Marginal Cost Based Pricing in Transport
Key Implementation Issues from the Economic Perspective

Erik Verhoef, Free University Amsterdam

Paper prepared for IMPRINT

1. Introduction

Marginal cost pricing in transport is a 'hot' topic, in at least two senses. First, as is well known, over the last decade(s), sophisticated pricing policies in transport have evolved from a primarily academic, theoretical construct, to a realistic and seriously considered option for many areas – urban and non-urban – around the world. This is due to (at least) two simultaneous, interacting developments, *viz.* the steady growth in transport related problems such as congestion and emissions on the one hand, and the development of technologies enabling automated charging on the other. So, marginal cost pricing in transport is 'hot' in the sense that many governments, at different spatial levels, seriously explore the possibilities for implementing some form of pricing policies aimed at the containment of transport-induced externalities. At the same time, such proposals are rarely met by great public enthusiasm, making it a 'hot' topic from the political viewpoint in that policy makers might easily burn their hands when proposing to drastic pricing reforms in transport. A very common result is that proposals for pricing schemes often end up in the proverbial wastebasket long before a first penny was to be actually charged.

Apparently, the implementation of marginal cost based pricing in transport is not as straightforward as it may seem after calculating the net social benefits that can be realized with it, in theory enabling the government to make everybody better of with the policy than without it. Transport analysts have acknowledged the implied paradox, and much research has recently been done – and is still underway – that aims at identifying, explaining, and sometimes proposing solutions to, the key barriers that may prevent a smooth and easy implementation of pricing policies. At the same time, transport analysts to an increasing extent have acknowledged that indeed, the conceptual black-board models often used for illustrating the motivations for and 'blessings' of transport pricing certainly may be of great pedagogical and illustrative value, but nevertheless often leave the policy maker largely empty handed when it comes to the design of pricing schemes for realistic situations, on real transport networks. A second, also voluminous stream of recent research is devoted to the latter set of questions: how should a transport pricing scheme be designed when the unrealistic, idealized conditions underlying the black-board exposition 'happen' not be fulfilled – as will typically be the case in reality?

This contribution aims at discussing these and other implementation issues for marginal cost based pricing in transport. The contribution will try to provide a concise overview, and many ideas and discussions will be based on various previous publications, including deliverables prepared for the projects AFFORD and MC-ICAM from the EU's RTD-programme and recent literature reviews, (co-)authored by the author of this paper. For reasons of readability, reference to these publications will not be made – the interested reader may consult the small bibliography at the end of this paper to find the publication details for these often somewhat more technical expositions and use them as a reference. Also for reasons of readability, explanations of some of the more technical economic terms that may sometimes prevent the non-economist reader from following the discussion are likewise relegated to a separate section at the end of the paper. The paper starts by providing a very brief discussion of the economic backgrounds of pricing and revenue use in transport. Next will follow a discussion of the various types of constraints that prevent textbook theory to be readily applicable in reality in Section 3. Second-best pricing, the logical result of such constraints, will be addressed in Section 4, while Section 5 discusses some principles relevant for combining second-best pricing and revenue use measures in 'policy packages'. Section 6 discusses 'policy packaging' from a dynamic perspective, when discussing implementation paths. Section 7 concludes.

2. Economic backgrounds and motivation of marginal cost based pricing

The basic economic motivation for applying marginal cost based pricing in transport is relatively straightforward to explain. The standard exposition considers automobile use as an application, and uses the well-known market diagram that identifies the free-market equilibrium at the intersection of the marginal benefit (or demand) and marginal private cost function, and the optimum at the intersection of the marginal benefit and the marginal social cost function. When the existence of marginal external costs (not borne by the decision maker herself, but imposed on others instead) drives a wedge between marginal private and marginal social costs, the free-market equilibrium and the optimum will not coincide. A (so-called 'Pigouvian') tax could then be used, equal to marginal external costs, which makes sure that the consumer is confronted with both the private costs *and* (a tax equal to) the external costs of her decisions. This is often called the 'internalization' of external costs. As a result, the market equilibrium with taxes would then correspond with the optimum, which is why this situation is often referred to as the 'decentralized optimum'. The external costs of road transport that cause such an analysis to be relevant comprise congestion, accidents, environmental pollution and noise annoyance. Indeed, although external congestion costs and at least a part of external accident costs will remain ‘internal’ to the collective group of automobile users, the relevant level of aggregation when distinguishing between internal and external costs is that of the individual, implying that the value of time losses and accident risks that the one car driver through her presence on the road imposes upon another are indeed to be considered as external costs. A primary motivation for the use of marginal cost based pricing is thus that individuals, when confronted with prices that reflect the marginal social cost of their choices, would exhibit behaviour that not only maximizes their own utility given the prices they face, but at the same time that the resulting equilibrium would be the one that maximizes the social objective in the sense that any deviation from that equilibrium would imply that either the extra social costs would exceed the extra social benefits (when increasing the level of consumption), or that the social benefits foregone would exceed the social costs foregone (when decreasing the level of consumption).

In addition to this primary motivation, other important considerations have been put forward for the use of prices over different policy measures, such as command-and-control policies. One of these – discussed in greater detail in AFFORD – concerns the fact that under optimal pricing, optimal incentives are given in achieving certain goals in a socially cost-effective manner when individuals have multiple ‘margins of behaviour’ they can adjust. In other words, marginal cost pricing would not only make sure that the short run optimum for a market with given demand and cost functions is attained, but in addition that in the long run, these functions themselves would be pushed to their optimal ‘shape and position’. A simple example may clarify this: suppose that automobile users could reduce their emissions by either driving less given the vehicle they posses, or by choosing a more environmentally friendly vehicle. Under marginal cost pricing, incentives are given such that the automobile users make a socially optimal trade-off between these two possibilities, in the sense that every unit of emissions avoided will be avoided in a way that minimizes the associated cost (or foregone benefits). Hence, the short-run marginal external cost function will – through the choice of optimal technologies – be pushed to its optimal position. Although this may seem a somewhat theoretical advantage, it is in fact extremely important in reality. A derived advantage would namely be that when using marginal cost based pricing policies, the government does not have to predict what the exact properties of the long run optimum would be in order to achieve it. Clearly, if the government would avoid the use of pricing and would instead try to achieve the optimum by a package of direct regulatory policies that should ‘imitate’ its properties (including for instance technological standards for vehicles, prescribed patterns of spatial organization, and allowed levels of mobility), the informational requirements would be enormous and the risk of government failures would be huge. It is certainly not inconceivable that such long run advantages of marginal cost pricing would by far outweigh the short run advantages as predicted by empirical models that ignore long run impacts of pricing in terms of locational behaviour, technology choice, *etc*.

A third motivation for the use of pricing in general has often been the need to raise the revenues necessary for the financing of infrastructure capacity. It is on purpose that here we do not speak about a motivation for the use of 'marginal cost based pricing', but of 'pricing in general', instead. For public transport, for instance, it is believed that its specific cost structure – when exhibiting scale economies – may often cause strict marginal cost pricing to raise insufficient revenues to cover all costs involved. It is even questionable whether the variable operating costs alone could be covered from the revenues generated with marginal cost pricing, let alone the often substantial fixed costs involved in the supply of infrastructure capacity. For road transport, a famous result – originally formulated by Mohring and Harwitz – states that the revenues from optimal congestion pricing should be just sufficient to cover the capital costs involved in the supply of the optimal level of infrastructure when certain technical conditions are fulfilled. Here, too, however, a balanced transport budget may still be distorted as soon as either these conditions are not fulfilled (*e.g.* when capacity can only be supplied in discrete quantities, such as numbers of lanes), or – also not inconceivable – when a part of the capacity is considered to be already fully financed, with no (interest) costs being to be covered from current toll revenues. Therefore, the raising of revenues would be a motivation for pricing in general, but the application of marginal cost pricing does certainly not necessarily mean that given revenue targets will indeed exactly be met. This, however, is not a specificity of transport. For any economic market, marginal cost pricing would lead to an exact coverage of social costs only when certain technical conditions are fulfilled, an important one being the occurrence of constant returns to scale, reflected in constant average costs which then by definition are equal to marginal costs. Evidently, as soon as efficiency of prices would require prices that deviate from those required for full cost recovery, a choice has to be made as to which goal is more important. Economic theory cannot help in defining the goal that should be dominant – this is a political choice – but it can help in identifying ways in which certain goals can be achieved in a socially cost-effective, efficient manner (*e.g.*, finding other ways in raising necessary funds at a socially cost minimizing manner).

As a fourth motivation, we mention the fact that governments need to raise tax revenues also for goals outside the transport sector, such as the provision of public goods like national defence. The taxes currently used for this purpose will typically be distortive, implying that one motivation for the use of taxes in transport could be that it allows to lower taxes that are, at the margin, more distortive than the transport taxes that would be set if such revenue raising aspects were ignored. We will return to this issue in Section 4, when discussing second-best pricing in transport.

Related to this, a final and fifth motivation that we wish to address concerns the fact that other taxes are typically motivated by the government's desire to redistribute wealth over different groups in society (often from richer to poorer groups). As any transport tax would have differentiated effects over different social groups, the same redistributive objectives can be pursued at least to some extent using transport taxes. Clearly, the idea here is not to propose that transport taxes and the use of the revenues should be used as the only instrument in pursuing redistributive goals. The argument is much more subtle than that: if the government does currently use distortive taxes and subsidies to redistribute income, small changes in purely externality based transport taxes may give the opportunity to take over part of this task, and when the existing distortive taxes and subsidies are reduced by corresponding small amounts, the same redistribution of income may be achieved against a smaller overall social cost in terms of efficiency losses.

The final two motivations in fact already brought us to the area of second-best pricing, to be discussed in greater detail in Section 4. The practical relevance of second-best pricing – for which optimal prices will typically deviate from the marginal social costs on at least some of the market (segments) priced – is the reason that in this note, we have used the phrase 'marginal cost *based* pricing', rather than simply 'marginal cost pricing'. This reflects that under second-best circumstances, prices set equal to marginal external costs will typically not be the same as the prices that optimize the social objective, and that require second-best distortions to be taken into account.

We will end this section by making a final remark on the motivation of marginal cost based pricing. Regardless of whether we consider first-best or second-best pricing, marginal cost based prices are motivated by the fact that these optimize the social objective under the constraints that happen to apply. It is important to realize that analysts in general and economists in particular who advocate marginal cost based pricing do so on the grounds that it is taken for granted that the government seeks to realize its objective(s) in a socially cost-effective way; in other words, social benefits (or utility in general) and costs should be the main arguments in the social objective for marginal cost based pricing to be a relevant concept. The choice of the objective function itself is a political choice, not at the discretion of analysts. The reason why it is important to raise this point here, is that marginal cost based pricing in transport is sometimes criticized for the 'unacceptable' redistributive impacts it may have. Such criticisms are impossible to reject by economists, as the simple reply would be that if income distribution is indeed important, this should be reflected in the specification of the social objective in the first place. Given such a social objective, reflecting the desired income distribution for instance using appropriate welfare weights, the policy conclusion would not be any different than sketched above; *i.e.* its maximization would again require the use of marginal cost based transport taxes. In other words, objections against marginal cost based transport prices on the basis of equity arguments can best be interpreted as objections against the social objective used when deriving these prices, and not as an objection against the concept of marginal cost based pricing *per se*.

3. Constraints on pricing

It is one thing to draw up a transport model – be it a conceptual small model or a large realistic network model – and compute optimal prices; it is quite a different thing to design a pricing scheme for a real-world situation, as required for actual implementation. In the former case, prices or taxes are a model instrument, that can fully be controlled by the modeller. In the latter case, all sorts of constraints and barriers may exist that prevent a regulator from charging the prices that it ideally would like to charge. Under such conditions, the regulator has to resort to 'second-best' pricing: setting the prices that are available optimally, under the constraints applying. Such constraints may take various forms. We mention a few important ones, in arbitrary order.

Technological and practical constraints

Ideally, prices should be allowed to vary according to the same dimensions as do the marginal external costs that are to be internalized using these prices. For instance for automobile users, this would involve charges that vary continuously over time, place, route chosen, driving style, type of vehicle and its technical state, driving style, etc. It is clear that the resulting pricing scheme may be too complex to be understood by car drivers, and requiring too sophisticated a pricing and monitoring technology than currently available. For instance, for realistic road pricing schemes, one would expect differentiation over users classes to be possible only for a crude distinction into passenger cars, vans and trucks; over time up to the level of a few steps during the peak and one level outside it, at a maximum; and tolls to be charged on a few key-roads (highways only?) in the network, only.

Acceptability constraints

A move towards marginal cost based pricing may induce so much social resistance and (technological) uncertainty, that it may seem preferable to start with a few small-scale demonstration projects. These typically create constraints that reflect that prices will be in operation only on a small part of the network, leaving other links unpriced. Alternatively, acceptability concerns may create a situation in which for instance a constraint on the maximum level of charges, or limits on charge variability and/or predictability, is pre-specified.

Institutional constraints

Various types of institutional constraints may limit the possibility for a regulator to set the prices it would ideally like to charge. One example is where a local or regional government either cannot affect some transport charges that are set by a higher level government (*e.g.* fuel taxes), or has to accept lower and/or upper limits on charges allowed, set by a higher level government. Another example is when the government in one jurisdiction cannot affect the prices charged by a neighbouring jurisdiction, while trans-boundary traffic and/or externalities are relatively important. The two governments may then end up in some form of tax competition. Other problems may arise when public transport is operated by a private party who is relatively free in choosing prices and service levels but does so in a non-optimal way, or when transport pricing would become more efficient if other government bodies – for instance, the Ministry of Finance – would adjust taxes and subsidies (*e.g.* on commuting cost tax deductions) that are outside the control of the transport authority in charge of the transport prices – for instance, the Ministry of Transport.

Legal constraints

It may not always be possible to charge the ideal prices on the basis of legal arguments. For instance, suppose that the law implies that the level of taxes should be predictable to the tax payer. If congestion is to some extent unpredictable (which it is in reality, *e.g.* due to weather conditions or road work), also the optimal congestion charge would vary not only over the day but also between days. A legal constraint could then exist that prevents the latter type of variation from being implementable.

Financial constraints

Another type of constraint limiting the freedom of choice in setting prices could be the prior definition of minimum or maximum (tax revenue) sums to be collected, or similar financial constraints. For instance, for public transport, the government may require full cost recovery (implying minimum charges) while at the same time not allowing large profits (implying a maximum). Similar types of constraint may be imposed on road taxes; *e.g.*, the governmental budget surplus (possibly after making reservations for investments an maintenance) would not be allowed to exceed a certain percentage of total revenues.

Market interaction constraints

Transport has – arguably more than any other economic activity – direct links with numerous other economic sectors and markets. This means that the transport regulator typically will have to take into account that a (change in) transport prices may affect the equilibria on numerous other markets, many of which will be distorted and might thus call for an upward or downward adjustment of these transport prices. Among the most important of these other markets would typically be the labour market, where relatively high marginal tax rates may lead to a serious under-supply of labour, calling for a downward adjustment in transport taxes when its price effect dominates the labour supply decision, or an upward adjustment when the revenues are used to lower labour taxes and this revenue effect appears to dominate. Another example would be freight transport for goods that create environmental pollution in their production process, typically calling for an upward adjustment in transport prices as this might offer an indirect way of taxing a polluting firm's output. The existence of inefficiencies in the various markets 'served' by transport thus create constraints that should be taken seriously in setting transport prices (see the discussion of second-best pricing in Section 4 below).

Political constraints

As a final type of constraint, it may be possible that in some democracies, especially where coalition governments are in office, the level of charges becomes a political issue much more than an economic question. In such cases, deals between political parties accepting the implementation of marginal cost based pricing may create limitations on the types of charges and the flexibility that can be implemented.

4. Second-best pricing in transport

Under conditions such as described above, the regulator has to resort to 'second-best' pricing: setting the prices that are available optimally, under the constraints applying. It is safe to state that second-best pricing will be the rule for the implementation of marginal cost based pricing in reality. As stated before, under second-best pricing, the 'tax rule' to be applied will no longer be the simple Pigouvian rule that can be derived from textbook expositions of marginal cost pricing in transport, and that would dictate that the price should be equal to the marginal social cost (meaning that the tax should be equal to the marginal external cost). To see why second-best pricing indeed will be the rule, it is instructive to list some of the key conditions that should hold for Pigouvian taxes to be optimal (these key conditions are of course closely related to the various types of constraints identified in Section 3 above):

1. All markets other than transport (including labour) operate under first-best conditions, being perfectly competitive and causing no externalities, with all prices equal to marginal social costs
2. There are no imperfections arising due to the existence of different jurisdictions and countries, *i.e.* all governments cooperate perfectly so that no externality spill-overs nor any other 'distorted' economic spill-overs (*e.g.* due to trade policies) exist
3. The government can use lump-sum taxes to finance any deficits that may occur in transport budgets with prices equal to marginal social costs, and lump-sum subsidies to dispose of any surpluses

4. The government can use lump-sum taxes and subsidies to pursue any redistributive goals it may have
5. Transport taxes can be put into effect on all links in all networks (*i.e.* all modes), and can be varied freely over time and place, and can be differentiated perfectly over users according to vehicle used, its technical state, the driving style, etc.
6. There is complete certainty on the levels of marginal social costs as they vary over time place, and users; and users perfectly understand the pricing scheme imposed upon them
7. There are no other constraints on pricing; neither for institutional, legal, political, nor for any other reasons.

Hence the assertion that second-best pricing will be the rule.

A substantial technical literature has emerged over the last decade, that addresses various types of second-best pricing, and considers questions towards the optimal design of second-best pricing schemes (*i.e.*, what do the tax rules look like, how do they deviate from the Pigouvian rule?), and towards the relative efficiency of these schemes (*i.e.*, which share of the theoretically attainable efficiency gains – without the constraint(s) – can be achieved with the constraint(s) applying?). Much of this literature is reviewed in, for instance, Lindsey and Verhoef (2001) and in MC-ICAM deliverable 2, in which for instance the following types of second-best transport are considered: network issues arising in uni-modal networks; network issues arising in multi-modal networks, second-best issues in transport nodes where operators have market power, second-best issues stemming from heterogeneity of users, second-best issues arising from different levels of government; and second-best issues arising from the pre-existence of distortive taxes elsewhere in the economy.

Although this is not the place to discuss the technical details of such studies, it is instructive to mention a few (necessarily general) lessons that can be drawn from such studies:

1. The second-best optimal tax rules are a lot more complicated than the simple Pigouvian 'tax-equals-marginal-external-costs' rule, which is due to the fact that all sorts of indirect effects of the second-best taxes are taken 'on board' in the tax rule
2. The informational requirements for the regulator, necessary for applying second-best taxes optimally, are considerable higher than for Pigouvian rules, for which 'only' knowledge of the levels of marginal external costs is required. Second-best tax rules typically require these, too, *plus* a host of other variables, often including demand and cost elasticities within the transport sector and elsewhere, and measures for the discrepancy between prices and marginal social costs in other markets
3. Elements 1 and 2 together imply that the risk of 'government failures' – *i.e.*, the setting of wrong prices – becomes considerably larger under second-best pricing

4. The welfare gains that can be achieved with second-best pricing may be substantially lower than those arising from theoretical first-best pricing, but may also be nearly as high
5. A naïve use of second-best tax instruments, defined as ignoring the second-best distortions and applying Pigouvian rules instead, would however lead to even lower efficiency gains, and may in some cases even lead to a reduction in overall social welfare
6. Second-best pricing often lacks the ‘convenient’ property of first-best pricing that optimal incentives are given to adjust behaviour efficiently in all relevant behavioural dimensions
7. In a second-best world, the use of revenues is not an issue that only affects the eventual distributive impacts of the scheme – as often conveniently assumed in public debates on transport pricing – but instead will also have efficiency impacts that may be substantial and could even outweigh the primary effects of the pricing part of the scheme itself, and that may be negative or positive.

Indeed, a somewhat alarming message arises from these 7 points. Once we have entered the world of second-best – which we undoubtedly will when implementing marginal cost based pricing in reality – we are apparently in a world where little can be predicted with certainty without solid prior investigation, and where we have to be careful – to put it mildly – that price instruments are indeed used optimally given the constraints that apply. To put it less mildly, we even have to be careful not to make things *worse* using the instruments available. Fortunately, the body of research on this topic is growing fast, and many models are currently available with which the detailed effects of various types of second-best pricing instruments can be studied in great detail. So, the above list should much sooner be seen as a warning that careful research on direct and indirect effects of pricing and revenue use is necessary before implementing proposed pricing schemes, than as a plea for the abandonment of the idea of marginal cost based pricing altogether.

One final remark seems appropriate. In discussing second-best pricing above, we have implicitly considered the relevant constraints as ‘hard’ ones. This often will indeed be the case, at least in the short run. In the longer run, however, constraints will typically become more flexible. Why is this a relevant issue? Because it has been argued that it may be unwise to adjust transport taxes too much to reflect tax distortions elsewhere in the economy (including in other transport systems), as doing so would minimize the incentive for tax reforms in these other sectors. In the event that tax reforms elsewhere would indeed be contingent on whether or not transport prices are adjusted to reflect the price distortions elsewhere, the longer run efficiency effect of using second-best taxes instead of other pricing rules could then be negative. Whether or not this is a theoretical argument or one that bears some empirical relevance seems to be still an open question.

5. Policy packaging

Let us now return to the 7 'general lessons' listed in Section 4 above. It is important to emphasize that these lessons emerged largely from exercises involving isolated second-best instruments. Project AFFORD explored one particular way of dealing with the 'risks' involved in the use of second-best price instruments. This concerns the use of so called 'policy packages': combinations of second-best price instruments that in isolation may be prone to all sorts of efficiency losses due to unproductivity or perhaps even counterproductivity with regard to some behavioural dimensions, but that in combination may be capable of replicating the full set of incentives given by an imaginary first-best pricing scheme more accurately.

The general idea can be illustrated using automobile use as an example, and focussing on the charging side (not the revenue use side) of a charging schemes only, and ignoring general equilibrium issues for the sake of keeping the example manageable. The general strategy would then be to first identify the different *dimensions of behaviour* that road users could adjust in response to hypothetical first-best pricing (the notion 'cost drivers' refers to the same phenomenon).

Table 1 distinguishes the following broad dimensions of behaviour: car use, car ownership and spatial behaviour. 'Spatial behaviour' refers to the choice of residence and the location of other activities. These broad categories are further subdivided into more refined dimensions. In particular, the inclusion of the 'number of trips' in addition to 'vehicle kilometers' indicates situations where, in a given 'market', a larger number of cars making proportionally shorter trips would cause higher total external costs (practically: cold starts, queuing up, etc). The purpose of the inclusion of the item 'place of driving' is to distinguish between externalities that are local in the sense that they vary over space (*e.g.* some pollutants, noise), and externalities that are route-specific (*e.g.* congestion).

Regarding the congestion externality, Table 1 makes an important distinction between two archetypes, bottleneck congestion and flow congestion, both of which may be relevant for congested transport networks. The main distinguishing feature is that bottleneck congestion is caused by the existence of physical bottlenecks in the network, such as bridges, tunnels, ramps, etc. Flow congestion, on the other hand, refers to the capacity of roads in general, and can occur also in the situation where no link has a relatively speaking smaller capacity than the other links. In real networks, observed congestion will often represent a mixture of bottleneck and flow congestion. As shown in the table, however, the contribution to pure bottleneck congestion depends only on the question of whether a user wants to pass the bottleneck, and is independent of the total length of the trip. The more closely we approach the other extreme of pure flow congestion, the more closely the overall marginal external congestion costs become dependent upon total distance travelled.

	Car use					Car ownership		Spatial behaviour[b]
	Vehicle kilometres	Number of trips	Time of driving (peak or off-peak)	Place of driving (area or route)	Driving style	Fleet size[a]	Vehicle technology	
Intra-sectoral externalities:								
– Flow congestion	*	-	**	**	**	*	-	**
– Bottleneck congestion	-	**	**	**	-	*	-	**
– Infrastructure damage	**	-	-	-	-	*	*	**
– Accidents	*	-	*	*	**	*	*	*
Inter-sectoral externalities:								
– Noise	*	-	*	**	**	*	**	**
– Local emissions	**	*	*	**	**	*	**	**
– Global emissions	**	*	-	-	**	*	**	**

** particularly strong and direct relation
* possibly strong indirect relation, or moderately strong direct relation
– no particular strong or direct relation
[a] Also allows for car size.
[b] Location of residence vs work and leisure activities.
Source: AFORD Deliverable 1

Table 1 Dependence of various external costs of road transport on behavioural dimensions.

The table indicates the relevance of each dependence on a three point scale. The assigned stars are merely indicative and debatable; one could evidently find counter-examples. That is also the reason for using a three-point scale only. However, the table is illustrative in drawing explicit attention to the dependence between various externalities and behavioural dimensions. (A similar illustration could be given for freight transport and public transport.)

The next step would be to consider, for a selection of possible second-best pricing instruments, the extent to which they are capable of discriminating according to the behavioural dimensions identified in Table 1. This is shown in Table 2 for some second-best tax instruments.

	Car use					Car ownership		Spatial behaviour
	Vehicle kilometres	Number of trips	Time of driving (peak or off-peak)	Place of driving (area or route)	Driving style	Fleet size	Vehicle technology	
First-best bench-mark pricing	***	***	***	***	***	***	***	***
Some second-best taxes								
ERP per km	**	**	**	*	-	**	-	**
Toll booths	*	**	*	*	-	**	-	*
ERP Cordon	*	**	**	*	-	**	-	*
Peak permits	*	*	*	*	-	**	-	*/-
Area licences	*	*	*	*	-	**	*	**
Parking fees	*	**	*	**/*/-	-	*	-	*
Fuel taxes	**	**	-	-	*	**	**	**

*** optimal (first-best) impact
** likely direct impact, possibly approaching first-best standards
* possible direct impact
– no particularly strong direct impact, or at least unlikely in practice
Source: AFORD Deliverable 1 (adapted)

Table 2 Impact of various second-best pricing instruments on different behavioural dimensions.

The qualifications made in Table 2 are again rough, and the scores assigned assume that the particular instrument is used optimally, given its inherent practical restrictions. Tables 1 and 2 together, however, can be used to construct policy packages of second-best taxes that together could be designed so as to replicate the full set of incentives given by hypothetical first-best pricing as closely as possible. For instance, if in a certain situation the dominant externalities would be bottleneck congestion at a few 'black spots', localized environmental externalities in a city centre, and global externalities, a package of a toll cordon, inner-city parking charges (provided through-traffic and private parking are no major problems) and fuel taxes might offer a relatively efficient second-best policy package, and may become attractive in particular when time- and place- differentiated electronic kilometre-charges are not available.

The example is simple, but the general point is important: the possible weaknesses that can arise from second-best taxes when applied in isolation can often be reduced when constructing a policy package of second-best measures, that is designed to cover the most important externalities and dimensions of behaviour relevant for the particular case considered. What *is* of importance here, of course, would be that the various charge levels employed for the individual instruments in the package be carefully chosen, simultaneously, so as to maximize the package's eventual efficiency.

6. Implementation paths

There is a close link between the issue of second-best pricing, policy packages, and the question on the design of an implementation path – the definition of the various phases during the transition from the current situation to some preferred end-state in which marginal cost based pricing is implemented in the preferred way. The current project MC-ICAM is concerned especially with such implementation paths, and although this is novel and still ongoing research, it is perhaps worthwhile to spend a few words on this issue here.

From a policy perspective, it would often be intuitively plausible that the move from the current situation towards one with some preferred scheme of marginal cost based pricing cannot be a one-step, 'overnight' transition. However, from a purely economic viewpoint, there is no *a priori* reason to assume that an implementation path should be followed. That is, of course it may be the case that optimal prices would vary during the first years after the initial introduction of a scheme, for instance due to longer-run responses materializing gradually, but to state that therefore flexibility in pricing is necessary during some 'implementation phase' would unjustly create the suggestion that after such a phase, optimal prices would remain constant to eternity.

Often, therefore, implementation paths will be motivated by other considerations than strictly economic efficiency. The link with issues of second-best pricing and policy packaging, from the perspective of economic analysis of implementation paths, would thus be that an implementation path can be considered as a sequence of second-best optima, which arise as the set of constraints on pricing changes over time (typically, the number of constraints and/or their 'tightness' can be expected to decrease during the course of an implementation path). This is also the type of interpretation of implementation paths that is currently considered to be employed in MC-ICAM.

The types of constraints considered may reflect for instance the following motivations for implementation paths:

- to gain public acceptance over time
- to teach the public to understand increasingly complex pricing schedules
- to reflect that capacity, too, cannot be optimized instantaneously
- to help the regulator get used to pricing – *i.e.*, set up toll collection agencies, acquire experience in automated billing, *etc.* – in a small scale project or using simple pricing schedules
- to reflect that the degree of policy coordination between vertically or horizontally ordered governments will change (typically increase) over time
- to reflect that practical and/or technical consideration may prevent simultaneous implementation across modes

- *etc.*

It will be clear that different considerations from the above list may often imply different *types* of implementation paths. What these paths will have in common is that, in economic analysis, the constraints will be assumed to be exogenously determined. If anything, this is necessary to prevent ending up with trivial results such as the advice that all binding constraints should be removed immediately.

The implementation paths will of course differ in the types of constraint imposed on the pricing schemes. One can imagine that at least the following three archetypes of implementation paths can be distinguished, which will have a different exact interpretation depending on whether the analysis considers a uni-modal transport network (case A), a multi-modal transport network (case B), or a transport network in relation to the (spatio-)economic system to which it belongs and with which it interacts (case C):

Type 1: Stepwise expansion over sub-markets

1A The number of priced links within a mode increases over time
1B1 The number of priced modes increases over time, or:
1B2 The number of priced links in a multi-modal network increases over time
1C The number of priced sectors or activities in an economy increases over time

('priced' here means second-best optimally priced)

Type 2: Stepwise convergence to optimal prices over all sub-markets simultaneously

2A All links in a mode get prices, which move to (second- or first-best) optimal levels in a discrete number of steps
2B All links in a multimodal network get prices, which move to (second- or first-best) optimal levels in a discrete number of steps
2C All sectors or activities in an economy get prices, which move to (second- or first-best) optimal levels in a discrete number of steps

(to 'get prices' here means that the regulator can affect these prices from the first period onwards, but that the charges allowed are constrained in some sense)

Type 3: Stepwise (further) differentiation of second-best prices

3A The degree of differentiation of prices (over links, user types, time periods, *etc.*) within a mode increases over time
3B The degree of differentiation of prices (over links, user types, time periods, *etc.*) in a multimodal network increases over time

3C The degree of differentiation of prices (over links, user types, time periods, sub-markets, *etc.*) in a multi-activity economy increases over time (whether this one is useful remains to be seen – it is now here for the sake of completeness and symmetry only)

The above classification, admittedly, is rather 'off the cuff', and it remains to be seen whether a different classification will eventually be used in MC-ICAM. For the present paper, the classification suffices in the sense that even after narrowing down the interpretation of an implementation path to 'a sequence of constrained second-best optima', there is till a wide variety of types of implementation paths that can be distinguished, and that will undoubtedly have markedly different welfare and efficiency properties – depending, to a considerable extent, also on the exact design of the implementation path (given its type). It is fair to say that an exciting area of challenging research, with high and immediate policy implications, seems to be lying bare right in front of us.

7. Conclusion

This position paper aimed to give an overview of key implementation issues surrounding the introduction of marginal cost based pricing in transport, taking an economic perspective. We started by discussing the concept of Pigouvian taxes and the prime motivations that exist for the use of marginal cost based pricing in transport. We then moved on to those issues that would be relevant in the actual implementation, where we chose to focus on the following three topics: second-best pricing, policy packaging and implementation paths.

The paper aimed to emphasize the importance of second-best issues, which for instance may arise because charging mechanisms are not perfect, because revenue goals and efficient pricing may require diverging charge levels, or because distortions exist elsewhere in the economy. It was stressed that the relatively straightforward black-board exposition of marginal cost pricing in transport, ignoring such considerations, may be misleading in creating the suggestion that things are not so complex. Instead, one has to be aware of the fact that the efficiency gains of marginal cost pricing may be limited under second-best conditions, and even negative if second-best pricing is performed according to naïve policy rules. It was however also argued that packages of second-best instruments may reduce much of such risks, provided they are composed according to some economic principles. Finally, we spent a few words on 'implementation paths', which appears to be a very important research area that is currently being explored, for instance in the context of the MC-ICAM project. A first set of policy conclusions therefore involved conclusions motivating the necessity of considering second-best issues in designing transport pricing schemes, warning against a naïve use of pricing, describing ways of combining second-best pricing instruments in policy packages that would reduce the risk of using second-best pricing scheme and at the same time would increase the efficiency of the packages when synergy between pricing instruments can be exploited, and finally sketching ways of using the same sort of methodology when designing implementation paths for transport pricing policies.

A second set of policy conclusions is more implicit, and involve conclusions that the reader may draw as regards the desirability of starting such implementation paths now or in the very near future. An anonymous referee of this paper observed in this respect that "[o]ne of the conclusions that might be taken from the paper is that implementing pricing reform based on marginal cost principles is very complicated and getting it wrong might actually lead to things getting worse rather than better. Therefore, we should undertake much more research before trying to implement anything. On the other hand, one might argue that things are so complex we are almost bound to 'get it wrong' initially but until we start to implement the reforms we can't start learning from experience and moving towards a less congested, more efficient transport system, and while we postpone implementation in favour of further research the system continues to get worse." There indeed appears to be an important trade-off here. One way of getting around this would be to introduce pricing but to design the system such that sufficient flexibility in terms of price setting is allowed. This means that if future research – or an assessment of study of the scheme itself – should yield insights that strongly suggest that the prices charged should be adjusted, this indeed would be possible. A second approach would be to try and identify 'no-regret' pricing strategies or implementation paths, the welfare gains of which are relatively robust to small mistakes by the regulator that either arise from an imperfect understanding or application of complicated second-best pricing rules, or from an imperfect measurement of the various variables entering the associated second-best tax expressions. These and related issues may prove to offer fertile research ground for future research efforts, including the ongoing project MC-ICAM.

Small bibliography

AFFORD, deliverables 1-3, http://data.vatt.fi/afford/reports-index.html:

Fridsrøm, L., H. Minken, P. Moilanen, S. Shepherd and A. Vold (2000) "Economic and Equity Effects of Marginal Cost Pricing in Transport" AFFORD Deliverable 2A, VATT, Helsinki.

Milne, D., E. Niskanen and E.T. Verhoef (1999) "Operationalisation of Marginal Cost Pricing within Urban Transport", AFFORD Deliverable 1, VATT, Helsinki.

Milne, D., E. Niskanen and E.T. Verhoef (2000) "Legal and Institutional Framework for Marginal Cost Pricing in Urban Transport in Europe", AFFORD Deliverable 2B, VATT, Helsinki.

Schade, J. and B. Schlag (2000) "Acceptability of Urban Transport Pricing", AFFORD Deliverable 2C, VATT, Helsinki.

Niskanen, E. *et al.* (2000) “AFFORD, Final Report”, AFFORD Deliverable 3, VATT, Helsinki.
Button K.J. and E.T. Verhoef (eds.) (1998) *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* Edward Elgar, Cheltenham.
Lindsey, C.R. and E.T. Verhoef (2000) "Congestion modelling". In: D.A. Hensher and K.J. Button (eds.) (2000) Handbook of Transport Modelling, Handbooks in Transport 1 Elsevier / Pergamon, Amsterdam, pp. 353-373.
Lindsey, C.R. and E.T. Verhoef (2001) "Traffic congestion and congestion pricing". In: D.A. Hensher and K.J. Button (eds.) (2001) Handbook of Handbook of Transport Systems and Traffic Control, Handbooks in Transport 3 Elsevier / Pergamon, Amsterdam, forthcoming.
MC-ICAM, deliverable 2, http://www.mcicam.net/:
Rouwendal, J., De Palma, A., De Borger, B., Lindsey, R., Niskanen, E., Pels, E., Proost, P., Shepherd, S., Van Dender, K., Verhoef, E.T. (2001) “Relevant Optima: First-best and Second-best”, MC-ICAM Deliverable 2 (forthcoming).
Small, K.A. (1992) *Urban Transport Economics* Harwood, Chur.
Verhoef, E.T. (1996) The Economics of Regulating Road Transport Edward Elgar, Cheltenham.
Verhoef, E.T. (2000) "The implementation of marginal external cost pricing in road transport: long run vs short run and first-best vs second-best" *Papers in Regional Science* **79** 307-332.

Glossary of technical economic terms used in this note

This section gives an explanation of some of the technical economic terms used in this paper. The intention is not to give the most solid technical definitions possible, but much more to give an intuitive explanation of the various concepts, that should be useful to the non-economist reader of this paper.

Constant returns to scale A cost structure for which an increase or decrease in the output level would lead to a proportional change in total costs, so that average costs (per unit of output) are independent of the output level, and marginal costs (for an additional unit of output) are always equal to these average costs.

Equilibrium A situation where nobody wants to change their consumption (or production) levels given the price(s) prevailing on a (set of) market(s).

External costs Costs not borne by the person causing these costs, but imposed on others instead, which can occur if no market for the ‘good’ or ‘bad’ involved exists. A classic example is the ‘good’ clean air (or the ‘bad’ air pollution). Absent environmental taxation, polluters do not have to pay for air pollution as no market for clean air exists.

General equilibrium The counterpart of ‘partial equilibrium’, where a market in isolation is considered, ignoring interactions with other markets. A general equilibrium setting (model / analysis) does take interactions with other markets explicitly into account. An example from road transport would be a study on the desirability of peak charging, taking into account that it would to some extent discourage labour supply which may be less than optimal already due to existing taxes on labour, but that at the same time the revenues from road pricing could be used to reduce labour taxes, which may stimulate (initially too small) labour supply.

Marginal benefits The benefits attached to a unit of the extra consumption (*e.g.* an extra vehicle-kilometre) that would take place if the price or costs to the user would be lowered by a (very) small amount.

Marginal private costs The costs associated with a unit of extra consumption (*e.g.* an extra vehicle-kilometre) that would be borne by the person consuming that unit.

Marginal social costs The costs associated with a unit of extra consumption (*e.g.* an extra vehicle-kilometre) that would be borne by society at large, including those costs borne by the person consuming that unit.

Marginal external costs The difference between marginal private and social costs as defined above.

Optimum A situation where any feasible change in production and/or consumption levels would lead to a reduction in the social objective.

Pigouvian tax A tax equal to the 'marginal external costs', aimed at the internalizing these costs.

Scale economies or *Economies of scale* A cost structure for which an increase or decrease in the output level would lead to a less than proportional change in total costs. Average costs (per unit of output) are falling with the output level, and marginal costs (for an additional unit of output) are also falling with the output level – and are always smaller than to the average costs.

Social objective (function) A (mathematical) formal expression of the set of goals that the government wishes to pursuit, with appropriate indicators also of the relative importance of these goals. It is typically assumed that – absent 'transaction costs' and under perfect information – the government would try to maximize the social objective given the constraints under which it operates.