European Regional Science Association  
36th European Congress  
ETH Zurich, Switzerland  
26-30 August 1996

Erik Verhoef  
Department of Spatial Economics, Free University Amsterdam,  
The Netherlands  
Fax: ++31 20 4446005  
everhoef@econ.vu.nl

ECONOMIC EFFICIENCY IN SECOND-BEST  
REGULATION OF ROAD TRANSPORT  
EXTERNALITIES

ABSTRACT: This paper studies the regulation of road transport externalities from the perspective of economic efficiency. The growing concern for the environment has induced a large body of literature on the economics of the environment and the regulation of externalities. Although many of the principles developed equally hold for transport as for any other sector, some particular features of road transport deserve closer consideration. Two of these will be addressed in this paper. These include the fact that road transport takes place in a network environment, so that regulation on a part of a network may give rise to spill-over effects on an unregulated part; and secondly the fact that transport demand often is a derived demand, necessitating consideration of interdependencies between transport and the spatial economy in formulating regulatory transport policies. Therefore, the paper provides insight into the efficiency of second-best regulation of road transport externalities, taking the theoretical first-best Pigouvian solution as a bench-mark for evaluating these more realistic second-best pricing strategies.
1. Introduction

The regulation of road transport externalities has come to the forefront as one of the most important issues in contemporary transport policy debates (Banister and Button, 1993; Button, 1993a, 1993b; Nijkamp, 1994; Verhoef, 1994a). This is, in part, a direct consequence of the increasing general environmental awareness that has arisen since the 1970's. Transport affects the local and global environments in many ways, and for a number of pollutants, the transport sector, in particular road transport, is one of the most important contributors to environmental externalities. Furthermore, transport is one of the main causes of noise annoyance, and congestion has become one of the main problems facing urbanized areas nowadays. Last but not least, accidents and fatalities due to road transport can be mentioned as the fourth main negative side effect of road transport.

Therefore, notwithstanding the central role that transport plays in modern societies, it is increasingly recognized that current and predicted road transport activities pose potentially excessive pressures on social and ecological environments, and more stringent regulation of road transport seems necessary. This is reflected in policy plans at various spatial levels, such as the European level (EC, 1992a, 1992b, 1996), the national level [for The Netherlands, the most recent national policy plan in this context is the so-called Second Structure Scheme on Traffic and Transport, abbreviated in Dutch as SVV-II (Tweede Kamer der Staten-Generaal, 1989-90)], and the local level. In such policy plans, pricing policies are often mentioned as indispensable elements in the regulation of road transport externalities.

The growing concern for the environment has induced a large body of literature on the economics of the environment and the regulation of externalities (see Baumol and Oates, 1988; Pearce and Turner, 1990; Cropper and Oates, 1992; and Tietenberg, 1994). Although many of the principles developed equally hold for transport as for any other sector, some particular features of road transport deserve closer consideration. As a matter of fact, already back in the 1920's, economists like Pigou (1920) and Knight (1924) recognized in the context of road transport that marginal external cost pricing offers the first-best solution for optimizing an external cost - congestion in their examples. Also for environmental externalities, such marginal external cost pricing offers the theoretically optimal regulatory policy. After seventy years therefore, economists’ answers to questions of externality regulation, also in road transport, typically still rely heavily on this concept of Pigouvian taxes.
Under the rather stringent assumptions of first-best conditions elsewhere in the economic system and perfectly flexible regulatory policies for coping with road transport externalities, there would indeed not be much scope for improving on the Pigouvian solution to the problem of external costs of road transport, at least not when the aim of the policies is merely to achieve Pareto optimal levels of road usage. Unfortunately however, such assumptions are usually not satisfied in reality. This paper studies the efficiency of regulating road transport externalities, concentrating on two particular second-best situations that are likely to occur in the practice of policy making.

The first second-best case, to be studied in Section 2, concerns the implications of the fact that road transport takes place in a network environment, so that regulation on a part of a network may give rise to spill-over effects on an unregulated part. Next, Section 3 acknowledges that transport demand is often a derived demand, necessitating consideration of interdependencies between transport and the spatial economy in formulating regulatory policies.

In this way, the paper provides insight into the efficiency of second-best regulation of road transport externalities, taking the theoretical first-best Pigouvian solution only as a bench-mark for evaluating these more realistic second-best pricing strategies. The paper concentrates on efficiency; issues of equity and social feasibility in the regulation of road transport externalities are discussed in, for instance, Verhoef (1995) and Verhoef, Nijkamp and Rietveld (1995a, 1996a, 1996b). 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 (1990a) on uniform or step wise pricing of a bottleneck; and Arnott (1979), Sullivan (1983), and Fujita (1989, ch. 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. Seminal papers on the theory of road pricing as such are Walters (1961), Vickrey (1969), and Arnott, De Palma and Lindsey (1993). Two recently edited books on this topic are Johansson and Mattsson (1995) and Button and Verhoef (1996). Finally, 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). Mayeres and Proost (1995a,b) in fact study optimal taxation and marginal tax reforms in the context of transport externalities.
2. Network effects and the regulation of road transport congestion

A specific problem in the regulation of road transport externalities is that the activity takes place in a network environment. This may give rise to unforeseen and sometimes unwarranted spill-over effects when regulation only takes place on a limited part of the network (see also Nijkamp and Shefer, 1996; and Bernstein and El Sanhouri, 1994). Clearly, as people will generally prefer free (or at least cheap) alternatives, the resulting choice processes are particularly interesting when considering congestion. There are various reasons why untolled alternatives may be present in practice. First, the regulator may leave a route untolled for equity reasons; for instance to protect low-income groups from having to pay fees, or to increase the social feasibility of road pricing (see Starkie, 1986). Alternatively, untolled alternatives may be present when (electronic) toll experiments are being undertaken. Furthermore, the same type of situation prevails with the occurrence of so-called ‘rat-running’: drivers using escape routes in order to avoid certain toll-points. On the other hand, the cost of toll collection may actually justify the choice of not regulating an entire road network, but only some of its major links instead.

Pigou (1920) and Knight (1924) were probably the first to discuss congestion pricing with two routes (one of which, in their case, was assumed to be uncongested). Later on, Lévy-Lambert (1968) and Marchand (1968) were the first to derive the optimal one-route toll with an untolled congested alternative. Recent discussions of two-route problems in the context of the dynamic bottleneck model with inelastic demand can be found in Arnott, De Palma and Lindsey (1990b, 1992) and Bernstein and El Sanhouri (1994) - although the latter actually also do consider elastic demand, but not with the second-best optimal one-route toll. In the present analysis, elasticity of demand, for instance resulting from the presence of alternative transport modes, can easily be considered, and will actually turn out to be of crucial importance for the efficiency of one-route tolling. On the other hand, dynamic departure time decisions will be ignored, which renders the present models supplementary to the latter three mentioned above.

2.1 Second-best one-route tolling

This section discusses some basic welfare economic properties of congestion pricing with an untolled alternative. A simple network is considered, with two competing, congested routes: one tolled (route T), and one untolled (route U). It is assumed that the regulator wishes to set the fee on the tolled route so as to maximize efficiency under the inherent limitation of not tolling the other route. In doing so, some sub-goals related to overall efficiency have to be traded off. These are (1) an overall demand (‘modal split’) effect, being the extent to which road users efficiently
leave the road system altogether due to congestion pricing; and (2) a route split effect, being the extent to which the remaining road users divide themselves efficiently among both routes. Generally, as one of the routes remains untolled, it will be impossible to realize the first-best situation where both effects are optimized (see also Bernstein and El Sanhouri, 1994).

In the problem’s purest form, the public regards the two alternative routes as perfect substitutes. The model therefore contains one single demand function $D(N)$, where $N$ denotes the total number of road users (on both routes); and two average user cost functions $c_T(N_T)$ and $c_U(N_U)$, where naturally $N=N_T+N_U$, and where average user cost and also the value of time are assumed to be equal for all road users. In line with Wardrop’s first principle (Wardrop, 1952), at any equilibrium should then the average cost on route $U$ be equal to the average cost on route $T$ plus the one-route fee $f$; otherwise people would shift from the one route to the other. Furthermore, both should be equal to marginal benefits $D(N)$.

Assuming that the regulator aims at maximizing total benefits, as given by the area under the demand curve, minus total costs, he has to solve the following Lagrangian:

$$L = \int_0^N D(n) \, dn - N_T \cdot c_T(N_T) - N_U \cdot c_U(N_U)$$

$$+ \lambda_T \cdot \left( D(N) - c_T(N_T) - f \right) + \lambda_U \cdot \left( D(N) - c_U(N_U) \right)$$

with: $N = N_T + N_U$  \hspace{1cm} (1)

which can be solved for the following optimal one-route fee:

$$f = N_T \cdot c_T'(N_T) - N_U \cdot c_U'(N_U) \cdot \left( \frac{-D'(N)}{c_U'(N_U) - D'(N)} \right)$$

The first term in (2), equal to the marginal external congestion costs on route $T$ in the second-best optimum, captures the direct impact of the fee on congestion on the tolled route itself. However, the second term indicates that, for optimal use of the fee, one should also take account of the ‘spill-over’ effects on the untolled route by subtracting some non-negative term, which is a fraction of the marginal external congestion costs on the untolled route in the second-best optimum. This fraction, which may range between 0 and 1, is given by the term between the large parentheses, and depends on the (relative) values of $D'(N)$ and $c_U'(N_U)$; that is, on the slopes of the demand curve and the average cost curve on the untolled route in the second-best optimum.
2.2 A simulation model

Below, the outcomes of some simulations will be discussed that were performed in order to arrive at some more explicit results than the general specification in (2) allows. When switching towards explicit functions, one is soon confronted with very tedious expressions, depending of course on the functional forms chosen for the demand and cost functions. As there is no theoretical reason to prefer any of the functional forms possible, and in order to keep the analysis manageable and the outcomes tractable, the simulation model is kept as simple as possible by assuming that these functions are linear over the relevant ranges considered (that is, the range containing the non-intervention, second-best, and first-best levels of usage). Although the use of linear functions may be criticized, they are in any case sufficient to serve the general goal of the simulations, being the assessment of the influence of some key factors related to demand and cost structures on the relative performance of one-route tolling. Also, it is worth mentioning here that Arnott, De Palma and Lindsey (1990b, 1992) have pointed out at several occasions that the linear congestion cost function is not necessarily unreasonable, since it can be interpreted as a reduced form representation of the Vickrey (1969) bottleneck model.

The simulation model then consists of one joint linear demand function, characterized by slope $\alpha$ and intersection $\delta$:

$$D = \delta - \alpha \cdot (N_T + N_U)$$  \hspace{1cm} (3)

Next, for both routes ($i=T,U$), the marginal private cost $\text{MPC}_i$, equal to average social cost $\text{ASC}_i$, consists of a free-flow cost component $\kappa_i$, and a congestion cost component which is assumed to be proportional to total usage $N_i$ with a factor $\beta_i$:

$$\text{MPC}_i = \text{ASC}_i = \kappa_i + \beta_i \cdot N_i; \hspace{0.5cm} i = T, U$$  \hspace{1cm} (4)

All parameters are non-negative; and only 'regular' networks will be considered, where both routes are at least marginally used under non-intervention, first-best, and second-best regulation. For the 'base case' of the model, the following parameter values were chosen: $\alpha=0.01; \delta=50; \kappa_T=\kappa_U=20; \beta_T=\beta_U=0.02$. So, both routes are assumed to be identical in the base case. No surprise then, that equilibrium usage under non-intervention is equal on both routes: $N_T=N_U=750$. Marginal private cost amounts to 35.00 on both routes; marginal social cost to 50.00. Under first-best regulation, optimal road prices of $r_T=r_U=10.00$ are found for both routes, with marginal private cost then amounting to 30.00, and marginal social cost to 40.00 on both routes. The optimal road prices for both routes are therefore equal to the difference between these two, as theory dictates. Optimal usage is 500 on both routes. Under second-best one-route
tolling, the second-best optimal fee for route T is 5.45. Marginal private cost is 30.91 on route T, and 36.36 on route U; the difference is exactly equal to the additional fee on route T, so that user equilibrium is indeed guaranteed. Marginal social cost is 41.82 on route T, and 52.73 on route U, readily demonstrating a non-optimal route split: road usage is 545 on route T and 818 on route U.

By varying the model’s respective parameters, it is possible to gain insight into the conditions under which one-route tolling is a relatively (un)attractive option. Some of the most outstanding results are presented below. In doing so, the index of relative welfare improvement \( \omega \) will be used, defined as the ratio of the overall welfare gain under second-best regulation compared to non-intervention, and the overall welfare gain under first-best regulation compared to non-intervention. A value of 1 therefore implies that second-best one-route tolling is as efficient as first-best two-route tolling, whereas a value of 0 implies that no welfare gains can be realized with two-route tolling, leaving it as inefficient as non-intervention. In the base case, \( \omega \) is equal to 0.273, indicating that 27.3% of the potential efficiency gains under two-route tolling will be realized with one-route tolling.

2.3 Varying cost parameters

First, consider the free-flow cost parameters, for instance reflecting the length of the links. Figure 1 gives the course of the index of relative welfare improvement \( \omega \), and of the optimal first-best and second-best fees, for an increasing difference between these parameters. Instead of varying just one of the parameters, \( \kappa \), was raised while simultaneously lowering \( \kappa \); both by 1.5 each step, the base case of \( \kappa = \kappa = 20 \) being in the centre. First of all, although the congestion parameters are equal for both routes, first-best tolls on both routes do not coincide. This may be surprising at first sight, as one might expect the level of free-flow cost to be a purely ‘internal’ cost component, without any impact on optimal fees. However, due to the fact that road users distribute themselves among the two routes such that marginal private cost, including the internal congestion cost component, are equalized, there is a direct effect of differences in free-flow costs on the optimal tolls. Generally speaking, the lower the free-flow cost, the higher the internal congestion cost, and hence also the higher the external congestion cost will be. This is illustrated in Figure 1 by the courses of the first-best fees.
Due to this effect, also the optimal one-route toll and the index of relative welfare improvement show an interesting pattern in Figure 1. When the free-flow costs on route T grow sufficiently high compared to those on route U, on the right-hand side of the figure, the optimal one-route toll may actually turn into a subsidy: $f$ is negative. Marginal social costs on route U are then so much higher than those on route T that it is even worthwhile attracting some new traffic as a negative side-effect to the main aim of diverting traffic from route U to route T. For obvious reasons, first-best congestion tolls will never be negative. At the turning point, where $f$ changes sign and is equal to zero, the index of relative welfare improvement therefore reaches its theoretical minimum of zero. Sticking to the range where the optimal one-route toll is a tax, it is clear that second-best regulation becomes more attractive the lower the free-flow costs on route T compared to those on route U. This is due to the fact that the regulator then controls that route which is in the first place more important in terms of usage, and secondly, where market forces tend to give rise to the largest congestion externalities. The first of these two reasons is illustrated by the fact that the optimal one-route toll approaches the optimal first-best toll on route T in these cases.

Also for varying differences in the congestion cost parameters $\beta_i$ while keeping $\kappa_T=\kappa_U=20$ [not presented graphically here, but the interested reader is referred to Verhoef, Nijkamp and
Rietveld (1994)], one-route tolling is more efficient, the higher $\beta_u$ in comparison to $\beta_T$. Apart from the fact that the regulator then again controls the more important route, route $U$ then becomes an increasingly unattractive alternative for route $T$ because of the internal part of the congestion cost, which makes the occurrence of adverse spill-overs due to regulation on route $T$ less likely. In contrast with Figure 1, for any difference in the congestion cost parameters, the first-best fees will be equal for both routes. The reason for this again perhaps counter-intuitive result becomes clear after considering the equilibrating effects of user behaviour. In the first-best optimum, marginal social cost should be equalized between the two routes. Given the fact that road users distribute themselves over both links such that average social cost are equalized, and given the equality of free-flow costs and the linear form of the marginal cost functions, the result follows. Finally, with equal free-flow costs, $f$ will not fall below zero.

In conclusion, for both types of cost parameter, the regulator should preferably perform one-route tolling on the lower cost route.

2.4 Varying demand characteristics

Next, the demand parameters are considered. Figure 2 shows what happens when 'tilting' the demand curve around the original non-intervention equilibrium, doubling the slope at each step. Both the slope $\alpha$ and intersection $\delta$ therefore change simultaneously, in order to avoid ending up with very large (small) markets when demand approaches perfect (in-) elasticity. Both routes are again assumed to be identical in terms of cost functions, and it should therefore be no surprise that the optimal first-best tolls are equal for both routes in all cases.

On the left-hand side of Figure 2, low values for $\alpha$ are found, indicating high demand elasticities. The regulator may in the extreme case of a flat demand curve ignore spill-over effects from route $T$ to route $U$, as road use on route $U$ remains unaffected by changes in $f$ due to the completely homogeneous and sufficiently large group of potential road users. The best thing to do in this case is to optimize usage of route $T$. This is also reflected by the fact that the optimal one-route fee on route $T$ is equal to the first-best optimal fee. Logically, the index of relative welfare improvement is 0.5 in this case. However, as demand approaches complete inelasticity, regulation should more and more concentrate on route split effects than on overall demand, suggesting increasing scope for one-route tolling. It may in that light seem odd that $\omega$ decreases when moving towards more inelastic demand.
The reason, however, is that with identical routes, the market itself will take care of optimal route split. Since, in this case, any one-route toll will therefore only act distortionary in this respect, its beneficial effects in terms of overall demand reduction are largely eroded. Put differently, the property of one-route tolling affecting route split is only useful in those cases where the market itself does not lead to efficient route splits, which it actually does when both routes are identical.

As a matter of fact, it is not even enough to introduce a difference in the congestion cost parameter to make one-route tolling only slightly efficient at inelastic demand. As already noted, differentials in congestion cost parameters do not affect the efficiency of market based route split. However, when free-flow user costs differ between the two routes, the effect suggested above is obtained, where one-route tolling at inelastic demand yields the same welfare improvement as does two-route tolling. In Figure 3, the tolled route is assumed to have the higher free-flow costs (this case could correspond to two highways between two cities, with the tolled highway being longer than the untolled highway). The optimal one-route toll at perfectly inelastic demand is a subsidy, exactly equal to the difference between the two first-best tolls, and yielding exactly the same welfare improvement: \( \omega = 1 \). As in Figure 1, the turning point where
the optimal one-route toll turns into a subsidy, so that $f=0$ and $\omega=0$, marks that specific unfavourable combination of parameters where the sub-goals of route split and overall demand are equally important for overall efficiency, but require opposite incentives.

When setting $\kappa_T - \kappa_U = -10$ instead of 10 (that is the case corresponding to the reasonable real-world situation where an arterial road parallel to a toll-road is not priced), optimal one-route taxes (no subsidies), and higher values of $\omega$ result throughout. The optimal one-route toll again equals the first-best toll on the tolled route in the limit of completely elastic demand, and equals the difference between the two first-best tolls in the limit of perfectly inelastic demand, with again $\omega=1$. For reasons of space, the corresponding graph is not presented here.

In conclusion, the lower the two cost parameters considered (a free-flow cost parameter and a congestion cost parameter) on the tolled route, the less unattractive one-route tolling becomes from an efficiency point of view. With identical routes, one-route tolling becomes less unattractive the more elastic the demand; when free-flow costs differ between the two routes however, one-route tolling also becomes attractive at inelastic demand, as it is then route split that determines the efficiency of regulation.
3. Transport, spatial economy and global environmental targets

In the previous analysis, road transport and the regulation of its externalities were considered in isolation. This is in line with most of the literature on the regulation of (road) transport externalities. However, such a partial equilibrium approach to transport ignores the fact that transport demand is often a derived demand, critically depending on issues such as spatial patterns of economic activity, and spatial and modal patterns of infrastructure supply. Therefore, ‘optimal’ levels of transport and ‘optimal’ Pigouvian transportation taxes derived in such partial analyses may in fact often suffer from considerable second-best biases, as first-best policies may require adaptations in the phenomena just mentioned. Indeed, considering transport in isolation is equivalent to assuming first-best conditions to apply for the entire spatio-economic system. Apart from this, emissions from different, economically related regions or sectors may often infringe on the same global environmental goals (such as emissions of greenhouse gases). As a consequence, regulation on the level of a sub-system may often, indirectly, either benefit from synergetic side-effects, or suffer from counterproductive compensatory effects in related sub-systems. It is important to investigate the potential impacts of such interdependencies upon the effectiveness of environmental regulation aiming at global targets, especially when performed under second-best circumstances.

This section is concerned with these issues. The analysis focuses on interdependencies between transport, spatial economy and the environment in the context of regulatory policies aiming at a global environmental target. For this purpose, a joint modelling framework of course has to be used, in which each of these elements can be included. The spatial price equilibrium (SPE) approach offers such a framework.

3.1 The SPE methodology and the simulation model used

The SPE methodology, first presented by Samuelson (1952) and further developed by Takayama and others (Takayama and Judge, 1971; and Takayama and Labys, 1986), has the property that equilibrating transport flows between two nodes come into existence as soon as the difference between nodal prices exceed transport costs. Both nodes can be shown to benefit from such trade, and also overall efficiency increases. Usually, and also here, SPE models are used to analyze spatial interactions in terms of commodity flows, with flexible prices clearing spatial excess demands and supplies for given transport cost structures and local demand and supply structures. A more general interpretation of SPE can embrace flows of production factors and
intermediates, and even passenger transport. For the present purpose, an advantage of the SPE approach is its close relation to traditional welfare economic modelling, thus lending itself to formulations in terms of welfare maximization, and derivations of associated optimal policies.

We consider the most simple SPE model possible, with two nodes (A and B), one good and one transport node. Figure 4 illustrates the workings of such a model. The left panel depicts the local demand and supply curves D_A and S_A for a certain good in node A, where Y_A gives local consumption and Q_A local production (note that Figure 4 is a back-to-back diagram, so that Y_A and Q_A increase when moving leftwards from the origin). In autarky (denoted with superscripts A), equilibrium is given by Q_A^A = Y_A^A, and the local market price P_A^A would prevail. The right panel shows the same for node B, where the autarky equilibrium is given by Y_B^A, Q_B^A and P_B^A. Suppose that transport cost between the nodes is equal to t, which is less than the autarky price difference between the two nodes. Then it will become profitable to transport some goods from the lower price to the higher price region. In Figure 4, it is assumed that P_B^A > P_A^A and that P_A^A - P_B^A > t. In order to determine the after-trade equilibrium (denoted with superscripts T), for both nodes R an excess demand/supply curves X_R(F_R) is constructed by horizontal substraction of the supply curve from the demand curve. Hence, for each after-trade nodal price P_A^T > P_A^A, X_A gives the net export F_A that node A would supply to node B; for P_A^T < P_A^A, negative values of F_A (hence,
positive net imports) imply that node A would be a net demander. Considering a closed system, the same holds for node B, and the after-trade equilibrium is given by \( F_A = -F_B \) and \( |P_A^T - P_B^T| = t \).

In Figure 4, \( P_A^T - P_B^T = t \), and \( Q_B^T - Y_B^T = F_B = -F_A = Y_A^T - Q_A^T \): node B is the net exporter.

For the purpose of studying interactions between transport and the spatial economy in the regulation of environmental policies, we consider a simulation model consistent with Figure 4. The two nodes are assumed to be spaceless points, so that transport only occurs between, not within the nodes. In addition, we consider one single type of environmental degradation. Production in both nodes leads to pollution, and so does transport, and each source afflicts the same global environment target (for instance the greenhouse effect). Environmental technologies in production and transport are exogenously given [see Verhoef, Van den Bergh and Button (1996) for endogenous environmental technologies].

This is probably the most straightforward case for studying sectoral and spatial interdependencies in the context of environmental policies. We assume that there is an overall environmental target of a maximum allowable level of emissions, denoted \( E^* \). This target may represent what is sometimes called the 'environmental utilization space' (Opschoor, 1992). We will investigate two sorts of policies aiming at reaching this target: first-best 'integral activity regulation' (denoted IAR), where the regulator controls the entire system, and second-best 'transport volume regulation' (denoted TVR), where he can only regulate the transport sector, but is still concerned with the overall environmental goal.

For the simulation model, we specify for both regions the following affine nodal demand and supply relations, which are in line with the 'quadratic welfare approach' to spatial price equilibrium analyses as discussed by Takayama and Judge (1971):

\[
D_R = d_R - a_R \cdot Y_R \quad R = A, B \tag{5}
\]

\[
S_R = s_R + b_R \cdot Q_R \quad R = A, B \tag{6}
\]

where all parameters and variables are non-negative. The environmental emissions model is as follows. There is one type of emission, which depends in a linear, source-specific fashion (represented by parameters \( \epsilon_i \)) on production and transport, where the transport volume \( T \) equals total trade \( |F_A^T| = |F_B^T| \). Hence, total emissions are given by:

\[
E = \epsilon_A \cdot Q_A + \epsilon_B \cdot Q_B + \epsilon_T \cdot T \\
\text{with: } T = |F_A^T| = |F_B^T| \tag{7}
\]

The environmental utilization space \( E^* \) is exogenously given, and the global target is met when \( E \leq E^* \). The first-best (IAR) policy mix of optimal nodal production taxes \( \pi_R \) and the optimal
transport tax $\tau$ can be found by maximizing net social welfare (the sum of region specific gross benefits $GB_A + GB_B$ measured as the area under the Marshallian demand curves, minus the sum of region specific production costs $PC_A + PC_B$, minus total transport costs $TC$); subject to the constraint implied by the environmental utilization space; given the market behaviour of the actors involved under regulation; and subject to appropriate non-negativity conditions concerning prices, production, consumption and transport. This Kuhn-Tucker problem can be represented as:

$$\text{MAX } GB_A + GB_B - PC_A - PC_B - TC + \lambda_E (E^* - E)$$

s.t. $\lambda_E \geq 0$; $E^* - E \geq 0$ and $\lambda_E (E^* - E) = 0$

individual maximizing behaviour under regulation
appropriate non- negativity conditions

where $\lambda_E$ gives the Lagrangian multiplier associated with the environmental constraint. This multiplier will in the sequel conveniently be referred to as the 'environmental shadow price'. This maximization problem is presented formally in Verhoef and Van den Bergh (1996). The following optimal production taxes $\pi_R$ and transport tax $\tau$ can be derived for the present model:

$$\pi_R = e_R \cdot \lambda_E \quad R = A, B$$

$$\tau = e_T \cdot \lambda_E$$

(9)  

(10)  

The optimal second-best (TVR) transport tax, where the regulator cannot affect environmental regulation in the nodes of origin and destination but aims at meeting the environmental constraint in the most efficient way, is given by:

$$\tau = e_T \cdot \lambda_E + \frac{e_O \cdot \lambda_E \cdot \pi^0_O}{1 + \frac{b_O}{a_O}} - \frac{e_D \cdot \lambda_E \cdot \pi^0_D}{1 + \frac{b_D}{a_D}}$$

(11)  

(see Verhoef and Van den Bergh, 1996). In (11), subscript O (D) denotes the node of origin (destination), and $\pi^0_R$ denotes the now exogenously given level of producer taxation in node R.

For the 'base case' of the simulations, the following parameter values are chosen. First, the demand functions for the good are assumed to be identical for both nodes, with $d_R = 80$ and $a_R = 0.5$. The production side of the two nodes are different, with $s_A = 25$ and $b_A = 1.5$; and $s_B = 5$ and $b_B = 0.5$: production is more efficient in node B than it is in node A. In autarky, $Q_A^\wedge = Y_A^\wedge = 27.5$ with $P_A^\wedge = 66.25$; and $Q_B^\wedge = Y_B^\wedge = 75$ with $P_B^\wedge = 42.5$. Transport costs are equal to 5, which is smaller than the autarky price difference, and hence equilibrating transport flows will exist. As expected,
in the trade equilibrium, B will be the net exporter: \( F_B = 30 \), which exactly compensates for the nodal imbalances implied by \( Q_A^T = 20 \) and \( Y_A^T = 50 \) at \( P_A^T = 55 \); and \( Q_B^T = 90 \) and \( Y_B^T = 60 \) at \( P_B^T = 50 \).

In comparison with the autarky situation, total welfare in node A increases from 756.25 to 925, and in node B from 2812.5 to 2925. Note that welfare is a narrowly defined concept, measured as the sum of consumer and producer surpluses, and not including any environmental values because environment is treated as a constraint rather than as a temporal externality. Both nodes thus benefit from trade; as they would with voluntary trade. By setting \( e_A = e_B = 10 \), and \( e_T = 15 \), total emissions of 1550 result in the unregulated trade equilibrium: \( E_A = 200 \); \( E_B = 900 \) and \( E_T = 450 \).

Transport accounts for ±30% of the emissions. The environmental utilization space \( E^* \) is set at 1000.

3.2 **Optimal regulation of transport in a first-best spatio-economic environment**

To illustrate the SPE methodology in combination with the environmental model, in the first simulation the two production structures are gradually interchanged. On the right-hand of Figure 5, the base case is found, whereas on the left-hand \( s_A = 5 \) and \( b_A = 0.5 \); and \( s_B = 25 \) and \( b_B = 1.5 \). Moving right, \( s_A \) (\( s_B \)) is increased (decreased) by 1 each step; and \( b_A \) (\( b_B \)) is increased (decreased) by 0.05 each step. The simultaneous variation of the four parameters is summarized along the horizontal axis by considering their impact on the autarky price difference \( P_A - P_B \). Given the identical demand structures, this simulation will yield symmetric results, with completely identical nodes (as reflected by \( P_A - P_B = 0 \)) in the centre.

Figure 5 focuses on environmental issues. From the curvature of the non-intervention emissions (NIE), it is clear that the more the two nodes differ, the higher will these emissions be, due to the induced transportation flows. Alternatively, when the two nodes are identical, in the centre of the figure, NIE fall within the environmental utilization space \( E^* \) and the environmental shadow price \( \lambda_E \) is zero. The basic relation between NIE, \( E^* \), and emissions and \( \lambda_E \) under integral activity regulation IAR is illustrated. As long as \( NIE < E^* \), no regulation is needed, and emissions under IAR are equal to NIE. As soon as \( NIE > E^* \), regulation becomes necessary in order to prevent emissions to exceed \( E^* \). This is reflected in a positive environmental shadow price \( \lambda_E \). The larger the difference of NIE minus \( E^* \), the higher this value of \( \lambda_E \).
The spatio-economic impacts of regulation, as well as some typical SPE characteristics, are

Figure 5. Emissions in the market based spatial equilibrium and under integral activity regulation, and the environmental shadow price

Figure 6. Production and transport in the market based spatial equilibrium and under integral activity regulation

The spatio-economic impacts of regulation, as well as some typical SPE characteristics, are
shown in Figure 6. First, with identical nodes and autarky prices, no trade takes place; when the autarky price difference exceeds the transport costs, the node with the lower autarky price becomes the net exporter. When NEI exceed $E^*$, free market activity levels are excessive. Figure 6 shows that the more different the nodes are, the larger the discrepancy between non-intervention and optimal levels of trade and nodal specialization. For the optimal spatial configuration, production and transport have to be increasingly restricted. Given the identical demand structures, this implies a relatively stronger restriction in production in the exporting node than in the importing node, as can be seen on both ends of Figure 6.

3.3 Second-best transport volume regulation

We now turn to the case where the regulator is not capable of affecting regulation in the two production sectors, but can only conduct transport policies to meet the environmental constraint. This could, for instance, correspond to the situation of a relatively small transit region, concerned with the impact of its 'through-put' on some global environmental amenity, but unable to influence environmental policies in the nodes of origin and destination directly. The Netherlands are a good example. Although such a regulator cannot directly affect production and consumption, its transport policies will indirectly affect overall production and consumption. In the simulation discussed below, we focus on how the underlying spatio-economic system might affect the efficiency and effectiveness of such second-best transport policies - as given in (17) - in comparison with the first-best situation where the regulator can set an optimal policy mix of transport and production taxes.

Figure 7 considers emissions and environmental shadow prices under both types of regulation. Along the horizontal axis, the emission coefficient in the node of destination $e_\lambda$ is raised from 0 to 45 (with a jump from 18 to 36); in comparison with the base case, $e_\tau$ is set at 5 rather than 15. Furthermore, $\pi_\lambda=\pi_\tau=0$. The underlying spatio-economic structure has an enormous impact on the performance and potential of second-best regulation. On the left-hand side, this structure is relatively favourable for such policies. Not only does second-best transport volume regulation TVR have a favourable direct impact on emissions of transport itself, it also induces a shift from consumption of imported goods towards the purchase of locally produced goods in node A - which are produced in a relatively environmentally friendly way compared to production in node B. When moving right, however, this favourable indirect effect of transport policies increasingly erodes. Up to the point where $e_\lambda=9$, this shows in an increasing discrepancy
between in the environmental shadow prices $\lambda_e$ for both policies. This shadow price not only depends on the extent to which NIE exceed the $E^*$, as illustrated by the gradual increase of $\lambda_e$ for IAR. In addition, this shadow price also depends on the efficiency and effectiveness of regulation itself.

When $e_A$ exceeds the value of 9, we end up in the range where TVR is no longer capable of meeting the environmental constraint. Optimal TVR then consists of the corner solution of prohibitive taxation with zero transport (Figure 8). This explains the kink in the curvature of $\lambda_e$ for TVR: TVR is no longer capable of meeting the environmental target. The effectiveness and efficiency of TVR increasingly falls short of those of IAR, as shown by the increasing difference of emissions and $\lambda_e$ for both policies in Figure 7. With prohibitive transportation taxation, total regulatory tax revenues will be zero. With IAR, internal solutions will generally result, implying positive tax revenues for the regulator (Figure 8).

When $e_A$ increases further, a point is reached where TVR becomes fully ineffective and inefficient. In this simulation, $e_A=40$ creates that particular unfavourable combination of parameters where transport regulation has no effect whatsoever on total emissions. The direct environmental impacts on emissions from transport are exactly compensated for by additional

---

Figure 7. Emissions and environmental shadow prices under integral activity regulation and transport volume regulation
emissions from increased local production in the node of destination; an increase induced by the transport policy itself. Here, $\lambda_e$ for TVR approaches infinity, reflecting the complete inefficiency of the policy.

When moving beyond this point, we end up in a third regime, where second-best TVR is in the form of transport subsidization rather than taxation. Transport taxation would be counter-productive, as it induces more emissions from production in the node of destination than it reduces transport emissions. As shown in Figure 8, the best thing the TVR regulator can do is to subsidize transport in a way that local production in node A is completely reduced to zero (in this case, TVR transport subsidies should be direction specific, which is never the case for transport taxes). The simulation results show that in this regime, welfare under TVR falls considerably, and is below welfare under IAR even though the environmental constraint is not met with TVR. Such TVR subsidization creates severe distortions in the spatio-economic system.

Although one might argue that this situation of TVR subsidization is quite extreme and unrealistic because one would never expect a transport regulator in the sort of transit region considered to actually subsidize transport for environmental reasons, the simulation also has im-

---

**Figure 8.** Regulatory tax rates and revenues under integral activity regulation and transport volume regulation
Figure 9. The effectiveness of transport volume regulation in a spatio-economic setting

important implications for a less ambitious transport regulator. The underlying spatio-economic equilibrium processes leading to the pattern seen in Figure 7 simply cannot be ignored, and will affect the effectiveness of any form of transport regulation. This is illustrated in Figure 9, where the impact of four different levels of transportation taxes ($\tau = -4, \tau = -2, \tau = 2$ and $\tau = 4$) on total emissions is seen for various levels of $e_A$. On the left-hand side, taxes have a very favourable impact on total emissions because of their direct effect on transport, as well as the indirect impact of stimulating a production shift from node B to A; the impact naturally being higher the higher the tax. With an increasing emission coefficient however, these impacts decline, and beyond $e_A = 40$, the transport regulator would find total emissions increasing more, the higher the transport tax charged. Transport subsidization is necessary if TVR is to reduce total emissions. If he is not inclined to subsidize transport, the best thing a regulator can do is to keep transport taxes at zero.

These simulations clearly demonstrate the sometimes unexpected effects of regulation when considered in the context of a full spatio-economic setting, including the interdependencies between the transport sector and the spatial pattern of economic activities.
5. Conclusion

This paper addressed two particular second-best situations that are likely to occur in the regulation of road transport externalities, and that necessitate adoption of 'naive' Pigouvian tax rules derived under implicit assumptions of first-best conditions.

Section 2 explicitly acknowledged the network environment in which road transport externalities arise, by studying the situation where regulation on a part of the network causes spill-overs on an unregulated part. The regulator then has to trade off a number of 'sub-goals' contributing to the overall goal of efficient usage of the network, comprising overall demand as well as route split effects. As a consequence, the optimal second-best fee may turn out to be a subsidy. A simulation model showed the impact of a number of key parameters on the relative efficiency of such policies.

Section 3 discussed a spatial price equilibrium model for studying the interdependencies between transport, infrastructure and spatial economy in the formulation of environmental policies aiming at global sustainability targets. First-best and second-best policies were considered. An endogenous 'environmental shadow price' arose from this model, which not only depends on the extent to which non-intervention emissions exceed the target, but also on the 'quality' of the environmental policies pursued to meet this target. Hence, this shadow price partly reflects the social costs of inefficient regulation.

Consideration of such second-best situations is important now that pricing solutions for the regulation of road transport externalities are not unlikely to become practice in the near future. Such pricing rules are often more complex than is suggested by the basic economic exposition of the problem of externality generation. In some cases, second-best optimal taxes may even be negative (that is: subsidies). The importance, therefore, of the sort of analyses presented above in the formulation of regulatory policies can hardly be overestimated.

Notes
1 This section draws from Verhoef, Nijkamp and Rietveld (1995d).
2 This section draws from Verhoef and Van den Bergh (1996) and Verhoef, Van den Bergh and Button (1996).

References
Arnott, R., A. de Palma and R. Lindsey (1990b) "Departure time and route choice for the morning commute" Transportation Research 24B (3) 209-228.


Verhoef, E.T., P. Nijkamp and P. Rietveld (1995c) "The economics of regulatory parking policies" *Transportation Research* **29A** (2) 141-156.


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