this policy involves welfare losses as given by the shaded areas *ced* and *sqr*. Hence, in this case, we find the following expression for ω:

$$\omega = \frac{abc - ced + opqs}{abc + opqs + sq\ r}$$

Clearly, the smaller the surface of the shaded areas in comparison with the surface of the bold triangles, the closer to one ω is. In terms of the more general expression (7), ω is closer to one, the smaller the second term in the numerator and the third in the denominator are in comparison to the other terms.

Before turning to the factors determining the value of ω, however, we generalise (7) for the *m* group case. For that purpose, a distinction should be made between those groups which are *undercharged* and those which are *overcharged* under second-best policies. The critical (possibly fictional) group μ, which draws the distinction between these two, is determined by that particular value of *m* for which the optimal first-best fee is equal to the optimal second-best fee. For instance, we may assume the external cost function to be linear in some physical factor Γ: $E=\varepsilon\bullet\Gamma$, and Γ to be additively separable in group-specific factors γ_m:

$$\Gamma = \sum_{m=1}^{M} \gamma_m \bullet N_m$$

Groups may, for example, be classified according to trip length γ_m. The above equations then mean that the total distance driven is equal to the summation over all possible trip lengths γ_m times the number of trips of that length, while total external cost depends linearly on total mileage driven. Alternatively, Γ may give total noise emissions over a day, where the different weights γ_m indicate that the severity of noise annoyance varies over the day. In such cases, the marginal external costs can be written as $\varepsilon\bullet\gamma_m$, and μ is that particular group for which:

$$\varepsilon\bullet\gamma_\mu = \varepsilon\bullet \sum_{m=1}^{M} \frac{\gamma_m / D'_m}{\sum_{i=1}^{M} (1/D'_i)} \Rightarrow \gamma_\mu = \sum_{m=1}^{M} (\gamma_m \bullet w_m)$$

At any rate, we may generalise (7) as follows:

$$\omega = \frac{\sum_{m=1}^{\mu} I_m[N_m(c_m+r_m), N_m(c_m)] + \sum_{m=1}^{\mu} I_m[N_m(c_m+f), N_m(c_m+r_m)] + \sum_{m=\mu}^{M} I_m[N_m(c_m+f), N_m(c_m)]}{\sum_{m=1}^{\mu} I_m[N_m(c_m+r_m), N_m(c_m)] + \sum_{m=\mu}^{M} I_m[N_m(c_m+f), N_m(c_m)] + \sum_{m=\mu}^{M} I_m[N_m(c_m+r_m), N_m(c_m+f)]} \quad (8)$$

Equation (8) confirms the intuitive expectation that second-best policies are a better substitute than first-best policies, when the total population of car drivers is more homogeneous in terms of external cost generation. Obviously, given the demand structure, ω is closer to 1 the smaller the spread in the terms r_m (that is, the nearer the optimal common fee approaches both the minimum and maximum optimal individual fees). In the extreme where $f=r_1=r_2=\ldots=r_M$, $\omega=1$.

The effect of the demand structure on ω, given the values r_m, is a bit less clear-cut. Speaking in absolute terms, deviations of the common fee from the first-best fees are less distortive (and hence absolute welfare losses are smaller) the more inelastic all demands and the smaller all groups are. However, since ω is a relative measure, this observation does not get us much further. It may be noted, though, that the more concave D_m for groups $m<\mu$, and the more convex D_m for groups $m>\mu$, the smaller the relative welfare losses will be (represented by the two "distortion terms" in (8)), and hence the larger ω will be (see also Figure 1). Furthermore, ω will be closer to one, the steeper the demand curves of the more extreme groups (in the optimum) in comparison to those near to μ. As indicated in the previous section, these slopes jointly capture the effects of demand elasticities and group sizes. For instance, suppose that in the optimum the distributions of the factors r_m and of the slopes of demand curves are perfectly symmetric, with μ in the centre of both distributions. The optimal second-best fee will then be r_m. However, the value of ω will be larger, the more inelastic the demands and the smaller the relative importance of the groups in the tails are in comparison to those in the centre of the distribution. The extreme $\omega=1$ is only attainable if all but one of the demand curves are perfectly inelastic and f is therefore set equal to the optimal individual fee for that remaining group. On the other hand, $\omega=0$ only occurs if one of the groups (say group 1) has zero external marginal cost and a perfectly elastic demand curve (the group is infinitely large and completely homogeneous). Second-best regulation then becomes a very unattractive option as it is completely frustrated by $f=r_1=0$. For all other cases, $0<\omega<1$.

When considering the optimal common congestion fee (5), the indices m in (8) may simply be replaced by s, which also have to be assigned according to increasing optimal individual congestion fees, with σ (like μ) being the group for which the second-best fee equals the first-best congestion fee. Again, ω is closer to 1, the smaller the spread in optimal individual first-best fees and the steeper the demand curves in both tails of the distribution are (for groups with either very low or very high congestion levels) as compared with the slopes in the centre. The same holds for the slopes of the average cost functions, which also determine the welfare losses due to uniform taxation in this case. However, one should not ignore the interplay between these slopes and the optimal individual congestion fees, which determine the assignment of the indices s in the first place. In particular, with "comparable" group sizes, a group with a steep average cost function would never be in the left tail of the distribution.

3.2 The effectiveness of second-best environmental policies

A question which is especially relevant from an environmental point of view, is how the absolute reduction in emissions under second-best policies compares to the reduction under a scheme of first-best regulation. Intuition tells us that the former is likely to be smaller, since second-best policies involve certain welfare losses. Hence, the cost of reducing the externality is larger, and we expect a lower optimal reduction in total emissions for any $0\leq\omega<1$. In general, a policy change from second-best to first-best regulation leads to a change in total environmental cost which is equal to:

$$\Delta E = E[N_1(c_1+r_1),N_2(c_2+r_2),\ldots,N_M(c_M+r_M)] - E[N_1(c_1+f),N_2(c_2+f),\ldots,N_M(c_M+f)]$$

In order to keep the analysis manageable, consider a model with linear demand curves and linear, additively separable external cost (as discussed above). The change in total environmental cost, resulting from a policy change from second-best to first-best policies, amounts to:

$$\Delta E = \sum_{m=1}^{M} \Delta N_m \bullet \varepsilon \bullet \gamma_m = \sum_{m=1}^{M} (r_m - f) \bullet \frac{1}{D'_m} \bullet \varepsilon \bullet \gamma_m \tag{9}$$

For the relatively low-externality groups ($m<\mu$), $r_m<f$, and the relevant term in (9) is positive, indicating the effect of the *increase* in emissions of these groups due to a higher participation. Conversely, for $m>\mu$, $r_m>f$, and the relevant term in (9) is negative, indicating the effect of the *decrease* in emissions of these groups due to lower participation. The overall effect on total emissions can be found by substitution of the optimal second-best fee f given in (2) into (9), and using $r_m = \varepsilon \bullet \gamma_m$:

$$\Delta E = \sum_{m=1}^{M} \left[\left(\varepsilon \bullet \gamma_m - \varepsilon \sum_{i=1}^{M} \frac{\gamma_i / D'_i}{\sum_{j=1}^{M} (1/D'_j)} \right) \bullet \frac{1}{D'_m} \bullet \varepsilon \bullet \gamma_m \right]$$

$$= \varepsilon^2 \bullet \sum_{m=1}^{M} \left[\frac{\gamma_m^2}{D'_m} - \frac{\gamma_m}{D'_m} \bullet \left(\sum_{i=1}^{M} \gamma_i \bullet w_i \right) \right] \tag{9'}$$

It is not easy to determine the sign of (9') at first sight. After some manipulation, however, we find that this expression can be rewritten as:

$$\Delta E = 1/2 \bullet \varepsilon^2 \bullet \left[\sum_{i=1}^{M} \sum_{j=1}^{M} \left(\frac{(\gamma_i - \gamma_j)^2}{\sum_{k=1}^{M} (D'_i \bullet D'_j)/D'_k} \right) \right] \tag{9''}$$

For instance, for three groups, equation (9'') turns into:

$$\Delta E = \varepsilon^2 \bullet \left(\frac{(\gamma_1 - \gamma_2)^2}{D'_1 + D'_2 + \frac{D'_1 \bullet D'_2}{D'_3}} + \frac{(\gamma_1 - \gamma_3)^2}{D'_1 + D'_3 + \frac{D'_1 \bullet D'_3}{D'_2}} + \frac{(\gamma_2 - \gamma_3)^2}{D'_2 + D'_3 + \frac{D'_2 \bullet D'_3}{D'_1}} \right)$$

The expression in (9'') is always negative (or, in some extreme cases, equal to zero). Therefore, first-best differentiated regulation usually involves lower total environmental cost (and lower total emissions Γ) than do second-best policies — at least with linear demand and linear, additively separable, external cost. (However, it is not difficult to construct corner solutions in which second-best policies actually do lead to a lower level of total emissions than does first-best regulation.)

Since congestion is an intra-sectoral external cost, the benefits of optimising the externality are also to be reaped within the sector. Therefore, it is less useful to consider the consequences of both types of policy on the optimal level of the total external cost. From a welfare point of view, it is sufficient to recognise that, when ω is smaller than 1, the potential welfare gains from second-best instruments are smaller than those resulting from first-best regulation.[3]

[3] However, one would intuitively expect the total level of congestion generally to be smaller under first-best policies for the same reason that the cost of reducing congestion is larger under second-best regulation due to the associated welfare losses.

4. Group Choice and Modal Choice: Considering Demand Interdependencies

Because of the nature of the Lagrangians (1) and (4), some possibly important behavioural responses to the kind of regulation applied (that is, first-best versus second-best) are ignored. In particular, the postulated independent, stable demand curves for all groups, suggest that car drivers do not consider the possibility of switching group in response to the form of regulation applied. However, this possibility may additionally affect the performance of second-best policy instruments.[4] Analytically, the issues of *group choice* and *modal choice* (in terms of whether to leave the road system) are closely related: whereas the former involves exiting one group and entering another one, the latter merely involves exiting the initial group without entering another.

In transport markets such as those considered in this paper, it seems implausible to specify continuous individual demand functions, including continuous "cross-relations" between the different groups. Rather, individual behaviour is more adequately described by considering binary choices of whether to make a certain trip, based on the benefits and costs of doing so. Furthermore, the different groups are likely to be mutually exclusive for each individual driver; per trip, he chooses only one route (and therewith one trip length), one departure time, and one vehicle. Consequently, when considering the issues of modal and group choice in relation to the kind of regulation applied, the only assumption concerning individual behaviour that seems generally valid is that the (rational) individual car driver i chooses that group m where his (expected) net surplus (benefits $B_m{}^i$ minus (group-specific) private cost c_m minus any regulatory fee t_m) is maximised, provided it exceeds some threshold (money metric) indirect utility $Q_0{}^i$ associated with leaving the road system. This latter determinant depends on the utility associated with the best alternative available, which may be either an alternative mode such as public transport, or refraining from making the trip at all.[5] Therefore, individual behaviour may be characterised as:

$$MAX\{MAX_m\{B_m{}^i - c_m - t_m\};Q_0{}^i\} \tag{10}$$

Via the regulatory fee t_m (which may be either the first-best fee r_m or the optimal common fee f), the prevailing form of regulation to some extent determines the individual net surpluses enjoyed. By definition, under first-best policies, actors face the optimal incentives for group and modal choice. Actors choose a group by maximising the

4 The potential relevance of such behavioural responses increases, the easier they can be made. For instance, changes in departure times or routes may involve lower costs than changes in trip length (requiring changes in spatial activity patterns), change of mode or change of vehicle. However, we will ignore such transitional costs.

5 After the drivers have chosen a group, the discontinuous individual demand functions can still be aggregated to more or less continuous demand functions on the group level, simply by ordering the group members according to decreasing willingness to pay (as was done in Section 2). However, within each group, the individual willingnesses to pay to make a trip in one of the other groups may of course vary unsystematically among the group members. We therefore refrain from setting up general analyses such as those in Section 2 (which require continuous functions), and restrict the discussion to a comparison of incentives faced by the actors under first-best and second-best regulation.

difference between benefits $B_m{}^i$ and the generalised costs $c^g{}_m$, being the sum of private costs and the regulatory fee. Since under first-best policies these generalised costs are by definition equal to the social costs, individual utility-maximising behaviour maximises social welfare, and the optimal distribution of road users over the groups and the optimal number of road users within each group will result. Under second-best regulation, however, the incentives given for switching from high-externality groups to low-externality groups are far from optimal.

For the environmental externality, the optimal common fee provides no incentives at all to switch from one group to another, since the factors $t_m=f$ are equal for all groups. This fee reduces the net surpluses achievable in each possible group by equal amounts. Therefore, actors belonging to one group under non-intervention will either remain in that group or leave the system. Hence, the incentive to switch from a "high externality group" j to a "low externality group" i misses out a factor:

$$\varepsilon_j(\bullet) - \varepsilon_i(\bullet)$$

Therefore, second-best uniform taxation not only leads to a larger (smaller) than optimal reduction in the sizes of low (high) externality groups via the direct effect discussed in the foregoing sections. This is the "modal choice effect" that (in terms of equation (8)), for $m<\mu$, the excess reduction in the number of trips is $N_m(c_m+r_m) - N_m(c_m+f)$; and for $m>\mu$, the reduction in car use falls short with $N_m(c_m+f) - N_m(c_m+r_m)$. In addition, there is the more dynamic indirect effect (the "group choice effect") of actors facing insufficient (that is, *no*) incentives to change their behaviour in terms of switching to low externality groups.

For congestion, the modal choice effect is analogous to the effect mentioned above. That is, for $s<\sigma$, the excess reduction in the number of trips is $N_s(c_s+r_s) - N_s(c_s+f)$; and for $s>\sigma$, the reduction in car use falls short with $N_s(c_s+f) - N_s(c_s+r_s)$. However, the group choice effect may actually even be counterproductive for common congestion fees. Again, the uniform tax in the first instance leads to increases in generalised costs which are equal for all groups, suggesting a lack of incentive to switch from a "high congestion group" j to a "low congestion group" i equal to:

$$N_j \bullet c'_j(N_j) - N_i \bullet c'_i(N_i)$$

However, the common fee may very well lead to larger reductions in the average cost c_m for high congestion groups than for low congestion groups. If this is the case, the generalised costs for less congested groups increase more strongly than for more congested groups. This may lead to the adverse indirect effect of drivers who initially belonged to low congestion groups switching to more congested groups (see (10)). Whether this adverse "group choice effect" can occur depends on the slopes of the demand and average cost curves of the initial groups. Suppose there are only two congestion groups: a low congestion group i and a high congestion group j. The common fee f, by definition, leads to reductions in the number of trips made by both groups such that, in terms of the initial demand and average cost curves:

$$\int_{N_i(c_i)}^{N_i(c_i+f)} \left(\frac{\mathrm{d}D_i}{\mathrm{d}N_i} - \frac{\mathrm{d}c_i}{\mathrm{d}N_i}\right) \mathrm{d}N_i = f = \int_{N_j(c_j)}^{N_j(c_j+f)} \left(\frac{\mathrm{d}D_j}{\mathrm{d}N_j} - \frac{\mathrm{d}c_j}{\mathrm{d}N_j}\right) \mathrm{d}N_j$$

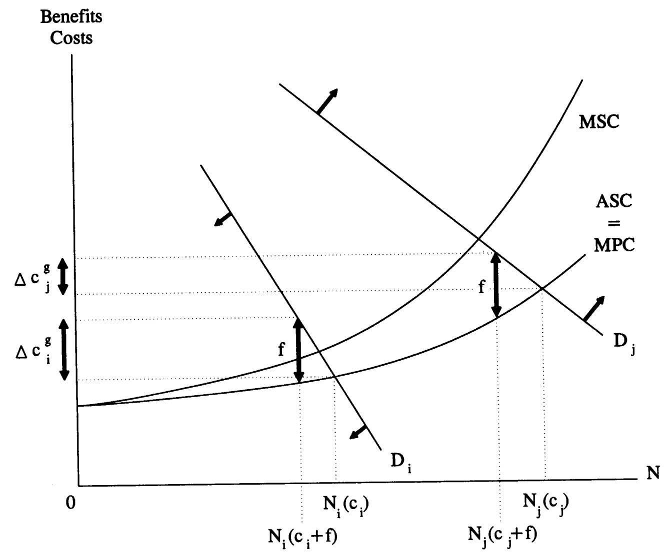

Figure 2
Adverse Group Choice Effects Resulting from a Common Congestion Fee

Since j is assumed to be the high congestion group, $N_j \bullet dc_j/dN_j > N_i \bullet dc_i/dN_i$, and $dc_j/dN_j > dc_i/dN_i$ if congestion rises progressively with group size. Then, the decrease in c_j exceeds the decrease in c_i unless the (absolute value of) the average slope (over the relevant range) of the demand curve D_j exceeds that of D_i, at least to the same extent as the slope of the average cost curve c_j exceeds that of c_i. This, of course, need certainly not be the case; especially since high congestion is often associated with large groups, other things being equal, leading to a flattening of the demand curve.

For instance, Figure 2 shows a road network used during a peak hour by group j and in the off-peak by group i. The curves D give demand, *MSC* is the marginal social cost and *ASC=MPC* gives average social cost = marginal private cost. With a common congestion fee f (such as a time-independent regulatory parking fee), which is drawn as some weighted average of what the optimal differentiated fees would be, the generalised cost for group i increases more than for group j: $\Delta c^g_i > \Delta c^g_j$. Therefore, some drivers, initially using the network during the off-peak, may actually be attracted to the peak hour, causing an inward shift of D_i and an outward shift of D_j (as illustrated by the arrows). As a consequence, the reduction in congestion during the peak (the off-peak) may be even further below (above) the optimal reduction than indicated by the modal choice effect alone. Of course, these shifts may in turn affect the optimal common fee, an effect from which we abstain.

5. Second-best Charging Mechanisms: A Simple Form of Cost Interdependencies

A last, somewhat different reason why second-best instruments may be an imperfect substitute for optimally-differentiated regulation occurs when a certain number of actors is not confronted with the second-best fee at all. Apart from sheer fraud (which may of course also occur under first-best policies), this may happen when second-best fees are levied via second-best charging mechanisms. For instance, road users who use private parking spaces will not be confronted with a regulatory parking fee.[6] Especially when the aim of the policy is to optimise congestion, the implications are not immediately clear. At first sight, one might argue that those who do park on public space should be taxed more heavily, in order to try and realise the optimal reduction in traffic flows as closely as possible. However, the taxation of the public parkers leads to a reduction in congestion, and therefore to a reduction in the marginal private cost of making trips, thus inducing additional traffic from private parkers. There is clearly a welfare loss associated with this process, as some road users (public parkers) with a certain willingness to pay who are "taxed off the road" will to some extent be replaced by other road users (private parkers) with a (necessarily) lower willingness to pay. This type of problem introduces a simple form of cost interdependency into our model; a more complex case is discussed in the Appendix.

When tolled road users (group 1) and untolled road users (group 2) jointly use a congested road network, where the average cost of driving depends on the total number of road users $N=N_1+N_2$ (implying that the marginal cost of adding an extra road user is equal for both groups), the regulator faces the following Lagrangian:

$$\begin{aligned} \mathcal{L} = &\int_0^{N_1} D_1(n_1)dn_1 + \int_0^{N_2} D_2(n_2)dn_2 - N \bullet c(N) \\ &+ \lambda_1 \bullet [D_1(N_1) - c(N) - f] + \lambda_2 \bullet [D_2(N_2) - c(N)] \end{aligned} \tag{11}$$

with: $N=N_1+N_2$.

The first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial N_1} = D_1(N_1) - c(N) - N \bullet c'(N) + \lambda_1 \bullet [D'_1(N_1) - c'(N)] - \lambda_2 \bullet c'(N) = 0$$

$$\frac{\partial \mathcal{L}}{\partial N_2} = D_2(N_2) - c(N) - N \bullet c'(N) - \lambda_1 \bullet c'(N) + \lambda_2 \bullet [D'_2(N_2) - c'(N)] = 0$$

$$\frac{\partial \mathcal{L}}{\partial f} = -\lambda_1 = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = D_1(N_1) - c(N) - f = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = D_2(N_2) - c(N) = 0$$

[6] The type of problem considered in this section is not specific to parking. Another example would be congestion pricing with the constraint that commercial vehicles be exempt, as in fact happened originally in Singapore (we owe this example to the anonymous referee).

Using $\lambda_1=0$, we find:

$$f = N \cdot c'(N) + \lambda_2 \cdot c'(N).$$

Solving for λ_2 yields:

$$\lambda_2 = -\frac{N \cdot c'(N)}{c'(N) - D'_2(N_2)}$$

Substitution of λ_2 then yields the following optimal second-best fee:[7]

$$f = N \cdot c'(N) \cdot \left[\frac{-D'_2(N_2)}{c'(N) - D'_2(N_2)} \right] \tag{12}$$

A comparison of (12) with the standard congestion fee (see Section 2) shows how the presence of untolled road users affects the result. The optimal second-best fee now depends on the slope of the demand curve of the untolled users, which in turn depends on their group size and demand elasticities, as outlined in Section 2.

In the one extreme, where the untolled users have a perfectly elastic demand in the optimum (group 2 is infinitely large and completely homogeneous), the term between the large brackets in (12) is zero, and therefore the optimal second-best fee is zero. Any positive fee will only lead to welfare losses, since any tolled driver who is "taxed off the road" will be replaced by an untolled driver with a lower willingness to pay for using the road network. Congestion therefore remains unaffected, while the total benefit decreases. Therefore, the best thing the regulator can do is avoid levying any fee at all. At the other extreme, where the untolled road users have a perfectly inelastic demand in the optimum,[8] the term between the large brackets in (12) is equal to 1, and (12) reduces to:

$$f = N \cdot c'(N) \tag{12'}$$

This optimal fee is similar to the standard congestion fee derived in Section 2. In this case, however, the optimal fee for the tolled users depends on the (fixed) number of untolled users. In a sense, the presence of untolled drivers can be thought of as being a mere restriction on the road network capacity, leading, other things being equal, to a lower optimal number of tolled road users and therefore to a higher second-best regulatory fee.

In any situation between the two extreme cases mentioned above, the optimal fee simply depends on the joint effects of the slope of the untolled drivers' demand curve (the flatter their demand curve, the lower the optimal fee) and, of course, on the overall second-best demand in relation to the capacity of the network.

Clearly, the presence of users who are not subject to regulatory policies seriously harms the performance of such policies. In the first place, not every road user responsible for (and subject to) congestion can be reached with such policies. Moreover, and perhaps less obviously, regulatory policies will be frustrated by the fact that the users who are "taxed off the road" may be replaced by untolled users, representing a lower willingness to pay for road use and consequently representing lower benefits. This process involves welfare losses which have to be weighed against the welfare gains of second-best regulatory policies in terms of reducing congestion.

[7] This result is equivalent to the one obtained by Glazer and Niskanen (1992); see their equation (18).

[8] An important reason why, in the case of regulatory parking policies, the untolled private parkers' demand could be perfectly inelastic in the second-best optimum is a binding restriction on private parking space.

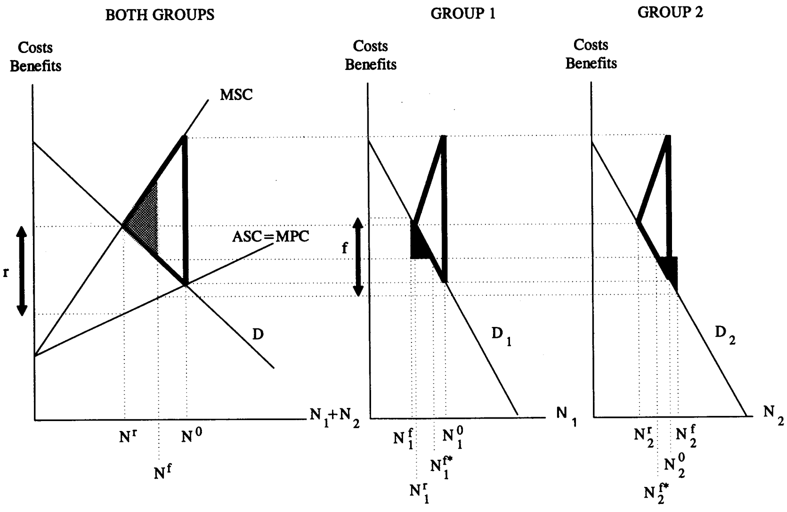

Figure 3

The Welfare Effects of First-best Congestion Fees versus the Optimal Second-best Fee in Presence of Both Tolled Users (Group 1) and Untolled Users (Group 2)

Figure 3 gives a diagrammatic representation of the situation considered above. In the left panel, the demand (*D*), average social cost (*ASC*), which equals marginal private cost (*MPC*), and the marginal social cost (*MSC*) curves are drawn for the entire population of road users. The middle and right panel give the demand curves for group 1 (the tolled users) and group 2 (the untolled users), respectively. Optimal first-best regulation implies a levy *r* for both groups, leading to a reduction in the total number of trips from N^0 to N^r and an increase in social welfare equal to the bold triangle in the left panel (or the sum of the two bold triangles in the middle and right panel). However, the use of the optimal second-best fee as given by (12) leads to a reduction in the number of trips made by group 1 from N_1^0 to N_1^f (rather than N_1^r) and an increase in the number of trips made by group 2 from N_2^0 to N_2^f (rather than the optimal reduction to N_2^r). The total reduction in number of trips is therefore only from N^0 to N^f, and the use of this policy misses out on the potential welfare gains as given by the shaded triangle in the left panel. Moreover, the reduction from N^0 to N^f is not accomplished in the most efficient way, which would have been reductions to N_1^{f*} and N_2^{f*}, respectively, where the marginal benefits for both groups are equalised. Therefore, the two black triangles give additional welfare losses of optimal second-best regulation in comparison with optimal first-best differentiated regulation. The index of relative welfare improvement ω will in this case be as follows:

$$\omega = \frac{I(N^f,N^0) - [B_1(N_1^f, N_1^{f*}) - D_1(N_1^{f*}) \bullet (N_1^{f*} - N_1^f)] - [D_2(N_2^{f*}) \bullet (N_2^f - N_2^{f*}) - B_2(N_2^{f*}, N_2^f)]}{I(N^r,N^f) + I(N^f,N^0)} \tag{13}$$

The function I is as defined before, but it is now specified for the entire population of road users; and $B_i(\bullet)$ gives the integral of the inverse demand function i over the specified range. From (13), it follows that ω is closer to 1, the steeper the demand curve of group 2, either due to a lower demand elasticity or a smaller group size. This reduces the value of the first term in the denominator, since f in (12) approaches r. Furthermore, the second and third term in the numerator then approach zero.

Finally, the total reduction in the number of trips under second-best policies is in this case obviously always smaller than under first-best regulation whenever the demand of group 2 is not perfectly inelastic. First, group 1 is confronted with a lower fee; and secondly, group 2 even enjoys a fall in the cost of driving.

6. Conclusions

In this paper, the welfare economic characteristics of second-best alternatives to first-best differentiated electronic road pricing were investigated. We considered economic instruments only. Second-best regulation will often suffer from the impossibility of optimal tax differentiation between different types of road users in cases where first-best policies do call for such differentiation — for instance, when external costs are trip-length-, time-, route- or vehicle-specific.

The optimal common fee turns out to be a weighted average of the optimal first-best differentiated fees. For an environmental externality, the optimal weight assigned to a group depends on the inverse of the slope of its demand curve in the optimum, and is therefore positively related to its (second-best) optimal group size and to its demand elasticity in the second-best optimum. For the optimal common congestion fee, the weight is also (inversely) related to the slope of the average cost function in the second-best optimum.

By application of an index of relative welfare improvement, we were able to compare the welfare effects of second-best policies to those resulting from first-best policies. It turned out that second-best instruments become a less favourable substitute, the larger the differences in marginal external costs, and the larger the relative importance and the demand elasticities of the groups in the extremes. It was also shown that the optimal reduction in the externality is usually smaller when second-best instruments are used, than under first-best regulation.

A second-best common fee results in non-optimal incentives in terms of modal choice, in that low externality groups are overcharged and high externality groups are undercharged. In terms of group choice, a common fee completely fails to provide (optimal) incentives to switch from high externality to low externality groups for an environmental externality. A common congestion fee may even create adverse incentives for group choice.

In the last section, it was shown that the performance of a second-best congestion fee, suffering from the incapability of charging all road users on a network, critically depends on the relative importance and the demand elasticity of the group which remains unaffected by the policy.

Armed with the results obtained in this paper, it is easier to form a judgement on the relative desirability of second-best alternatives to road pricing, whenever suggested.

References

Arnott, R. J. (1979): "Unpriced Transport Congestion". *Journal of Economic Theory,* 21, pp.294-316.

Arnott, R., A. de Palma and R. Lindsey (1990): "Economics of a Bottleneck". *Journal of Urban Economics,* 27, pp.11-30.

Arnott, R., A. de Palma and R. Lindsey (1991): "A Temporal and Spatial Equilibrium Analysis of Commuter Parking". *Journal of Public Economics,* 45, pp.301-35.

Atkinson, A. B. and J. E. Stiglitz (1980): *Lectures on Public Economics.* London: McGraw-Hill.

Braid, R. M. (1989): "Uniform versus Peak-Load Pricing of a Bottleneck with Elastic Demand". *Journal of Urban Economics,* 26, pp.320-27.

Borins, S. F. (1988): "Electronic Road Pricing: An Idea Whose Time May Never Come". *Transportation Research,* 22A, pp.37-44.

Button, K. J. (1992): "Alternatives to Road Pricing". Paper presented to the OECD/ECMT/NFP/GVF *Conference on The Use of Economic Instruments in Urban Travel Management,* Basel.

Dawson, J. A. L. and I. Catling (1986): "Electronic Road Pricing in Hong Kong". *Transportation Research,* 20A, pp.129-34.

Diamond, P. J. (1973): "Consumption Externalities and Imperfectly Corrective Pricing". *Bell Journal of Economics and Management Science,* 4, pp.526-38.

d'Ouville, E. L. and J. F. McDonald (1990): "Optimal Road Capacity with a Suboptimal Congestion Toll". *Journal of Urban Economics,* 28, pp.34-49.

Evans, A. W. (1992): "Road Congestion Pricing: When is it a Good Policy?" *Journal of Transport Economics and Policy,* 26, pp.213-43.

Fujita, M. (1989): *Urban Economic Theory: Land Use and City Size.* Cambridge: Cambridge University Press.

Giuliano, G. (1992): "An Assessment of the Political Acceptability of Congestion Pricing". *Transportation,* 19, 4, pp.335-58.

Glazer, A. and E. Niskanen (1992): "Parking Fees and Congestion". *Regional Science and Urban Economics,* 22, pp.123-32.

Goodwin, P. B. (1989): "The Rule of Three: A Possible Solution to the Political Problem of Competing Objectives for Road Pricing". *Traffic Engineering and Control,* 30, pp.495-97.

Hau, T. D. (1992): *Congestion Charging Mechanisms: An Evaluation of Current Practice.* Preliminary Draft, Transport Division, the World Bank, Washington (monograph).

Jones, P. (1991): "Gaining Public Support for Road Pricing through a Package Approach". *Traffic Engineering and Control,* 32, pp.194-96.

Knight, F. H. (1924): "Some Fallacies in the Interpretation of Social Cost". *Quarterly Journal of Economics,* 38, pp.564-74.

Lave, C. (1994): "The Demand Curve under Road Pricing and the Problem of Political Feasibility". *Transportation Research,* vol.28A, no.2, pp.83-91.

Lévy-Lambert, H. (1968): "Tarification des Services à Qualité Variable: Application aux Péages de Circulation". *Econometrica,* 36, pp.564-74.

Marchand, M. (1968): "A Note on Optimal Tolls in an Imperfect Environment". *Econometrica,* 36, pp.575-81.

May, A. D. (1992): "Road Pricing: An International Perspective". *Transportation,* vol.19, p.313-33.

Mohring, H. (1989): "The Role of Fuel Taxes in Controlling Congestion". In: *Transport Policy, Management and Technology Towards 2001*. Proceedings of the Fifth World Conference on Transport Research, vol. 1: "The Role of Public Sector in Transport", pp.243-57. Yokohama: WTCR.
Pigou, A. C. (1920): *Wealth and Welfare*. London, Macmillan.
Poole, R. W. (Jr) (1992): "Introducing Congestion Pricing on a New Toll Road". *Transportation*, 19, pp.383-96.
Sandmo, A. (1976): "Optimal Taxation: An Introduction to the Literature". *Journal of Public Economics*, 6, pp.37-54.
Small, K. A. (1992a): "Using the Revenues from Congestion Pricing". *Transportation*, 19, pp.359-81.
Small, K. A. (1992b): "Urban Transportation Economics". *Fundamentals of Pure and Applied Economics*, 51, Harwood, Chur etc.
Sullivan, A. M. (1983): "Second-Best Policies for Congestion Externalities". *Journal of Urban Economics*, 14, pp.105-23.
Verhoef, E. T. (1994): "The Demand Curve under Road Pricing and the Problem of Political Feasibility: A Comment". Forthcoming in *Transportation Research*.
Verhoef, E. T., P. Nijkamp and P. Rietveld (1995): "The Economics of Parking Management Systems: The (Im)Possibilities of Parking Policies in Traffic Regulation". TRACE discussion paper TI93-254, Tinbergen Institute, Amsterdam-Rotterdam. Forthcoming in *Transportation Research*, 29A, 2, pp.141-56.
Wilson, J. D. (1983): "Optimal Road Capacity in the Presence of Unpriced Congestion". *Journal of Urban Economics*, 13, pp.337-57.

Appendix
The Optimal Common Fee with Cost Interdependencies

In equations (4) and (5) (Section 2), the optimal common congestion fee is derived for S different groups having their own level of intra-group congestion, ignoring the possibility of inter-group congestion effects. When allowing for congestion effects not only *within* groups but also *between* groups (such as congestion at road crossings when considering route-specific congestion, or groups with different trip lengths using one joint network), the Lagrangian in (4) may be rewritten as follows:

$$\mathcal{L} = \sum_{s=1}^{S}\int_{0}^{N_s} D_s(n_s)\mathrm{d}n_s - \sum_{s=1}^{S} N_s \bullet c_s(N_1,N_2,\ldots,N_S) + \sum_{s=1}^{S} \lambda_s \bullet [D_s(N_s) - c_s(N_1,N_2,\ldots,N_S) - f] \qquad \text{(A1)}$$

The cost functions (in the second term) have changed so as to allow the average cost of group $s(c_s)$ to depend not only on the size of the group itself, but (possibly) on all group sizes. Consequently, optimal discriminatory road pricing involves road prices r_s equal to:

$$r_s = \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t(\bullet)}{\partial N_s}$$

The optimal common fee can be found by solving the set of first-order conditions to (A1):

$$\frac{\partial \mathcal{L}}{\partial N_s} = D_s(N_s) - c_s(\bullet) - \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t(\bullet)}{\partial N_s} + \lambda_s \bullet \frac{\mathrm{d}D_s(N_s)}{\mathrm{d}N_s} - \sum_{t=1}^{S} \lambda_t \bullet \frac{\partial c_t(\bullet)}{\partial N_s} = 0 \quad s = 1,2,\ldots,S \qquad \text{(A2)}$$

$$\frac{\partial \mathcal{L}}{\partial f} = -\sum_{s=1}^{S} \lambda_s = 0 \qquad \text{(A3)}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_s} = D_s(N_s) - c_s(\bullet) - f = 0 \quad s = 1,2,\ldots,S \tag{A4}$$

Contrary to the problem in (4), it is not possible to solve for each λ_s in terms of $\partial D_s/\partial N_s$ and $\partial c_s/\partial N_s$ only. In matrix notation, the set of first-order conditions can be represented as follows:

$$\begin{bmatrix} \frac{dD_1}{dN_1} - \frac{\partial c_2}{\partial N_1} & -\frac{\partial c_1}{\partial N_1} & \cdots & -\frac{\partial c_s}{\partial N_1} \\ -\frac{\partial c_1}{\partial N_2} & \frac{dD_2}{dN_2} - \frac{\partial c_2}{\partial N_2} & \cdots & -\frac{\partial c_3}{\partial N_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial c_1}{\partial N_s} & -\frac{\partial c_2}{\partial N_s} & \cdots & \frac{dD_s}{dN_s} - \frac{\partial c_s}{\partial N_s} \\ -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_S \end{bmatrix} = \begin{bmatrix} f - \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t}{\partial N_1} \\ f - \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t}{\partial N_2} \\ \vdots \\ f - \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t}{\partial N_S} \\ 0 \end{bmatrix} \tag{A5}$$

By leaving out the bottom row of the matrix and of the vector to the right of the equal sign, we can use Cramer's rule and solve for each λ_s in terms of the remaining elements of the vector on the right and elements of the matrix. Next, using (A3), we may rewrite one of the conditions (A2) as:

$$f - \sum_{t=1}^{S} N_t \bullet \frac{\partial c_t}{\partial N_s} - \sum_{t \neq s} \lambda_t \bullet \left(\frac{dD_s}{dN_s} - \frac{\partial c_s}{\partial N_s} + \frac{\partial c_t}{\partial N_s} \right) = 0 \tag{A6}$$

It is then a matter of substitution of the expressions found for the terms λ_s and rewriting (A6) to find the expression for the optimal common fee. For two groups, the expression for the optimal common fee is still quite straightforward:

$$\begin{aligned} f = {} & \left(N_1 \bullet \frac{\partial c_1}{\partial N_1} + N_2 \bullet \frac{\partial c_2}{\partial N_1} \right) \bullet \frac{\dfrac{1}{\frac{dD_1}{dN_1} - \frac{\partial c_1}{\partial N_1} + \frac{\partial c_2}{\partial N_1}}}{\dfrac{1}{\frac{dD_1}{dN_1} - \frac{\partial c_1}{\partial N_1} + \frac{\partial c_2}{\partial N_1}} + \dfrac{1}{\frac{dD_2}{dN_2} - \frac{\partial c_2}{\partial N_2} + \frac{\partial c_1}{\partial N_2}}} \\ & + \left(N_1 \bullet \frac{\partial c_1}{\partial N_2} + N_2 \bullet \frac{\partial c_2}{\partial N_2} \right) \bullet \frac{\dfrac{1}{\frac{dD_2}{dN_2} - \frac{\partial c_2}{\partial N_2} + \frac{\partial c_1}{\partial N_2}}}{\dfrac{1}{\frac{dD_1}{dN_1} - \frac{\partial c_1}{\partial N_1} + \frac{\partial c_2}{\partial N_1}} + \dfrac{1}{\frac{dD_2}{dN_2} - \frac{\partial c_2}{\partial N_2} + \frac{\partial c_1}{\partial N_2}}} \end{aligned} \tag{A7}$$

For more than two cohorts, the expression for the optimal common fee becomes increasingly complicated, due to the appearance of an increasing number of cross-effects. For instance, for three cohorts, it can be shown that:

$$f = \frac{\displaystyle\sum_{s=1}^{3} \frac{\displaystyle\sum_{t=1}^{3} N_t \bullet \frac{\partial c_t}{\partial N_s}}{\displaystyle\prod_{u,v \neq s; u \neq v} \left[\left(\frac{\mathrm{d}D_s}{\mathrm{d}N_s} - \frac{\partial c_s}{\partial N_s} \right) \bullet \left(\frac{\mathrm{d}D_u}{\mathrm{d}N_u} - \frac{\partial c_u}{\partial N_u} + \frac{\partial c_v}{\partial N_u} \right) + \frac{\partial c_v}{\partial N_s} \bullet \left(\frac{\mathrm{d}D_u}{\mathrm{d}N_u} - \frac{\partial c_u}{\partial N_u} + \frac{\partial c_s}{\partial N_u} \right) + \frac{\partial c_u}{\partial N_s} \bullet \left(\frac{\partial c_v}{\partial N_u} - \frac{\partial c_s}{\partial N_u} \right) \right]}}{\displaystyle\sum_{s=1}^{3} \frac{1}{\displaystyle\prod_{u,v \neq s; u \neq v} \left[\left(\frac{\mathrm{d}D_s}{\mathrm{d}N_s} - \frac{\partial c_s}{\partial N_s} \right) \bullet \left(\frac{\mathrm{d}D_u}{\mathrm{d}N_u} - \frac{\partial c_u}{\partial N_u} + \frac{\partial c_v}{\partial N_u} \right) + \frac{\partial c_v}{\partial N_s} \bullet \left(\frac{\mathrm{d}D_u}{\mathrm{d}N_u} - \frac{\partial c_u}{\partial N_u} + \frac{\partial c_s}{\partial N_u} \right) + \frac{\partial c_u}{\partial N_s} \bullet \left(\frac{\partial c_v}{\partial N_u} - \frac{\partial c_s}{\partial N_u} \right) \right]}} \tag{A8}$$

Comparison of (A7) and (A8) with (5) shows how the cross-effects associated with inter-group congestion effects affect the weights in the optimal common fee. In Section 2, it was concluded that a group receives a larger weight, both the flatter its demand curve and its average cost curve. The reason was that the optimal weights should be such that the (private) welfare losses due to deviations of the common fee from the optimal individual fees are minimised. In the presence of inter-group congestion effects, the welfare losses due to such deviations obviously also depend on the extent to which a group influences congestion (and therefore welfare) in other groups. For two groups, a group's weight decreases, the stronger the cross-effect it has on the other group; the terms reflecting the cross-effects have a sign opposite to those reflecting own-welfare effects. This is easy to understand once it is realised that the decrease of a group's weight simply implies a relative increase of the other group's weight. In fact, (A7) says that the regulator's concern with welfare losses in a certain group A due to second-best intervention should be reduced, and his concern with losses in group B should be increased, (a) the smaller the welfare losses within group A itself due to such second-best intervention, and (b) the more strongly group A influences welfare in group B compared to the opposite effect. Equation (A8) reflects a comparable message, albeit that it is even more difficult to put the content of (A8) into words, due to the appearance of all the possible cross-effects in the expression.

Date of receipt of final manuscript: July 1994