OECD

International Transport Forum

IMPLEMENTING CONGESTION CHARGES

ROUND TABLE

147

TRANSPORT RESEARCH CENTRE

OECD

International Transport Forum

IMPLEMENTING CONGESTION CHARGES

ROUND TABLE

147

TRANSPORT RESEARCH CENTRE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

ISBN 978-92-821-0284-8 (print)
ISBN 978-92-821-0285-5 (PDF)

Series: ITF Round Tables
ISSN 2074-3378 (print)
ISSN 2074-336X (online)

Also available in French: *Mettre en œuvre la tarification de la congestion*

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

INTERNATIONAL TRANSPORT FORUM

The International Transport Forum is an inter-governmental body within the OECD family. The Forum is a global platform for transport policy makers and stakeholders. Its objective is to serve political leaders and a larger public in developing a better understanding of the role of transport in economic growth and the role of transport policy in addressing the social and environmental dimensions of sustainable development. The Forum organises a Conference for Ministers and leading figures from civil society each May in Leipzig, Germany.

The International Transport Forum was created under a Declaration issued by the Council of Ministers of the ECMT (European Conference of Ministers of Transport) at its Ministerial Session in May 2006 under the legal authority of the Protocol of the ECMT, signed in Brussels on 17 October 1953, and legal instruments of the OECD. The Forum's Secretariat is located in Paris.

The members of the Forum are: Albania, Armenia, Australia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Canada, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FYROM, Georgia, Germany, Greece, Hungary, Iceland, India, Ireland, Italy, Japan, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Moldova, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, the United Kingdom and the United States.

The OECD and the International Transport Forum established a Joint Transport Research Centre in 2004. The Centre conducts co-operative research programmes addressing all modes of transport to support policy making in Member countries and contribute to the Ministerial sessions of the International Transport Forum.

Further information about the International Transport Forum is available on Internet at the following address:
www.internationaltransportforum.org

TABLE OF CONTENTS

SUMMARY OF DISCUSSIONS

SUMMARY CONTENTS

Paris, April 2010

1. INTRODUCTION

The Round Table addressed the broad question of what research and experience tell us about how to arrive at a successful introduction of congestion charging schemes. Attention was limited mostly to urbanised areas where road traffic congestion is or may become an issue. “Success” means (a) that a policy is implemented, (b) that it works, (c) that it is accepted by actual and potential users, and (d) that it generates benefits for society overall. In order to shed light on these dimensions of success, lessons are drawn from more and less successful attempts to implement charges (Section 2). In addition, we ask if and how the evolving understanding of the economics of road traffic congestion charging might affect the assessment of congestion charging policy (Section 3). The conclusions in Section 4 summarize the main recommendations for policy-makers who contemplate the introduction (or the removal) of congestion charges.

Transport economists have long lamented the lack of policy interest for a tool that to them is as obviously welfare-improving as congestion charges. However, what is obvious in principle is less obvious in practice. Questions regarding the desirability and feasibility of congestion charges become apparent when policy constraints and costs are taken into account. How to convince voters and their representatives that it is a good idea to make travel more expensive when traffic is bad? How to set charges and deploy revenues so that the distribution of gains and losses constitutes a marketable political proposition? Section 2 discusses these and related issues by taking a careful look at practice, in particular in Singapore, Stockholm and London. In studying these cases we try to distinguish idiosyncrasies of the particular cases from general concerns.

Section 2.1 ends with some remarks on cordon pricing, value pricing and parking pricing. Cordon pricing and value pricing both are approximations to ideal congestion charges, but reflect different basic strategies. Cordons focus on maintaining an acceptable average level of service, while value pricing is about offering variety (low quality at a low or zero price, higher quality at a premium)[1]. Parking pricing is part of the debate because parking charges could, to some extent, mimic congestion charges. Probably more important, however, is that removing subsidies to parking on public roads (as well as possibly through reducing relatively beneficial tax treatment of employer-offered parking) is likely to have major beneficial effects on the use of space, and on congestion levels where space and road capacity are scarce, while involving very low cost.

The principal question in Section 2 is how to improve the chances of getting congestion charges introduced. Focussing on implementation implies a presumption that congestion charging is a sensible policy instrument to combat excessive congestion. There is indeed widespread agreement among transport economists that, where there is congestion, charges are required to ensure that potential road users take the marginal social cost of their trips into account when deciding if, how, where and when to travel. When individuals decide on the basis of marginal social costs instead of marginal private costs, welfare increases because the marginal external cost of congestion is no longer ignored, so that the social benefit of an additional trip will be (more or less) equal to its social cost.

Translating this key principle into practical recommendations for policy is, however, less than straightforward. Section 3 deals with some complicating factors. Congestion charging systems are not costless. To ensure a welfare-improving outcome, the efficiency gain from charging needs to be greater than the cost of the system. The size of the efficiency gain depends on getting the prices right. Determining the level and structure of charges requires an understanding of the physical and behavioural aspects of congestion. The mechanisms involved are far more complex than is suggested by the argument that travel times increase with traffic volumes, as used in the basic rationale for charges. Charges are always an approximation to the theoretical ideal, so that decisions must be made on what approximation is best. Experience demonstrates that analytical approaches using disaggregated network models are more likely to produce an efficient result than prices based only on common sense. Such models are better placed to capture complex network activity across modes and reveal impacts on traffic flow that cannot easily be anticipated. A period of model testing and iteration, before charges are set, is indicated.

Charges are introduced in a world that is rife with other market imperfections and where other policy objectives than efficiency matter. Again, the question is if and how this affects recommendations for setting congestion charges. This is a subject of some controversy, but the emerging view is that charges should remain closely tied to marginal external congestion costs rather than adjusted to compensate for one or a range of other imperfections. There was some support for making the charge deductible from income taxes for commuters. However, this needs to be determined within the broader taxation framework and existing deductions for commuting expenses. Moreover, other forms of revenue use may have a better claim than tax deductions.

2. LESSONS FROM PRACTICE

2.1. Some general principles

2.1.1 Acceptance

A successful congestion charging scheme is one that works technically and reduces congestion, is acceptable, and generates net socio-economic benefits. Acceptability is the overriding concern for policy-makers, as without it no lasting implementation is possible. It follows that acceptability affects system design and that the way impacts, in particular benefits, are perceived is critical. It also tends to lead to congestion pricing systems being modified and expanded to contribute to a variety of other policy goals, including broader tax reforms and environmental protection. In extreme cases, there could be perceived benefits but social losses. This could be the case, for example, where "congestion" charges are introduced to improve the environment but where there is actually relatively little congestion. The improvement in the urban environment that results may well be real but smaller than the cost of running a congestion pricing system. In the absence of major congestion mitigation, the net benefits are likely to be negative. At the other extreme, a perception of unacceptably high system costs may prevent implementation where charges would in fact produce net benefits. This is the more common scenario because of the difficulty of presenting the benefits of congestion management to the public in tangible terms.

Modifying the design of congestion charging systems to promote acceptability, by targeting goals other than congestion management, usually results in a trade-off between efficiency benefits and acceptability. This can lead to a sterile divergence of views, with one side arguing that focusing on acceptability results in foregoing too much of the potential benefits of charging, and the other side saying that too much emphasis on the economic measurement of benefits jeopardizes acceptability, to the point of becoming counterproductive. A more productive intermediate position is that acceptability comes first, even though it may be costly in the sense that some benefits are given up.

Rather than taking perceptions as immutable, communication and marketing can help shape them, paving the way for the introduction of congestion charging and for increasing its benefits over time, and potentially reducing the need to sacrifice efficiency in system design. As the existence of effective congestion charging systems in Singapore, Stockholm and London shows, the tension between acceptability and welfare potential is not irredeemable. Certainly, it makes no sense to drop a feasible and productive solution because in principle a better one is available.

Acceptability evolves over time. Surveys of public opinion reveal a typical pattern whereby acceptance first rises as the general idea is discussed, then deteriorates as details become known and implementation approaches, but rises to its highest level once the system is operational. Higher acceptance of operational systems may relate to sharply increased awareness of now tangible benefits or to cognitive adaptation. At any rate, the dynamics of public acceptance make politics a critical factor. As is well known, the Mayor of London, Ken Livingstone, risked his political future by proposing the introduction of the London Congestion Charge in the run-up to a mayoral election. In Stockholm, conditions imposed by the junior party in the governing coalition drove reluctant political partners to introduce congestion charging for a trial period. It appears that no political champion is needed for continuing a scheme (most likely partly because support tends to rise after implementation). The citizens of Stockholm voted to reinstate congestion pricing on expiry of the charging trial. In London, although the Western Extension Zone to the congestion charge will be discontinued following a change in government, and other extensions are no longer under consideration, there are no plans to discontinue charging in the Central Zone.

2.1.2 Prime objective

The primary goal of a congestion charging system is to reduce congestion to a more efficient level. This level is determined by the cost of the scheme, by the behavioural response to it, and by the relation between external congestion costs and traffic volumes. This is straightforward but has some important implications. First, congestion can only be reduced where it is (perceived to be) excessive to begin with. Introducing a congestion charging system in anticipation of excessive congestion in the future is a very hard proposition to sell, because perceived benefits will be low in the absence of an acute problem (certainly after discounting). This notion that drastic policy changes are easier at times of crisis is familiar from environmental and safety policy, as well as from macroeconomic policy.

Second, setting a congestion charge low and gradually increasing it is a risky strategy for gaining acceptance. If the charge starts too low to have a visible impact on congestion the strategy will backfire. Similarly, when congestion charging is part of a larger reform of charging for use of the roads, the congestion element may be critical to the new system being perceived as a success; leaving the differentiation of a new charge to deal with congestion to a second stage of reform may undermine rather than ease acceptance.

The third implication relates to policy targeting. The introduction of congestion charges is often defended for reasons other than congestion, including environmental benefits and meeting a revenue need. The point is not that such ancillary benefits do not exist or are unimportant, but rather that they should not be turned into the first priority. Over-emphasizing environmental benefits may have contributed to the rejection of congestion charging proposals in New York in 2008, although equity concerns probably were at least as important. The primary goal for the proposed congestion charge in Gothenburg is raising revenue for investment in a road tunnel, as congestion in the city is relatively light. Oslo, Bergen and Trondheim in Norway successfully introduced toll rings to raise revenues for infrastructure investments but without confusing the objective. There is a risk that if a revenue-targeted system is implemented as a congestion measure, acceptance will decline rather than improve after implementation. A related issue concerns shifting from fuel taxes to road pricing (in the sense of distance charges rather than congestion charges) for the collection of public revenue, as discussed in the USA, mainly in relation to the revenue shortfall in the Highway Trust Fund. This may be a reasonable policy but, whatever its merits, it is not mainly a congestion charging policy and *if* marketed as such would probably result in problems of acceptance. Indeed, advocates of distance charges focus on revenue needs more than on congestion management, as the latter is seen to be particularly problematic in terms of acceptance.

As noted, there is a potential for ancillary benefits to congestion charges and these should be taken into account when designing, marketing and assessing the scheme. Some environmental impacts of car use are strongly correlated with congestion. The damage from noise and pollutants like particles as well as the general discomfort caused by dense traffic, serve as examples. It makes sense to take these effects into account when setting charges. For a local scheme, it does not make sense, however, to set congestion charges in relation to greenhouse gas emissions: the effects of emissions of greenhouse gases are decidedly non-local, and much better instruments than congestion charges exist for tackling them (notably, fuel taxes). This, of course, does not mean that congestion charges do not have an effect on CO_2 emissions: emissions will fall if traffic levels fall and will decline with smoother traffic flows at more optimal speeds.

2.1.3 Revenues

Pricing reforms in transport, including but not limited to congestion charges, are sometimes accompanied by commitments to revenue neutrality[2]. For example, at the time of introduction of congestion charges in Singapore, revenue neutrality was promised and (more than) achieved by reducing vehicle taxes; it became of lesser concern as the system became firmly implanted. The Dutch proposal to replace vehicle ownership-based charges with usage-based charges was designed to be revenue neutral.

Revenue neutrality is not a universal objective or constraint. Discussions on introducing usage-based charges in the USA are often inspired by prevailing revenue shortfalls in the Highway Trust Fund (fed by capped fuel taxes), so revenue neutrality is not a concern[3]. The London congestion charging scheme produced additional revenues, although its prime objective was to reduce congestion. Discussions on potential charges in Moscow look to increased revenues from transport. In all cases, however, discussions on what to do with revenues are at the core of system design and of public debate. The odd man out here is Stockholm, where revenue concerns were not crucial, and decisions on what to do with the new funds came almost as an afterthought[4].

Commitments to revenue neutrality are common when changes in the way taxes are collected are proposed. When the only aim is to *replace* one tax with another that is less costly in terms of efficiency or collection costs, or more difficult to avoid or otherwise preferable, opting for neutrality is relatively straightforward[5]. When the change involves internalising an external cost, the picture changes, as now a tax is *added*. Pricing the externality inevitably involves raising new revenues. The issue then becomes what to do with those revenues. They can be recycled into reducing other taxes, and if these are transport taxes then the overall change might be neutral in terms of transport tax revenue. There are, however, two important respects in which the change cannot be neutral for all users. First, some road users will be worse off financially from the change (those paying the highest rates of the congestion charge most often) and some better off[6]. Also, charging for congestion requires expenditure on the system for monitoring traffic and collecting and processing payments. Experience in Singapore, London and Stockholm suggests the costs of the system represent 15 to 30% of the gross charge income and that the figure cannot be compressed to below around 10%, given their definition of costs. The precise ratio is partly a function of the level of charges that bring congestion to the optimal level (the higher the charge the larger the revenue), but the key message is that congestion charges are a relatively expensive way of raising revenue.

Any calculation of neutrality should be based on net revenues and not gross income from congestion charges. This should not be too difficult to communicate. Achieving the large benefits of congestion relief requires a system that has to be paid for[7]; and if the objective is to maximize net revenues rather than manage congestion, taxes on fuels and vehicles are much cheaper to collect. The communications mission becomes more complicated when congestion charging is part of a wider reform of transport charges and taxes. The answer may be to put the emphasis on transparent use of revenues and to be precise about what is meant by revenue neutrality, if it is committed to[8]. Public support can be improved by a neutrality pledge that allays suspicions of hidden tax increases. Political support is increased by revenue neutrality as it alleviates concerns regarding public finance shortfalls.

2.1.4 Whatever the technology, running a congestion charging system is not cheap

Once it is clear what a congestion charging system is supposed to achieve, a technology that meets the requirements can be found. It deserves emphasis that setting goals and then choosing means to attain them is the logical order of things. Taking the opposite approach, aiming to make congestion charging an application for a technology seeking to develop a market, should be avoided. Choosing technology as a function of system design also avoids false choices between, for example, dedicated road-side communications (DRSC) and GPS/GSM-based systems[9]. These systems fit different contexts, e.g. depending on whether there are many or just a few charging points.

Experience with congestion charging schemes suggests they require costly investments and are expensive to operate. Operating costs generally outweigh start-up costs (often by a factor of 10, according to Bernhard Oehry) and should not be ignored in the design stage, as otherwise there is a risk of investing too little in capital. Acceptability increases costs for various reasons. First, as emphasized in the discussion of the Stockholm experience, policy-makers will not incur the risk of going ahead with a system unless they can be convinced that it will work as announced from day one. Minimizing the risk of malfunctioning leads to duplication of components or even systems, which inflates costs. A functioning system also is one that is enforceable. Legal constraints – for example, on what constitutes proof of non-payment when payment was due – tend to make enforcement expensive. Adaption of congestion charging systems can save costs over time, and much can be learned from the early adopters. Political risk is reduced when local policy-makers can argue that congestion charges are standard and good practice elsewhere. This helps avoid excessive risk aversion and over-specification in system design. Stockholm began with a system that installed DRSC transponders on all vehicles. But as the legal requirements for proof of identity for enforcement pushed development of the camera-based, automatic number-plate recognition system to levels of performance far beyond capabilities in the early stages of design, transponders became redundant, thus eliminating one cost element.

Acceptability requires that the congestion charging system accommodates occasional users[10]. These users are not familiar with the system and cannot be expected to subscribe to cost-cutting services in the same way as frequent users. Making occasional users pay less than frequent users poses an incentive problem, in that everyone will try to look like an occasional user, and raises fairness concerns. Making occasional users pay more may pose political or legal issues[11], although it is a frequent feature of public transport and many other service products. Accommodating occasional users is a key driver for costs and for the overall design of the system, in the sense of limiting system complexity and versatility. Occasional users may generate little revenue, particularly in relation to the large costs they impose on the system.

Costs are also affected by interoperability requirements. The costs are not so much driven by technological as by procedural requirements, i.e. who is responsible for what part of handling transactions[12]. Since incentives for interoperability are weak from the point of view of an individual system, progress has been slow in the EU, inducing European authorities to mandate it.

2.1.5 Differing approaches: value-pricing and area-pricing

The Singaporean and European examples of congestion charging use cordons to charge for entry or travel in a congested zone. This is different from the value pricing systems used in the US, where travellers on a particular (segment of a) facility are given a choice between using faster toll lanes and slower free lanes[13]. The introduction of value pricing was a response to shortages of funds to provide new capacity, but interest for applying it to fund maintenance or to make better use of existing infrastructure is rising ("value pricing 2.0", Poole, 2009).

One potential explanation for the different approaches lies in the typical spatial structure of urbanised (congestion-prone) areas on either side of the Atlantic. European cities correspond more closely to the monocentric, radial archetype, while US cities are more polycentric and grid-like[14]. The European pattern lends itself more to the introduction of cordons, as there are "natural" boundaries, which are less self-evident in the typical US context. While there is some truth to this argument, it is not complete. It is, for example, hard to see why value-pricing could not be introduced *if desired* on the A1 Highway from Charles de Gaulle Airport to central Paris, when it is possible to create a dedicated taxi lane. Similarly, the bus lane from Heathrow Airport to London could be turned into a HOT lane. Spatial structure does not prevent facility pricing from being introduced in Europe. Similarly, even in grid-like cities such as Los Angeles or Atlanta, centres do emerge, making the design of cordons possible in principle. The London Congestion Charge covers less than 3% of the area of Greater London and many cities have congested hubs.

If differences in spatial structure provide a very partial answer at best, what explains the differing approaches to congestion charges[15]? A widely-held view is that the philosophy underlying charges differs fundamentally between the US and the EU. Value-pricing reflects the view that the charge is in return for higher-quality service, whereas the cordon approach is more in line with an internalisation perspective. Of course, these points of view are not irreconcilable as such: value-pricing is an approximation to internalisation and cordons provide higher-quality travel. Nevertheless, there is a real difference: with a cordon there is no avoiding payment for car users, whereas at a value-pricing facility drivers can decide at the last second which lane to use. Public transport could be seen as a lower-quality alternative to car use within a cordon, and using interregional roads instead of tolled highways could be seen as offering a similar choice, but in these cases choices to use either alternative need to be made at an early stage in the trip, and involve more than just using one or the other lane on a particular facility.

One more real difference between cordons and value pricing is that value pricing up to now has been associated with providing new capacity or with providing access to spare capacity (on HOV lanes) at a premium, whereas cordons are about charging for the use of existing, congested roads. In sum, value pricing emphasizes the supply of premium service while retaining the option for an easy switch to a lower-quality alternative; cordon pricing emphasizes maintaining reasonable service levels overall. Economic assessment (e.g. Small and Yan, 2001) shows that both options produce better outcomes than when there is no pricing. Value pricing can perform better than cordon pricing by offering greater discrimination in the choices users can make. This can be used to maximize benefits as users differ in how much they are willing to pay to gain time. Evidently, in terms of making efficient use of the road network, value pricing (partial facility pricing) is outperformed by well-designed full facility pricing schemes.

2.1.6 Pricing parking

When not in use, cars are parked. Parking requires space that could be used for other purposes so there is a resource cost associated with it. Many drivers do not pay the resource cost of parking even when there is a charge for on-street parking. They may park for free at work or they may pay a fee that is below resource costs. Most stores with parking lots do not charge directly for parking. These parking policies lead to inefficiencies in the transport market, creating excessive demand for underpriced parking spots and resulting in more traffic than if parking was charged for directly. These inefficiencies can be very large: some evidence (Calthrop *et al*., 2000) suggests that the gap between private and social costs per kilometre is larger for the parking inefficiency than for the congestion inefficiency.

In debates on congestion charging this raises the following questions: what would be the effect on congestion if parking were priced differently, and could parking charges be used instead of, or in conjunction with congestion charges to tackle the congestion externality?

Direct charges for parking would have a number of effects. Congestion would fall, optimal congestion tolls would be lower[16] and the net benefits from congestion tolls would also be lower, as congestion is lower to begin with. This is of importance, given the high costs of running a congestion charging system compared to the lower cost of modifying parking charges.

That parking charges are relatively cheap suggests they might be used instead of congestion charges to manage congestion in particular locations. Parking charges then would rise above parking costs and congestion would fall because the demand for travel from those wishing to park in the centre would fall. How effective such an approach is depends on how much through-traffic there would be, i.e. how many drivers would use roads in congested zones but not park there, and on how much “cruising-for-cheaper-spots” would be induced[17]. With limited through-traffic and little extra cruising, parking charges can mimic congestion charges well and attain similar efficiency gains. However, excessive parking charges would also generate efficiency loses and equity concerns – very high charges for those that park and no charges for those that don’t – reduce the appeal of this approach.

The key message from the debate on parking pricing is not so much that parking charges could mimic congestion charges (that is possible, but very much context-dependent) but that restructuring the way parking is paid for would lead to better use of space and to less congestion where capacity is scarce.

2.2. Insights from Singapore and Stockholm

The cases of Stockholm and Singapore were debated extensively at the Round Table, on the basis of background papers on those charging systems[18]. This section presents highlights from the discussion. It does not contain complete descriptions of the Singapore and Stockholm schemes and does not summarize the background papers.

2.2.1 Singapore

The Singapore congestion charging scheme (see Chin, 2010, for a detailed description) is a demand-management system, not a revenue-generating device. It is one component of a broad transport policy that also relies on a vehicle licence quota, on infrastructure planning and on public transport provision to offer high-quality transport options to a growing number of users at a reasonable cost. The vehicle quota system is intended to keep the growth of the vehicle stock roughly in line with the planned expansion of available road space (with allowed growth of the stock recently reduced from 3% to 1.5% per annum). Surveys indicate that car users are mainly interested in high-quality road transport (more road investment, effective congestion management) than in cheaper public transport alternatives.

The approach to system design is pragmatic. The system has become gradually more refined over time, moving from manual to electronic fee collection and enforcement, and covering more of the city as roads become busier, with initially 33 and now 66 gantries for automated control. Charges differ between gantries and vary with the time of day. In 2009, active management of the morning peak through changes in charge levels was extended to cover the evening peak hours too. Rates are revised every three months in order to keep speeds between 45 and 65 km/h on the freeway links in the charged area[19]. Rate changes respond to perceptible changes in congestion levels. Extensions to the system and significant changes in rates are accompanied by major communication efforts. Changes in revenue and concerns over revenue neutrality are of secondary concern, although the scheme is intended to be revenue-neutral in principle. Revenues from congestion charges are dwarfed by those from vehicle licences, so that changes in congestion charge revenues are not a major policy concern. Revenue neutrality was ensured at the time of the introduction of congestion charges through reduced vehicle taxes (government actually lost revenue as congestion charge revenues were overestimated). Revenues are not earmarked and are in fact lower than road and public transport spending.

The incremental development of the congestion charging system has had the benefit of making the more sophisticated later evolution easier for users to adapt to. Starting with a simple system and making it gradually more complex might be expected to help improve acceptance but, at the same time, too simple a system may lead to too large a share of benefits foregone (cf. the discussion on Stockholm). The gradual expansion of the Singapore system reflects pragmatic responses to evolving circumstances more than an explicit strategy.

The Singapore case is atypical in a number of respects, some by virtue of the geographical situation of the city and some by design. First, the number of foreign vehicles (occasional users) is small and easily identifiable with foreign licence plates, allowing design choices geared towards more frequent local users that may be hard to duplicate elsewhere. Second, the ability to change prices every three months cannot be replicated everywhere - the example of Stockholm was given, where a rate change might easily take two years because of legal requirements. By contrast, tolls in some value-pricing systems in the US (e.g. the I-15 in San Diego) are adapted every six minutes. This is accepted both in Singapore and California, as the goal is to maintain free-flow speeds. Such a rule-based pricing system may be easier to accept than one where prices are determined by discrete decisions. More frequent changes allow better congestion management and they may also serve acceptability, as large rate changes can be avoided. Finally, the charging system in Singapore is just one component of a broader system that manages supply (road infrastructure) and demand (vehicle licences and charges) with a view to what performance levels need to be reached. Ownership policies are more restrictive than in many other countries[20]. In other words, if ownership policies were less restrictive, the congestion charges would have to be higher to attain similar service levels (travel speeds).

The Singapore approach can be read as one where congestion charges are used to fine-tune overall transport prices so as to obtain acceptable service levels throughout the system. Other proposed or existing congestion charging schemes are not limited to fine-tuning to the same extent. Given the weaker degree of integration of the various components of transport policy elsewhere, the mismatch between demand during peak hours and capacity may be larger than in Singapore in many cases.

2.2.2 Stockholm

The Stockholm congestion charging system (see Eliasson, 2010 and Hamilton, 2010 for an in-depth discussion) is effective in reducing traffic volumes and increasing travel speeds[21]. Traffic volumes have declined in all time periods, suggesting there is more trip elimination than rescheduling. Commuters tend to reschedule, but trip purposes are strongly diverse, with about 40% non-commuting trips and many occasional users. Cost-benefit analysis suggests that the system produces net benefits; 75% of gross benefits come from time savings, the remaining quarter mainly from better air quality. The Stockholm evidence also suggests only limited direct gains came from the expansion in public transport expenditure that was part of the overall transport policy reform.

Public acceptance at this time is sufficiently broad to expect indefinite continuation of the system's operation[22]. Acceptance is related to congestion reduction, but also to the improvement of the urban quality of life (e.g. less traffic, less pollution) and with the perception that support for the scheme reflects – and signals – green preferences. In this sense, the Stockholm case illustrates that perceived benefits are not necessarily entirely the same as benefits included in traditional appraisal.

With respect to the relation between acceptance, political risk and system costs[23], the Stockholm case is a clear example of the cost-inflating effect of strong risk aversion (cf. Section 2.1). In addition, it illustrates how political risk is transformed into risk for the administration responsible for designing and implementing the scheme, and how this administration shifts risk to the companies selected for executing the plans. The administration passes on risk by adding features to the charging system and assigning legal liabilities to contractors, who respond by "over-specifying" system components and building redundancy into the system, all of which inflates costs. It deserves emphasis that concerns about acceptance create these risks. It follows that if acceptance of a project can be won early, risk is lower throughout the design and implementation stages, and this allows costs to be cut. If acceptance is won late, as in the Stockholm case, then higher costs are incurred in early stages as a form of insurance. Once it becomes clear that risks do not materialise, insurance expenses can decline, and project costs can gradually be trimmed down[24]. However, higher costs were incurred in the past and irreversible design choices partly drive future costs, so that gradual cost cutting does not allow the recovery of all expenses associated with high initial risk.

3. NEW INSIGHTS FROM TRANSPORT ECONOMICS?

3.1. Refining the basic argument for charges

The basic economic argument for congestion charges is well established. In a nutshell, it says that since travel times increase with traffic volumes, an additional car on the road slows down all other cars, increasing time costs for all the occupants of all the cars. The decision to travel made by the occupants of an additional car is based on their own travel costs (their private or internal costs). They ignore any increase in travel costs for all other car users (the external costs)[25]. This is inefficient when private costs are below the full social cost of the decision to travel. When decisions are made on the basis of "underestimates" of costs, too much of a good (in this case, travel) will be consumed. A congestion charge is intended to confront users with costs imposed on other users, so as to align private costs with social costs. The charge will suppress part of demand, reduce congestion and increase surplus.

This simple rationale for congestion charges is based on a range of explicit or implicit simplifying assumptions. Research on what happens if these simplifications are dropped is progressing rapidly. The question here is what recent research insights tell us about congestion charging policy. In answering that question, Fosgerau and Van Dender (2010) focus on design issues more than on acceptability concerns. The key message is that allowing for more complexity in the analysis strengthens the economic case for congestion charging.

A first remark on the basic argument is that it relies on a flow model of congestion, where speed declines because distances between cars decline with increasing traffic density. An alternative and at least as relevant model of congestion focuses on bottlenecks, where queues appear when demand exceeds the capacity of some part of the road network. Bottleneck congestion models highlight the possibility of trip rescheduling. In basic bottleneck models, tolls are used to affect users' decisions on departure times, so that queues disappear. Drivers' costs of waiting in line are replaced by toll costs, which generate revenues that are not lost to society, in contrast to the costs of waiting. This model illustrates that the rescheduling of trips can generate very large social benefits, a dimension of social gains from pricing that is obscured by the standard model of flow congestion. A full appraisal of congestion charges should take explicit account of rescheduling effects.

A second research strand concerns heterogeneity among travellers in terms of their values of time. Here, empirical evidence has produced a stylised fact: there are typically many travellers with low values of time and fewer travellers with high values of time. The range of values is huge and the upper end of the distribution has a long tail. The presence of strong heterogeneity has some immediate implications. First, introducing a toll increases the value of time of the average road user by suppressing trips associated with low time values. It follows that the time losses imposed by one driver on other users increase, so the equilibrium toll is higher than would be derived on the basis of the pre-toll average value of time. Because the distribution of values of time among users is not symmetrical, the effect on the average value can be quite large. Second, congestion charging schemes that maximize the number of alternative responses, as is the case with value pricing schemes in the US, can be seen as forms of product differentiation that are strongly welfare-improving when people differ. The point here is not that value pricing would not produce gains in the absence of heterogeneity (it does) but rather that gains are stronger when there is heterogeneity.

A third body of research focuses on defining and measuring the value of reliability. Whereas the basic argument for congestion charges focuses entirely on how travel time increases as congestion rises, in practice, travellers care about expected travel time and about travel time risk. The expected cost of a trip, on which travel decisions are based, is higher when the trip is expected to take longer, and it is higher when the probability of deviations from the expected travel time is larger (i.e. reliability declines). Travel time risk is positively correlated with expected travel time, but it is not the same. Focussing on travel time and ignoring reliability implies underestimating time costs so, again, tolls are higher when reliability is taken into account even if the relationship is not simply linear (see ITF/OECD, 2010).

Fourth, the understanding of what congestion means in an economic sense is evolving. For example, the standard model of charging systems presupposes that more traffic flow leads to higher travel time. However, in practice, hypercongestion can and does occur (the stage at which traffic slows to such a degree that flow – and not just speed – decreases as the rate of vehicle arrivals at a bottleneck increases). Standard traffic models ignore the possibility of hypercongestion, rendering their relevance to the preparation of congestion policy problematic and implying in particular that these standard models underestimate how high tolls should be set to alleviate congestion by any given amount[26]. Again, this points in the direction of higher congestion charges for optimal outcomes.

It should be underlined that the shortcomings of standard traffic models do not mean they are useless. Improved understanding of congestion allows better use of conceptually simple models, e.g. by establishing the direction of error. More sophisticated models can replace the simpler ones when they become available but there is no need to postpone the use of models in policy design until then. It was pointed out repeatedly at the Round Table that model-based judgment on where to locate tolling points and what tolls to charge performs better than common-sense judgment.

3.2. Congestion charges in a broader economic context

The basic justification for congestion charges is that confronting travellers with costs they impose on others that they would otherwise ignore improves welfare. The charges remove an inefficiency. If congestion were the only inefficiency in the economy, the argument would be complete. However, as the economic theory of “second-best” suggests, congestion charges potentially trigger complicated interactions with other inefficiencies and these interactions might affect policy recommendations.

An extreme example of such interactions concerns labour markets: if all travellers are commuters who have no choice other than to use their car if they want to get to work, and flexitime is not an option, should a congestion charge be introduced if taxes on labour are already high? The answer is no, unless the labour tax were too low to cover marginal external congestion costs. A slightly more realistic model would allow different travel modes and different trip purposes. The core message remains the same: avoid increases of higher effective taxes on labour if that is possible, e.g. by making charges paid for commuting trips deductible from income taxes, as is effectively done in Stockholm[27]. Care should be taken, however, to treat different commuting modes on an equal footing, e.g. season tickets for commuting by public transport are also eligible for tax rebates. Tax systems where commuting tends to receive favourable tax treatment in general may be more suitable to allow tax deductions of congestion charges than systems where no such treatment exists.

Discussions about deductibility are part of a larger debate on how revenues from congestion charges should be used. As pointed out before, charges generate substantial amounts of revenues, and misusing them could easily dwarf the gains from reduced travel times and ancillary benefits. If revenues were simply burned, no welfare improvement would be possible from congestion charges – time losses would simply be converted to monetary losses. Revenue use should be at the core of the design process as it affects to what extent society is better off overall with charges. How revenues are used has an impact on how benefits from charges are distributed, which in turn has direct links with acceptability[28].

There is often tension between what economics suggests concerning revenue use and what is seen as practically feasible and desirable. For example, from a practical point of view it is often proposed to return revenues to car users by reducing other taxes. But car users gain from faster travel as a result of congestion pricing, so they would be overcompensated if all revenue also accrued to them.

4. CONCLUDING REMARKS

Next to the subsidization of parking, the failure to charge for the external costs of congestion is one of the main inefficiencies in metropolitan transport systems. Recent economic evidence strengthens the case for using charges to bring congestion closer to efficient levels. The key to successful implementation of congestion charges is to get the policy accepted. Acceptance is dynamic. It can be managed to an extent and depends on a number of factors, including reduced congestion. Ensuring acceptance may require giving up some of the benefits of a closer-to-ideal system; but less-than-ideal systems (simple cordons, value pricing schemes) can still be satisfactory. Rule-based systems for changing prices (e.g. maintaining predetermined levels of speed) appear to be more popular than those requiring political discretion. There can be some trade-off between perceived and assessed benefits of charges. The extent to which such trade-offs are made should not, however, be allowed to undermine the core objective of charges – which is to cut congestion. Ancillary benefits, including reduced environmental impacts, can in some cases have an impact on how much to charge, and should always be included in assessments, but they are not the principal goal of congestion charging mechanisms.

Congestion charges potentially raise substantial amounts of revenue, but the systems are costly to run as well. This renders statements about revenue neutrality with new congestion charges risky, as equalizing gross revenues implies lower net revenues when the unavoidable costs of congestion charging systems are taken into account. In general, emphasizing revenue neutrality may reduce policy flexibility. It may, however, be a requirement for getting public and political support. Transparency and accountability in revenue use is at least as important for acceptance. This, rather than revenue neutrality, was one of the keys to success in London.

NOTES

1. Though all forms of congestion charge offer a variety of response options (shifting time of travel, shifting mode, etc.), value pricing differs by allowing last-minute decisions in response to recent information on travel conditions.

2. We refer here to practical proposals, not to analytical work. Assuming neutrality is a useful analytical device, but this as such does not indicate whether it is justifiable or not.

3. The US discussions on usage-based pricing are closely tied to infrastructure expenditures. This is less so in Europe. In London and Stockholm, the introduction of charges did not change plans for road expansion, despite the expected effect that the need for infrastructure becomes less pressing.

4. There are earmarks on how revenues from the Stockholm scheme are spent, which might suggest revenue concerns are crucial. However, the earmarks are mainly an accounting issue, with little impact on prevailing revenue streams.

5. Even here, neutrality is not entirely straightforward, as it can be argued that more revenues should be raised in order to expand the supply of public services when raising revenues becomes cheaper.

6. Whether consumer surplus rises or falls is a broader question, discussed in Section 3.

7. A counter-argument here is that the revenue flows triggered by charges are very large compared to the net benefits from congestion relief. This argument has some merit but tends to be overstated, as discussed in Section 3.

8. Revenue neutrality is difficult to define in a dynamic sense (as it is uncertain how revenues would evolve were there no reform). It was noted that opposition to congestion charges may arise for fear of high future charges, even if current charges do not lead to higher average tax burdens.

9. The view that technology is neutral is supported by evidence that (a) operating costs dominate fixed costs over the life-cycle of the investment, and (b) the costs of an additional user in a system are largely independent of technology, being roughly equivalent to the cost of an internet or mobile phone subscription (i.e. 10 to 20€ per month in 2010).

10. The argument applies to passengers, not to trucks where competition concerns make equal payment for all users imperative.

11. This is mainly an issue that concerns truck-km charges, where discrimination on the basis of country of registration of the vehicles is not tolerated by agreements such as the European Union's treaties.

12. While overall technology costs do not dominate, there may be losers from standardization.

13. We use the term "value pricing" to indicate "partial facility pricing", because of the close connotation with product differentiation. The US Government's use of the term is different, as value pricing refers to both partial facility pricing and cordon pricing – value pricing is, in fact, the same as congestion pricing.

14. There are, of course, differences within both continents, with newer cities in the US being closer to the grid patterns, and spatial structures in Europe evolving away from the monocentric pattern to some extent. Nevertheless, as far as averages make sense, differences between both continents remain.

15. It could be argued that offering travellers a choice between tolled motorways and free but slower roads for interurban travel (as is done, e.g., in France) is a form of value pricing. The principle is not extended to urban areas because building alternatives is too costly. By consequence, congestion will need to be managed by area or by cordon pricing. In this sense, one difference between Europe and the US may be that extra capacity in Europe is more (and too) expensive.

16. Congestion would decline because there is less traffic and because fewer drivers would be cruising for parking (a type of driving behaviour that is particularly disruptive to traffic flows).

17. This in itself of course depends on how many cheaper spots would be available, an issue susceptible to public management.

18. http://www.internationaltransportforum.org/jtrc/roundtables.html#RTCongestion_Charging

19. It is likely that these speeds are below benefit-maximizing speeds (i.e. that tolls are below marginal external costs or below second-best cordon levels).

20. The aggregate level of government revenue raised from road users/car owners in Singapore also is unusually high, at around 20% of total revenue.

21. There is evidence for Stockholm that volume effects are highly persistent over time while travel time effects tend to wear off to some extent. This has happened because of an increase in road maintenance works and because some road space has been allocated to other use (neither development related to the pricing system), so it is not in itself evidence of declining welfare gains.

22. Support can decline over time, as evidenced by a Norwegian survey on toll roads. This is because public memory of why the system exists erodes, so that charges – and, more strongly, charge increases – may meet with increased resistance.

23. Assessing whether a system is expensive or not should not be done using revenue/cost ratios, as revenues are endogenous through toll levels.

24. The Stockholm system initially used ANPR as well as transponders. Once it became clear that ANPR was sufficiently reliable – which it became partly because of technology development related to the Stockholm project – transponders were no longer used. This avoided the costs of managing transponders, which were higher than expected.

25. It is sometimes argued that congestion is not an externality because transport users ultimately bear the cost of travel. While the latter is true, one can wonder whether the distinction between transport users and the rest of the economy is artificial. Of more direct importance is that who bears the cost is irrelevant: when individual users make travel decisions and they ignore costs imposed on others, there is an external congestion cost.

26. Ignoring reverse causality in estimation implies such underestimation.

27. Deductibility did not exist during the trial but was introduced afterward.

28. It is sometimes argued that relations between, e.g., labour taxes and transport taxes should not concern policy design too much, given the different political responsibilities. However, analysis points out that accepting such constraints may be costly.

REFERENCES

Calthrop, E., S. Proost and K. Van Dender (2000), Parking policies and road pricing, *Urban Studies*, 37**,** 63-76.

Chin, K.-K. (2010), The Singapore experience: The evolution of technologies, costs and benefits, and lessons learnt, JTRC Discussion Paper 2010-1.

Eliasson. J. (2010), So you're considering introducing congestion charging? Here's what you need to know: a FAQ based on Stockholm's experiences, JTRC Discussion Paper 2010-4.

Fosgerau, M. and K. Van Dender (2010), Road pricing with complications, JTRC Discussion Paper 2010-2.

Hamilton, C. (2010), Revisiting the cost of the Stockholm congestion charging system, JTRC Discussion Paper 2010-5.

ITF/OECD (2009), Improving reliability on surface transport networks – Summary document, OECD/ITF, Paris.

Oehry, B. (2010), Critical success factors for implementing road charging systems, JTRC Discussion Paper 2010-3.

Poole, R. (2009), When should we provide separate auto and truck roadways?, JTRC Discussion Paper 2009-24.

Small, K. and J. Yan (2001), The Value of Value Pricing on Roads: Second-Best Pricing and Product Differentiation, *Journal of Urban Economics*, 49, 310-336.

Critical Success Factors for Implementing Road Charging Systems

Bernhard OEHRY

Rapp Trans AG
Basel
Switzerland

SUMMARY

Basel, January 2010

1. FROM ROAD-USER CHARGING TO CONGESTION CHARGING

Road-user charging is used as an “umbrella” term to describe a wide range of applications for pricing roads and infrastructure. Road-user charging includes a number of charging measures that governments and other road owners use to:

- finance new or maintain existing road infrastructure;
- manage traffic (e.g. reduce congestion);
- minimise environmental impacts of transport;
- internalise the external costs of road transport caused, e.g., by pollution and noise emissions.

Historically, the common approach to charging for road use is some form of general taxation rather than differentiated road-user charging. Road-user charging has long been proposed as an efficient and equitable method to pay for road use and to fund road infrastructure projects. However, there is an important distinction between charging for revenue generation purposes, as opposed to pricing roads to provide congestion relief. The two basic objectives, revenue generation and congestion management, differ in several ways, as shown in the following table.

Table 1. **Road-user charging objectives: revenue rising *vs.* congestion management**

Revenue generation	***Traffic management***
▪ generate funds ▪ rates set to maximise revenues or recover specific costs ▪ revenue often dedicated to road infrastructure projects (construction and maintenance) ▪ traffic diversion to alternative routes and modes not desired as it reduces revenue collections.	▪ reduce peak-period vehicle traffic ▪ used as a travel demand management strategy ▪ revenue not dedicated to road infrastructure projects ▪ requires variable charging rates (i.e. higher during congested periods) ▪ travel shifts to other modes and times considered desirable.

Source: *TDM Encyclopaedia* – http://www.vtpi.org/tdm/tdm35.htm

Road-user charging can improve transport efficiency by rationing road capacity, including influencing the demand for road capacity as it applies to various road classes, vehicle classes or peak traffic conditions. It can be a useful travel demand management tool. Effective pricing schemes can change travel patterns by exposing road users to the marginal social cost of their travel choices. Pricing affects all stages of travel decision making, from choosing to make a trip to destination choice, mode choice, time of travel choice and route choice.

Electronic pricing and related Intelligent Transport Systems (ITS) technologies have matured considerably in recent years. Improvements in coverage, ease of implementation, cost and public acceptance are occurring at a rapid pace.

The importance of road-user charging as an effective instrument against congestion is increasing. The well known London Congestion Charge, the Stockholm City Ring and the Cordon Pricing in Oslo are European examples of successful realisation of such a charging policy.

2. AN INDISPENSABLE CONGESTION MANAGEMENT TOOL

Traffic is increasing, and there is no limit in sight. In many areas of the world, congestion will figure among the most pressing problems of this century.

Supply-side solutions alone do not suffice. Demand management, including the use of price incentives, is also needed. The need to combat traffic congestion and the desire to find new revenue sources for transport investment have stimulated the interest in schemes where charges for road use are introduced, such as parking fees and charges to allow vehicles to use certain roads. The theoretical advantages of charging for road use have long been discussed in the economic literature. Practical experience with such schemes is more recent. A key element of demand management in urban transport is the allocation of road space. As this space is a finite resource, an absence of regulation can lead to overuse, which appears in the form of congestion. Reserving road space for public transport vehicles or private vehicles with high occupancy are two ways of allocating road space. Another way is to restrict access to certain areas of the city. This latter kind of control measure does not make a distinction between trips of different value. Conversely, if travellers are faced with a road user charge, they will be encouraged to make their own judgement on the value of their trip. Charging for road use, often referred to as road pricing, has long been advocated by economists on the grounds that it is socially advantageous. Congestion occurs as every additional trip made forces those vehicles already on the road to slow down. The introduction of a corrective charge will make each driver aware of the cost he imposes on other drivers. This may help reduce traffic volumes and have positive effects in terms of reduction of congestion and overall travel time on the network. In reality, the efficiency effects of road charging policies are the consequences of both the actual behavioural responses of the travellers as well as the way the revenues from charges are spent. In addition to travel time savings, benefits include increased travel time predictability, reduced pollution and noise, reduced accidents and improved travel conditions for public transport (from Ref. [1]).

Economic theory and mathematics of road pricing also dictate that the charge should be equal to the monetary value of the additional travel time imposed by each driver on other drivers. This is the "marginal cost pricing" principle. The principle extends into "marginal social cost pricing" if, in addition to time, other costs, for example pollution, are considered when setting the charge. Economists have gone on to extend the welfare maximization framework to cases where the price of public transport is set together with road charges. The other motivation for introducing road charges is to provide financial resources for infrastructure investment programmes. Revenues from road charges can be spent as well to improve transport at large. Both demand management and fund raising make charging for road use of interest to policymakers.

Research efforts, as well as the introduction of toll rings and congestion charging schemes in a few European cities, in the last two decades provide considerable experience to draw from and support the translation of theoretical principles into practical policies. This paper looks at factors that make a pricing measure a success. The approach taken is not a theoretical one, but looks at practical implementation aspects.

Access to inner cities is becoming a scarce resource and has to be managed. According to market theory, pricing is the most effective means to allocate scarce resources. Congestion pricing will become an indispensable tool for managing the challenges of urban mobility.

3. WHAT MAKES A CHARGING MEASURE A SUCCESS ?

Simply put, a charging measure is a success if it works in all relevant aspects:

3.1. Technical implementation

Obviously, an important precondition of success is a technical implementation that works smoothly, i.e. is implemented in time and at budgeted costs, and operates without hiccups. The choice of technology is one important aspect of the technical system. **Charging technologies** as such are mature but other system aspects remain demanding. One such aspect is the special technical challenge imposed by the **urban surroundings**, where many design constraints apply. Another aspect of core importance is the handling of **occasional users**, who may enter the charging system without being properly informed, unequipped and unwilling to spend undue time and money to become compliant. Enforcing **compliance** of all users is a further critical aspect where no ready-made solution is available. Finally, the technical system has to be implemented at acceptable overall **system costs** and must encompass a **security concept** that defends the system against all conceivable threats.

3.2. Scheme layout

Successful technical operation is only a precondition for a charging measure to achieve its true goals, which are namely, to manage traffic and to generate revenue in the desired way. This requires a proper scheme layout, where above all the **scheme principle** is decisive: namely, whether the charge is simply a flat fee, a distance-based area charge, cordon pricing, or a combination of such basic principles. Besides the principle, the right choice of **charging perimeter** is also decisive, i.e. which geographical areas are subject to the measure and therefore which types of trip become subjected. Whether the size of the effect is as expected will depend on the right choice of **tariff level**. Often disregarded at first, but actually one of the core decisions to make in system layout, is the **legal nature of the charge**, i.e. whether the charge is legally constructed as a public levy, say a tax, or as a private levy, like a fee for use.

3.3. Acceptance

In fact, the single most important aspect for the successful introduction of a charging measure is its acceptance. Without acceptance the system will simply never become a reality. Acceptance needs to be actively planned and **public perception has to be managed**. A precondition to achieve acceptance is that the public perceives a **pressing traffic problem**. No majorities can be found in favour of a pricing measure thought to relieve congestion without sufficient daily annoyance. No less important is the **use of revenue**. If the pricing measure appears to be "just another tax", political failure is guaranteed. Finally, acceptance by the general public requires a **transparent and understandable system** and careful treatment of the user's **privacy**. Acceptance can be increased if **accompanying measures** are designed into the system from the very beginning, with the intention to bring additional benefits to the individual users.

In the following, this paper looks into the three core success factors in turn.

A successful measure

- works smoothly regarding the technical implementation,
- achieves its traffic management goals, and
- most importantly, becomes accepted.

4. TECHNICAL DESIGN ISSUES

4.1. Technology

The choice of technology is less critical than it appears at first glance. Actually, the technologies employed in road-user charging devices are standard technologies; for example, short-range communication (DSRC), satellite localisation (GPS) and mobile communications (GSM).

More important than mere technology is the system concept that is made out of the technical building blocks. Here we see four basic concepts that can all be considered mature, including practical proof in real systems:

– Manual, manned and barrier-controlled **toll stations**. Such stations are the traditional means to collect charges on interurban motorways, especially on networks operated by concession companies. Due to the tight space constraints in an urban environment, such toll stations are rarely an option for congestion charging.

– User **self-declaration systems**. In such a system concept, road usage has to be declared by the user either before or after the trip. An on-board device is not required. The operational processes for such a manual charging system strongly depend on the details of the charging scheme. Users may book and pay trips at kiosks or fuel stations, with their cell phone via SMS or Internet. An indispensable part of any manual system is an Automatic Number Plate Recognition system that can check whether users have actually paid for their trip.

– In **DRSC systems**, on-board equipment communicates with road-side beacons for an electronic charging process. The on-board equipment works as an electronic tag that will be recognised when passing specific beacons, which have to be installed at pivotal points of the road network. DSRC on-board equipment is low price, can be battery-powered and does not need a proper installation.

– **GPS/GSM-based systems** use GPS for localising the vehicle. The GPS positions are then identified on a map and matched to the most likely roads. This map matching can occur in the on-board equipment or in a central system. Cellular communication via GSM or similar transfers either the raw GPS position data or the map-matched road usage data to the back office system. This technical concept requires comparatively sophisticated on-board equipment that needs at least some minimum installation since it requires electrical power supply from the vehicle, either via permanent installation or from the cigarette lighter.

GPS/GSM systems are the ultimate dream of any traffic manager since the capabilities for differentiated charging are nearly unlimited. The price for this is the need for capable on-board equipment, which is not only a cost issue but implies a number of complex challenges, e.g. the logistics of distribution and installation, the handling of occasional users, and means of protecting the privacy of users in a convincing way.

The choice of technology nowadays can be considered a mere engineering decision and should under no circumstances become a political issue ("we want the technically most advanced satellite system"). Technology is mature and choices should be made for technical reasons only.

Several countries have examples of projects where the choice of technology has been made even before other critical questions, such as the determination of the charged road network, of liable vehicles and of basic charging rules, have been made. In particular, satellite technology appears to be magically attractive but should not become an implicit pre-requisite of system design since it is not by necessity the best choice. Especially for congestion charging in the typical city environment, with a small geographical perimeter, many occasional users and limited financial room for manoeuvre, more mundane technologies usually make for a better overall outcome, with lower costs, lower risk and faster implementation.

Technology is mature and available, and as such not a critical success factor – at least as long as optimal choices from an engineering perspective are made.
Technology should not become a political issue or a matter of national pride.

4.2. Urban constraints

In an urban context, a successful technical system design must address the challenges imposed by the urban environment with high priority. Charging technologies have initially been designed for the inter-urban motorway environment and do not naturally blend well into an urban situation. Several issues have to be considered [2]:

4.2.1. Aesthetic impact

The physical appearance of roadside installations is a much more politically sensitive issue in urban areas than in interurban contexts. There are generally tighter restrictions, as well as existing visual, environmental and historical contexts. Street furniture needs to be sympathetic to such contexts, including colour, style, size and location. Only rarely will it be possible to even contemplate the use of gantries and thick structural elements in these locations. This reflects the fact that many more people live and work in urban areas and they have some degree of ownership of that landscape. As a policy, road-user charging is sensitive enough without the controversy associated with physical changes to the local built environment. Therefore, any system that is deployed in the urban environment must be discreet, have minimal impact and be sympathetic to the surrounding environment.

4.2.2. Chaotic traffic behaviour

The traffic characteristics in urban areas are different from interurban contexts, including much more chaotic patterns of movement and behaviour. The urban road can be just as much a destination as it is a through-route, for a wide range of people and goods. With roadworks, building works, parked or static objects, contra-flow bus lanes, slow traffic, overtaking and generally chaotic driving behaviour in urban areas, there is often no real concept of a left- or right-hand "side" of the road, no real concept of a lane, and the potential for unusual manoeuvres (e.g. u-turns and reversing) at any location on the road at almost any time. Unlike interurban roads, urban thoroughfares have a very diverse range of traffic restrictions and traffic management measures on them, including segregated lanes, traffic islands, chicanes, barriers, rising bollards, road humps, textured surfaces, pedestrian crossings and roundabouts. Finally, charges may be applied which are direction-dependant. Therefore, any system that is deployed should provide complete carriageway coverage for the monitored directions and have the capability to determine the direction of travel.

4.2.3. Diversity of road users

The range of objects using or adjacent to roads in urban areas is different from interurban contexts, reflecting the greater diversity of travel activities taking place in urban areas. This includes a much wider range of powered and unpowered vehicles, pedestrians and static objects (e.g. refuse skips, parked vehicles, trees). It is also reasonable to expect that for any particular urban charging scheme there will be a mixture of equipped and non-equipped vehicles legally using the road, with potentially a relatively high proportion of non-equipped users. Again, this reflects the fact that, as a destination, the urban area cannot always be by-passed, in contrast with most interurban routes, and it is unlikely that all objects using the road will be subject to a charge.

42.4. Highly variable road topology

The topology of a road in an urban area is much more likely to vary between different charge point locations than in an interurban context. Road widths are highly variable, ranging from as little as three metres through to five or six lanes in each direction at busy intersections. Footways, narrowing roads, bends, skew junctions and roundabouts all reflect the extent to which urban roads are as much multi-purpose spaces between the buildings (and the subject of historical precedent and shared usage) as they are a thoroughfare designed to move traffic.

Figure 1. **Examples of variable road topologies**

Source: Courtesy of Transport for London.

4.2.5. Challenging installation

With lower traffic speeds, urban roads are much more likely than interurban roads to have other physical structures immediately adjacent to, over and below the road surface. This will include railway lines, tram lines, power lines, telephone lines, buildings, sewers, ducts, water, gas and electricity supplies. The works involved in constructing charge points may require a degree of consultation with the owners of such assets in terms of disruption and future access. This creates straightforward physical as well as logistical and administrative challenges in trying to erect structures, tune performance and maintain systems. Ultimately, this may limit the range of locations where charge points can be erected.

> The special challenges of the urban context need to be addressed from the very beginning and are not restricted to purely technical issues but also a matter of aesthetics and perception of the required technical gear.

4.3. Occasional users

The handling of occasional users strongly determines charging system design. If charging requires the use of on-board equipment, then it can be assumed that frequent users will equip themselves in order to take part in the scheme. But, according to the UN Convention on Road Traffic [3], all users, including non-equipped ones, must equally be admitted to the road network.

The UN Convention on Road Traffic of 8 November 1968, also known as the "Vienna Convention", is an international treaty that facilitates international road traffic by standardizing uniform traffic rules among the contracting parties. It defines that signatory states have to unconditionally admit vehicles to their territories if they fulfil the requirements of the convention.

Charging on-board units are not a requirement listed in the convention. Hence, also vehicles without installed charging devices have to be admitted to the road network. It is possible, though, to require vehicles to carry something non-permanently on board, such as a simple sticker or a windscreen mounted tag.

Occasional users will not be prepared to equip themselves with permanent equipment. If the technical solution is based on complex on-board equipment with a high degree of effort for integration into the vehicle – requiring, e.g., several hours of installation in a dedicated workshop – then a second solution must be offered to cover the needs of occasional users. This second solution must ensure that unequipped or occasional users can have access to the system in a simple way and with minimal effort, for example, with a ticket-based solution.

In charging, equal treatment means that all user groups pay the same when using the same roads under the same conditions. System design must ensure that equipped users essentially pay the same as non-equipped ones. Especially with vehicles coming from abroad, it is usually not possible to give equipped users better tariffs or to treat occasional users with a simplified charge without violating international treaties, such as European Community law.

Therefore, even the most sophisticated technical charging solution is limited in its charging capacity (tariff modulation, complex tariff structure, flexible extensibility, etc.) to the capacity of the solution for the occasional user. Only what can be done for the occasional user can also be done for the equipped frequent user. One solution to the problem of occasional users, who are often foreign users in transit, is to construct the road-user charge as a national tax which foreign users do not have to pay.

In city schemes, a high number of users will be occasional visitors from the wider surroundings of the city. Costs to treat these users adequately are usually rather high, and revenue will be very limited. In many systems, the handling of occasional users is the single most important cost driver. Hence, a proper solution for the occasional user has to be seen as a core problem of system design, and is by no means a fringe problem.

Occasional users require high consideration in system design. A simple and cost-effective solution for this user group is decisive for the overall success of the system.

4.4. User compliance

A charging system is all about compliance. There are traditional toll stations where the user stops in front of a barrier, then throws coins into a basket and the barrier will open. If there is no coin in the basket the barrier does not open. The barrier, and not the basket, has made the user pay.

In an electronic fee collection system, the on-board equipment will automatically, via some modern technology, “throw coins into the basket” and free the user from this task. But there is a catch: the on-board equipment replaces the basket but does not replace the barrier. The vehicle is not stopped by the on-board equipment if the coin is not thrown. Hence some kind of virtual barrier is required to make the user pay. This virtual barrier is the compliance checking system. In essence, a charging system is as good and reliable as its compliance regime.

Figure 2. **Compliance control by barrier**

The aim of compliance activities is that, on average, a compliant user pays less than a non-compliant one. There are two extreme options to achieve this target. The first is to have many compliance checks and a low penalty. The second is to have only occasional compliance checks but high penalties. There is a limit to penalties which is given by political and legal acceptance. This limit defines the minimum density of compliance-checking activities required in a system.

Advanced technical solutions exist that allow for charging and enforcement in free, unconstrained traffic. Usually, there is a mix of permanently installed automatic enforcement stations combined with a certain number of mobile enforcement units, i.e. manned patrol cars with special enforcement equipment. The fixed enforcement stations check the correctness of payment of vehicles passing by without any obligation to stop. This is achieved by a combination of technologies, including laser scanners to determine vehicle and tariff class, automatic licence-plate reading equipment to check for user registration, and video cameras to create pictures for proofing the case.

The processes of compliance checking and prosecution cannot fully be transferred to private organisations because of legal constraints. Especially for a public levy, enforcement is under the overall responsibility of, and has to be controlled by, an agency equipped with the necessary legal powers.

Enforcement of correct toll payment is amongst the most important single issues in a charging system. The reliability and the strictness of the enforcement system is the basis for ensuring the revenues and the overall acceptance of a system. Only an enforcement system which detects violators in a proper way can assure the integrity and acceptance of the system. In a system where incorrect payment or non-payment is tolerated, confidence in the system decreases and acceptance by the paying majority may become critical.

The compliance system is also a major cost driver. For every system, there is an optimum balance between enforcement effort and revenue loss due to violations.

> Every charging system is only as good and reliable as its enforcement system.

4.5. System costs

Much has already been said about cost drivers in charging implementations. Experience shows that people have a tendency to focus on capital expenditure at implementation time and often pay less attention to operational costs.

In fact, one should consider overall costs of ownership over the whole operational lifetime. In our experience from many implementations, the operational costs over a typical system lifetime of ten years exceed the implementation costs by about a factor of ten.

Cost drivers are all non-automated processes needing manual intervention:

- Handling occasional users creates little revenue but requires costly infrastructure and a dense network of points of sale. In addition, every single payment transaction comes at some cost.
- Operation of the compliance system requires a considerable number of staff, either on the road for compliance checking patrols or in the back office for checking automatically generated evidence.
- Back-office operations are costly, especially if there is high number of non-standard activities, such as user data mutations and user requests, or complaints via the user front desk.
- Payment transactions consume a fraction of the revenue. Every loading of an account and every credit card transaction has an associated cost. Costs of payment are usually in the order of 2-5% of total revenue.

A successful system must focus on good solutions for the cost drivers. Neither policymakers nor users easily accept a charging system that itself consumes an appreciable fraction of the revenue. It is rather important, though, to communicate quite early in the political process that a complex measure such as a time-distance-place differentiating charging system comes at a cost. It is close to impossible to design a system where operational costs are below 10% of revenue. Operational costs of 15 to 20% of gross revenue are a more typical goal for a city-size charging system. Even these figures have proven to be difficult to achieve in practice.

> Operational costs by far outrun implementation costs. Complex charging systems come at a cost. Realistic costs have to be communicated as being an intrinsic element of capable road pricing solutions.

4.6. Security

Charging systems are payment systems with large money flows. System security is an important question that has to be considered in the design of every aspect of the system. The issues regarding the security of a road-user system can be grouped into:

- Security of the data, payment and information flows against fraud and external interference;
- Protecting the operations against system breakdown;

– Overall system security regarding the correctness of the charged toll and its solidity in case of legal disputes.

External and internal threats to the correct flow of data, payment and information must be analysed and appropriate countermeasures designed. High attention has to be paid to all conceivable attempts by users to fraud the system. Available countermeasures are physical protection of critical system elements and cryptographic protection of communication channels. In particular, fraud by internal users (tolling personnel) has to receive high attention.

In the case of system breakdown, it is necessary to have a back-up solution ensuring that no loss of data occurs. The probability of an extended system breakdown strongly depends on the technical approach chosen, the dependence on external entities and the system distribution into central and decentralised components. Core central components have to be designed fully redundantly.

Ultimately, system security relies on the availability of data that allow monitoring of the system. Errors and frauds can only be detected if sufficient data is available. It is essential that security requirements are designed into the system from the very beginning in order to be able at any time to have correct data that can stand detailed scrutiny in front of a court of law, in case the need arises.

Road-charging systems are large payment systems. System security is a critical design issue. Security is an end-to-end problem, where the complete, long path from use of the road to booking of the payment in the central system needs to be protected.

5. SCHEME LAYOUT ISSUES

5.1. Scheme principle

In theory, the charging scheme, as defined by its rules for charge determination, can take many forms. Basically, we distinguish cordon-oriented schemes, where crossing of a defined border triggers the charge, and area schemes where being in a certain area defined by a border triggers the charge, which might be a fixed sum per day, time-dependent such as a parking fee, or distance-dependent. Schemes can be designed with many complex variations: e.g. where one charging cordon with a high tariff is situated within another, larger cordon with a lower tariff, or where a distance-dependent system is simultaneously time-dependent, with tariffs varying according to the time of day.

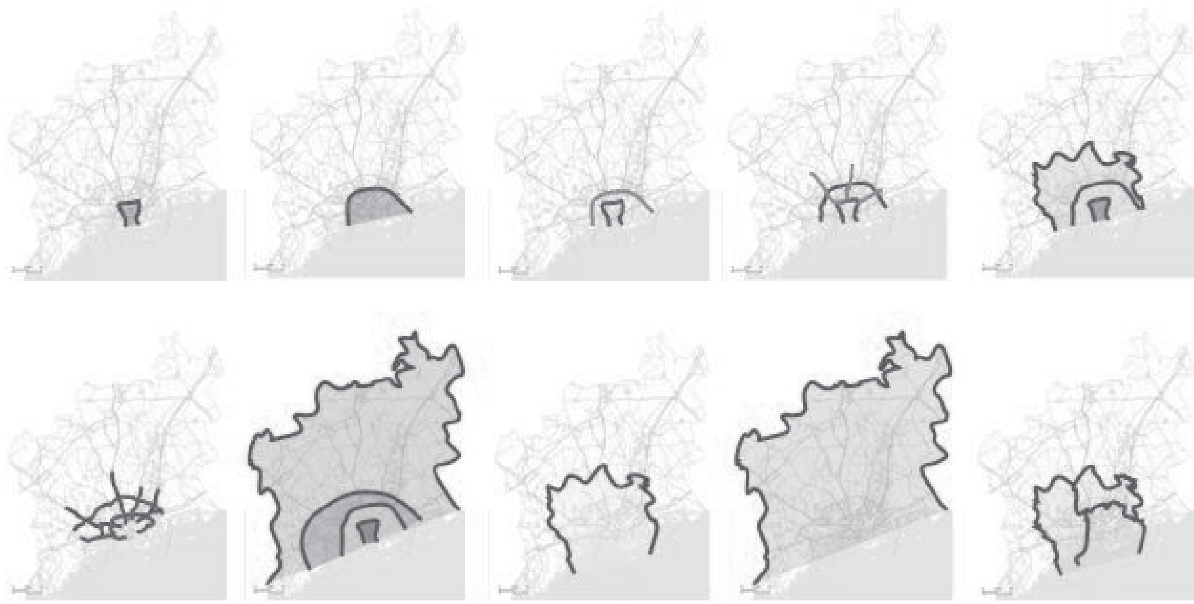
This is not the place to contemplate all available possibilities. Modern charging equipment, especially of the GPS/GSM type, will enable all conceivable scheme principles. In practice, only the most simple scheme principles stand a chance of realisation, for the following practical reasons:

– In order to have an effect on the user, the system must have simple rules, like “driving in the rush hour is costly” or “driving off the motorway is costly” or “using an ‘old banger’ is costly”. Users must be able to react to the price incentives given. This means that the rules must be understandable and must give clear indications on how to best react.

– In fact, the system component caring for the occasional user will limit system complexity. As discussed above, it is difficult to equip occasional users with sophisticated equipment. Occasional user solutions typically are simple booking systems that cannot support a complex, highly differentiated and variable charge.

– As laid out in the next section, every city has a structure leading to natural perimeters for charging zones. Theoretical considerations work in idealised surroundings. In practice, historically established structures of spatial development or of traffic flow will themselves lead to a very limited choice of feasible options.

Figure 3. **Ten options discussed for a congestion charging scheme for the city of Helsinki** [4]



Marginal social cost pricing is a theoretical concept. Practical implementations will try to consider these results but will in practice be limited by very mundane considerations.

In theory, charging principles of any complexity can be designed that would lead to optimum outcomes and efficiency. In practice, systems must have simple rules in order to be understandable, must support simple solutions for occasional users and must work in grown urban patterns. "Keep it simple" is the best advice one can give in this respect.

5.2. Geographic perimeter

When road pricing is being planned as a tool to combat congestion, it is vital to choose the right border of the area where the charging sets in. This is true irrespective of the choice of scheme principle, e.g. cordon pricing or area pricing. It is essential to set the border such that a large fraction of the commuter traffic is being captured, without disrupting modes of traffic that are less relevant for congestion, such as delivery traffic, transit, occasional trips and – especially critical for some stakeholders – shopping traffic.

Again, theory will provide for excellent guidelines on which line crossing should trigger a charge, or at least start liability or an increased tariff. In practice, always there is very limited choice. Many cities have a natural structure that is already visible from a cursory glance at the city map, and that is also reflected in the mental representations, the "personal maps", of the city's inhabitants. Such natural structures might be a ring road as, for example, the perimeter of the London congestion charge, or a natural barrier, such as the waterways surrounding Stockholm's inner city.

Figure 4. **Charging perimeters of the London and Stockholm congestion charges**

Although such natural choices for the charging perimeter are usually not the ones that would be optimal from theory, they are very effective in practice. Over a long period of time, such natural structures have been shaping the spatial development of living and working areas, of shopping streets and quiet neighbourhoods and, to an amazing extent, even cultural identity. For a Londoner born south of the Thames it is close to unthinkable to live in the north, and *vice versa*. These grown spatial and behavioural patterns are reflected in the daily traffic flows. These considerations may explain why "logical" or "natural" choices of charging perimeters, even if not necessarily optimal in theory, will normally be quite effective in practice.

Perimeters cannot be chosen in a clean environment. Since the western extension of the London Congestion Charge has been implemented, an uncharged transit route cuts through the congestion charging zone. In Stockholm, inhabitants of the Lidingö island cannot travel anywhere without crossing the charging zone. This situation requires a special arrangement. The inhabitants of this special area may cross the inner city at any time without being charged. The administration of this special measure represents an appreciable portion of the operating costs of the Stockholm congestion charge.

When setting the perimeter for the charging measure it is also important to consider the availability of other transport modes. In many cities there is a rich choice of public transport means in a radial direction, i.e. in and out, but much less convenient choices for going from one peripheral destination to another. Paris is a very good example of such a centralised city. If one is living in the west but working in the north, there are far less good public transport connections as compared to living west and working centrally. Congestion charging will only flourish if good transport alternatives are available to commuters. These considerations are to be taken into account when designing a pricing measure, and especially when defining the charging perimeter.

Most cities have historically developed structures that lead in themselves to natural choices for charging perimeters. It is wise to benefit from them rather than to contemplate "optimal" but artificial boundaries.

5.3. Tariff level

It is difficult to find the right tariff level for a congestion charge. The charge must be felt sufficiently such that people will react, and at the same time be acceptable enough that people will not heavily oppose it.

Theory offers several approaches to tackle the problem. One can look, for example, into the value of time in order to see what the monetary equivalent of a quarter hour lost in congestion might be, or one can look at marginal congestion costs to society. In practice, these deliberations may give some indication of where the right tariff level should be, but ultimately the tariff is a political decision that will rather be oriented towards creating acceptance than at creating an optimal traffic management outcome.

The political promise is often of "revenue neutrality". This sounds like a good deal for the citizen: before introduction of the pricing measure, everybody pays some fixed tax or fee, such as an annual registration fee, vehicle tax, or the like. If this fee is replaced by a variable congestion price that is revenue neutral, it is likely that the citizen will "buy in". "If I have to pay anyway, why not pay for something intelligent, which has a positive traffic effect?" In addition, most citizens will drive less than average and are most likely better off with the new regime, thereby creating a natural majority of winners.

This sounds like a clever strategy to create acceptance, but in fact carries some inherent risks. In most communities, tax income from the road sector cannot decrease, under any circumstances. Revenue from the road sector is essential for financing many public duties, like road maintenance and construction, or even as a major contribution to the general community budget. Hence, whatever the new pricing measure, net revenue must remain at least the same. Neutrality in terms of net revenue means that gross revenue has to increase – but gross revenue is what is being collected from the users. The difference between net and gross revenue is the collection costs, which are close to zero for annual registration fees and similar levies but, even under very favourable circumstances, can hardly lie below 20% of revenue for city road pricing schemes.

Hence it is politically dangerous to promise revenue neutrality when in fact gross revenue needs to increase in order to keep budgets balanced. Public acceptance might well be improved if, from the very beginning, the pricing measure is designed to create additional income which is earmarked for projects widely accepted as beneficial, such as the construction of a new bridge or tunnel or the introduction of new tram or bus lines.

The right tariff level is a political decision rather than the result of theoretical considerations. "Revenue neutrality" sounds like an attractive way to go, but has proven to be difficult to achieve in practice.

5.4. Legal status of the charge

A road charge can have the legal status of a tax, which is a public levy that is owed unconditionally, or a fee, which is a private levy and the price of a particular service consumption.

In the case of a tax, only an authority has the fiscal sovereignty to collect it. Daily operations may be outsourced but control and governance must remain in the public sector. A tax is usually defined as a levy that does not give the right to a certain service. Taxes are simply instruments for financing community budgets, and strict regulations and procedures apply: there exists, for example, a well-defined appeal process. Toll charges with a tax status usually rely on a user's self-declaration of road usage, and the technical equipment is merely considered as a tool to help the declaration. For a tax, compliance checking and prosecution are legally quite simple to implement, since public servants can be sworn and receive the necessary powers, e.g. to stop a vehicle, or to access the national licence-plate database.

In contrast, a fee is usually defined as a charge that is directly related to the use of a service. A fee can be collected by a public authority or a private organisation alike. With fees, the freedom in system design is higher since the environment is less regulated. For a fee it is, for example, acceptable that the road-side system measures road usage without a user declaration. Compliance checking and enforcement is more of an issue, though. A private company cannot prosecute ordinary citizens, as could a sworn public servant. For a fee, special legal or organisational measures are required to set up a tight enforcement regime.

Often, a simple way to distinguish between tax and fee is through Value Added Tax (VAT). If a road usage charge is legally considered a tax, it is not subject to VAT. If the charge is considered a fee, VAT normally applies. VAT adds some complexity to handling and processing payments, but is not an issue for the road-user charging scheme itself.

The legal status tax or fee might appear to be a minor issue, but in practice it is of crucial importance to get the legal status right. For the Stockholm scheme, for example, this single issue has caused many months of delay. The legal status reaches deep into many aspects of the charging processes; from the right to collect, the need for fully equal treatment over the legal significance of the user "declaration", compliance and enforcement powers, and applicable court processes for legal recourse, to the applicability of VAT.

> The legal status "Tax" or "Fee" is not a minor design issue but of crucial importance and requires early and careful consideration for a successful implementation.

6. ACCEPTANCE ISSUES

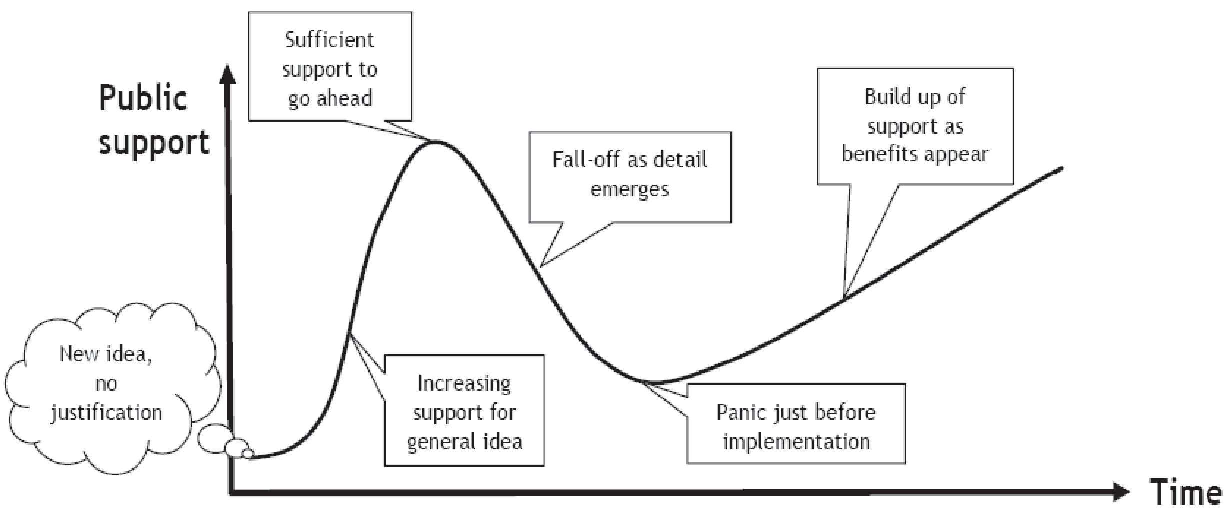
6.1. Managing public perception

Acceptance for a measure is not constant but changes with the development of the public discourse. When looking at the available examples, one recognises a pronounced temporal development.

Obviously, people first focus on their individual costs, rather than expecting positive personal effects; costs are conceivable, but not benefits. Thus, fierce initial rejection is often experienced and discussion is generally polemic and undifferentiated.

It is the task of a well-managed public dialogue to address, rather than repress, preconceptions and outright fears. A precondition is that the public indeed perceives a problem, as discussed below. Over time, emotions will calm. This is the point where opinions are being made.

Figure 5. **Public acceptance changes over time** [from Ref. 7]



From observation, a critical success factor appears to be that this process is actively managed. A positive factor is strong political leadership. Leadership and decidedness appear to support identification with the measure and also lead to a deeper involvement in the discussion.

This is the "window of opportunity", when there is sufficient support to trigger deployment. From then on, support will fade away again as issues emerge in detailed planning and law-making. Support is lowest just before introduction, since fears will peak and benefits are not yet visible. However, after introduction, "seeing is believing" and support will build up.

> Public opinions need to be managed through leadership, decisiveness and skilful timing.

6.2. Pressing traffic problems

Studies have shown – and this comes as no surprise – that in all successful implementations it has been essential for road pricing to be perceived as a true solution to a problem [see, e.g., Ref. 5]. This finding entails three elements:

– There has to **be a major traffic problem**. People are only prepared to pay for mobility if there is a massive congestion issue without simple and common solutions. Interestingly also, financial issues are perceived as relevant problems – such as a lack of resources to fund the building of important infrastructure, e.g. a relief road or an important tunnel or bridge, to combat congestion in the inner city.

  Unfortunately, these findings mean that pricing cannot be used to combat congestion *before* the problem occurs. Pricing as a pre-emptive measure will find neither political nor public acceptance. Consensus in society appears to be a precondition for no longer tolerating the traffic situation – irrespective of whether individuals are concerned by the problem in their daily lives.

– It is also important that people **trust that charging will make a genuine contribution** to alleviating the problem. Acceptance can only be expected if the pricing measure "works" according to the judgment of the people concerned. Since pricing measures are new instruments in traffic management, lack of personal experience is a hindrance in this process. Humans learn from experience and the effects of a pricing measure cannot be felt beforehand: effects are predicted but cannot be experienced before the measure is introduced. Tests and pilot projects can make an important contribution by proofing effects and even making them perceptible for the individual. This concept was exceptionally successful in introducing the Stockholm charging system.

– Finally, the traffic problem needs to be **perceived** as severe, with pricing providing a suitable measure. The open dialogue in London and Stockholm has moved congestion into the public's awareness. At the beginning of the discussions in both cities, congestion has been accepted as unavoidable and "god-given". Only public discussion has created awareness about the extent of congestion and its disruptive influence on a city's quality of life. In the UK, phrases like "we cannot build our way out of congestion" have become common wisdom. This perception is an essential basis for widespread acceptance of pricing as a traffic management tool.

Worldwide, all studies – see [Ref. 6] for a typical example – underline the importance of the perceived necessity and effectiveness of accepting a road pricing measure. In the cited study, the authors compare systems across the globe and conclude that publicly perceived necessity is the single most important success factor.

Prerequisites for success are: a massive congestion problem that is perceived as such; and that pricing is believed to be effective.

6.3. Use of income

As discussed in the section about tariff levels, revenue neutrality is probably not the best approach to win the public over. If a new pricing measure is introduced, it is wise to abolish or reduce some other taxes and fees, but creating extra net revenue is nothing to shy away from. Under no circumstances should extra revenue be part of a hidden agenda – e.g. "We will be grossly revenue neutral. Later tariff increases cannot be excluded as this point in time."

Use of revenue is critical. Initially it was believed that broad acceptance of a measure requires that revenue is earmarked to return into the road sector and that public transport in particular is an acceptable destination for spending the money. More recent experience, as collected e.g. in Ref. 5, shows that earmarking is the essential element. As a destination for the money other generally "good" placements are acceptable to the public. It is important that those who pay also benefit. Revenue from city pricing needs to be spent in the city, for road infrastructure, public transport, or even to subsidize ecological projects. "My money should be used in my neighbourhood."

Revenue is an essential element of pricing. Instead of promising "revenue neutrality", which is difficult to achieve and to maintain in practice, it might be well worth being courageous and using the extra revenue for something that is widely welcomed.

6.4. A transparent, fair and understandable system

Acceptance requires trust. Trust can only be won through transparent, open communication. Transparency implies that the system rules are clear and simple. Everybody must understand where he has to pay, when and how much. There should be simple rules to follow, without undue complexity and without many exceptions.

The successes of the charges in London and Stockholm are to a large extent due to the simple and transparent rules. The charge perimeter is easy to recognise and well communicated. It is simple to understand at what time of day the different tariffs apply. It is clear where one can pay and what to do if one did not manage to pay before use.

It is essential that people feel treated fairly. Nobodies wants to be fined just because he did not know some detail or because of a minor neglect. This implies clear and simple rules but also a simple access to the system. I personally do not mind to pay a few Euros for parking or for road tolls, but I do not want to be hassled. I do not want to search for the right coins, or to have to stop and leave the vehicle, or to fill in long forms, and the like. The "front desk" of the charging system must be user friendly. National charging systems for heavy vehicles, such as the ones deployed in Germany, Austria or Switzerland, and as currently being procured for France, have no need to be overly simple to use. Truck drivers are professionals used to pay fees, tolls and taxes of all sorts. One can reasonably expect that commercial drivers are informed and can manage more complex tasks. This is not true for a city charging system that has to accommodate everybody, not just the interested and informed, young and active citizen.

Fair treatment also implies a credible compliance regime. It is essential that people trust that everybody pays and not only the “honest idiots”. Compliance checking has to be recognised as proportionate. It must have a certain visibility but should not be perceived as constant surveillance, and must not be intrusive.

A transparent, fair and understandable system is achieved through clear rules, simple access to the system and a credible compliance regime.

6.5. Privacy

Privacy aspects influence the technical solution for a road-user charging system and its public and legal acceptance. In a charging system, privacy must be guaranteed regarding:

- Personal user data;
- Payment and contract data;
- Movement data.

The privacy requirements regarding personal user data, payment and contract data of a road charging system are basically the same as in similar systems such as telecommunications, banking or retail, and must be handled in the same way.

A special property of a road-user charging system is that a vehicle is identified as being at a certain location at a certain time. Depending on the charging concept and the technical solution, this may imply that data on the daily movement profile of a user is available in the system. System design must ensure that such data are treated according to privacy laws and in a transparent and understandable way for the public. Privacy is more of an issue for private vehicles than for commercial fleets, since private vehicles can be linked to a certain person (owner/driver) whereas commercial vehicles cannot (changing drivers).

There is a tendency to treat privacy issues by technical means. Various measures based on cryptographic techniques are being proposed. There is also a discussion on keeping data within certain environments, such as processing all data in the on-board equipment, and only transferring daily aggregates to the back-office. Such privacy-protecting concepts usually imply some costs and a certain inconvenience in system design. The solutions are fine in principle and are normally sufficient to fulfil the legal requirements.

The perception is not influenced by design, however. It is probably better to look into the institutional arrangements. For example, I have no problem with the fact that my mobile phone provider knows every movement I make, or that my bank knows a lot about my financial capabilities. People like to have a choice regarding who can access their data, and they are more relaxed about it if they receive a clear benefit. I would like to be able to pay my road-user charge through some non-suspect body, such as my automobile club, the supermarket or my mobile phone company. Authorities usually outsource operation of the charging system anyway. Why not use this also to provide for effective user privacy, especially as perceived by the public?

Movement data is sensitive and has privacy implications. Again, perceived privacy is decisive. The feeling of having a "spy in the car" has to be avoided.

6.6. Accompanying measures

A price is only acceptable if there is something of the correct value offered in return. As discussed before, the benefits of congestion relief cannot be experienced beforehand, and are just vague promises to most citizens.

It is therefore wise to bundle the pricing measure with more returns for the individual. To the author's personal amazement, the focus in this discussion has been on providing "value added services". The message is this: "Dear user, you will need a box in your car that charges you, but it can also provide some useful services for you". A long list of such value-added services has been proposed in the past, especially with the introduction of heavy-vehicle charging systems, which required viable on-board equipment. Proposed services included navigation, traffic information, fleet management, emergency calls, automated payment for parking, payment at the pump, and even McDonalds' drive-through payment.

The proposals have been very technology-driven and have not achieved market success. When it comes to useful applications in a city environment, modern mobile devices such as the iPhone are hard to beat. I do not need a charging box to tell me which way to go to find the next restaurant or to pay for my burger.

What the author would love, though, is a proper integration of the whole charging measure into daily mobility in the city. If the congestion charge, e.g., was the same as a day pass for public transport, or had an option to include city parking, or at least had the same cordon layout as the public transport tariff zones, we would feel that someone has invested some thought and care into what touches our daily lives. If we feel that we are not simply being charged but that this is part of a package that improves our city mobility, we are less opposed to paying a price.

As discussed before, it is advisable to plan for some extra revenue and not go for full "revenue neutrality". Extra money is well spent on accompanying measures such as improving public transport, providing park-and-ride facilities or introducing a city mobility card, which can be used to access all transport offers, from public transport, through parking to road pricing.

Strictly speaking not an accompanying measure, but quite akin to it, is the provision for interoperability. Currently, city pricing schemes are confined to larger metropolitan areas and are lone "islands in the sea". As soon as more regions employ charging for demand management purposes and as soon as interurban motorway tolls increasingly become electronic, users must be offered a single device with a single contract to pay for all road-use related charges and fees. The focus today is on commuters within city limits with usually little need for interoperable equipment. With the increasing success and inevitability of pricing measures, interoperability will become a must, and not a nice accompanying goodie.

Pricing should be part of a package. Individual users will more likely accept a new charge if there are clear benefits that accompany the pricing measure.

7. CONCLUSIONS

In order to make charging a success, three core aspects need to be planned and implemented skilfully:

– Technical design;
– Scheme layout;
– Acceptance.

The order of priorities is in fact bottom-up: without sufficient acceptance and support, a new charging measure will not stand a chance to pass early political hurdles, and implementation will not even start. A clever scheme layout comes next, since the charging rules determine whether or not the desired traffic effects will be seen. Proper technical design is certainly also an important success factor, but appropriate technology is available and should not become a driver of the project.

Unfortunately, the level of difficulty is bottom-up: acceptance is hard to steer. The report gives some hints on where the focus of activities should be according to experience from projects worldwide, but there is no guarantee. The subject is emotional, non-linear, time-dependent and multi-faceted. The good message is that knowledge and understanding about proper scheme layout has increased over the last years, and especially that technical issues are no longer an obstacle. Good solutions and products exist and only require proper engineering and project management.

Yet, the main conclusion of this review of critical success factors is that pricing as a traffic management measure has become a tool that is now available to tackle the problems of our growing cities. Congestion pricing is not yet a tool as readily usable as any other traffic management tool. Success is not guaranteed, but the effects are unsurpassed. Apart from a driving ban, no other traffic management tool has a better effect on congestion than pricing.

Designing a nation-wide high-tech charging system for trucks needs good engineering and project management – it requires good craftsmanship.

Designing a congestion pricing system for a city needs careful handling of a number of acceptance-critical implementation issues – it requires vision, inspiration, leadership and luck.

City congestion pricing is art, not craft.

REFERENCES

1. European Commission (2006), "Urban Pricing – From Theory to Reality", DG Energy and Transport.

2. Adapted from: Wedlock, Mark, Jeremy Evans, Nick Patchett, Paul Mellon, Jesper Engdahl and Ian Wilkinson (2006), "The Urban Environment: The New Challenge for DSRC", ITS World Congress, London.

3. Convention of the United Nations on Road Traffic, 8 November 1968, Vienna.

4. Kristian Appel and Bernhard Oehry (2008), Helsinki Road-User Charging Workshop, Helsinki, November (available from the authors).

5. Rapp, M., B. Oehry, Ch. Egeler (2007), *Mobility Pricing Synthesebericht* (Mobility Pricing Synthesis Report), Rapp Trans AG; published by Schweizerischer Verband der Strassen- und Verkehrsfachleute (VSS), October.

6. Report to Ministry of Transport, New Zealand (2005), *Congestion Charging: Critical Success Factors*, Booz Allen Hamilton.

7. P. Goodwin, keynote speech at the IET Seminar on Road Pricing, 22-23 May 2007, Birmingham.

The Singapore experience: The evolution of technologies, costs and benefits, and lessons learnt

Kian-Keong CHIN

Land Transport Authority
Singapore

SUMMARY

Singapore, January 2010

1. INTRODUCTION

Singapore is an island-state with a land area of about 710 square km, measuring 42 km across and 23 km from north to south. Densely populated with more than 4.8 million people, its transport needs are served by an infrastructure of 147 km of MRT/LRT[1] lines and 3 300 km of roads catering to more than 900 000 vehicles.

Given its land constraints, Singapore's overall transportation strategy cannot rely on building more and more roads to serve its population's travel needs. It needs a comprehensive and affordable public transport system and sustainable demand management tools. Hence, its recently launched Land Transport Master Plan is based on making public transport a choice mode, while continuing to manage road usage and meet the diverse needs of its travellers. A key element in attaining these objectives is the continued use of road pricing.

Road pricing has long been associated with Singapore, starting way back in June 1975. Many changes have been made to the road pricing scheme since that time. Started as a manual scheme, based on paper permits and hence using little technology, it has evolved to become a sophisticated system today, involving various technologies. The economic principles for road pricing, however, continue to be valid but the charging structure has evolved to keep the scheme effective and derive benefits to the community as a whole.

2. MANUAL ROAD-PRICING SCHEME

The manual scheme, called the Area Licensing Scheme (ALS), was based on the display of paper licences that were purchased prior to their entering the part of city defined as the Restricted Zone (RZ). This scheme was in operation for 23 years before it was replaced with the Electronic Road Pricing System (ERP) in 1998.

To enter the RZ during the restricted periods, non-exempt vehicles had to purchase an ALS area licence from roadside sales booths located on approach roads to the RZ, petrol stations, post offices or convenience stores. These are available as daily and monthly ALS area licences.

Enforcement personnel were stationed at the control points to ensure that non-exempt vehicles displayed valid ALS area licences on their windscreens, or on their handle-bars in the case of motorcycles and scooters. Violating vehicles had their vehicle licence numbers noted down and their owners received summonses for entering the RZ without a valid licence. Vehicles were free to move around or leave the RZ without having the ALS area licences.

2.1. Operating hours

The ALS started in June 1975 with the restricted hours of 7.30-9.30 a.m. daily, except on Sundays and public holidays. Three weeks later, the restricted hours were extended to 10.15 a.m., in order to restrain the surge in vehicle entries immediately after the lifting of the ALS at 9.30. The ALS operated for 2¾ hours each weekday during the morning peak period until June 1989, when fundamental changes were made. The restriction period was extended to cover the evening peak hours of 4.30-7.00 p.m. on weekdays.

In January 1994, more fundamental changes were made to the ALS scheme. The restricted hours were further extended to cover the 10.15 a.m.-4.30 p.m. time period on weekdays and the post-peak morning period of 10.15 a.m.-3.00 p.m. on Saturdays. The Saturday restriction period was subsequently cut back to 2.00 p.m. due to improved traffic conditions within the RZ.

2.2. Affected vehicles and charges

When the ALS started in 1975, taxis, public transport buses, goods vehicles, motorcycles, and passenger cars carrying three or more passengers (car-pooling) were exempted from the scheme. A few months later, in August of the same year, the exemption on taxis was removed.

In June 1989, motorcycles and goods vehicles were also required to purchase ALS licences prior to their entering the RZ during the restriction period. Exemptions for car-pools were also removed because private cars were picking up bus commuters instead of forming genuine car-pools.

Starting at $3[2] per day for an ALS car licence in 1975, this crept up to $5 per day in 1980. However, with the major review in 1989, there was a reduction in rates because more vehicles were then required to purchase licences. The daily licence fee for a car was reduced back to $3.

The January 1994 changes also saw two fee levels for two different types of licence – one allowing usage throughout the day, and the other for use during the inter-peak period only. The car fees were $3 and $2, respectively.

2.3. Traffic impact

Traffic entering the RZ dropped by 44% initially but returned to a 31% drop by 1988. However, this was despite growth by one-third in city employment, and by 77% in the vehicle population during the same period. The drop in traffic arose from the decanting of motorists whose destinations were not in the city but who had been using the city roads as a bypass, as well as of those who entered the city earlier to avoid having to buy ALS area licences.

2.4. Going from manual to electronic

Being manually operated, the ALS required substantial manpower. About 60 enforcement personnel and another 60 officers were required each day at dedicated licence sales booths. The enforcement duties were demanding, given the long hours spent under the sun and rain, not to mention the dust and the noise. In addition, there were 16 different types of licence in use at its peak, and much concentration by the enforcement officers was required to ensure that they identified them correctly.

It was inevitable that there were errors made sometimes, resulting in the occasional wrongful issue of summonses.

There was always a rush to enter the Restricted Zone just before or after the restricted hours because of the significant change of licence fee from nothing to $3 or *vice versa*. This resulted in sharp and short peaks in the volume of traffic entering the city. Having intermediate or shoulder rates would have smoothed out the peaks, but it was difficult to implement, given the need for various types of paper licence which had to be distinguished by enforcement officers.

3. ELECTRONIC ROAD PRICING (ERP)

With the shortcomings of the manual road pricing scheme, the search for a more efficient technology began in earnest in the early 1990s. Technology for an electronic road tolling system was emerging at that time, and after several years of prototype testing with potential suppliers[3], a contract was awarded in 1995 for the installation of a Dedicated Short-Range Communication (DSRC) electronic road pricing system.

The ERP system has three major groups of components. The first centred around the in-vehicle Unit (IU) and the stored-value smart-card. While the IUs were produced specifically for the ERP system, the smart-cards were marketed by a consortium of local banks for multiple uses.

The second group comprises equipment installed in the field – at the ERP gantries. These include the antennae, the vehicle detectors and the enforcement camera system. Data collected is transmitted back to the Control Centre continuously through leased telecommunication lines.

The third group of components is at the Control Centre, and includes various back-up computers and monitoring systems, as well as a master-clock to ensure that the timing at all the ERP gantries is synchronised. All the financial transactions and violation images are processed here.

The ERP system is designed to be simple to use. With the smart-card inserted into the IU, the appropriate ERP charge would automatically be deducted whenever the vehicle passes through the ERP gantry. There would be a short beep to signify a successful transaction. Should there be insufficient cash in the smart-card or should there be no smart-card in the IU, the enforcement cameras in the gantry will take a picture of the rear of the vehicle. A similar enforcement picture would also be taken of any vehicle that had no IU installed. The vehicles' registration numbers would be automatically read using OCR techniques and the vehicles' owners issued with letters asking for payment of outstanding charges, inclusive of administrative fees. Failure to pay the charges and fees could result in the offender being called to appear in court.

3.1. Pre-ERP launch programmes

There were two major programmes launched prior to the start of the ERP. The first was the installation of in-vehicle units on the then 680 000 eligible vehicles, while the second was on publicity, to make motorists and motorcyclists aware and ready for the ERP system.

The publicity programme was an important element in the success of the Electronic Road Pricing launch. It began even before the start of the IU fitting programme and was in place for more than a year, all the way up to and beyond the launch date of the ERP system. All vehicle owners were sent brochures, detailing the ERP system, how it works and the differences with the then working ALS/RPS. Advertisements were also placed in the print media as well as on television to drum up awareness of the new road pricing system.

With the ERP system, all vehicles are required to have IUs fitted if they intend to pass through the ERP gantries. ERP charges are applicable for all types of vehicle during the operating hours, with the exception of emergency vehicles (ambulances, fire engines and police cars). When the system was first launched the charges varied from $0.50 to $3.00 per passage through the ERP gantries.

3.2. Traffic impact

Traffic volume into the Central Business District (CBD) had reduced by about 10-15% during the ERP operating hours, when compared to the ALS scheme. While the charge payable by most motorists was lower, ranging from $0.50 to $2.50, as compared to the ALS, there was one major change in the pricing structure that brought about a change in travel behaviour. With the ERP system, there was no longer an unlimited number of entries into the CBD for a single payment. It was estimated that about 23% of trips that entered the CBD during the ALS days were repeat trips, i.e. whose marginal road pricing charge was zero. Hence, with ERP, many of these multiple trip-makers cut down their number of trips, e.g. office workers no longer used their cars to attend mid-day meetings or lunches – instead, more relied on the public transport system.

3.3. Varying the road pricing charge

The ERP system, being less dependent on manpower, allowed more frequent changes to be made to the road pricing charges. This helped to better optimise usage of road space in the network. The rates are set to ensure that flow levels are kept as high as is practicable (thereby allowing the maximum number of road users to benefit), and this is measured using average speeds as the proxy. On urban roads, the average speeds should be between 20 km/h to 30 km/h while for expressways, the speeds should be between 45 km/hr to 65 km/hr[4]. When speeds exceed the upper threshold, too few vehicles are deemed to be using the roads and, hence, the road space available is not being optimally used. Hence, the road pricing charge can be reduced to allow more vehicles to use the roads. Conversely, if the speed falls below the lower threshold, too many vehicles are on the roads and this is a signal that the road pricing charge can be increased.

This ERP rate review is conducted every three months, and over the years the rates have stabilized, with only a handful of gantries having their rates adjusted each time the ERP rate review takes place.

3.4. The economics of Singapore's ERP system

The original ERP contract was worth about S$200 million, with half the amount representing a million in-vehicle units. There were originally 33 ERP gantries, but this has now expanded to 66. The cost of each IU including installation was S$150, but for the 680 000 existing vehicle owners at that time, this cost was borne by the Government.

Given that the Electronic Road Pricing scheme charges for each passage through the ERP gantries, it was expected at that time that motorists would be paying more than what they had collectively been paying under the Area Licensing Scheme. Hence, the annual road-tax structure was reviewed to have each motorist paying lower road taxes. In addition, there was a one-off road tax rebate for each vehicle owner, so that overall the introduction of the ERP would be revenue-neutral to the Government. This was also to make the ERP scheme more palatable to the motoring public.

Although the ERP revenue collected goes to the Government Consolidated Fund and not hypothecated for transport, much effort was made to stress that the ERP system is a traffic management tool and not for revenue collection. As it turned out, the revenue collected when the ERP scheme started operations turned out to be significantly lower than what was collected with the ALS[5], about 30% lower. However, given that the ERP system is a traffic management tool, this lower-than-anticipated ERP revenue collection was not an issue. With the increased number of gantries, revenue has naturally increased and presently is marginally more than what was collected under the ALS. Given that revenue was not the reason for ERP, there were no explicit cost-benefit assessments, although there was much effort to contain the costs of implementing and operating the system. Instead, to gauge the effectiveness of the ERP, it relied on the outcome in terms of observed travel speeds on the roads.

Naturally, the cost of each ERP gantry has increased over the years and presently, each typical ERP gantry costs about 1½ times as much as in 1998. While the cost of each IU has also increased, installation costs have lessened due to the various installers having more experience and having sunk their overheads. Hence, the cost of installing an IU today has remained the same as before, i.e. S$150 per IU, including installation. The cost of managing and maintaining the ERP system has increased over the years, consistent with the increase in the number of gantries and IU numbers, but this has remained at about 20-30% of total revenue collected.

3.5. Extending the coverage of ERP

With the launch of the ERP in 1998, transport planners started to investigate its use to manage congestion outside the city. An Outer Cordon was planned, charging motorists who passed through it, but it was decided that each of the roads entering the Outer Cordon would only have ERP gantries implemented when travel speeds on each of these roads fell below the threshold. Hence, the Outer Cordon gantries were introduced gradually, starting from 1999. As at end-2008, 15 ERP gantries, of a total of 21 needed to complete the Outer Cordon, are in use to manage traffic during peak periods.

3.6. ERP to manage home-based trips

The ERP gantries that replaced the ALS, and the subsequent ones on the Outer Cordon had been effective in managing traffic flows into the city. However, these ERP gantries do not affect traffic flows out of the city and, consequently, some of the major corridors taking traffic away from the city towards major residential areas became congested. Hence, the pricing strategy had to be extended to deal with home-bound trips during the evening peak period. In August 2005, an ERP gantry was introduced on a major expressway leading from the city to the north, on its most congested stretch. Traffic was redistributed to other roads and other time periods, with traffic volume dropping by around 25% initially on this major expressway during the evening peak period, from 6-8 p.m. on weekdays.

However, over time the traffic came back onto the expressway, and a significant portion of the traffic actually left the expressway just before the gantry. This behaviour meant that on this expressway, upstream of the ERP gantry, congestion became prevalent. Hence, two years later in November 2007, another ERP gantry had to be introduced upstream of the existing one. Based on the traffic flow profile on this expressway, this new ERP gantry has to operate from 5.30-10.30 p.m. on weekdays. Traffic volume on this stretch of expressway reduced by about 20% during this five-hour period, and congestion cleared up.

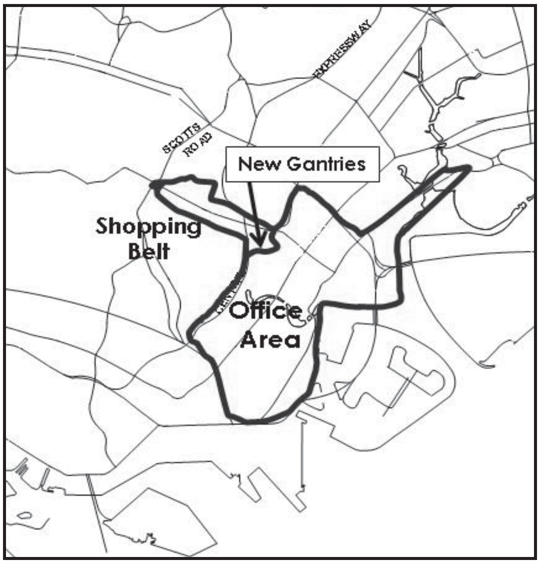
3.7. Electronic Road Pricing to manage intra-city traffic in the shopping belt

The CBD pricing cordon covers a part of the city that is predominately shopping-based in nature, and has traffic characteristics somewhat different from the office-based city roads. This is the Orchard Road shopping belt and since most of the shops opened after 10.30 a.m., there was low traffic flow there during the morning peak period. However, from midday onwards, heavy traffic with recurring congestion was normal, on weekdays and on Saturdays.

About 35% of the traffic on the main shopping thoroughfare was found to be through-traffic - i.e. crossing the shopping area to the office-dominant area in the Central Business District.

Hence, the ERP scheme for this area was refined in 2005, with the shopping belt made into a sub-zone, with additional gantries at its boundary with the office-dominant CBD area (see Figure 1). These new gantries allow three pricing strategy changes to respond to the traffic flows in the shopping belt.

Figure 1. **Separating the shopping belt from the office-dominant area**



First, the ERP charges for traffic entering this shopping zone during the morning peak period were removed, since there was no congestion there at that time of the day. Second, from mid-day onwards, the ERP charge for traffic crossing the shopping sub-zone into the office-dominant area had to go through two ERP gantries, and the total charge was set to be higher than that for entering the office-dominant area directly via other roads. This cut down the amount of through-traffic on these shopping streets, from about 35% to 20%. To ensure that this change would not discourage shopping trips, the ERP charge for those destined for the shops was actually reduced. As a result, the volume of destination traffic remained unchanged. Third, the shopping sub-area was now priced on Saturdays without the office-dominant area (which had no congestion on weekends) being charged as well.

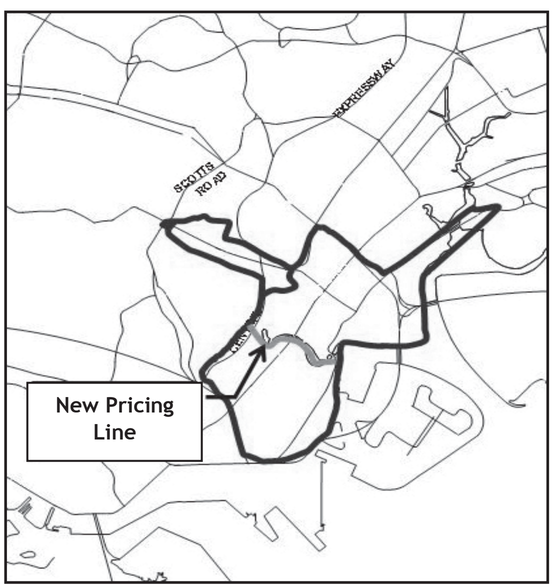
The outcome of this pricing strategy change was in line with expectations. Hourly traffic on the shopping streets was reduced by 14-36% during the priced periods on weekdays and reduced by 19-34% on Saturdays, and this came predominantly from a reduction in through traffic. Indeed, a survey of vehicles entering the car-parks in the shopping area showed a slight increase on both weekdays and Saturdays. In addition, the car-parks there on Sundays also showed increased usage. One can argue that the pricing strategy change has influenced some of the weekday/Saturday shoppers to do their shopping on Sundays, and the relief of traffic congestion on weekdays/Saturdays has attracted new shopping traffic into that area.

3.8. ERP to manage intra-city traffic during the evening peak

With increasing traffic in the city, in 2008 the pricing strategy was revised to deal with intra-city traffic because the ERP scheme then did not impose any charges on vehicles travelling solely within the city roads.

The office-dominant CBD was divided into two parts, with a new pricing line along the Singapore River involving just five ERP gantries. Traffic crossing this line in either direction was subjected to ERP charges (see Figure 2) but this was only during the evening peak period from 6-8 p.m., since it was only during this time that congestion was severe enough. Following this change in pricing strategy, the hourly traffic across the city passing this line dropped by between 28-37% during the evening peak period. This was a significant drop in traffic and there were concerns from the shopping community which had also seen a drop in their businesses.

Figure 2. **Managing intra-city traffic congestion with a new pricing line**



However, it was not conclusive that this drop in business was a direct result of the pricing strategy change, because there was also a decline in the general economic activity brought about by the worldwide financial climate. Nevertheless, given that the travel speeds on the roads were increased significantly, there was an adjustment in the ERP rates three months later, in line with the established pricing adjustment criteria. The net reduction thereafter in traffic flow became smaller as expected following this pricing adjustment, with a reduction in hourly traffic now of up to 30% (instead of 37%), in relation to traffic conditions before the introduction of the new pricing line. Traffic conditions on the city roads remained relatively good.

3.9. Other Issues

Road pricing schemes and the issue of privacy are never separable. Hence, there was much done to allay the fears of motorists. Being an active system, there was no need for the central computer system to keep track of vehicle movements, since all charges were deducted from the smart-card inserted at the point of use. Records of such transactions were kept in the memory chip of each individual’s smart-card. The authorities also took a further step to assure the public that all records of transactions required to secure payments from the banks were afterwards erased from the central computer system, typically within 24 hours.

The issue of the ERP system as a revenue tool for the government was also raised. However, ERP has always been positioned as a traffic management tool and revenue was not and never has been a consideration. Indeed, when the ERP system replaced the ALS in 1998, the revenue collected was only about 60% of what it used to be. Nevertheless, there had to be a continued effort to publicise the fact that ERP is not a revenue-generating tool and, to drive home this point, whenever there were major changes in the ERP scheme, there were reductions in vehicle up-front taxes and recurring annual licence fees.

4. LESSONS LEARNT

One of the major lessons learnt from Singapore’s experience must be the importance of remaining adaptable and ready to make changes to the congestion pricing scheme. In that way, those specific groups which contribute to traffic congestion on the roads can be targeted through, e.g., the Orchard Cordon to deal with shopping traffic and the New Pricing Line that passes through the city to manage intra-city traffic.

The rationale for congestion pricing should be robust and supported by motorists’ real-life experiences as they travel on the roads. In Singapore, travel speeds experienced on the roads under the influence of the ERP pricing gantries are used to decide on the introduction of ERP and the adjustment of rates. The provisions to make adjustments to the rates six times a year, in order to deal with changing and seasonal traffic patterns when they occur, is useful for convincing motorists that the road pricing scheme is a traffic management tool and not a revenue-generating one. Ultimately, the scheme should be technically and logically robust. It should also be kept simple, so as to allow all users to easily understand why and how road pricing schemes work.

While a congestion pricing scheme may be justified from a technical perspective and can be rationally argued for, utmost importance must be given to communicating the rationale of the scheme to road-users and the communities, including businesses. There is never too much communication and publicity when it comes to road pricing schemes. In Singapore’s case, even with intense publicity and communications, there are still instances where motorists claim ignorance or continue to challenge the validity of the scheme.

There should always be viable alternatives for motorists who decide not to pay the congestion pricing charges – it might be an alternative route or time of travel. For those who decide not to drive, there has to be a viable public transport alternative. In the latest revisions to the ERP pricing strategies in 2008, significant effort was expended to increase public transport capacity, with premier buses (private buses offering seated services almost from door-to-door during peak periods), reduced headways on public buses and on the underground trains and expanded bus lanes and bus priority schemes.

In addition, congestion pricing schemes should not be seen as a means to increase government revenue. There should be corresponding reductions in taxes or expenditures by the vehicle users, which should be equal or greater than the expected revenue from the road pricing scheme. In the recent road pricing changes made in 2008, vehicle taxes were reduced by about $110 million per year, and this was much higher than the expected $70 million increase in ERP pricing revenue. The challenge is to get the motoring public to retain this idea of a reduction in vehicle taxes being a beneficial trade-off for them, as, unlike the ERP charges which recur almost on a daily basis, this reduction in vehicle taxes is either a one-off occurrence (at the time of purchase of the vehicle) or annual or semi-annual (the road tax is payable once a year or once every six months).

Congestion pricing schemes are not the ultimate solution to traffic congestion in urban areas. Ultimately, it has to be a combination of schemes. Travel demand has to be managed in other ways too, e.g. through proper land-use planning and decentralisation policies, parking policies, car-ownership policies (in Singapore, this is the Vehicle Quota Scheme) and an increasingly more effective (public) transport alternative. Road network capacity must still be improved but perhaps more selectively, and technology used to continuously optimise the available capacity. This includes appropriate resources to deal with obstructions caused by accidents and traffic incidents on the road network, as delays in clearing these obstructions will only mean more congested roads in the network, with or without congestion pricing.

5. CONCLUSION

Since congestion pricing was introduced to manage traffic on Singapore's roads in 1975, there has been a recognised need for the pricing strategies used to keep traffic flowing to continually evolve in order to meet changing traffic conditions. The pricing strategies also have to be sensitive to the specific needs of the community, be they shoppers or office workers. Over the past 35 years, Singapore has been refining and adjusting its pricing strategies to meet the changing circumstances, but the principle of congestion pricing as a traffic management tool to tackle traffic congestion has remained unchanged.

NOTES

1. Mass Rapid Transit/Light Rapid Transit.

2. The fees mentioned in this paper are in Singapore dollars and, as at December 2009, the exchange rate is USD 1 = S$ 1.40.

3. GPS-based technology for road pricing was not available at that time, and hence the systems considered then were based on the use of electronic transponders, each with its own unique identifier code. Even then, the transponders available required vehicles to be travelling in a single queue, while the requirements we had were for a free-flowing multi-lane solution. Three contractors were short-listed and their solutions tried on newly-completed and unopened stretches of road. The Government provided funding of S$1million to each of the three contractors, to mitigate their risks and cover part of their expenses. The transponder, or IU, was designed to deduct congestion charges directly from smartcards, primarily to deal with the issue of privacy, since transaction records need not be kept by any central computer system, only stored in the inserted smartcard.

4. The computation of these optimum speed ranges are based on speed-flow curves, derived from empirical data collected on expressways and arterial roads. The lower speed threshold is a value close to the optimum point of the speed-flow curve that gives the maximum traffic flow. On expressways, this is 45 km/hr and on arterial roads, 20 km/hr. The lower value for arterial roads is due to the presence of traffic signals and side-friction caused by various road-side activities, such as on-street parking and passenger pick-up or drop-off. The upper speed thresholds were chosen to allow stability in the ERP rates, as too narrow a range is likely to give oscillating ERP rates each time they are reviewed.

5. About S$100 million a year were collected with the ALS, just before it was fully replaced by the ERP system in September 1998.

BIBLIOGRAPHY

Menon, A.P.G. and Kian-Keong Chin (1998), The Making of Singapore's Electronic Road Pricing System, in: *Proceedings of the International Conference on Transportation into the next Millennium*, Singapore.

Menon, A.P.G. and Kian-Keong Chin (2004), ERP in Singapore – what's been learnt from five years of operation? *Traffic Engineering and Control* (tec), Vol. 45, No. 2, February.

Li, Michael Z.F. (2002), The role of speed-flow relationship in congestion pricing implementation with an application to Singapore, *Transportation Research Part B*, 36.

Chin, Kian-Keong (2005), Road Pricing – Singapore's 30 Years of Experience, CESifo DICE Report, *Journal for Institutional Comparisons*, Vol. 3, No. 3, Autumn.

Harun, Halim, Kian-Keong Chin, Wai Yan Leong (2009), Managing Traffic on City Roads: A Demand Management Approach, 2nd World Roads Conference, Singapore.

So you're considering introducing congestion charging? Here's what you need to know

An FAQ based on Stockholm's experiences

Jonas ELIASSON

Royal Institute of Technology
Stockholm
Sweden

SUMMARY

Stockholm, January 2010

1. INTRODUCTION

Congestion pricing has been advocated for a long time as an efficient means to reduce road congestion, without much success in practice. In the last few years, however, congestion pricing has been introduced in various forms, with London and Stockholm attracting the most international attention, starting in 2003 and 2006, respectively.

The Stockholm experiences have attracted widespread attention from many cities that consider introducing similar schemes, and the people involved in the development and evaluation of the Stockholm system (including the author, who was responsible for the design of the charging system and forecasting its effects, and who subsequently chaired the Expert Evaluation Panel) have acted as advisors to many policymakers from cities and countries all around the world. This paper is based on these experiences. A number of questions repeatedly surface in discussions with policymakers, politicians and planners. A number of facts and insights repeatedly turn out to be especially relevant, interesting and, to a certain extent, surprising. The purpose of this paper is to summarize these questions, facts and insights, for the benefit of cities considering congestion charging schemes.

The paper draws from already published material (see Section 6 - References). In fact, a reader already familiar with the congestion charging literature will find few completely new findings or insights. The contribution of the paper is rather the selection of the most relevant, interesting, important and sometimes surprising facts, insights, findings, advice and conclusions for policymakers, out of a vast literature on congestion charging in general and Stockholm in particular.

It should be stressed that the paper is not meant to be a comprehensive survey of theory or experience. In fact, the aim has been to keep the paper as short as possible, in order to focus on the most relevant facts: the key issues and advice that should be emphasized to policymakers as early as possible.

It is necessary to consider three areas for a successful congestion charging implementation. First, the charging structure needs to be well designed, in order to achieve large social benefits. This is the topic of Section 3. Second, investment and operational costs must be kept low, as discussed in Section 4. Third, one must obtain public acceptance for the system. This is covered in Section 5. The paper starts (Section 2) with some fundamental facts about congestion charges – why they are needed, how they work and their general effects.

2. BASIC FACTS: WHY AND HOW IT WORKS

This section covers issues related to why congestion charges are needed, what they can and cannot do, and what effects charges have on traffic.

Congestion cannot be eliminated by investments alone

It is a well-established fact, both theoretically and empirically, that investments in road and/or public transit is not sufficient to eliminate road congestion in the centres of large cities. There are several reasons for this: two of the most important from a practical point of view are the inevitable eventual scarcity of urban land and public resources.

Congestion charging will not solve everything

Introducing congestion charging will reduce the need for transport investments but, generally speaking, not eliminate it. Normally, a growing urban region will need both congestion charging and transport investments (both roads and public transit). Obviously, cities are different as to what investments are the most cost-efficient and the most needed.

Only introduce congestion charges when they are needed

This may seem obvious – but in fact, one is sometimes confronted with cities with virtually no road congestion that are nevertheless considering "congestion charges" (there are several current Swedish examples of this). The purpose may instead be to raise revenues or reduce traffic emissions. Generally speaking, there are usually more cost-efficient ways to do such things than introducing congestion charges[1] – not least because a charging system is typically a fairly expensive investment.

Car drivers are cost-sensitive

At least a sufficient majority of them. This means that increasing the cost to drive a car at certain places at certain times will decrease the number of drivers choosing to drive there and then. How large a decrease depends on the ease of adaptation – in other words, how many good alternatives there are. Alternatives may be other time periods, modes, routes, destinations, etc. It is imperative to keep as many options open as possible to achieve good traffic reduction effects – but it is up to the drivers themselves to choose how to adapt.

There are many ways to adapt

Not all car drivers who change due to the charges will switch to transit, nor to other routes or time periods. Route and mode changes are far from the only adaptation strategies. Trips, especially discretionary trips (shopping, leisure, etc.) are not "replaced" in a simple one-to-one fashion. Many people, traffic experts not least, seem to be unconsciously stuck with the assumption that there is a more or less fixed number of trips to be made, and that the effect of the charges should be possible to sort neatly into categories like "mode change", "destination change" and "departure time change". Especially for discretionary trips, adaptations are much more multi-faceted. This means that commonly encountered statements, such as "congestion charging won't work in our city because our transit system is too bad" or "…because we have no ring road", are oversimplistic: there are many more ways to adapt than changing mode or route.

Traffic isn't just work trips

Work trips only make up a fraction of car traffic – a typical figure could be 40%, with the rest being discretionary trips and professional traffic (where a typical figure could be 15%). Discretionary trips are easier to affect, because there are more ways to adapt in the short run, and represent a significant fraction of traffic, especially during afternoon peak hours, when congestion is often just as severe as during the morning peak. Professional trips are very heterogeneous: some types are very difficult to change, while some are not. Typically, values of time are very high, which means that time savings for professional traffic will constitute a significant part of the travel time benefits. Despite all this, it is common that the discussion focuses exclusively on work trips, both among planners, policymakers and the general public. This is a mistake that suprisingly often confuses the discussion about what congestion charging can do and how it may work.

Many car drivers won't even know that they have adapted

In fact, many car drivers will not even know if or in what way they adapted. This is because travel patterns are much less repetitive and stable than many people think. Many of the affected drivers are "occasional car drivers", who drive on the charged road perhaps a couple of times each month. Other days, they use other modes, times or routes. These drivers will change "on the margin", and it will often be impossible to tell if and how they changed – they often won't know themselves. Moreover, there are many other changing processes going on. People move houses and change jobs, for example. After just a few years, it may be pointless to ask how a given person has "adapted" - because the entire situation where travel choices are made has changed.

Retail effects are generally small

Fear of an adverse impact on retail inside a cordon is common. Large efforts have been made to track such effects, only to conclude that they are very small or non-existent. There may be effects on particular stores, especially if they lie close a cordon, but the average effect in an urban centre is usually small. This should be evident, especially in the long term: if the retail market inside the cordon becomes less attractive, then floorspace rents will, in equilibrium, decrease to counteract this, making the effect on the number of stores even smaller.

3. GETTING LARGE BENEFITS: EFFICIENT CHARGE DESIGN

This section covers issues related to the design of the charging system – decisions such as where and when to levy charges and how to forecast effects.

The goals have to be explicit and relevant

First, the system needs a goal. The goal may be to reduce congestion, improve air quality, yield revenues, or a combination of such goals. Whatever the goals are, they need to be explicit. Moreover, they should be quantified, at least to some extent. This quantification usually has to be done in co-operation between policymakers and traffic experts: setting up relevant goals and targets is harder than most policymakers realise. Goals must above all be relevant and consistent. Specifically, one should at this stage *not* specifically strive to make them easy to communicate to the public. Communication is important, but comes later. The goals set at this stage are the ones that will be used during the design process, and they need to be consistent and relevant, not necessarily easy to explain or "sound good". An example of a consistent and relevant goal that happens to be rather difficult to communicate is: "to achieve maximal social benefit from congestion reduction" (perhaps given some restriction on charge levels). A common example of a goal popular among policymakers and communicators is: "getting more people to choose public transit". This is *not* a relevant goal for congestion charges, which should be obvious. (The relevant goal is to make less people choose the car during congested hours. If they instead choose transit, that's fine; if they prefer to adapt by cancelling trips or changing departure times or destinations, then this is just as fine.) Choosing ill-formulated goals and targets will very likely cause problems during the design process, at the very least causing confused discussions.

Designing the charges is a job for experts

Designing a charging system is, as a rule, a very difficult task – how difficult depends on the topography of the city. (For example, in Stockholm it is reasonably easy, with the worst congestion problems located along a natural cordon, while Gothenburg is difficult, with congestion problems spreading out from a complicated multiple-arterial junction.) It is absolutely necessary to have sufficient time, and access to a reasonably good transport model. If one has access to design optimisation tools, this can come in very handy. Even given this, it will be difficult. In particular, intuition and prior knowledge in general will not be sufficient, even for experienced traffic planners: transport systems are simply too complex.

There will almost certainly be surprises, and the first attempt at a charging system design will most likely not be optimal or even good – it may even make congestion *worse* overall by "moving congestion around". The system design is an iterative process, where involvement of politicians does not help. This is why it is so important that goals are stated clearly at the outset. Design and goal-formulating, ideally, are part of an interactive process as well: it is likely that some goals or (more likely) some design restrictions were forgotten at the outset. But this does not change the basic premise that, while formulating goals and restrictions is a job for policymakers, designing the details of the system – locations and levels of the charges – is a job for experts. An ill-designed system may not only be "sub-optimal"; it may likely cause *more* problems than before.

You need a good transport model

(This section is important for experts designing a charging system, but policymakers and general planners can skip this.)

Most transport models are constructed for other purposes than modeling the impact of congestion charging. Certain shortcomings of most current models become especially important in the context of congestion charges, and one needs to be aware of them. First, the value of time differs between vehicles, but this is often neglected in the assignment step. Technically, one must use multi-class assignment, i.e. divide traffic into several "classes", each with a value of time of their own. The value of time of each class will decide whether it is worth taking a detour to avoid a charge. Depending on the topography, the value-of-time distribution over classes may affect results strongly. Often, there is little evidence on the value-of-time distribution, so sensitivity analyses will play an important role. Second, departure time choices and scheduling considerations are often sketchily implemented, if at all. Obviously, this will underestimate the impact of a time-differentiated charging system, since the opportunity to adapt by changing departure time is not reflected in the model. Less obvious, one may underestimate traffic decreases during non-charged hours – since those trips are partly made up of "return trips", i.e. the second leg of a trip whose first leg was during the charged time period. Third, static assignment models, in general, will underestimate travel times in the presence of severe congestion. Among other things, they will, by definition, neglect the effect of "spillback congestion". This means that during the design process, it may be better to focus on traffic decreases in known bottlenecks, rather than to focus on actual travel times from a static traffic model (although travel times need to be used as well).

Try to get political and legal capacity to adjust the system once it is in place

Even with careful planning, surprises are likely to appear when the system starts. In the best case, surprises are positive (in Stockholm, travel time improvements were larger than anticipated, for example). But there may be negative surprises as well: e.g. unexpected "rat-running". Because of this, it is good if one can obtain political and legal leeway to make minor adjustments to the system with a minimum of delay and hassle. Politically, this will be easier if goals are clearly formulated: if so, then it will be easier to see if they are met or not, and if not, the system can be changed. The legal problem may be harder to solve. In Sweden, for example, the charges (which are formally a state tax) have to be decided by parliamentary decisions, which involves a lot of time and political effort.

There is a conflict between "effective" and "easily communicated" design, but erring towards too simple seems more common

Policymakers often stress that they want a design that is "simple to understand". While it is an important consideration that the system must be sufficiently simple for the presumptive users to understand, policymakers often seem to underestimate people's cognitive ability. The Singapore system and the US "value pricing" roads, for example, appear complex at first glance. The charge is finely differentiated by time and location and, on top of that, may change fairly often. Despite this, it turns out that users are able to grasp and adapt to the system. Forcing the system to be too simple too early in the design process is likely to cause design restrictions that are difficult to solve. The reluctance of many politicians and planners to consider "over-complicated systems" can lead to the point where the system becomes so simplified that it will not deliver the promised congestion reduction. This will not only be a waste of resources – it will also lead to low acceptability of the charges (we return to this below).

4. GETTING LOW COSTS: EFFICIENT PROCUREMENT AND TECHNOLOGY

This section covers issues related to the investment and operations costs of the technical system, in particular achieving a procurement process that keeps these costs down. Designing a cost-efficient technical system amounts to more than just establishing a system that identifies cars: the difficult part may be to construct cost-efficient payment channels, sort out the legal status of the charges, etc.

Clarify the legal conditions early

Early in the techical design process, one must know the legal conditions. For example, what is acceptable proof that a vehicle has passed a gantry? What possibilities to appeal must exist? The answers to such questions will have important repercussions on the technical design; for example, whether transponders can be the sole means of identification or not.

In Stockholm, a problem occurred that hopefully should be rare: midway in the procurement process, the legal status of the congestion charge changed from a "municipal environmental charge" to a state tax (a legal investigation concluded that it was illegal for a city to charge moving vehicles on existing roads). This had many effects, including that the responsibility of the procurement had to be changed from the City of Stockholm to the National Government. This considerably increased the cost for establishing the system.

Consider the cost-efficiency of service level targets

Consider what the cost-efficient targets of service levels are, given the goals of the system, and how different service levels affect the intended function of the system. Going from, say, 95% to 99% or from 99% to 99.9% on any given service level may be a significant cost driver. In Stockholm, the "up-time" of the system (measured as the share of "lane-minutes" when the system was actually registering passages) was required to exceed 99.9%. To meet this high requirement, the prime contractor designed a system where (almost) every component was duplicated, spare parts were obtained in large quantities, trained staff were made available to do on-site service with short notice, and technical IT support was initially on standby 24/7. Obviously, this increased investment and operations costs. Moreover, it should be obvious that lowering the up-time requirement to, say, 95%, would not affect the traffic-reducing effect of the charges. After all, travellers make their decisions as to whether to drive or not based on the fact that they are highly likely to have to pay if they do so. From this perspective, availability requirements could have been relaxed substantially, and thereby also system build costs, without losing any of the ultimate effect on the traffic situation. This illustrates the principle of having cost-efficiency in mind when formulating technical system requirements.

Choose cost-efficient payment channels

Each payment transaction comes at a cost, both in terms of convenience for the user and as a fee from the financial service provider. Hence, allowing for aggregated monthly payments rather than paying each passage individually will reduce operating costs. Cash over the counter (in shops, for example) might be necessary for user acceptance, but it is probably the most expensive form of payment.

Handling transponders is expensive

Transponder (or "tag-and-beacon") technology is efficient in many ways, not least because it allows complex charging structures and makes it easy for the driver. The production of many transponders may be a significant cost driver, though, but less well known is that it is often a major cost to *administrate* transponders. New cars need new transponders, cars change owners, and transponders are lost, stolen and broken. In Norway, where over 40 different road toll schemes are in operation, transponders are used in some, while others are managed by manned tollbooths. And even there, where the comparison technology is highly manual, there is a slight productivity advantage for those *not* using transponders (Odeck, 2008). With today's technology, cameras and Automatic Number Plate Recognition (ANPR) can potentially reach a very high identification ratio, which offers ample competition for any transponder-based solution. The Stockholm system started out as a transponder-based system, with ANPR as an add-on for legal reasons, but it has relied on ANPR exclusively for several years.

When doing a functional procurement, make sure to align cost and risk responsibilities

In Stockholm, the call centre, answering questions about the charges, was initially vastly oversized, which was a major cost driver at first. Part of the reason for this was that if the call centre would not meet its service quality targets (e.g. maximal answering times), then the prime contractor would be financially penalised, while it was the buyer who carried the cost of call centre staff. Hence, risk and cost were borne by different parties[2], and the contractor had no incentive to increase its own risk by cutting down on resources. If procuring a system as a function, one should make sure that the party carrying the risk is also the one carrying the cost for risk mitigation, in all areas of the operation.

High political risks will weaken the public negotiating position, and will increase costs by having the contractor require a "risk premium"

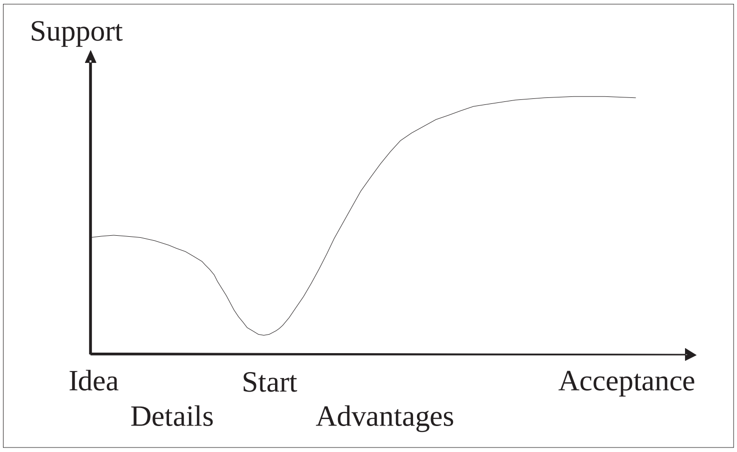
In Stockholm, the stakes were high for almost all the actors involved. Individual careers as well as the prosperity of private firms and political coalitions was at risk, or at least perceived as being so. This dominated the context in which the project was carried out, and it was under the influence of this risk environment that decisions were made. There were many unknown factors that were thought to kill the project on their own. Above all, if the system did not work, or was perceived not to work right from the start, it would almost certainly be abolished immediately. This is at least a partial explanation of cost drivers such as the oversized call centre, the excessive service level requirements, etc. It all goes back to intense political pressure and high political stakes: the outcome of the next election would depend on the outcome of the trial, perhaps not only in the city but also on the national level. This meant that the public negotiating position was weak – the system *had* to work, and it *had* to be finished on time. Obviously, such a situation creates opportunities for a contractor to charge more money. For the contractors, a failure – even if not due to mistakes of their own – could be potentially disastrous for future business. This means that contractors will require a risk premium to even engage in the work of constructing the system. Hence, the lesson is that a stable political environment and ample time to plan and implement the system will keep costs down.

5. GETTING PUBLIC ACCEPTANCE

This section covers issues related to public and political acceptability. It draws from the conclusions in Eliasson and Jonsson (2009).

Acceptability decreases with detail but increases with familiarity

Support for congestion charging often follows a typical pattern. The figure below shows the principle.



A fairly large fraction of the population is generally willing to support the idea of congestion pricing. How large this fraction is depends on how the question is formulated and framed – for example, revenue use, the purpose of the charges and what policy alternatives it is contrasted against, all matter. But once a detailed proposal is worked out, support generally decreases. There may be several reasons for this – for example, that the disadvantages suddenly become more evident than the potential advantages, or fears that the technical system will not work or become very expensive. This is sometimes summarized in the formula: "acceptability decreases with detail". But once the system is in place, support will generally increase, which is often summarized as: "familiarity breeds acceptability". There are probably several reasons for acceptability to increase through familiarity with a real system. One oft-quoted reason in Stockholm was that the positive effects on road congestion and the urban environment were much larger than most people expected. A second, also oft-quoted reason is that the public fear that travel costs will increase more, and/or their travel patterns will have to change more, than is actually the case. Once the charges are in place, many people may realise that the charges do not in fact affect them as much as they had thought. A third reason is the psychological effect known as cognitive dissonance, a phenomenon that can be simply summarized as: "accept the inevitable". In other words, once the charges are in place, it is less worthwhile to spend energy on opposing them. A fourth reason may be decreased reluctance towards pricing a previously unpriced good. There is evidence that people in many cases do not like prices as an allocation mechanism. But once familiar with the thought that road space is in principle a scarce good that can be priced – much like parking space or telecommunications capacity – this reluctance may tend to decrease.

Plan the political process accordingly

The general acceptability curve above has implications for the political process. For a successful implementation, one needs to avoid having elections when support for the charges can be expected to be at its lowest. In London, the election was held before the details were completely worked out; in Stockholm, a referendum was held once the charges had been in place for a little more than half a year (in fact, support then continued to increase, so a later referendum would probably have been preferable from the point of view of charging supporters). This can be contrasted with Manchester and Edinburgh, where detailed proposals were worked out (support fell, just as could have been expected) but charges were not implemented before a referendum was held. In both cases, the referenda rejected the charges.

The system has got to deliver benefits

In Stockholm, the most important factor explaining attitudes to the charges turns out to be the perceived effect of the charges – in particular less congestion. People agreeing that the charges have had positive effects are much more likely to support them – which is of course to be expected. Even if one should not confuse "perceived" effects with "objective" effects – since attitudes influence which effects are actually perceived – it seems clear that achieving objective effects is necessary to reach acceptance. This underscores the importance of designing the system carefully and only using congestion charges when congestion really is a problem. Moreover, it seems likely that measuring effects and communicating the results through, for example, the kind of scientific evaluation carried out in Stockholm will increase the awareness of positive effects – provided, of course, that there are, in fact, positive effects.

"Branding" matters

In Stockholm, the charges were, to a certain extent, marketed as "environmental" charges. The charges did certainly have environmental effects (in particular on emissions in the inner city), but this effect was dwarfed by the very large effects in terms of congestion reduction. It turns out, however, that voters' environmental concerns were an important factor for the acceptability of the charges. (Interestingly, data shows that it seems as if it is not environmental *behaviour per se* that is important, but the *self-image* of being an environmentally concerned person.) This is in line with findings in the literature that social norms of this type influence acceptability in general, and that support depends not only on the "objective" characteristics of the measure itself, but also on the defined *objective* of congestion charges. Moreover, several authors have found that it is not just perceived individual benefits that determine acceptability: perceived *social* costs and benefits can also strongly affect acceptability. Hence, the "branding" of the charges matters – how they are marketed, explained and perceived. In Stockholm, "re-labeling" congestion charges to "environmental charges" and emphasizing their positive effects on air quality may very well have had an impact on acceptability. Other cities may employ different strategies, but the general conclusion remains: it is important how the charges are "branded". A condition for this to be possible is that the system design is well aligned with the stated purpose of the charges. A system branded as a "congestion charges" system, for example, should not levy charges where or when there is no congestion.

Car dependence and transit satisfaction seem to matter less than many believe

Analysis of attitude data from Stockholm shows that, as expected, car dependence decreases support for the charges while transit satisfaction increases it. Somewhat surprisingly, it is a less important factor than environmental attitudes and perceived effects of the charges. Econometric simulations on this data show that even with a fairly important level of car dependency, support stays relatively high. Moreover, residential location zones hardly matter once the other factors are controlled for. This is surprising, since the consequences of the charges in terms of, e.g., tolls paid differ quite a lot depending on residential area. Other analyses, though, have shown significant impacts on voting behaviour from the changes in travel times and travel costs of voters' residential zones. The evidence is not conclusive, but it seems that "self-interest" variables such as car and transit use matter less than one might think initially.

"Fairness" can mean many things

The "fairness" or "equity" issue is always important. Initially, the dominating perspective is often "before-after". In other words, how do travel costs (and perhaps times) change for different groups - high- *vs*. low-income, men *vs*. women, inner city *vs*. suburb residents. At least in cities with decent transit shares, it is often the case that the "rich" will pay more than the "poor", with middle-income groups "suffering" the most, relatively speaking. But once the charges are in place, another perspective becomes more important – "fair pricing". In other words, what price is "fair" to charge? From this perspective, it is "fair" that one pays more to drive on a congested road or to cause emissions in densely populated areas – irrespective of income or place of residence, for example.

"Winners" and "losers" become increasingly difficult to identify over time

As discussed above, travel patterns are not static. Not even the context – workplace and residence location, for example – is constant over the span of a few years. This means that, over time, identifying "winners" and "losers" will become increasingly meaningless. The charges will have changed from being experienced as an "external shock" to being a factor considered when making choices of workplace, travel mode, etc. Hence, the question of who "wins" and "loses" is only relevant in the short term.

Power issues may be decisive for political acceptability

From the politicians' perspective, often a decisive question is who has the power over revenues and charge levels. If it is at the national level, then regions and cities will obviously be much more reluctant to introduce charges. But even if the region keeps the revenues, another issue is important: how the existence of this new revenue stream affects the complicated negotiation between national and regional levels about national infrastructure grants. In Norway, the state "matches" income from regional charges with national funding. A recent trend in Sweden is that regional funding is "leveraged" with national funding. The regional funding often comes from charges called "congestion" charges (which in most cases is a terrible misnomer). This has made "congestion" charges much more popular among politicians, which shows the importance of the institutional context and incentives.

NOTES

1. It should be pointed out, however, that road user charging may be a cost-efficient way to improve air quality in urban centres.

2. To be fair, it should be pointed out that this misalignment of costs and risks was an exception in the Stockholm procurement.

BIBLIOGRAPHY

This paper draws heavily on already published material. In this draft version, references have not been cited in the text. Instead, we will list the most important published sources of the statements made in the paper.

Bamberg, S. and D. Rölle (2003), Determinants of People's Acceptability of Pricing Measures: Replication and Extension of a Causal Model, in: J. Schade and B. Schlag (eds.): *Acceptability of Transport Pricing Strategies*, Elsevier, Oxford.

Bonsall, P., J. Shires, J. Maule, B. Matthews and J. Beale (2007), Responses to complex pricing signals: Theory, evidence and implications for road pricing, *Transportation Research A*, **41**, 672-683.

Brundell-Freij, K. and L. Jonsson (2009), Accepting charging – a matter of trusting the effects? Presented at the 2009 European Transport Conference.

Eliasson, J. (2009a), Lessons from the Stockholm congestion charging trial, *Transport Policy*, doi: 2008.12.004.

Eliasson, J. (2009b), Expected and Unexpected in the Stockholm Trial – A personal view, in: A. Gullberg and K. Isaksson (eds.), *Congestion Taxes in City Traffic. Lessons learnt from the Stockholm Trial*, Nordic Academic Press.

Eliasson, J., L. Hultkrantz, L. Nerhagen, L. Smidfelt-Rosqvist (2009), The Stockholm congestion-charging trial 2006: Overview of the effects, *Transportation Research A*, 43, 240-250.

Franklin, J., J. Eliasson and A. Karlström (2008), Traveller Responses to the Stockholm Congestion Pricing Trial: Who Changed, Where Did They Go, and What Did It Cost Them?, in: Saleh and Sammer (eds.), *Demand Management and Road User Pricing: Success, Failure and Feasibility.*

Frey, B.S. (2003), Why Are Efficient Transport Policy Instruments So Seldom Used?, in: J. Schade and B. Schlag (eds.), *Acceptability of Transport Pricing Strategies*, Elsevier, Oxford.

Gullberg, A. and K. Isaksson (2008), The Stockholm trial, in: Isaksson (ed.), *Stockholmsförsöket – an unlikely story*, Stockholmia, Stockholm. (In Swedish; forthcoming in English.)

Hamilton, C. (2010), Revisiting the cost of the Stockholm congestion charging system. Working Paper, Centre for Transport Studies, Royal Institute of Technology, Stockholm.

Jaensirisak, S., A.D. May and M. Wardman (2003), Acceptability of Road User Charging: The Influence of Selfish and Social Perspectives, in: J. Schade and B. Schlag (eds.), *Acceptability of Transport Pricing Strategies*, Elsevier, Oxford.

Jones, P. (1991), Gaining public support for road pricing through a package approach, *Traffic Engineering and Control*, April.

May, A.D, R. Liu, S.P. Shepherd and A. Sumalee (2002), The impact of cordon design on the performance of road pricing schemes, *Transport Policy*, 9, 209-220.

Schade, J. and B. Schlag (2003), Acceptability of urban transport pricing strategies, *Transportation Research Part F*, *Traffic Psychology and Behaviour*, 6 (1), 45-61.

Revisiting the Cost of the Stockholm Congestion Charging System

Carl HAMILTON

Royal Institute of Technology
Stockholm
Sweden

SUMMARY

Stockholm, January 2010

ABSTRACT

In January 2006, a system for congestion charging was introduced in the City of Stockholm, Sweden. The charging scheme was run in the form of a full-scale trial for seven months, after which it was deactivated, awaiting its evaluation and an advisory public referendum. Several parties, including representatives of the scientific community as well as media and special interest groups, have analysed and evaluated the system. A recurring theme in several of these analyses is that the cost to build and operate the system was excessive, compared to costs for other road charging installations.

This study revisits some of the key project participants and archive data, to provide a deeper understanding of what were the major cost drivers and whether it can be lower in future installations. The approach taken is to emphasize understanding of the particular circumstances rather than comparing aggregates with other seemingly similar systems. A main conclusion is that the political context, with a tight time plan and very high political risks for all involved, were key factors for the eventual costs of establishing the system.

1. INTRODUCTION

It is well known that optimal road pricing on a congested road will yield a social surplus. From this surplus, however, costs for investment and operation of the road pricing system must be subtracted. Since these costs can be large, it is necessary to carefully investigate the cost side of road pricing.

The Stockholm charging system was introduced as a trial for the first seven months of 2006. After a subsequent referendum, the system was reintroduced permanently in August 2007. The system has been portrayed as an expensive installation in the media and public debate. The initial cost, including the seven-month trial, was indeed 1 800 million SEK (approx €180 million) for the technical system alone. But does that make for an expensive endeavour? If so, compared to what? The aim of the paper is to provide for a deeper understanding of what were the major cost drivers, why the cost ended up the way it did and whether it can be lower in future installations. Thereby, the question "what is expensive?" can be better answered, and more importantly still, the lessons learned can be better harvested by those who consider investing in congestion charging in the future.

The question of costs can be seen from two different perspectives. One is the actual situation when the system was constructed and implemented, given the then-current context and knowledge. Another is the *ex post* perspective – whether costs are high given the knowledge and context available afterwards. The first question is interesting from a historical perspective, while the second is relevant for other cities considering road pricing. By analysing how the specific Stockholm context affected the system costs and the implementation process, the intention is to shed light on both of these questions.

The methodology and sources are described in Subsection 1.1. Section 2 briefly outlines the story of the Stockholm congestion charging system and its design, with a particular focus on some claims of cost and costliness in the public debate at the time. In Section 3, various cost items are examined, beginning with those most often highlighted in the media and public debate, and then those identified by the stakeholders involved in the project. Each cost item or cost driver is discussed, both in the then-current context and viewed from an *ex-post* perspective, and a lesson is formulated. In Section 4, observations are synthesized and conclusions drawn.

1.1. Sources and methodology

The initial investment and the operation during the seven-month trial was managed as one unified project, with one prime contractor managing the technical design, implementation and operation of the system during the entire trial. Therefore, it is dubious to make any clear distinction between investment and operational cost.

Discussing costs in general, and what is expensive in particular, requires quantitative measures. Wherever possible, costs are presented in monetary terms, or in some comparable resource quantification. However, the focus is not on determining quantities exactly, but rather on understanding the significance of each cost driver and what can be learned from it for future cases.

The two main types of source used are archive documents from the Swedish National Road Administration (Swe: *Vägverket*) and the Swedish Transport Agency[1] (Swe: *Transportstyrelsen*), as well as interviews with key stakeholders. The archive data includes tendering documents, contracts and official communication between the contract parties. In accordance with Swedish public law, any document received byo a government agency must be recorded and made available for the public to view, unless it has been classified as secret. In this way, the author has been able to obtain and examine a large number of documents, from the early tendering through the entire project delivery.

Interviews were carried out with key stakeholders representing the City of Stockholm, who initiated the procurement and wrote the specification requirements for the system, the Road Administration, the ultimate buyer and owner of the system, and IBM, the prime contractor. Each interview took between one and two hours. The representative from IBM was interviewed via telephone without sound recording, while the other interviews were made face-to-face and sound-recorded.

By going through the documented material and interview comments in chronological order, it is possible to recreate much of the situation as it was when each decision was made, including what information was available at the time. This has been key in offering any judgement as to whether taking on a particular cost item was called for or not. The summary of each area presented here is based on a synthesis of all interviewees and the supporting documentation. Each interviewee has been presented with the full text, and been asked to comment not only on direct citations, but also on the interpretations made.

During the implementation phase of the Stockholm congestion charges, the author was employed by the prime contractor, IBM, and involved in the system design of the system. Therefore he has a basic understanding of how the system works, and may also have some interest in portraying the project as more successful than if he had a different background. It has been his ambition to counter any such tendencies, firstly by requesting frequent and critical examination of the work in progress, and secondly by ensuring that any information presented, even if known to him beforehand, is either stated officially by an interviewee, or confirmed by publicly available project documentation.

2. BACKGROUND

2.1. The Stockholm congestion charges

Between 3 January and 31 August 2006, congestion charges were being tried out in Stockholm, Sweden. After the 2002 general election, the Green Party ended up with the balance of power both in Stockholm's municipal parliament as well as in the national assembly. The Green Party wanted to introduce congestion charging permanently, and the Social Democrats had made an explicit promise not to introduce them. (Instead they wanted to have a referendum based on a detailed suggestion prior to any implementation.) Implementing the charges in the form of a trial was the result of a compromise, determining whom the Greens were to support to form the next Government. After the trial was ended, it was agreed, after some pressure from the opposition, that there would be an advisory referendum, where the people of Stockholm would have their say in making the charges permanent or not. The referendum was scheduled on the same day as the next general election, 17 September 2006. Before that day, a scheme was to be designed, a system procured and installed, and then run in full-scale operation for "several years", according to the original agreement. (See Gullberg and Isaksson, 2009, for a detailed record of the political proceedings.)

After a series of legal delays, the "several years" were reduced to seven months, but in most other key aspects, the trial was executed as planned. After the yes-side had won the referendum held in the city of Stockholm, the system was made permanent, and has been in operation since August 2007.

Practically, the congestion charges were implemented as a cordon around the city, with gantries across all entries and exits. Cars passing in and out were identified using a combination of cameras and transponders. The price for a passage varied between 10 and 20 SEK depending on the time of day (approximately 1 and 2 €), with the highest charge during peak hours. Nights and weekends were free of charge.

Users had to pay the charge within five days after their passage, and they could choose from direct debit from their bank account (this required that they use a transponder), manual bank payment, and cash-over-counter at convenience stores from two widely-spread chains.

2.2. Project environment

Information technology projects in general have a bad reputation for delivering late and above budget, and IT used for road tolling is no exception. The German truck tolling system, TollCollect, was initially planned to be launched in 2003, but after a series of technical, managerial and political problems, it faced delays of more than two years. During that time, the German Federal Government lost toll revenues corresponding to €3.5 billion, in addition to receiving a flood of bad press, badly hurting both the German Government and the contractors responsible for building the system (*The Economist*, 2004; Wieland, 2005; Deutsche Telecom, 2008).

The TollCollect story was an ongoing drama in parallel with the political process leading up to the Stockholm congestion charges, underlining the magnitude of the risk for anyone involved. Politically, a failure of any sort had the potential of tipping the next general election, locally as well as nationally. According to Gunnar Söderholm (2009), who at the time was one of the leading civil servants involved at the municipal level, the centre-right political opposition in Stockholm was so certain that the congestion charges would be a failure, that the decision to introduce them was considered the "biggest political suicide in history" (a view they shared with several people on the left as well) and all they had to do, they thought, was to "stand back and watch the Left-Green coalition commit it".

However, risk was not limited to the political sphere. The prime contractor selected, IBM, was under pressure as well. The procurement contract included stiff penalties, in order to align the buyer's incentives with those of the supplier (Road Administration, 2004, pp. 19, 23, 30). According to Gunnar Johansson (2009) of IBM, however, the large penalties for project delay and performance losses detailed in the contract, were still considered a smaller risk to the firm than the negative effects on the corporate brand following a possible failure. The weekly engineering newspaper, Ny Teknik (2005), saw the risk, and gave their verdict even before the system was launched. They dubbed it "Sweden's least profitable IT investment in history", three weeks before its implementation. Not delivering on time, or launching a system that produced incorrect tax claims, would hurt the company nationally among customers of all industries, and globally as a supplier of road tolling solutions, according to Johansson (2009).

> [IBM's] *future as a player in the international road user charging arena was at stake. If we had failed in Stockholm, we would not have been able to compete for any road charging bids in the future.*
>
> Johansson (2009)

Adding fuel to the fire, the political and media debate was intense and aggressive. Birger Höök (2009) of the Road Administration and Gunnar Söderholm both stress that this was far beyond what is normal in the Swedish debate on transportation infrastructure. The Automobile Association (Motormännen) and the Stockholm Chamber of Commerce (Handelskammaren) took the lead among the non-party-tied associations and issued pamphlets, a campaign website (tullvalet.se), and discussion articles in the newspapers. The campaigns routinely resorted to hyperbole and scare-mongering in their argumentation. For example, the Chamber of Commerce claimed that the design of the system was to lure people to vote yes, after which the charge levels would "soon be both doubled and tripled", and that new charging stations would appear all over the city (Stockholm City, 2006).

> *I don't know where [the Chamber of Commerce] got their numbers […] there is nothing that is correct about them. I think this is more a question of feelings than of facts.*
>
> Höök (2009)

During the early phases of the project, failure seemed almost inevitable, among authorities as well as in the media. The Tax Authorities actively went to the press, warning that congestion taxes would lead to children being indebted by an assumed widespread tactic, where parents were expected to register their children as owner of their vehicle and then refuse to pay the tax (DN, 2005a). The National Collection Agency (Swe: Kronofogden) estimated that some 6 000 cases would be passed over to legal collection every day (DN, 2006; SvD, 2005). And the daily *Aftonbladet* (2005) ran on their first page their own estimates that 85% of people would attempt to cheat the system so as to avoid paying.

Next to those who hoped and believed that the project would be a dismal failure, were some who were in favour of congestion charging in general, but who feared that this particular implementation would do more harm than good. The reasoning was that if the system failed here, then no politician would dare touch the idea again for decades to come. So pressure on the project mounted not only from those who opposed it in general, but also from those who supported it in principle (Söderholm, 2009; Johansson 2009; DN 2005b).

In reality, none of these fears materialised. The system started as planned on 3 January 2006. The cost per passage is still unchanged as of 2010. Fewer than 600 cases were sent to collection during the entire seven-month trial combined (Söderholm, 2009), and the level of cheating was barely measurable (Höök 2009; Johansson 2009). But it is not under the soothing influence of hindsight that the project was carried out, but under the pressure of the fear and defeatism present at the time.

In summary, the stakes were high for almost all of the actors involved. Individual careers as well as the prosperity of private firms and political coalitions were at stake, or at least perceived as being so. This is an important aspect to bear in mind, as it dominated the context in which the project was carried out, and it was under the influence of this risk environment that decisions were made.

2.3. Costs and cost estimates in research and media

The morning daily, *Dagens Nyheter* (DN), ran a series of articles in 2008, with the common theme of the high costs of the congestion charging system. DN (2008a) compares early estimates of 800-900 million SEK for investment and 100 million SEK for annual operation, with their own estimates of final costs totalling at 1 800 million SEK in investment and 380-400 million SEK per year in operation. These "early estimates" are found in a report from Transek (2003), outlining a possible system design and its consequences. A doubled investment cost and a quadrupled increase in operating costs surely seems a noteworthy cost overrun.

To understand this discrepancy, let us start with the consultancy report, where the first estimates were made in April 2003. The main focus of this report is to suggest a high-level scheme design. Cost estimates make up only one paragraph of the report, and they are labelled as "very uncertain" (Transek, 2003, p. 15). Transek used two sources for preparing those early estimates, both of which are likely to underestimate real costs: First, there were data from Norwegian road charging systems, which are developed and operated in a stable political and legal context, with cost minimization as their main focus, and many of which had had more than a decade in which to trim their operation. Second, were data from equipment manufacturers, who have an incentive to underestimate such costs as part of their pre-sales process.

Furthermore, in the time between the publication of the first consultancy report (April 2003) and the eventual parliamentary budget decision (16 June 2004), the scope of the system changed in important ways, from being a municipal environmental fee, similar in function to the Oslo toll ring, to a national congestion tax with a partially different view of how the technology (especially the transponders, see below) was to be used. Hence, the early cost estimates were not only uncertain, they were also estimating a different system than the one later decided upon.

In the same DN article, the budgeted 380 million SEK for operational costs for the fiscal year 2008 are seen in relation to an estimated revenue of 750 million SEK. It is then concluded that about 50% of the revenue is spent on collection of the congestion tax. This kind of *cost ratio* is a commonly used way of indicating the efficiency of a road charging system, i.e. its total operating cost divided by the revenue. If a system is spending a large portion of its revenue on the actual collection process, it is argued, then the system is inefficient.

A similar line of argument is also found repeatedly in the debate about the congestion charging trial. In a report financed by the Royal Automobile Association, Prof. Ilja Cordi (2006) reaches a similar estimate for the cost ratio of 50% for the Stockholm system. The same way of reasoning is used by the Stockholm Chamber of Commerce (2005), which relates the 3 800 million SEK in total cost for the entire trial (including improved transit and park-and-ride), to the 500 million SEK estimated revenues during the seven-month trial, thereby implying a 760% cost ratio for the trial. In a later report, they put an estimated 800 million SEK of projected income in relation to 1 200 million SEK in projected annual costs, implying a cost ratio of 150% (Chamber, 2006). For comparison, a corresponding cost ratio for the Norwegian toll rings is in the range of 9-10% (Amdal *et al*., 2007).

By these cost ratio measures, the Stockholm system seems very expensive. Using the cost ratio as a measurement of the efficiency of a congestion charging system is not completely relevant, however. Firstly, the charge amounts collected bears no relation to the cost of collecting them. If a charging scheme has a complex tariff structure with many exemptions and price levels, the cost of collecting the charge will be higher without any related changes in the charge income level, and the cost rate will go up. Likewise, if the charge level is doubled, there is no additional cost for collecting it, and the cost ratio will go down. In neither of these cases is the cost rate a fair reflection of the expensiveness of the charging system.

Secondly, the Norwegian toll rings are primarily revenue-generating schemes, where the value sought after is the money collected. To measure efficiency is to relate quantity of *outcome* to its required quantity of *input*. Amdal *et al.* (2007) therefore rightly choose cost ratio as their primary illustration of the toll operators' efficiency. But the desired outcome of the Stockholm system was not primarily to collect funds, but to improve the traffic situation. Thus, the relevant ratio when measuring the efficiency of a congestion charging system is to place the social benefits of congestion reduction in relation to the costs.

Outside of the debate in the media, a more serious attempt to measure efficiency has taken place. Eliasson (2009) presents a social cost-benefit analysis, where the tax paid shows up first as a cost (as taxes paid by car drivers), and second as a benefit (as tax revenue to the government). Instead of seeing the tax itself as part of the outcome, positive effects are balanced against negative, where the cost to establish and operate the system is the single largest line item among all costs, which is then to be compared to the value generated by reduced congestion, etc. In addition, Eliasson (2009) uses the expected long-term operational cost, assuming that economies of learning will further reduce costs. Using Eliasson's figures to generate a metric similar to the cost rate, yields 220/654 (34%) as a measure of operational cost per social benefit (in monetary terms). Reversely, each SEK spent on technical system operation yields a return of 3 SEK in social benefit, although these are not ratios explicitly used by Eliasson. The assumed 220 million SEK of long-term costs have since been beaten by reality, and the system is likely to operate in the 180 million SEK range from 2010, reducing the societal cost ratio to 27% and the financial cost rate to 21% (Lissel, 2009).

Unlike Eliasson (2009), another evaluation of the Stockholm system, carried out by Prud'homme and Kopp (2007), shows a negative social surplus, primarily from using a different way to calculate time benefits (see Eliasson, 2008). The principal view of costs and benefits, however, is similar to Eliasson's, and the authors also agree by emphasizing three factors necessary for a successful congestion charging implementation; (1) a high degree of congestion, (2) low system installation and operation costs, and (3) low marginal costs in public transit.

Even when ignoring the most politically biased interpretation of costs, it is clear that the cost to build and run the system is an important factor in how it is to be evaluated. System cost is among the most important line items, regardless of whether one is measuring the social or financial cost ratio. To understand why the cost ended up the way it did, it is necessary to know more than system size in terms of charging stations or number of vehicle passages. For the full picture, we need to factor in the project environment, and the information available when the decisions were made.

The cost for the seven month trial period included items not related to the congestion charging system *per se*, such as increased transit capacity, purchases of new buses and new park-and-ride facilities. The total of all these costs amounts to 3 800 million SEK. This sum is sometimes quoted as the "cost of the system", although it apparently includes many other measures. The congestion charging system and its supporting functions alone, which is what is included in this study, was about half of that. The Road Administration's budget for the trial, set soon after the legislation was finished in July 2004, was 1 926 million SEK (Höök, 2009). For this amount, they obtained a system designed, built and operated for seven months. Contrary to the impression given in media, this budget was never completely used up (Höök, 2009). According to Eliasson's (2006) estimates, 1 050 million SEK of this was used prior to going live, which could serve as an indication of investment costs excluding operation. Since the congestion charges were made permanent, operational costs have gradually dropped, to 200 million SEK in 2009 and an estimated 180 million SEK in 2010 (Lissel, 2009). Eliasson (2006) uses an older estimate of 220 million SEK/year in his analysis. It should also be noted that included in the operational cost for the Stockholm system during 2007-2010 is the cost of transferring the operation from the contractor to the agency's own data centre.

For comparison, these investments (including the operations during the trial) and long-run operational costs, rough as they may be, can be put side to side with those of Oslo and London. Oslo's toll ring has been in place for a much longer time, and may not be fully comparable. However, it is estimated, by Ieromonacho *et al.* (2006), to have cost only 208 million SEK to establish (using 1.23 SEK to the NOK) and 148 million SEK per year to run (see also Fjellinjen, 2004, and Eliasson, 2009). It is difficult to compare with the costs of the London system, since costs for operations and implementation were combined in a different way than in Stockholm. Oehry (2006) estimates London's investment costs to amount to 1 495 million SEK and its operational costs for the year 2005 to be 1 530 million SEK (at 11.50 SEK to the pound sterling).

Worth noting is the asymmetry between Stockholm and London. The Stockholm system is larger in scope, and was slightly more expensive to put in place (taking the operation as included in the investment cost), but costs only a tenth to operate. This can partly be explained by the higher degree of automated identification and payment processes in Stockholm, and partly by the fact that the London system was procured differently from the Stockholm one, with the supplier carrying more of the capital investment.

3. COST ITEMS

This section lists a number of cost items and cost drivers, with a particular focus on circumstances or cost items that may have caused investment and/or operations costs to be unnecessarily high. Some of the claims of "excessive" or "unnecessary" costs have been put forward by external examiners, such as media or interest groups, while others have been put forward by project stakeholders in interviews.

For each suggested cost driver, the claims made are presented, followed by the author's own conclusions, based on comparison of the arguments presented, comparison cases and the documentation available. Where relevant, an order of magnitude of the impact of each cost driver is indicated. Finally, "lessons learned" for future road user charging systems are formulated.

3.1. Oversized back office

A new call centre was set up to support the public with information about the charges, when the system was active, who was liable for the charge, how to pay, as well as with user-specific payment information and account status. Even though the Road Administration already had a call centre in place for other purposes, a new, separate centre was built and staffed for the congestion charging trial. Based partly on comparisons with the call centre used for the London congestion charges, it was decided that the centre should be dimensioned to manage 30 000 calls per day, reaching a total of 400 seats. Managers of the London congestion charging system then advised the Road Administration that this might be too small (Söderholm, 2009). As it turned out, the numbers were vastly overestimated. The number of telephone calls during the first few days of the trial reached about 10 000, and then dropped to a steady rate of less than 2 000 calls per day. In hindsight, the cost for the call centre could have been between 50% and 75% lower.

Similarly, the Tax Authority had staffed a new department for dealing with appeals and complaints. Two managers, one legal advisor, four administrators and 27 clerks were assigned to deal with an expected inflow of 1 000-1 500 cases per day. In reality, appeals and complaints barely made up a tenth of the estimates, and the department was quickly downsized again [*Dagens Nyheter*, 2005a; *Dagens Industri* (DI), 2006]. Yet another call centre was established in the City of Stockholm, staffed with 15 people, to deal with political questions. Virtually nobody called, and that too was dismantled (Söderholm, 2009).

The oversized back office is one of the cost items singled out for mention in the cost-benefit analysis by Eliasson (2009). *Dagens Nyheter* (2008b) claims that this could have been foreseen, and that it should have been obvious that the London comparison was not relevant, since the London system used the call centre as a payment channel, unlike the Stockholm system, where payments were managed by other means.

Mr Höök at the Road Administration agrees that they were probably pushed by the prime contractor to employ somewhat more call centre staff than they would otherwise have done. This, he reasons, had to do with the way the contract was designed: If the call centre would not meet its quality of service targets, then the prime contractor would be financially penalised, while it was the buyer who carried the cost of the call centre staff. In this area, the design of the contract meant that risk and cost were borne by different parties, and the contractor had no incentive to increase its own risk by cutting down on resources (Höök, 2009).

It should be noted, however, that this separation of risk ownership and cost carrying in the case of the call centre is atypical for the contract, which in all other major areas manages to assign the prime contractor both the burden of the risk and the responsibility (and cost) for risk mitigation (Road Administration, 2004).

DN (2008c, 2008d) even claimed that the prime contractor had deliberately inflated their invoices to exploit the temporary monopoly they enjoyed. However, this is not something that the Road Administration believes to have had any significant influence on the final cost level, referring here not just to back office, but to all disputable cost items:

> *If we had put pressure on the supplier we could have saved a hundred or a couple of hundred millions. [...] but it takes time and […] our platform was not so good to negotiate from. We wanted to have a functional system in operation for the trial. A ‘money-fight’ would most probably only have ended up with delays and no system in operation for the trial. One can consider whether it is worth entering such a fight.*

Höök (2009)

Supplier incentives, however, are only one part of the answer to why the call centre was oversized. At the time nobody knew how many people would call, and with what kind of inquiries. One of the risk scenarios considered was that large numbers of disgruntled citizens would call and appeal every tax decision made, and thereby flood the citizen service channels. This would then lead to long waiting times, which would be portrayed as a project failure in the media. Given the heated tone in the public debate at the time, this scenario was generally believed to be plausible, and such attempts might actually have taken place. But if they were, with a high capacity call centre in place, they were unfruitful and discontinued.

Both Höök (2009) and Söderholm (2009) are of the opinion that poor service levels in the call centre, and the bad publicity expected in response, would have been more than the project could bear. In such an event, key people would quickly have pulled out their support for the project. Thus, a potentially *undersized* call centre had on its own the potential to kill the entire project, they argued, and therefore, it was preferable to err on the high side.

Yet another way of interpreting the overcapacity of the call centre is to see it as a sign of a successful information campaign, scheme design and system functionality. The users of the system knew more, and had to ask less than expected, about how the charge worked, and there were fewer disputes and appeals than expected.

Cost consequence: It is likely that the back office would have been equally good at keeping service levels up and nuisance calls at bay, even if the resources spent would have been cut down to half during the trial period.

Lessons learned: If procuring a system as a function, make sure that the party carrying the risk is also the one carrying the cost for risk mitigation, in all areas of the operation.

3.2. Transponders were not really necessary

As part of the tendering documents, where suppliers were invited to bid for the design and construction of the system, it was not explicitly stated that the offer had to be based on transponder technology. It is, however, clear that this was the expectation of all key stakeholders (Höök, 2009; Johansson, 2009; Söderholm, 2009) and that any bid not featuring transponders would not have been favourably evaluated. So transponders were part of the winning bid, and a total of 700 000 units were bought and 450 000 distributed to the users, and radio beacons were mounted on gantries at the 18 charging stations.

During the project's implementation phase, a change order of major importance was issued (a *change order* is an instruction to the prime contractor to adjust the specifications). The Ministry of Finance, together with the Justice, Enterprise, Energy and Communications ministries, came to the conclusion that under Swedish law, a transponder signal was not a sufficient basis for making a tax decision. Instead it was found that a photo of the licence plate was necessary. Transponders were still kept in the system, as the first tests with automatic licence plate recognition (ANPR) were only able to interpret around 60-70% of the photos taken without manual assistance (Höök, 2009; Söderholm, 2009).

Transponders, which have a ratio of automatic identification (ID ratio) close to 100%, could not be made mandatory for legal reasons. Hence, the contractor initiated a focused development effort to increase the identification ratio for ANPR and, after a few months of experimenting, they were able to push the ID ratio of ANPR well above 90%, which was the effective level of identification when the system went live in January 2006. After the end of the trial, when the system was re-launched in July 2007, the transponders were only promoted for users who wanted to be absolutely sure to get the benefit of the Lidingö exemption (see below). Meanwhile, the ID ratio of ANPR had been pushed another few percentage points, so that with a small amount of manual support the total ID ratio was steadily between 95% and 99%. Finally, in late 2008, support for the last remaining transponders was discontinued (Höök, 2009; Johansson, 2009; Söderholm, 2009).

The system had evolved from being transponder based at the time of the contract into gradually being more and more reliant on photo-based identification. Since the charge is legally defined as a tax, making the vehicle owner liable, there is no need for a separate payment account, which a transponder might be used to represent. The Road Administration already had in their possession a complete registry with the names and addresses of all vehicle owners liable to pay, so identification of the vehicle was all that was required.

Even though the transponders were already bought and paid for, and the radio beacons installed over the roads, there were still costs associated with managing transponders. New cars needed new transponders, cars changed owners, and transponders were lost, stolen and broken. To save the cost of managing transponders, the Road Administration decided to discontinue its support of them and rely solely on photographic identification (Höök, 2009; Söderholm, 2009).

In hindsight, the entire transponder investment can be seen as an excessive cost item. But then it must also be understood that relying solely on cameras, and reaching such high ID ratios via ANPR in a real situation, was unheard of in the industry at the time. Nobody could have foreseen that this would be possible, and even if some supplier had proposed such a system, it would have had a slim chance of being selected (Johansson, 2009). (See also "Lidingö exemption" below.)

Perhaps the discovery of the high relative cost of using transponders was not so surprising. In Norway, where over 40 different road toll schemes are in operation, transponders are used in some of them, while others are managed by manned tollbooths. And even there, where the comparison technology is highly manual, there is a slight productivity advantage for those *not* using transponders (Odeck, 2008).

Cost consequences: The cost for the radio beacons part of the installation is not separated out in the contract and invoices examined, so their cost can only be estimated by an order of magnitude as a share of the cost for roadside equipment. The cost for transponders is easier to single out in the archival material studied. All in all, 150-200 million SEK were spent on transponders and transponder-related equipment and services.

Lessons learned: For anyone considering a congestion charging system in the future, it should be clear that cameras and ANPR can potentially reach a high ID ratio, and offer ample competition for any transponder-based solution. One remaining advantage of transponders is that they offer other possibilities, such as the ability to use one payment account, represented by the transponder, between different cars.

3.3. Excessive transaction costs

In the same series of *Dagens Nyheter* articles as cited above, two were spent on the issue of the costs for payment transactions, paid via the prime contractor to the suppliers of financial services. In extreme cases, DN (2008e, 2008f) showed that the transaction fee could even exceed the charge itself, making a negative net for the government.

To understand the arrangement of payment transactions, one must first look at a remarkable aspect of the congestion charge. The charge itself was defined as a tax decision, summarizing one whole day of driving. Such tax decisions were made overnight, and were available to the drivers the following morning, after which they had to be paid for within five days. Factoring in a little bit of delay for making bank transactions, this meant that a frequent commuter had to make payments several times per week, to be absolutely sure that the money was available at the Road Administration’s account on the morning of the sixth day; otherwise a sizeable penalty would be applied.

The reason for this cumbersome requirement was, at least in part, that the legislator had made an overly ambitious interpretation of the concept of marginal pricing. The drivers should experience the cost of using the road in direct relation to their decision to do so. And experiencing the cost was interpreted as making the actual payment (Johansson, 2009; Söderholm 2009).

This rule, unique in the context of tax payments, had two negative consequences in terms of costs. The first was simple: making one payment per day instead of, for example, one per month multiplies the number of bank transactions with a factor of about 5^2. Even if the Road Administration obtained a bulk discount in the per-transaction fee from the bank, the total cost of dealing with such a large number of transactions is deemed to be significant.

The second consequence was more elaborate: in order to make the most of the five days available until the payment was due, and at the same time offering a convenient way to pay for occasional travellers, the Road Administration exercised an option in the contract, to be able to pay cash over the counter in retail stores. Two store chains, totalling over 400 outlets, were tied to the scheme, some of which were open 24/7. A new point of sales application was developed, to make it possible for the driver to look up the current tax claims just by stating a licence plate number, and to complete the payment in less than a minute. The store chains gathered the total amount paid during a day and transferred it with an absolute minimum of delay to the Road Administration.

The retail chains thereby gave the congestion charging system two important, positive qualities: it was possible to use almost the full five days to make the payment, and doing so was generally a swift and convenient experience. But for the Road Administration, this speed and convenience did not come for free. Taking up time at the cashier's desk in a busy convenience store means taking time away when something else could have been sold, and the retail chains charged a considerable transaction fee for it. The lowest tax decision was for a day with only one low-traffic passage, equivalent to about €1.00, and the fee charged by the retail chain was slightly above that, creating the bizarre situation where the Government made a negative net profit on any such payments.

Later, after the trial, the five-day period was increased to fourteen, but still with one payment per day's use of the roads, and then again in August 2008 the payment routines were changed to monthly invoices. Thereby, the number of transactions dropped by 80% and are no longer a major cost component (Transportstyrelsen, 2009).

Cost consequences: Bank payment transactions cost 10-20 times more by being charged by the day instead of by the month. DN estimates the transaction costs for the retail chain cost the Road Administration an excessive 50-60 million SEK per year.

Lessons learned: Marginal cost pricing is not the same thing as marginal cost payment. Each payment transaction comes at a cost, both in terms of convenience for the user and as a fee from the financial service provider. Cash-over-counter might be necessary for user acceptance, but it is probably the most expensive form of payment.

3.4. The Lidingö exemption

East of Stockholm lies the island of Lidingö, connected to the mainland by bridge. The only passage between Lidingö and the rest of Sweden crosses the toll cordon twice. Hence, Lidingö's inhabitants would be facing significantly increased driving costs – especially those commuting by car to destinations outside the cordon, who would have to pay the charge twice per single trip. Many considered this to be unfair. Several possible solutions were aired in the debate, ranging from letting them pay anyway, as they are in fact using congested roads, to entirely exempting Lidingö residents from the congestion tax. Finally, the compromise settled upon is to make journeys through the congestion charging area to or from Lidingö free of charge. For this exemption to be valid, the journey must be registered at one of the charging points facing Lidingö and any other charging point within an interval of 30 minutes (Söderholm, 2009; Gullberg and Isaksson, 2009).

In principle, this is an elegant solution, requiring no pre-registration of users eligible for the exemption. All motorists will have equal access to the exemption, no matter whether they are making a visit to Lidingö from the mainland, or *vice versa*. Also, the rule decided upon makes sure that Lidingö residents are still charged when Stockholm is the destination of their journey – two passages, both registered at the charging stations by the Lidingö bridge, does not trigger the exemption. Elegant as it is, this rule turned out to be a significant cost driver in the implementation of the technical system, because it made the effective service level requirements higher than formally stated.

In the contract between the Road Administration and the prime contractor, it is established that the minimum ID ratio is 95% (99.9% for transponder passages) and the maximum error ratio is 0.0001 (i.e. a maximum of one incorrect charge per 100 000 passages) (Road Administration, 2004, pp. 53-54). The design of the Lidingö exemption shortcuts these measurements, as it makes failed identifications turn into overcharging: If a vehicle is driving from Lidingö through the city and out on the other side, it generates two passage records. If both are correctly detected and identified, they should cancel each other out, according to the Lidingö exemption. But if one of them is not identified, the other passage will turn into a liable charge. Thus, to reach the low error ratio target, the ID ratio must also be very high. This may be a major cost driver, since identifying the remaining few per cent often requires considerable manual labour efforts.

The scale of the problem rapidly escalates. If 5 in 100 passages are not identified, as allowed by the ID ratio requirement, only 90.25% of exempted Lidingö trips are correctly identified at both charging stations, 0.25% are captured by neither of the charging stations and that in 9.5% of the cases one of the two passages is identified and a charge is incorrectly generated. There are more than half a million exempted Lidingö trips each month, which would mean more than 50 000 erroneous tax decisions every month. Clearly, the ID ratio could not be allowed to be nearly as low as 95%.

Cost consequences: It is not possible to separate out how much of the system redundancy was caused by this particular rule, but it is generally agreed that this aspect of the Lidingö exemption is one of the key reasons for adding a camera to capture the rear of passing vehicles (in addition to the cameras already taking pictures from the front), and it requires the system to run for longer hours each day. Höök (2009) estimates the increased project cost as “a couple of hundred million SEK”.

Lessons learned: When designing the price mechanism part of a charging scheme, make sure that each passage is priced based on information from that passage alone and other information available at the time, such as time of day and vehicle characteristics. Any pricing scheme using combinations of passages for pricing will immediately be more sensitive to errors and drive up the performance requirements of the system.

Additionally, service level metrics used in a contract benefit from being orthogonal, so that a failure to meet one of them does not automatically spill over to another one.

3.5. Appeal of procurement and project standstill

Public procurement of large-scale projects is strictly regulated by a directive in the Europe Union. Chapter 7 of the Swedish Public Procurement Act (the version relevant at the time) details how a procurement decision can be appealed, if one of the bidding parties experiences losses due to a biased or unprofessional decision (SFS, 1992). When the Road Administration awarded the contract to construct and operate the congestion charging system to IBM, their decision was sent to appeal by the runner-up consortium, dominated by the Austrian group, Kapsch, a long-time system provider in the road tolling business (Gullberg and Isaksson, 2009).

This led to a period of legal processing, where the decision went all the way through each level of the three-tiered public court system. During a period of almost two months, 8 February to 30 March 2005, the contract award decision was inhibited, placing the project in legal limbo – the award was not reversed, but nor was it affirmed. Awaiting a final decision, the project had to be put on hold, although the development work was already well under way. (For a more detailed description of the legal wrangling, see Gullberg and Isaksson, pp. 65-148.)

This put the Road Administration in a dilemma: The supplier’s team of people working on the project had grown large and got up to speed, people involved had got to know each other and what to do. If the team was to be dissolved, its members would be reassigned to other projects, and it would take a long time to start the project up again. So the Road Administration decided to keep hiring the full team, awaiting the final court decision, so that development could recommence exactly where it had left off, once there was a legal go-ahead.

During this period of standstill, there were to be no contacts made between the buyer and the supplier. No project work was to be carried out during this period, at least not on instructions from the Road Administration. But it is likely that a project with a tight time plan, suddenly given a ten-week "holiday", gains other benefits than merely some rest. Even without contacts with the Road Administration, the prime contractor had an opportunity to catch up any delays already accumulated, and to prepare for expected continued efforts.

In addition to the standstill period, the time needed for court proceedings also led to a delay in the planned start date. From its already delayed date of 15 August 2005, the launch date for the system was pushed all the way to 3 January 2006. From a project risk point of view, getting a paid catch-up period of ten weeks, and then getting the final deadline postponed for more than four months, is a gift. Ironically, the appeal filed by the submitter of the runner-up bid came to serve the prime contractor and increase the chances of project success. (See also "Project owner and scope changed" below.)

Cost consequences: Including the work ongoing with subcontractors, more than one hundred people were allocated to the project managed by the prime contractor, keeping them on standby mode for ten weeks, at a direct cost paid to the prime contractor of 140 million SEK. In addition to that, costs increased in the public administration, and the total congestion-charge revenue was lowered as operation time was shortened. All in all, the losses due to delays have been estimated at 600 million SEK (Gullberg and Isaksson, p. 121).

Lessons learned: The sequence of political and legal events that led to congestion charging first being decided upon, which seemed highly unlikely, and then finally happening against all odds, is so specific to the local circumstances and random events, that there are few comparable experiences. One key aspect, however, which seems to have influenced many of the events, is the significance of the election cycle. The entire process, from decision through planning, implementation and trial period to referendum, had to fit inside a four-year election cycle. As the decision to go for congestion charging arose as an outcome of government negotiations immediately following the 2002 election, there was no opportunity to plan ahead and start the new election cycle with a completed plan.

3.6. Excessive service level requirements

In 2002, immediately after the announcement of a congestion charging trial, the City of Stockholm defined their procurement strategy. Dennis Bring was assigned to be in charge of procurement planning. The core principle of the contract was defined as a functional procurement. That meant that the bidders were not asked to build a system according to a detailed design specification, but rather a functional system, where processes were explained at a high level, and their outcomes defined and quantified. It was up to each bidder to suggest what technologies to use, and even where to put the charging stations. An important component in this setup was the list of key performance indicators, originally nine of them, and their target levels. In principle, the bidders had to accept all of them as they stood, in order to be considered compliant in the evaluation. Each indicator was also, albeit somewhat ambiguously, tied to penalty clauses, which could have a severe impact on the eventual revenue to the supplier if targets were not met (Road Administration, 2004, pp. 53-54).

Over time, as a consequence of legal deliberations, the functional design changed from a municipal "environmental charge" to a national "congestion tax". This led to a vast range of changes to the specifications, which rendered some of the key performance indicators no longer applicable. Eventually, four remaining indicators prevailed, governing the ongoing evaluation of the system's performance: an identification ratio (number of vehicles identified divided by number of vehicles passing per month); an error ratio (number of incorrect charges divided by number of passages per month); an availability indicator (lane minutes of uptime divided by total uptime required, times the number of lanes), and a calls-taken ratio (number of calls answered within a particular time limit divided by the total number of incoming calls) (Johansson, 2009).

Each of these performance indicators was assigned numeric targets, and especially the error ratio and system uptime stood out as extraordinarily high (Johansson, 2009). In the case of the error ratio, it was stated that no more than one passage in 100 000 may lead to an incorrect charge, which means that either the automatic number-plate recognition system had to perform far better than any other system existing at the time, or that a large share of the passage photos had to be verified manually, or that transponders (which generally do not generate incorrect identifications) had to be used by almost all vehicles passing the gantries.

Transponder usage eventually reached about 50% of all passages at the time of launch, which was not enough to guarantee that the error ratio target was met, and there were both practical and economical limits to how many clerks could be assigned to read licence plates from photos. Recognising this dilemma, IBM initiated an internal research and development effort to improve the system's performance in terms of automatic licence plate recognition, as mentioned above, eventually leading to meeting and exceeding the requirements, both in terms of ID and error ratio (Höök, 2009; Johansson, 2009).

Total system uptime was measured taking partial availability into account, based on how many lanes were affected by a system outage, so that the metric used "lane minutes" of availability divided by total number of lanes, times the number of minutes of expected uptime. Thereby, the measurement would approximate the revenue lost from the downtime – a 50% drop in availability meant about half of the passages being recorded. The availability ratio calculated like this was to meet or exceed 0.999, or the supplier would be financially penalised.

To meet this high requirement, the prime contractor designed a system where (almost) every component was duplicated, so that a service outage would only occur if two parallel components malfunctioned at the same time. Additionally, spare parts were obtained in large quantities and trained staff were made available to perform on-site service with short notice. For the core IT system, technical support was initially on standby, 24/7. A system design with so much redundancy is obviously more costly than one where a larger degree of failure is accepted, but which needs only one of each component, and where a longer response time can be accepted (Höök, 2009; Johansson, 2009).

It can be reasonably argued that a congestion charging system delivers an equal amount of traffic reduction whether it is operating at 99.9%, 99%, or even 95% availability. After all, travellers decide whether to drive or not based on the fact that they are highly likely to have to pay if they do so. From this perspective, availability requirements could have been relaxed substantially, and thereby system building costs, without losing any of the ultimate effect on the traffic situation.

But there is another side to this availability rather than just lost revenue and behavioural influence. Gunnar Söderholm (2009) is certain that if the system had been anything but perfectly available and functioning from day one, public support would have disappeared quickly, fuelled by media attention. It just had to work perfectly from the beginning, according to him. Höök (2009) agrees that in order to win legitimacy for the system, the accuracy had to be high, and compares with Norway, where a lower degree of accuracy could probably be tolerated, as the systems there are already in place and accepted by the public.

In this respect, the service levels might have been not at all excessive, but just about right. The system was thus not designed only to work, but also to win and retain the trust and support of its users.

Cost consequences: It is not possible to state how much could have been saved if the performance indicators had been relaxed, without making far-reaching hypothetical assumptions about what alternative design decisions would then have been taken. But it is safe to say that a significant portion of the hardware and connectivity cost stemmed from redundancy requirements. It is, for example, likely that the system could have run with half the number of cameras, taking pictures only from the front instead of both from front and back (Johansson, 2009; Höök, 2009).

Lessons learned: Consider what the cost-efficient targets of service levels are, given the goals of the system and how different service levels affect the intended function of the system. Going from, say, 95% to 99% or 99% to 99.9% on any given service level may be a significant cost driver.

3.7. Project owner and scope changed mid-way

System procurement started with a prequalification during the autumn of 2003, followed by a request for a proposal, including a detailed requirements specification issued in November, with bids to be submitted in February 2004. Prospective suppliers were asked to commit to a long and detailed list of functional and non-functional requirements, all in all describing an "environmental charge system" (Swe: miljöavgiftssystem). As the political and legal process occurred, the environmental charge became a congestion tax, and the laws to underpin it were different than what had been assumed when the request for proposal was issued.

When the prime contractor was selected, and only the signing of the contract remained to be carried out in July 2004, the Congestion Tax Law (SFS, 2004a) was passed in parliament. This made clear what most people involved had long understood: that the congestion "charge" was in fact a "tax", from a legal point of view (Söderholm, 2009; Höök, 2009). But it was only now that it began to become clear what this change actually meant in terms of procuring practical changes to the system (Johansson, 2009). Among the new features required by the system was the ability to appeal against any tax decision. It was also clear that the Tax Authority and the Royal Collection Agency had to become administrative users of the system.

In November 2004, the Congestion Tax Ordinance (SFS, 2004b) was issued, with an additional range of requirements added. All in all, more than 200 change requests were handled by the system development project from the time the contract was signed to the time of execution. Having absorbed such a quantity of changes, the system delivered was vastly different from the system initially designed. Since the changes were issued while building, parts of the system were first built according to the original specifications, and then rebuilt according to the changes, which obviously adds to the cost of construction (Johansson, 2009; Höök, 2009).

It is a truism for any information technology project to state that it would be less costly if all was known upfront and no changes made. But it is hardly reasonable to expect this in reality. On the contrary, one could argue that one should expect a project of this magnitude to evolve with time, and make sure that it is planned for in a way that can cope with change. Höök (2009) lists among the success factors the fact that the prime contractor was a single company and not a consortium, since that makes for less internal quarrelling and better chances of dealing quickly with changing circumstances.

In addition to leading to higher costs, the project delays caused by the changes in project ownership and legislation may very well have contributed to the success of the system finally put in place. According to the contract (Road Administration 2004, pp. 16, 37), the prime contractor committed to delivering eleven months after the contract was enacted. Had everything happened according to the initial plan, this would have meant entering live operation in June 2005. Thanks to the legal processes that followed, the final date set to go live was not until 3 January 2006. These additional six months may very well be an important factor standing between successful operation and failure on day one.

Cost consequences: It is impossible to give a definite figure for the costs associated with the total set of changes made. Some changes reduced the scope and made the system simpler, while others made it more complex. All of them, however, interfered with the already established design and plan, and some had cost consequences. A tentative estimate is that 15-25% of implementation costs were derived from changed requirements and managing the consequences of those changes.

Lessons learned: It should have been possible to identify the charge/tax distinction earlier, and to foresee the consequences in terms of changing requirements. Generally speaking, not having the legal setting clear is likely to induce costs in unforeseen ways. For anyone contemplating congestion charging, fitting the project inside one election cycle might be needed. Accepting that this makes it impossible to foresee all requirements that come out of the political process, it is still likely that this type of fundamental legal requirement can be identified and planned for in advance.

4. CONCLUSIONS

All things considered, was the Stockholm congestion charging trial unnecessarily expensive? It did not exceed its budget, and there are other comparable projects that were more costly: still, there are several cost items to indicate that it was. One way of interpreting “unnecessarily expensive” is to ask the following question: “Would it have been possible, knowing what is known today, to establish a system such as the Stockholm congestion charging system, for a lower cost?” The answer is clearly “yes”. IBM even claims that half the cost would be possible, assuming that the specifications were describing a present-day system, and there were no major change orders to deal with during construction time (Johansson, 2009).

Note that this question is not just a philosophical counterfactual, but the relevant question to ask for anyone considering investment in a similar system today. But if the question is interpreted another way: “Given the knowledge and political circumstances at the time, is it reasonable to expect that any other combination of planners, government officers and suppliers would have been able to provide the same or better results to a significantly lower cost?”, then the answer is, at least, not as clear.

The criticism raised against the project, represented in this study by *Dagens Nyheter*, the Automobile Association and the Chamber of Commerce, fail in two important aspects. First, they do not compare the cost of establishing the system to the value it generates, in terms of reduced traffic congestion, only in terms of the revenue collected, which is an irrelevant measurement. Congestion charging is by no means the most effective way to collect tax revenue. Rather, tax revenue is a positive by-product, while congestion relief is the major value generated.

Second, they fail to recognise the extraordinary performance involved in getting a fully functional system in place in time for launch – politically, administratively, commercially, organisationally and technically. Complex IT projects, using new technologies in a previously untried configuration, have a well-earned reputation for failing to deliver on time and within budget, if at all. Listing all the factors that, in hindsight, could have been done better, or at a lower cost, moves the focus away from the remarkable fact that the Stockholm congestion charging system was set up quickly enough and was of sufficient quality to swing public opinion in time for the referendum.

> *The basis for the yes [in the referendum] was that it worked technically, it gave visible and tangible effects that people liked, and one was paying for a just cause. […] it is a necessary but not sufficient condition for a successful trial that it works really well technically.*

Söderholm (2009)

For a full understanding of the project, it is necessary to remember the high risks, real or perceived, at all levels. With the high risks involved, the order from the politicians was not to build the most cost-effective system, but to build one that worked. Gunnar Söderholm summarizes the overall attitude from the political sphere:

> *'It may cost whatever it costs. This shall be executed.' This doesn't mean we were sloppy with public money. Just that function was the over-arching priority.*

Söderholm (2009)

Risk reduction was at the centre of the procurement strategy. The Road Administration used a contracting form, where a single prime contractor is trusted to deliver a turnkey solution. This is the Road Administration's preferred way, in comparison to managing a consortium, in a situation like this, as the risk of delays through conflicts among the consortium members, where one party can take the whole project hostage, is lower (Höök, 2009). This way, the risk that was delegated from the politicians to the Road Administration, was delegated further to the prime contractor, with both the rewards and risks that go with it. Höök was quite clear with IBM about the responsibilities:

> *I told IBM several times: 'It is fully possible that this all goes to hell. But if it does, I will make sure you are going down with me.'*

Höök (2009)

4.1. Insurance comes at a cost

The way the Road Administration and other agencies dealt with risk can thus be comparable to an insurance premium. When a future event risks having a large negative effect if it goes through, it can be rational to pay for insurance, even if the event is unlikely to materialise.

Throughout the execution of the congestion charging project, the prime contractor, the Road Administration and other agencies were faced with unknown factors, where the potential downside was unacceptably large. Their actions can be interpreted as rational insurance policies. Those insurances came in the form of more staff in the call centres, more technical staff on stand-by, more cameras, more spare parts, redundant network connections, and in the form of not pushing every negotiation over a change order to the utmost in order to reach the lowest price.

Since a breakdown in any part of the system was perceived as potentially crucial, such extra layers of safety were added in many places, leading to a high total cost. And like any insurance, it risks looking unnecessary and expensive in hindsight, when the event insured against did not materialise.

However, unlike a typical insurance, which is a sunk cost after the insurance period, some of the "insurances" purchased by the Road Administration have a residual value, available long after the referendum. While the cost for the first months of overstaffing is never recovered, the insurances built for technical redundancies are still there. By building the system more solidly than required, the Road Administration today operates a system with close to zero downtime, and with a unique world share in automatic photo identification. With these benefits still in the system, the operational costs are gradually moving down to more moderate levels as well.

NOTES

1. Parts of what was the Road Administration have been transferred to the newly-formed Swedish Transport Agency. During the congestion charging trial, all of those activities were carried out by the Road Administration.

2. If all drivers were regular drivers, paying the charge each day, the multiplier would be around 22. But a majority of car drivers are in fact "occasional" drivers, who only pay the charge a few times each month. Hence, the multiplier drops to around five.

BIBLIOGRAPHY

Amdal, E., G. Bårdsen, K. Johansen and M. Welde (2007), Operating costs in Norwegian toll companies: a panel data analysis, *Transportation,* 34:681-695.

Chamber (2005), Trängselskatten – vägval för Stockholmsregionen. *Stockholms Handelskammare Report* series, 2005:3.

Chamber (2006), Trängselskatten – en tickande bomb. *Stockholms Handelskammare Report* series, 2006:3.

Cordi, I. (2006), *Stockholmsförsöket – en granskning av utvärderingen av projektet*. Report published by Motormännen, 29 August.

Deutsche Telekom (2008), *Annual Report*. Available at: http://www.annualreport2008.telekom.de/en/konzernabschluss/konzern-anhang/sonstige-angaben/index.php?page=180.

DI (2006), Ingen trängsel hos "Trängselskatteverket", *Dagens Industri*, published 11th September. http://di.se/Nyheter/?page=/Avdelningar/Artikel.aspx%3FArticleID%3D2006\04\26\184942

DN (2005a), Fem dagars frist med trängselskatt oroar. *Dagens Nyheter*. Published on December 9th. Available at http://www.dn.se/sthlm/fem-dagars-frist-med-trangselskatt-oroar-1.421437.

DN (2005b), Vägavgifter, nya spår och vägar är rätt medicin. *Dagens Nyheter*. Published on 12 December. Available at http://www.dn.se/sthlm/vagavgifter-nya-spar-och-vagar-ar-ratt-medicin-1.422464.

DN (2006), Skatteverket: Betala vägskatten i tid!. *Dagens Nyheter*. published 4 January, available at http://www.dn.se/ekonomi/skatteverket-betala-vagskatten-i-tid-1.498409.

DN (2008a), Trängselskatten går back till år 2011. *Dagens Nyheter*. Published on 20 February. Available at http://www.dn.se/ekonomi/trangselskatten-gar-back-till-ar-2011-1.564740.

DN (2008b), Vägverkets hjälptelefon kostade 500 kronor per samtal. *Dagens Nyheter*. Published on 20 February. Available at http://www.dn.se/ekonomi/vagverkets-hjalptelefon-kostade-500-kronor-per-samtal-1.507001.

DN (2008c), Fakturorna var oklara. *Dagens Nyheter*. Published on 21 February. Available at http://www.dn.se/ekonomi/fakturorna-var-oklara-1.655677.

DN (2008d), IBM fick betalt flera ganger av Vägverket. *Dagens Nyheter*. Published 21 February. http://www.dn.se:13025/ekonomi/ibm-fick-betalt-flera-ganger-av-vagverket-1.788339.

DN (2008e), Dyra avtal äter upp vinsten. *Dagens Nyheter*. Published 20 February 2008. Available at http://www.dn.se/ekonomi/dyra-avtal-ater-upp-vinsten-1.680366.

DN (2008f), Ren förlust att driva in en tia i skatt. *Dagens Nyheter*. Published 20 February 2008. Available at http://www.dn.se/ekonomi/ren-forlust-att-driva-in-en-tia-i-skatt-1.758628.

Economist, 2004. Road tolls in Germany - Trucking hell. *The Economist*. Published 19 February http://www.economist.com/businessfinance/displaystory.cfm?story_id=E1_NQVTSRR.

Eliasson, J. (2009), A cost-benefit analysis of the Stockholm congestion charging system. *Transportation Research Part A*, 43, pp. 468-480.

Eliasson, J. (2008), Lessons from the Stockholm congestion charging trial, *Transport Policy*, 15:395-404.

Fjellinjen (2004), Årsrapport [Annual Report] 2004.

Gullberg, A., K. Isaksson (ed.) (2009), *Congestion Taxes in City Traffic - Lessons learnt from the Stockholm Trial*, Nordic Academic Press, Stockholm.

Höök, Birger (2009), Representative of the Road Admninistration. Interviewed live. 23 October, sound recorded, 47 minutes.

Ieromonacho, P., S. Potter, J.P. Warren (2006), Norway’s urban toll rings: Evolving towards congestion charging? *Transport Policy*, 13-5, pp. 367-378.

Johansson, Gunnar (2009), Representative of IBM. Interviewed 23 October via telephone. 52 minutes.

Lissel, L.-O. (2009), E-mail correspondence, December.

Ny Teknik (2005), Leader page published 14 December. http://www.nyteknik.se/asikter/ledare/article38475.ece.

Odeck, J. (2008), How efficient and productive are road toll companies? Evidence from Norway. *Transport Policy*, 15, pp. 232-241.

Oehry, B. (2006), Presentation at the Conference on Road Charging Systems: Technology Choice and Cost Effectiveness, 1 June, Paris.

Prud’homme, R., P. Kopp (2007), Le péage de Stockholm: Evaluation et enseignement, *Transports*, 434, pp. 346-359.

Road Administration (2004), Contract regarding Equipment, system, and services for managing congestion taxes in Stockholm City, Object No. 4509. Diarienr AV50A-2004:9539.

SFS (1992), Svensk Författningssamling, Lag om offentlig upphandling 1992:1528.

SFS (2004a), Svensk Författningssamling, Lag om Trängselskatt, 2004:629.

SFS (2004b), Svensk Författningssamling, Förordning om Trängselskatt, 2004:987.

Stockholm City (2006), Debate page of daily newspaper Stockholm City, 5 September.

SvD (2005), Skatteträngsel hos kronofogden. Svenska Dagbladet. Published on 29 December 2005. http://www.svd.se/opinion/brannpunkt/skattetrangsel-hos-kronofogden_492241.svd.

Söderholm, Gunnar (2009), Representative of Stockholm City. Interviewed live on 27 October, sound recorded. 1 hr 12 minutes.

Transek (2003), Försök med miljöavgifter i Stockholm, Underlag för utformning och genomförandeplan. Consultant report to Stockholm City. Published in April. Accessed on 30/11. http://www.stockholmsforsoket.se/upload/MAK/Bakgrundsdokument/MAKunderlag.pdf

Transportstyrelsen (2009), Statistics web site, accessed on 10 December http://www.transportstyrelsen.se/sv/Vag/Trangselskatt/Statistik-trangselskatt2/

Wieland, B. (2005), The German HGV-Toll, *European Transport*, No. 31, pp. 118-128.

Road Pricing With Complications

Mogens FOSGERAU
DTU Transport, Denmark &
Centre for Transport Studies
Sweden

Kurt VAN DENDER
OECD/ITF Joint Transport Research Centre
Paris
France

SUMMARY

Lyngby and Paris, January 2010

ABSTRACT

Standard textbook analyses of road pricing tend to assume that users are homogeneous, that there is no travel time risk, and to view congestion as static. The simple analysis also ignores that real pricing schemes are only rough approximations to ideal systems and that the general economic context may also have implications for optimal pricing. This paper reviews these issues and discusses how taking them into account may affect estimates of optimal tolls.

1. INTRODUCTION

Road pricing can be used as a tool to reduce demand for travel when and where that is thought to be beneficial. It has some important advantages over other ways of reducing demand. By adding a toll to the cost of a trip, it removes just those trips that travellers themselves think are not worth the toll. So tolling ensures that the least beneficial trips are reduced first, with drivers themselves assessing the benefits of their trips. Tolling allows the individual decisions about whether, when and where to travel to remain decentralised. A toll equal to the total cost of delay imposed on other drivers by one additional car, the marginal external congestion cost, ensures that the socially optimal level of congestion results[1]. Precisely those trips that are worth the full cost are then undertaken. The toll payment is a loss for drivers, but the money does not disappear. The cost to drivers who choose to pay the toll is offset by the gains of those who receive the payment[2]. Overall, the gain from pricing congestion can be considerable.

An obstacle to the introduction of road pricing is the fact that all drivers have to pay in order to deter the least beneficial trips. This transfer from drivers is only fully compensated by reduced travel times in some special circumstances. Drivers as a group will therefore tend to lose when the use of the revenues from road pricing is not taken into account.

The efficient toll is in effect a price on time. It reflects the value of the delays an additional driver imposes on other drivers. To design an efficient toll the first requirement is then to find the value of travel time (VTT)[3]. This leads to the concept of generalised travel cost (GTC) that includes both monetary and time costs. For a certain trip, the GTC may be calculated as the monetary costs of the trip plus the duration of the trip, times the VTT; in symbols $GTC = c + \alpha t$, where c is the monetary trip cost, t is the travel time and α is the VTT.

Armed with an estimate of the VTT, the basic idea behind congestion pricing is simple: a supply curve relates traffic volume to delay. This can be used to compute the increase in travel time resulting from an additional car. Multiplying the increase in travel time by the traffic volume, the demand yields the total delay caused by an additional car. Next, multiplying by the VTT yields the associated marginal external cost of congestion. A demand curve relates traffic volume to GTC. This can be used to find the efficient toll which is equal to the marginal external cost of congestion when the toll is implemented.

The point of this paper is to present an overview of what happens when one opens the door for some of the complications that are present in the real world.

The first complication is that travellers are not identical. On the contrary, they seem to be very heterogeneous. The evidence indicates that the value of travel time in a population can easily range over several orders of magnitude. There is variation in the VTT among individuals. There is even variation for the same individual depending on the exact context of a trip. The paper explores some consequences of this heterogeneity.

The second complication is that travel time is not a deterministic function of traffic flow. Travel times increase on average as flow approaches capacity. Inherent randomness in traffic conditions also gains importance and causes travel times to become more variable and less predictable. The associated cost is significant in comparison to the cost associated with the average travel time.

The third complication is that the relationship between traffic flow and travel time is not simple. There are important dynamics that must be taken into account. It is important to recognise that congestion is time-specific. At any location there is sufficient capacity to allow traffic to flow, if only it were equally distributed over time[4]. Congestion occurs largely because people tend to want to travel at the same time. It is therefore important to recognise that the time of travel is a choice and that congestion is about time-specific preferences: scheduling preferences.

The fourth complication is that actual tolling systems are imperfect approximations to efficient tolls. For example, it may only be possible to toll some links or it may not be possible (or desirable) to achieve the theoretically optimal spatial and temporal differentiation of tolls. Some implications of such limitations are explored.

Finally, tolls cannot be evaluated in isolation from the economy in general. In particular, commuting, a large part of private travel, is complementary to working. Hence the interaction between tolls and the labour market and labour income taxes must be considered.

Section 2 reviews the standard textbook analysis of tolls. Section 3 considers some implications of the fact that travellers are heterogeneous. Section 4 discusses issues that result since congestion also causes travel times to become variable and unpredictable for travellers. Section 5 discusses the need for a sophisticated understanding of the dynamics of the transport system. Section 6 discusses second-best issues related to constraints on the design of road pricing schemes and interactions with the wider economy. Section 7 concludes.

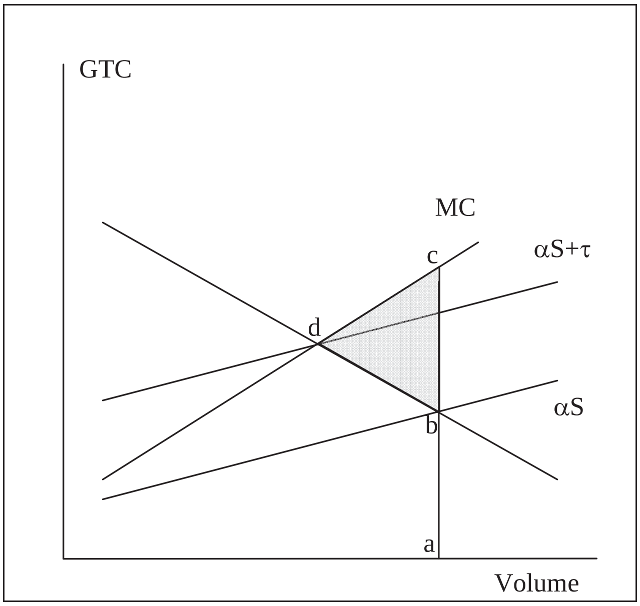
2. THE STANDARD TEXTBOOK ANALYSIS

From an economist's perspective, the core problem of congestion is not that there is congestion but that there is too much of it. It is not congestion itself that is the problem, but the associated externality. Each driver takes only his own costs into account when deciding whether to travel. The individual will choose to drive when his benefit of doing so outweighs his cost of travelling such that he receives a personal net gain. However, he does not take into account the delay he imposes on others. When that delay is large then the trip entails a net loss. There is thus a potential gain that can be found if traffic can be reduced in such a way that only those travel whose personal gain from travelling outweighs the delay imposed on others.

2.1. A static model

The standard analysis under simplifying assumptions is presented in the following figure. The horizontal axis is the traffic volume; the vertical axis is the GTC. The curve αS is the supply curve. The supply S is the travel time as a function of traffic volume; this is multiplied by the VTT α to convert travel time to monetary cost[5]. The supply curve is increasing to reflect a situation with congestion. It will generally be convex. All travellers experience the same travel time and hence the supply curve can also be understood as an average cost curve. The curve MC indicates the marginal cost. It is the change in total cost arising from an additional traveller. When the supply curve is increasing, the marginal cost curve will always be increasing, have a larger slope and lie above the supply curve.

Figure 1**. A static model**



The demand curve is decreasing to reflect that demand decreases in price. The market equilibrium occurs at the intersection of the demand curve D with the supply curve αS, at point b. The marginal traveller at this point faces a cost corresponding to the line segment a-b and a corresponding benefit of the same size. The group of travellers as a whole, however, have a cost associated with the marginal traveller given by the MC curve. For the marginal traveller at point b, this cost corresponds to the line segment a-c. So the last traveller imposes a net loss corresponding to the line segment b-c on the group of all travellers. The loss is zero at point d where the MC curve crosses the demand curve. The total loss is then represented by the shaded triangle b-c-d on the figure. The optimal toll implements the equilibrium at point d.

Expressed symbolically, the marginal external cost of congestion (mecc) is $mecc = S'(D) \cdot D \cdot \alpha$. This arises since the total cost, given realised demand *D*, is $TC(D) = S(D) \cdot D \cdot \alpha$, the marginal total cost is $TC'(D) = S'(D) \cdot D \cdot \alpha + S(D) \cdot \alpha$, and the mecc is the difference between the marginal cost and the average cost.

We shall refer to this analysis as the “simple static analysis”. It is useful as a basis for discussing more complicated models that take more aspects of reality into account.

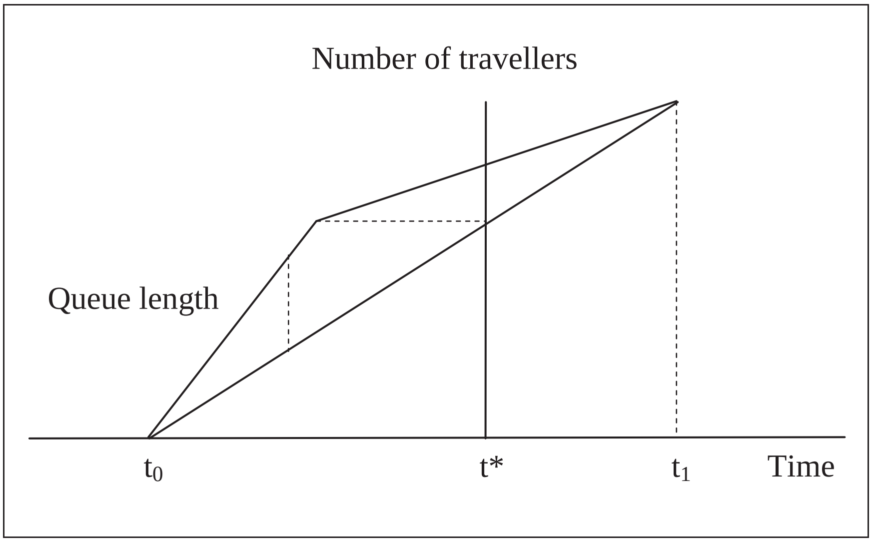
2.2. A dynamic model – the bottleneck model

The static analysis ignores the trip timing aspect of travel demand. Time is regarded merely as a resource of which a traveller can consume more or less, whereas in reality time at 8 a.m. is different from time at 9 a.m. As has been noted, congestion arises because travellers prefer to travel around the same times. If they could be spread evenly over time, there would be no congestion. Moreover, congestion is inherently dynamic, since adding a vehicle to a queue at some instant will affect the evolution of the queue until it is gone. These aspects are captured in the bottleneck model [1, 36], which describes the time dimension more explicitly. This model generates a number of important insights regarding the pricing of congestion. A main insight is that large efficiency gains may be found through the effect of pricing on trip timing.

In the simplest incarnation of the bottleneck model, travellers are viewed as homogeneous with VTT α. They have a common preferred arrival time t^* and prefer not to be early or late at the destination relative to this time. The cost of earliness is usually denoted by β and the cost of lateness by γ and the generalised travel cost is then $GTC\,(a,t) = \alpha \cdot (t-a) + \beta \cdot \min\left(0, t^* - t\right) + \gamma \cdot \max\left(0, t - t^*\right)$. This specification of scheduling preferences was introduced by Vickrey [36] and later by Small [28] in a different context[6].

A total of D travellers have to pass through a bottleneck in order to reach their destination. The bottleneck has a limited capacity of ψ users per minute. Assume for simplicity that travel time is zero before and after the bottleneck. Then the first traveller, departing at time t_0 also arrives at the destination at this time, since there is no queue yet. Denote the cumulative arrival rate at the bottleneck by *R*. Then $R(t_0) = 0$. Let the time of the last departure be t_1, such that $R(t_1) = D$.

The model assumes an equilibrium where no traveller can reduce his cost by changing departure time. This implies first that there is always a queue during the interval $[t_0, t_1]$ and, second, that the queue is gone precisely at time t_1, such that $\psi \cdot (t_1 - t_0) = D$. Third, equilibrium requires that the *GTC* is constant during the interval where travellers depart, and higher outside. A queue builds up immediately as the first traveller departs, since travellers initially depart at a higher rate than capacity. The queue has maximum length at the departure time when a traveller would be at the destination exactly at the preferred time *t**. From that point, the departure rate is lower than capacity such that the queue gradually dissipates and the queue is exactly gone at the time of the last departure. This is illustrated in Figure 2:



The first and last travellers experience no queue in this model. The first traveller is early at the destination while the last traveller is late. They incur the same *GTC* in equilibrium, which implies that $-\beta \cdot (t^* - t_0) = \gamma \cdot (t_1 - t^*)$. This fixes the interval $[t_0, t_1]$ such that the equilibrium travel cost can be computed. The total travel cost in equilibrium for all travellers is given by:

$TC = \frac{\beta\gamma}{\beta+\gamma}\frac{N^2}{\psi}$, such that the marginal external congestion cost is $mecc = \frac{\beta\gamma}{\beta+\gamma}\frac{N}{\psi}$.

This is the marginal external congestion cost associated with the addition of a marginal user to the equilibrium. It increases in the number of users and decreases in capacity.

We may regard the number of travellers D as being a function of the equilibrium generalised travel cost. Then, connecting the bottleneck model with the simple static analysis, we would find that the optimal static toll is equal to the mecc. This toll would not remove congestion. The number of travellers would be reduced by the toll, but there would still be a queue during $[t_0, t_1]$.

This static toll does not vary over time. It is the same for all travellers, regardless of when they arrive at the bottleneck. It is hence only optimal if the toll must be constant over the whole day. It is possible to do better by letting the toll vary over time. The optimal time-varying toll is zero at time t_0. It increases until the preferred arrival time t^*, then it decreases again, until it is zero at time t_1. The average toll is equal to the optimal static toll. The optimal time-varying toll does remove congestion completely, since it ensures that travellers arrive at the bottleneck exactly at the rate ψ, which is the bottleneck capacity.

3. HETEROGENEOUS TRAVELLERS

Where the simple analysis assumes that users are heterogeneous, it is clear that in reality also users are heterogeneous. This section explores first the measurement of heterogeneous value of travel time, and second the consequences of heterogeneity of the VTT for the optimal congestion charge.

3.1. Measurement of heterogeneous VTT

How can the VTT of travellers be inferred? If a traveller has a choice between two options for making a trip, where one is faster but more expensive than the other, then the traveller faces a trade-off between money and time. Assume that travellers have a generalised trip cost expressed as $GTC = C + \alpha \cdot T$, where C is the monetary cost of the trip, T is the travel time and α is an individual specific VTT. The difference in GTC between two trip options for a traveller with VTT equal to α is $\Delta GTC = \Delta C + \alpha \cdot \Delta T$. Holding everything else constant, travellers with $\alpha < -\Delta C/\Delta T$ will choose the slow option while travellers with $\alpha > -\Delta C/\Delta T$ will choose the fast option. The trade-off thus entails an implicit price of travel time, namely $v = -\Delta C/\Delta T$. Through his choice, the traveller reveals whether his VTT is larger or smaller than the trade-off price, i.e. whether $\alpha < v$ or $\alpha > v$.

Label by Φ the cumulative distribution function describing the distribution of the VTT among users. Observing many travellers at a trade-off price v allows assessment of the share of travellers with $\alpha < v$. This share is the cumulative distribution evaluated at the point v, i.e. $\Phi(v)$. Observing travellers in different choice situations with different v allows assessment of Φ over the range where v varies. To assess Φ completely, it is necessary to observe choice shares for values of v ranging from a point where travel time is cheap and all travellers choose the fast and expensive option to a point where time is expensive and all travellers choose the cheap and slow option[7].

In general, revealed preference data are preferable by virtue of relating to real choices where travellers actually feel the consequences of their choices. Suitable revealed preference data could come from situations where travellers face a choice between a slow, cheap route and a fast, expensive one. A notable example is routes combining free-access, regular freeway lanes with tolled express lanes [31]. However, it is often difficult to achieve the necessary variation in the price of time needed to reveal the distribution of the VTT. This is a reason for relying on stated preference data, where travellers make choices between hypothetical options. It is possible to construct stated preference choice situations to meet many of the demands of econometric modelling. Stated preference data are, however, tainted by doubt as to whether they represent actual behaviour well enough.

Figure 3. **Confidence band for the cumulative distribution of VTT based on Stated Preference (SP) data**

The unit is Danish Crowns (DKK) per hour: 1 EUR ≈ 7.5 DKK

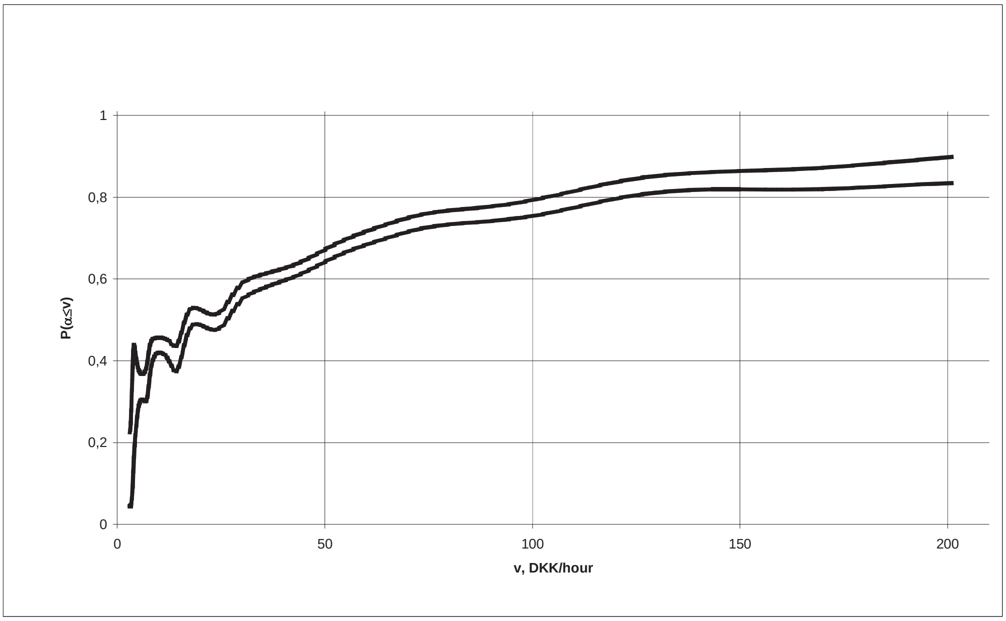


Figure 3 shows an estimate of a VTT distribution obtained from stated preference data [12][8]. The shape is broadly typical of many studies. It shows a right-skewed distribution with many travellers having low VTT and few travellers having large VTT. The median VTT is about 25 DKK/hour while the mean is considerably larger.

This estimate of the cumulative distribution of the VTT does not show the maximum of the VTT distribution. The largest trade-off price of time that was offered to respondents in the SP exercise was about 200 DKK/hour. A significant share of respondents, about 15%, indicated that they were willing to pay this amount per hour of travel time saved and hence that their VTT was higher than 200 DKK/hour. How much larger is impossible to say, based on the figure. It is not possible to calculate the mean VTT without information about the right tail of the VTT distribution. Such information can be found using data that allows the right tail to be observed. Otherwise, it is necessary to resort to more restrictive assumptions which may be hard to justify. One popular approach is to assume a specific form for the VTT distribution. The resulting estimate of the mean VTT is extremely sensitive with respect to such an assumption [12].

It is clear that there are enormous variations in the value of travel time, with several orders of magnitude from low to high. The VTT depends on observable factors. It is generally found to increase with income but the size of the income elasticity is debated. The VTT is generally thought to vary substantially between individuals but also with individuals, depending on the context. In general, a large part of the variation in VTT remains after controlling for observable factors. For example, Fosgerau [12] controls for gender, income, trip duration, time difference between alternatives, share of delay time due to congestion in travel time, age and trip purpose, and finds that the remaining variation in VTT has more than a factor of 50 between the 20% and 80% quantiles.

3.2. Road pricing with heterogeneous travellers

Consider now a situation where the population of potential travellers differs in their VTT. Groups of travellers are indexed by their value of travel time α. The group with VTT α has a demand function D(p(α)| α), where p(α) = τ + α t is the generalised cost for travellers α. The aggregated demand function is then D = E(D(α)), the average over groups of travellers. The average VTT in the population is E(α). This is not the same as the average VTT of those who actually travel, which is the weighted average $\overline{\alpha} = E[\alpha \cdot D(\alpha)]/E[D(\alpha)]$. If travel time as a function of demand is $t = S(D)$, then the change in travel time resulting from the marginal traveller is S'(D). Multiplying this by the number of travellers and by the average VTT among travellers indicates the marginal external cost of congestion. That is, $mecc = S'(D) \cdot D \cdot \overline{\alpha}$.

This shows that the difference from the case of homogeneous travellers is the VTT used to compute the marginal external cost of congestion. There is just a single VTT in the case of homogeneous travellers. With heterogeneous travellers, this single VTT is replaced by the average VTT in the group of actual travellers. Hence the *mecc* depends in general on the toll, since the introduction of a toll will change the composition of travellers and this affects the average VTT of road users.

This effect may be large. Consider the case where the value of travel time for a population follows a standard log normal distribution. Then the mean VTT is E(α) ≈ 1.65. Imagine now a toll that causes a reduction in traffic of 10% and that it is those travellers with VTT above the 10% quantile who remain. Their average VTT is then about $\overline{\alpha}$ ≈ 1.81, which represents an increase of about 10%.

A toll will discourage some from travelling. If travellers with low VTT are discouraged more, as might be expected, then the average VTT $\overline{\alpha}$ increases with the introduction of a toll. Then the optimal toll will be larger when travellers are heterogeneous than when they are homogeneous[9].

We are not aware of empirical evidence concerning the likely size of this effect. It is reasonable to suppose that the effect is noticeable, since the distribution of VTT is generally thought to have a shape similar to that presented in Figure 3, with many travellers having a relatively low VTT.

Consider still the situation with heterogeneous travellers and imagine that the optimal toll $\tau = mecc = S'(D) \cdot D \cdot \overline{\alpha}$ is in operation. Imagine then that road capacity is split in two halves and that travellers have to choose which half of the capacity they want to use. They will then divide equally among both halves of capacity. If the toll is increased slightly for one half of the capacity, then demand will decrease slightly there and shift to the other half or stop for both. Then the first half will be faster but more expensive than the other. This will cause rational travellers to sort into those with their VTT above some threshold who will use the first half and those with VTT below the threshold who use the other. As a consequence, the average VTT is high for the first half and low for the second. Then the toll can be raised for the first half and reduced for the second to produce a net welfare gain.

Verhoef and Small [35] consider differentiated tolls in a static network with serial and parallel links and with heterogeneous users. They are particularly concerned with second-best policies whereby only a part of the network is tolled. Such policies, they find, are in danger of losing much of their potential effectiveness if heterogeneity is ignored when setting toll levels. Ignoring heterogeneity in VTT may cause the welfare benefits of second-best policies to be drastically underestimated.

A real-world counterpart to such examples is value pricing as implemented in various places in the US. In these schemes, drivers can choose between lanes that are free but congested or tolled but less congested. Clearly, as shown by Small and Yan [29], heterogeneity is important for assessing the welfare impacts of such policies, as drivers with high VTT may be expected to choose the tolled lanes.

Pricing is not the only means by which to increase the efficiency with which existing road capacity is used. A recent paper [9] shows that there are unexpected benefits from car pool lanes that do not have to do with pricing. The benefit arises because the car pool lanes reduce disruptive vehicle lane changing. Even a severely underused carpool lane can in some instances increase a freeway bottleneck's total discharge flow[10].

In a dynamic setting there can be benefits from differentiation, even with homogeneous travellers and without pricing. Ongoing research (Fosgerau and de Palma) uses the bottleneck model to analyse a situation in which travellers are arbitrarily divided into groups. The immediate cause of congestion in the bottleneck model is that travellers initially depart at a rate higher than capacity. Congestion is reduced by a toll that makes travellers decrease the initial departure rate. This effect may be induced in other ways. One way is to divide travellers into, say, two groups. A more than proportional share of capacity is allocated to the first group. The second group can use the remaining capacity. The second group can also use the share of capacity allocated to the first group when the first group is not using it. The first group would find a new equilibrium where departures occur during a shorter interval of time and would hence experience a cost reduction relative to the situation without grouping. The cost of the second group would not increase since it is determined by the length of the interval during which departures take place and this is unchanged relative to the situation without grouping.

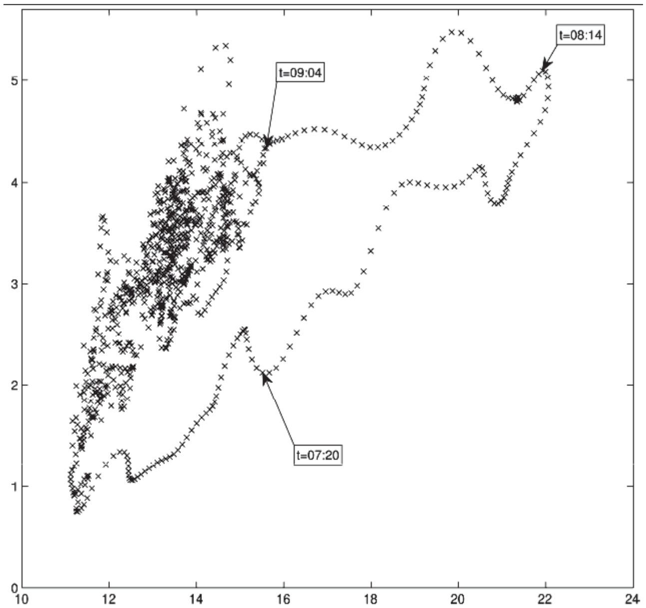
4. TRAVEL TIME RISK

Increasing travel demand leads to congestion and increasing travel times. As demand approaches capacity, travel times also become increasingly variable and unpredictable for users. This travel time variability (*TTV*) may add significantly to the generalised travel cost.

Figure 4 shows a scatter plot of the standard deviation of travel time against the mean travel time for a congested urban road, with a distinct morning peak. Each point on the plot corresponds to a time of day. The numbers have been computed using data covering a period of three months. Both the mean and standard deviations are small in the morning. They increase and then decrease over the peak. The standard deviation peaks later than the mean, indicating that there is no constant relationship between them. This creates the loop that is evident in the figure. It is a characteristic pattern that has been observed many times[11]. An implication is that the mean and standard deviations of travel time must be accounted for separately.

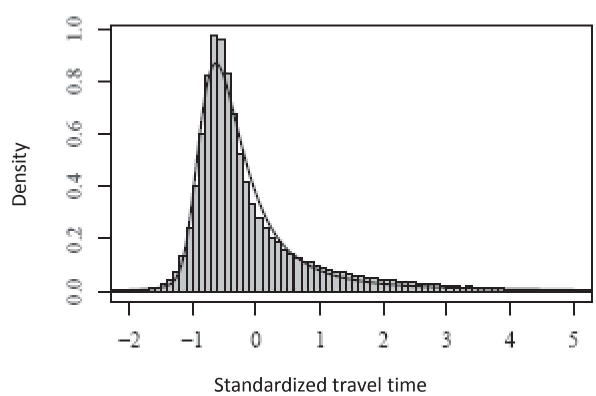
The definition of GTC must be extended in order to account for the cost associated with travel time variability. Some authors have simply added a term to the definition of GTC to reflect the increase in cost due to TTV; for example, using $GTC = c + \tau + \alpha T + \eta X$, where X is some measure of TTV. Various measures of TTV have been employed, such as the standard deviation or the variance of travel time, or a range between two quantiles [31]. Studies have then proceeded to estimate a value η of X based on revealed or stated preference data.

Figure 4. **Scatter plot of the standard deviation of travel time (minutes) against the mean travel time (minutes) for a congested urban road**



This is, however, not a completely satisfactory solution without some arguments to indicate why one measure of TTV should be preferred to another. The problem is complicated, since a travel time distribution is a shape rather than a number. Figure 5 shows an example of a travel time distribution. There are (infinitely) many possible such shapes and they cannot be described completely by a few numbers.

Figure 1. **An empirical travel time distribution**



At this point, it can be noted that all measures of scale are essentially equivalent when the shape of the travel time distribution is constant. In this case, the standard deviation is proportional to the range between any two given quantiles. A change from one scale measure to another is then reflected in an inverse change in the value of TTV given by η.

Intuitively, the cost associated with travel time variability is related to scheduling considerations. Compare two situations, one in which travel time is variable and one in which it is constant. The mean travel time is the same in both situations. Travellers have to decide when to embark on a certain trip. When travel time is constant, travellers choose an optimal time of departure which is directly associated with an optimal time of arrival at the destination. Faced with TTV, travellers may embark on the trip earlier than they would have otherwise done. On average, they will therefore arrive earlier than otherwise. Sometimes they arrive later than they would have chosen with constant travel time.

To make this more formal, economic theory generally assumes that travellers have preferences that encompass scheduling considerations, regarding when they depart from the trip's origin and when they arrive at destination. Travellers are pictured as knowing the travel time distribution and choosing the departure time optimally. A specification of scheduling preferences then leads to a relationship between the travel time distribution and the generalised travel cost. This relationship is not in general tractable, and there is generally no obvious candidate for defining a measure of TTV. There are, however, a few special cases, where simplifying assumptions enable a simple measure of TTV to be defined.

Fosgerau and Karlstrom [15] consider the departure time choice of a traveller facing travel time variability. The traveller cares about travel time and about being early or late at the destination, according to the Vickrey/Small scheduling preferences described above. The distribution of random travel is independent of the departure time, such that a change in departure time does not affect the shape of the travel time distribution but only shifts it earlier or later. Similarly, the monetary trip cost does not depend on the departure time.

When the traveller knows the travel time distribution and chooses the optimal departure time, it turns out that the expected GTC becomes linear in the mean and the scale of the travel time distribution, regardless of what that travel time distribution is[12]. More specifically, when the traveller chooses the optimal departure time, then $GTC = c + \alpha \cdot E(T) + \eta \cdot \sigma \cdot H$. In this expression, c is the monetary cost of the trip and $\alpha \cdot E(T)$ is the VTT multiplied by the mean travel time. The last term captures the effect of TTV: η is the value associated with TTV and is given in terms of scheduling parameters β and γ, σ is a measure of the scale of the travel time distribution and H depends on scheduling parameters and on the shape, rather than the scale, of the travel time distribution.

Fosgerau & Karlstrom provide an example of a congested urban road in Copenhagen, where the cost of travel time variability comprises about 15% of the time cost to travellers. Including TTV in the GTC is likely also to lead to an increased estimate of the mecc. This is because TTV tends to increase with demand just as does the mean travel time.

There is thus a basis for including a measure of the scale of the distribution of travel time as a measure of TTV. Given estimates of the scheduling parameters and the shape of the travel time distribution it is possible to calculate the contribution of TTV to the GTC. It is not necessary to know the preferred arrival time of travellers since this does not appear in the GTC when the departure time is optimally chosen.

While the Fosgerau-Karlstrom result has some advantages for application, there are also some drawbacks. First, the value of TTV depends on the shape of the travel time distribution. It may therefore vary across different contexts. It still remains to gather sufficient empirical evidence to be able to judge whether this is a serious drawback in practice, given that there are many other uncertainties and approximations in play.

Second, the result does not apply to scheduled services. The issue is that users of scheduled services are, by definition, not able to choose their departure time freely. Therefore they are not able to choose the optimal departure time, as required by the Fosgerau-Karlstrom model.

Third, the scale of the travel time distribution can be hard to compute in networks comprising many links. This is not an issue with the mean travel time, since the mean travel time may be computed at the link level and then summed over links to obtain a trip-level, mean travel time. The standard deviation is not additive in this way and so the GTC cannot just be computed at the link level and summed.

In a broader perspective, it is important that the specification of scheduling preferences is consistent with empirical evidence. The Vickrey/Small specification of scheduling preferences entails the prediction for an individual traveller that an isolated increase in travel time will cause a proportional change in departure time that leaves the arrival time unchanged. An isolated increase in the standard deviation of travel time would lead to earlier departure and earlier arrival on average. This may or may not be an adequate description of actual behaviour.

There is an alternative formulation of scheduling preferences that also leads to a tractable expression for the value of TTV. It is based on a less known paper by Vickrey [37], in which he defines scheduling preferences in terms of time-varying utility rates at the trip origin and destination. The traveller receives utility at some rate specific to the trip origin until he departs. When he arrives, he begins receiving utility at a rate specific to the destination. The cost of the trip is an opportunity cost associated with the foregone utility at the origin or at the destination. When the utility rate at the origin is decreasing and the utility rate at the destination is increasing, then there is a time at which the individual would optimally travel from the origin to the destination. This view of scheduling preferences is attractive, since it treats the origin and the destination of the trip symmetrically. In general, it is hard to argue why timing at one trip-end should be more important than at the other, as implied by Vickrey/Small scheduling preferences.

Using a simplified version of these scheduling preferences, Fosgerau [14] finds an expression for the value of travel time variability that does not depend on the shape of the travel time distribution. The related measure of travel time variability is the variance of travel time. This applies equally to travellers who can freely choose departure times and to travellers who use a scheduled service with fixed headway. Depending on parameters, travellers may be risk-averse or risk-seeking, and the value of travel time may increase or decrease in the mean travel time.

This model has some advantages over the Fosgerau and Karlstrom model. First, the value of TTV does not depend on the shape of the travel time distribution. Second, the result applies equally to scheduled services. Third, the variance of travel time is additive over links in a network, provided travel times on links are independent[13].

Ultimately, the choice between formulations of scheduling preferences and the associated measures and value of travel time variability should not be based on convenience but on conformity with observable behaviour.

Randomness is a lack of information; so information provision is a natural policy measure in the context of TTV. Consider a situation in which travel time is variable from day to day but perfect information about tomorrow's conditions is provided to travellers. Then every day they can choose the optimal departure time, and the GTC in the Fosgerau-Karlstrom model reduces to $GTC = c + \alpha E(T)$, which omits the term relating to TTV.

The information does not have to be perfect in order to reduce the GTC. In general, it can be just a signal that contains some information about tomorrow's travel time, i.e. it must have some dependence on tomorrow's travel time[14]. This reduces the risk that travellers face. The value of this information may be assessed with the same models that are used to assess the cost associated with random travel time variability.

In using these results, it is important to keep in mind that no consideration has been given to equilibrium. The departure time choice of a single traveller is considered, taking the choices of all other travellers as given. The random distribution of travel time affects each traveller's choice of departure time. But there is also a causal relationship in the other direction, whereby the combined departure time choices of travellers affect the distribution of travel time [3].

5. MEASURING AND MODELLING SUPPLY

The basis for efficient congestion pricing is the marginal external cost of congestion. It involves, essentially, the VTT and the supply relationship. So, it is clearly crucial for the design of congestion pricing schemes to have an adequate understanding of the supply side. The description of supply relationships for road travel has traditionally been considered the domain of engineering or physics, and economics has tended to ignore the complexities involved. This might have been reasonable in times when the main issues were to do with the design and capacity of road networks. When it comes to the design of road pricing schemes it becomes necessary to have a deeper understanding of the supply side. In particular, it is essential to be able to estimate the effect on travel times of changing demand. Perhaps economics should get involved in this. Small and Verhoef [30] discuss congestion from this point of view.

In economic models, supply is generally taken as given in the form of a supply curve for a road or a simple network; in the case of the bottleneck model, it is simply the bottleneck capacity, which is a single number. In reality, congestion is a hugely complicated phenomenon. Research into the relationship between travel demand and travel time has been active for at least 75 years and we shall make no pretence of being able to summarize the state of the field. However, it is important to point out some of the issues involved. There are still very many open questions that appear when we ask what is the consequence for mean travel time and variability of adding an extra traveller.

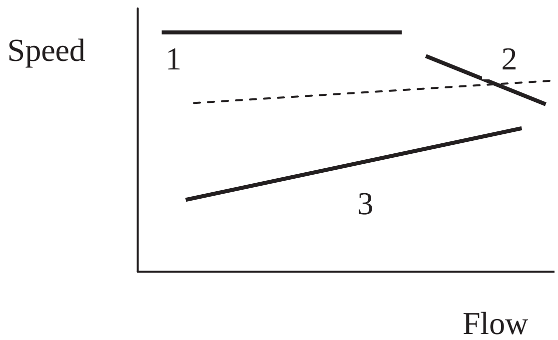
There are many causes and corresponding types of congestion. For example, flow congestion arises as traffic slows down due to increased density, independently of upstream or downstream links. Flow congestion is related to the microdynamics of traffic; it may arise due to small random fluctuations in flow and may be involved in the phenomenon of hypercongestion. The importance of flow congestion is debated, however, with some arguing that congestion is more likely to be related to bottlenecks such as intersections and merge lanes and to accidents that create temporary bottlenecks. Congestion involves spill-backs such that delays on a link may be due to upstream delays. A particularly clear instance of spill-backs is when a queue behind a bottleneck blocks upstream intersections for crossing traffic. Delay for the crossing traffic is then unrelated to that demand.

Consider now that the objective often is to price urban networks, comprising thousands of links and nodes, and even more combinations of origins and destinations. It is clearly a daunting task to try to describe such systems in detail. Here rescue may come from the existence of urban-scale volume-delay relationships [17] that allow the complexities of the network to be compressed into a single expression.

In general, it is not sufficient to consider only the mean travel time, when travel time variability accounts for a significant share of GTC. It is also necessary to be able to predict the impact of TTV on changes in demand. Generally, there is no simple relationship whereby TTV can be expressed as a function of mean travel time. This is evident in Figure 4, where the standard deviation of travel time is rising at times when the mean travel time is decreasing.

The example of the loop in Figure 4 also indicates the dynamic nature of congestion, whereby even a small event at one time and place may have significant effects later and elsewhere in the system. Such dynamic phenomena also create problems for the empirical measurement of speed-flow relationships.

Figure 6. **A stylised speed-flow relationship**



The fundamental diagram of traffic flow incorporates a relationship between speed and traffic flow on a road link, depicted as a backward-bending curve, as shown in Figure 6. It depicts a situation with three regimes. First, a free-flow regime in which speed is about constant, with flows ranging from zero to a certain point. Second, a congested regime in which higher flows are associated with lower speeds. Third, a hypercongested regime, in which speeds are lower than in the other regimes at the same flow levels and where higher flows are associated with higher speeds.

Consider a scatter of observations from a speed-flow curve. Then consider a regression of speed against flow. If there are many observations from the hypercongested regime, then it can happen that an increasing mean relationship seems to be present. This would imply the perverse prediction that increasing flow would lead to an increase in speed. The problem is that there is causality in two directions. The causal effect of interest is the effect of flow on speed. However, there is also a causal effect in the opposite direction, whereby low speed creates a blockage which causes flow to be low. The problem of reverse causality is a classical econometric problem and a range of econometric techniques exists to tackle it. In the present context, it is important to realise that the measurement and modelling of supply should be taken seriously and requires sophisticated methodology.

6. SECOND-BEST ISSUES

The key idea of congestion pricing is that it reduces waste by alleviating the misalignment of the private and social costs of travel, caused by the congestion externality. The previous sections have pointed out how hard it is to establish the magnitude of this misalignment, making an abstraction of the broader context in which pricing might be introduced and assuming that sophisticated instruments for charging are available. Second-best analysis asks the broad question of how the basic analysis is modified when these simplifications are abandoned. The literature on the subject is vast and we make no attempt at providing an overview, referring the interested reader to a concise discussion in Small and Verhoef [30], pp. 137-147. Instead, we limit ourselves to discussing some examples of second-best reasoning, and draw conclusions on how second-best analysis can help improve the practical implementation of congestion charging systems. Section 6.1 investigates the consequences of the fact that practical systems are approximations to the ideal charging system, and section 6.2 asks what are the consequences of the fact that congestion charging – even if potentially ideal – is implemented in an imperfect economy. Section 6.3 presents some guidelines for practical analysis.

6.1. Imperfect implementation

If the second-best issues discussed below in section 6.2 are ignored, then an ideal congestion charging system is one that charges the marginal external congestion cost at each time and place in the road network. A glance at existing and planned systems shows this ideal is not reached: charges are levied on only a very limited number of network segments and those charges differ from the marginal external cost. If the first-best welfare gain from congestion charging is taken as a benchmark (which, on the basis of the arguments of section 6.2, may be a disputable choice), this means lower welfare gains but not necessarily negative gains: congestion charging is still likely to be worthwhile, although the risk of counterproductive systems does exist and the loss of efficiency compared to ideal systems can be large.

The problem of which links to charge and what tolls to set, when only some links in a network can be charged, has a conceptual solution that is difficult to translate into a practicable one (e.g. Verhoef [34]). Simulations using detailed network models suggest that reasonably performant pricing schemes can be designed, even with a small number of tolled links or cordons, e.g. by choosing high volume and high-speed links with poor substitutes [27]. While this is in line with common sense, no links may fit the bill perfectly, so that choices can be hard in practice. Furthermore, the question of how much to charge remains unresolved. Also, when the choice is where to place a cordon instead of what link to charge, deciding where to place one or more cordons and what charge to levy appears to be particularly challenging, with the results from simulation work sometimes differing from common sense judgment [32].

In sum, it seems reasonable to conclude that imperfect implementation is unavoidable but that, nevertheless, good results can be obtained. However, systematic assessments of where and how much to charge can improve considerably on common-sense judgment or at least help avoid big mistakes. Detailed analysis using traffic models is likely to have considerable payoffs.

6.2. Distortions in the wider economy

The first-best analysis focuses on the congestion externality in transport. It implicitly assumes that there are no distortions (deviations from efficiency) in the rest of the economy, or at least none that should be taken into account when thinking about charging the external cost of congestion. In reality, the economy is rife with distortions that potentially do matter. One example, which has attracted considerable research attention because of its quantitative importance, is that labour income is taxed. A tax on labour income is a distortion because it reduces the supply of labour, resulting in a less than efficient level of employment. How might this distortion affect prescriptions for a congestion charge?

Consider the simplest possible case, where all road users are commuters and all commuters are road users: In order to work, it is necessary to commute. Say that the marginal external congestion cost per round-trip (or, equivalently, per work-day) is $20, whereas the tax on labour income per work-day is $50. In this situation, the labour tax and the congestion charge have the same effect since they both affect the same choice margin, whether to work or not. Since the labour tax more than covers external costs, no congestion charge is justified based on the argument of untaxed externalities. This polar case highlights the importance of interactions with other distortions.

Next, consider the same proportionality between labour supply and travel volumes, but introduce a second transport mode, "transit". Assume that this mode is free of congestion and other external effects, and it is priced at marginal cost. A congestion charge for car commuting now is potentially useful, as it helps attain the socially desirable modal split. In fact, setting the congestion charge equal to the marginal external congestion cost guarantees that commuters face the marginal social cost of each mode, so the optimal modal split will result. At least, this is the case when the revenues from the charge are used to reduce labour tax revenues by the same amount, through a modification of the labour tax rate. For if revenues were used differently, the effective labour tax – and with it the labour market distortion – would increase. This example shows that revenue use is of crucial importance in evaluating the effects of congestion charging.

These examples are extreme: in reality, there is no strict proportionality between labour supply and commuting, the labour market is not perfectly competitive, not all transport users are commuters, and commuters are a heterogeneous group. In general, this means that the strong results of the examples will play out weaker in reality. Some more specific insights follow.

De Borger [11] investigates the consequences of replacing the assumption of a competitive labour market by one where unions and employers bargain about wage levels. Arguably, this is a closer approximation of prevailing labour market conditions in some countries or some segments of the labour market, and it is seen to matter for the impact of the interaction of labour market distortions and the congestion externality. As long as congestion tolls are not differentiated between those who are and are not working, it makes sense to raise the toll above marginal external congestion costs as this shifts part of the tax burden to non-workers. Given the large distortion from labour taxes at the margin, such a shift is efficiency-improving. However, when differentiation of tolls between workers and non-workers is feasible, the toll for workers is best set below marginal external congestion costs and that for non-workers exceeds it.

The importance of distinguishing between work and non-work trips is also highlighted by Van Dender [33], who extends the Parry and Bento [20] analysis by considering two trip types that are strict complements to labour supply and to leisure respectively, and by assuming that transit and cars share a congested road. One result is that introducing a congestion toll can make sense even if the labour tax does not change, on the condition that the toll can be differentiated between both trip types. But if no differentiation is possible and the labour tax cannot be changed, then the scope for efficiency-improving tolls is quite limited (in a model where no possibilities for welfare-improving revenue use exist apart from reducing labour taxes). These analyses suggest that differentiation of congestion charges can be very useful not only because road users differ (as highlighted in previous sections), but because of the efficiency costs of distortionary taxation. In particular, it makes sense to charge less for commuting trips.

Introducing heterogeneity into the analysis forces us to abandon a strict focus on efficiency. If individuals differ, we need to be explicit on how we care about the distribution of welfare among them. In second-best, the strict separation between efficiency and equity cannot be maintained because there are no lump-sum taxes suitable for supporting that separation. This tells us that the equity impacts of congestion charges and the use of revenues need to be considered.

However, one should not forget that the labour income tax is there for more than one reason; a notable motivation is a desire for redistribution. Kaplow [22] argues this point of view and suggests that cost-benefit tests should not be corrected with the marginal cost of funds. In the context of road pricing, Kaplow might similarly argue that income taxation should not be counted as distortionary.

There are several important market imperfections, other than labour income taxes, that interact with congestion charges. First, search-unemployment occurs because it takes time for separated workers to match with new employment. The search duration depends on the number of job openings within reach of the unemployed, which in turn is affected by transport costs. Thus, increasing transport costs can lead to increased duration of unemployment spells and a lower employment rate. This effect is likely to be significant [26]. Agglomeration effects constitute another relevant externality to take into account when thinking about congestion charges. There is substantial evidence that production in cities is subject to agglomeration economies, i.e. economies of scale that are external to firms. This is relevant to the analysis of congestion, as workers contributing to and suffering from congestion also contribute to the positive externality of agglomeration. If all workers contribute equally to congestion and agglomeration, this suggests congestion tolls should be reduced by the value of the agglomeration externality, unless a separate instrument is available to stimulate agglomeration. But if workers differ in their values of time and their effects on agglomeration [as is the case according to Graham (2009)], tolls should differ among workers [19]. Of course, agglomeration effects are far more complex than suggested in these simple models and our understanding of them is limited [18]. In addition, policies to improve the benefits from agglomeration can be separate from congestion management policies. Nevertheless, the warning that congestion should not be considered in isolation from the productivity of a city's economy is valid.

6.3. Implication for implementation

One response to the reality that several potentially important market imperfections interact with the congestion externality, is to construct a model that encompasses the main imperfections (as judged by who builds the model) and derive a rule for the assessment of charges in this context. For example, Calthrop *et al.* [8] propose a rule for transport infrastructure investments, which could be modified for transport pricing, which includes many of the interactions discussed above. Their framework emphasizes the role of distortions, which comes at the cost of a strongly simplified representation of the transport markets. Fosgerau and Pilegaard [16] take the opposite route, showing how some general equilibrium interactions related to tax distortions can be integrated into cost-benefit analyses based on traffic models. This has the advantage of allowing a detailed model of the transport market (relying on traffic models that are often used in the practice of transport project appraisal), but the range of general equilibrium interactions is more narrow.

The use of such sophisticated rules is sometimes thought to be superior to the simple first-best rule, on the grounds that the latter is easily shown to imply large mistakes in some cases, i.e. it is not a very good approximation. However, given our imperfect understanding of the broader context in a conceptual sense, and even more in an empirical sense, it is not obvious that the approximation of the sophisticated rule is necessarily better. The fact that analyses of theoretical rules are typically illustrated by highly stylised numerical illustrations suggests that not too much faith should be put in the practical relevance of the numbers, but rather that they should be considered as rough indications of orders of magnitude associated with the mechanisms contained in the theory. This view is similar to the one expressed by Böhringer *et al.* [7], who discuss second-best analysis of climate change policies:

"The above results should be treated with caution. The numbers are neither accurate nor precise. They are ballpark estimates. What really matters are the insights. Climate policy need not cost a lot, but imperfect implementation implies excess costs. The excess costs are substantial relative to the costs of the first-best policy."

In addition, even the most comprehensive models on general equilibrium interactions are (by definition) highly stylised representations of reality. They highlight distortions thought to be of particular importance (with judgment ideally based on evidence) while ignoring others, and they miss features that matter in the applied analysis of proposals for congestion charging. For example, few models contain a sufficiently detailed representation of the capacity and usage of multimodal transport systems that would allow a comparison of revenue use to reduce labour taxes or to improve the supply of public transport. Clearly, such comparison would be relevant to the design of charging proposals. Parry and Small [25] present a detailed analysis of rationales for subsidizing public transport. They do not focus on the interaction with broader distortions, but do suggest that their impact on the optimal subsidy is limited.

Our view, then, is that the models generate insights that ought to be part of debates on the implementation of congestion charging. Applications of simple models can help clarify the importance of second-best concerns, but we should not expect a fully-fledged general equilibrium analysis to be carried out.

7. CONCLUSION

Unregulated congestion entails an efficiency loss and a corresponding possibility for obtaining a welfare gain. This gain can be achieved through road pricing, decentralising the decision about who should travel when and where. In the first step of the analysis, the toll should equal the value of the delay that a marginal car imposes on other travellers. The size of the total delay associated with the marginal car is determined from traffic models, ranging from simple supply curves to complex traffic models. The transformation to monetary value is accomplished through the value of travel time (VTT), which can be measured in various ways.

Congestion arises because people tend to travel at the same times. With an even distribution of traffic over time, there would be no congestion. As the example of the bottleneck model shows, there are potentially large benefits to be achieved if the trip timing aspect of demand is taken into account by varying tolls over peaks, thus inducing travellers to distribute departure times more evenly.

The simplifications involved in textbook analyses of congestion pricing allow the central insights to be easily communicated. There are, however, a number of complications that must be taken into account when this theory is put in practice. First, people are different. The stylised facts state that many people have a low value of travel time but some have a very high VTT. The mean VTT is larger than the median. There is much variation between people but also between seemingly identical individuals and even within the same individual in different contexts. Recognition of this heterogeneity will tend to lead to higher suggested tolls and will tend to reveal larger benefits from price differentiation between roads.

Travel times are inherently random. As congestion increases, travel times become not only longer but also increasingly variable and unpredictable. This travel time variability contributes significantly to travel costs. Taking travel time variability into account will generally lead to higher suggested tolls. A theory exists whereby the value of travel time variability can be expressed in terms of the standard deviation or the variance of travel time, in a simple and readily applicable way.

Traffic systems are hugely complicated and the complexities of measuring and modelling supply should not be underestimated. It is difficult to establish the size of the delay associated with a marginal vehicle and even more difficult to establish the consequences for travel time variability. The simple descriptions of the supply side often employed in economic papers are not adequate for a real-world assessment of road pricing systems.

Constraints on charging instruments lead practical systems to fall short of ideal (in a first-best or second-best sense) congestion charging, and this implies lower benefits than would be obtained in that ideal. Charging can, of course, still produce net benefits (after subtraction of investment and operational costs). Careful assessment of charging systems becomes crucial, however, because of the multitude of design options available and the large differences among them in terms of benefits produced. Such assessment should tackle the second-best aspects explicitly, and not evaluate a system as if it were first-best. Changes in travel speeds are not a sufficient indicator of the benefits of charging.

Congestion charging should not be considered in isolation, as there are economic interactions that have large potential effects on the benefits of charging. As indicated, we are sceptical about the possibility of capturing these interactions in one elaborate model, but do think it valuable to discuss explicitly how any charging system might interact with other taxes and other externalities, with the aid of models and data, however sketchy and imperfect they are.

Tolling may be suggested as part of a policy package that also specifies how revenues are to be used. However, such earmarking of revenues runs the risk of being wasteful. The justifiability of a package that combines tolling with earmarking depends on the way revenues are to be used. An otherwise justifiable tolling scheme may become unjustifiable in combination with a use of revenues that is unjustifiable. It is therefore helpful to evaluate the parts of such policy packages in isolation in order to make clear the merits of each component.

One specific recommendation that emerges is that charges for commuting should be lower than for non-commuting. Since a trip purpose cannot be observed at the time charges are levied, a correction through the income tax system could be envisaged.

The second-best analysis suggests that in many cases the best tolls are lower than the marginal external cost of congestion, because tolls match external costs imperfectly and they exacerbate pre-existing distortions. These, however, are partial results, depending to some extent on the interactions considered. If, for example, congestion tolls could be designed to fall mainly on non- or lower-taxed activities ("leisure") and the revenues used to reduce highly distortionary taxes, then this would be an argument for raising them above the marginal external cost of congestion. What matters then is to figure out which activities would be affected by the introduction of a charge.

In summary, it seems reasonable to use the basic analysis of congestion charging as a first approximation when considering its implementation, but to check whether important interactions with other market imperfections can be expected, and revise the analysis when there are concerns about large, indirect effects. Roy (2008) goes a step further, arguing for reliance on the simple rule and focuses on removing second-best obstacles rather than taking them as given. We are less optimistic about the possibility of removing at least some of the obstacles. Nevertheless, our view and Roy's are not mutually exclusive if charging mechanisms are seen as malleable through time.

NOTES

1. This is true in a first-best world, where all other prices equal marginal social cost.

2. This is also true in a first-best world.

3. Becker [6] related the value of time to the wage rate by considering the allocation of time between work and leisure. This was later extended in various ways. Jara-Diaz [21] explains the history of the development of this theory.

4. The alternative would require queues that keep growing indefinitely.

5. Other monetary travel costs are ignored.

6. [38] develops the bottleneck model under more general scheduling preferences, while allowing for random queue sorting.

7. This may sound easier than it is. Fosgerau [12] discusses issues related to the identification of the distribution of VTT. De Borger and Fosgerau [10] discuss an extension to take behavioural anomalies into account.

8. The estimate is computed using a non-parametric technique, which does not impose the restriction that the cumulative distribution should be increasing. It is therefore evidence of the internal validity of the stated preference (SP) data that an increasing function does result.

9. Arnott, de Palma and Lindsey [2] discuss pricing with heterogeneous travellers in the context of the bottleneck model. Arnott and Kraus [4] discuss marginal cost pricing when travellers are heterogeneous but the differences are not observed.

10. A theoretical investigation of these issues is undertaken in [23].

11. This pattern is generated by the random capacity bottleneck model for any distribution of capacity [13].

12. Previously, this had only been shown for some special cases [5, 24].

13. Travel times are not likely to be independent since delays on different links may have common causes. Still, additivity must be considered an improvement over no additivity.

14. There are cases where equilibrium effects imply that imperfect information is not necessarily welfare-improving [3].

REFERENCES

[1] Arnott, R.A., A. de Palma and R. Lindsey (1993), A structural model of peak-period congestion: A traffic bottleneck with elastic demand, *American Economic Review*, 83, 161-179.

[2] Arnott, R.A., A. de Palma and R. Lindsey (1994), The Welfare Effects of Congestion Tolls with Heterogeneous Commuters, *Journal of Transport Economics and Policy*, 28, 139-161.

[3] Arnott, R.A., A. de Palma and R. Lindsey (1999), Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand, *European Economic Review*, 43, 525-548.

[4] Arnott, R. and M. Kraus (1998), When are anonymous congestion charges consistent with marginal cost pricing?, *Journal of Public Economics*, 67, 45-64.

[5] Bates, J., J. Polak, P. Jones and A. Cook (2001), The valuation of reliability for personal travel, *Transportation Research Part E: Logistics and Transportation Review*, 37, 191-229.

[6] Becker, G.S. (1965), A theory of the allocation of time, *Economic Journal*, 75, 493-517.

[7] Böhringer, C., T.F. Rutherford and R.S.J. Tol (2009), The EU 20/20/2020 targets. an overview of the EMF22 assessment.

[8] Calthrop, E., B. De Borger and S. Proost (2009), Cost-Benefit Analysis of transport investments in distorted economies, *Transportation Research Part B*, forthcoming.

[9] Cassidy, M.J., K. Jang and C.F. Daganzo (2009), The smoothing effect of carpool lanes on freeway bottlenecks, *Transportation Research Part A*, forthcoming.

[10] De Borger, B., M. Fosgerau (2008), The trade-off between money and time: a test of the theory of reference-dependent preferences, *Journal of Urban Economics*, 64, 101-115.

[11] De Borger, B. (2009), Commuting, congestion tolls and the structure of the labour market: optimal congestion pricing in a wage bargaining model, *Regional Science and Urban Economics*, 39, 434-448.

[12] Fosgerau, M. (2006), Investigating the distribution of the value of travel time savings, *Transportation Research Part B: Methodological 40*, 688-707.

[13] Fosgerau, M. (2009), On the relation between the mean and variance of delay in dynamic queues with random capacity and demand, *Journal of Economic Dynamics and Control*, forthcoming.

[14] Fosgerau (2009), The value of travel time variance, Working Paper.

[15] Fosgerau, M. and A. Karlstrom (2009), The value of reliability, *Transportation Research Part B: Methodological*, in press.

[16] Fosgerau, M. and N. Pilegaard (2007), Cost-benefit rules for transport projects when labour supply is endogenous and taxes are distortionary, Munich Personal RePEc Archive 3902.

[17] Geroliminis, N. and C.F. Daganzo (2008), Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological* , 42, 759-770.

[18] Graham, D. and K. Van Dender (2009), Estimating the agglomeration benefits of transport investments: some tests for stability, JTRC Discussion Paper 32.

[19] Graham, D. and K. Van Dender (2009), Pricing congestion with heterogeneous agglomeration externalities and workers, Working Paper.

[20] Parry, Ian W.H. and Antonio Bento (2001), Revenue Recycling and the Welfare Effects of Road Pricing, *Scandinavian Journal of Economics*, 103, 645-671.

[21] Jara-Diaz, S.R. (2007), *Transport Economic Theory*, Elsevier, Amsterdam.

[22] Kaplow, L. (1996), The Optimal Supply of Public Goods and the Distortionary Cost of Taxation, *National Tax Journal*, 49, 513-533.

[23] Menendez, M., C.F. Daganzo (2007), Effects of HOV lanes on freeway bottlenecks, *Transportation Research Part B*, 41, 809-822.

[24] Noland, R.B. and K.A. Small (1995), Travel-Time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes, *Transportation Research Record*, 1493, 150-158.

[25] Parry, I. and K.A. Small (2009), Should urban transit subsidies be reduced?, *American Economic Review*, 99, 700-724.

[26] Pilegaard N. and M. Fosgerau (2008), Cost-benefit analysis of a transport improvement in the case of search-unemployment, *Journal of Transport Economics and Policy*, 42, 23-42.

[27] Safirova, E., K. Gillingham, I. Parry, P. Nelson, W. Harrington and D. Mason (2004), Welfare and distributional effects of road pricing schemes for metropolitan Washington DC, in: Georgina Santos (ed.), *Road pricing: theory and evidence*, Elsevier JAI, Amsterdam, pp. 179-208.

[28] Small, K. (1982), The scheduling of Consumer Activities: Work Trips, *American Economic Review*, 72, 467-479.

[29] Small, K.A. and J. Yan (2001), The Value of "Value Pricing" of Roads: Second-Best Pricing and Product Differentiation, *Journal of Urban Economics*, 49, 310-336.

[30] Small, K.A. and E.T. Verhoef (2007), *Urban transportation economics*, Routledge, London and New York.

[31] Small, K.A., C. Winston and J. Yan (2005), Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability, *Econometrica*, 73, 1367-1382.

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

[33] Van Dender, K. (2003), Transport taxes with multiple trip purposes, *Scandinavian Journal of Economics*, 105, 295-310.

[34] Verhoef, E.T. (2002), Second-best congestion pricing in general networks: heuristic algorithms for finding second-best optimal toll levels and toll points, *Transportation Research Part B*, 36, 707-729.

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

[36] Vickrey, W.S. (1969), Congestion theory and transport investment, *American Economic Review*, 59, 251-261.

[37] Vickrey, W.S. (1973), Pricing, metering, and efficiently using urban transportation facilities, *Highway Research Record*, 476, 36-48.

[38] de Palma, A. and M. Fosgerau (2009), Random queues and risk-averse users, Working Paper.

LIST OF PARTICIPANTS

Mr. Patrick DeCORLA SOUZA **Chairman**
Tolling and Pricing Program Manager
Federal Highway Administration
1200 New Jersey Ave SE, Room 64-302
WASHINGTON D.C. 20590
UNITED STATES

Dr. Kian-Keong CHIN **Rapporteur**
Chief Engineer, Transportation
Land Transport Authority
Road Operations & Community Partnership Group
No. 1 Hampshire Road
Block 9 Level 1
219428 SINGAPORE
MALAISIA

Prof. Jonas ELIASSON **Co-rapporteur**
Professor Transport systems analysis
Director, Centre for Transport Studies
Royal Institute of Technology
Teknikringen 78 B
SE-100 44 STOCKHOLM
SWEDEN

Mr. Carl HAMILTON **Co-rapporteur**
PhD Student
Centre for Transport Studies (KTH)
Royal Institute of Technology
Teknikringen 78 B
SE-100 44 STOCKHOLM
SWEDEN

Mr. Mogens FOSGERAU **Co-rapporteur**
Danish Transport Research Institute
Knuth-Winterfeldts Allé
Bygning 116 Vest
DK-2800 KGS. LYNGBY
DENMARK

Dr. Kurt VAN DENDER **Co-rapporteur**
Chief Economist
OECD/ITF Joint Transport Research Centre
2 rue André-Pascal
F-75775 PARIS CEDEX 16
FRANCE

Mr. Bernhard OEHRY **Rapporteur**
Head of Traffic Telematics Department
Rapp Trans AG
Hochstrasse 100
CH-4018 BASEL
SWITZERLAND

Mr. Nils-Axel BRAATHEN
Principal Administrator
Environment and Economy Integration Division
Environmental Policy Instruments
OECD
2 rue André-Pascal
F-75775 PARIS CEDEX 16
FRANCE

Mr. Ian CATLING
Partner
Ian Catling Consultancy
Ash Meadow, Bridge Way
GB-CHIPSTEAD CR5 3PX
UNITED KINGDOM

Mr. Francis CHEUNG
Senior Advisor/Economist
Centre for Transport and Navigation – Rijkswaterstaat
Netherlands Ministry of Transport
Van der Burghweg 1
P.O. Box 5044
NL-2600 GA DELFT
NETHERLANDS

Prof. Harry CLARKE
Professor of Economics
Department of Economics and Finance
La Trobe University
MELBOURNE 3086
AUSTRALIA

Mr. Alberto COMPTE
Director of the Transport Research Centre
CEDEX
Autovia de Colmenar Viejo Km 18,2
28760 EL GOLOSO (Madrid)
SPAIN

Mr. Ramon CRUZ
Climate Policy Specialist
Institute for Transportation and Development Policy
127 W. 26th St., Ste. 1002
NEW YORK, NY, 10001
UNITED STATES

Dr. Vadim DONCHENKO **Observateur**
Director General
Adviser to Vice Transport Minister
Scientific and Research Institute of Motor Transport (NIIAT)
Geroev Panfilovcev str., 24
125480 MOSCOW
RUSSIAN FEDERATION

Dr. Daniel GRAHAM
Centre for Transport Studies
Department of Civil Engineering
Imperial College London
GB-LONDON SW7 2BU
UNITED KINGDOM

Dr. Timothy D. HAU
Associate Professor
School of Economics and Finance
Office: KK 905
The University of Hong Kong
Pokfulam Road
HONG KONG

Mr. Stewart JONES
General Manager
Road Transport Policy Reform
Department of Infrastructure, Transport,
Regional Development and Local Government
111 Alinga Street
CANBERRA, ACT 2600
AUSTRALIA

Mr. Charles KOMANOFF
Director
KEA
11 Hanover Square
15th Floor
NYC 10005
UNITED STATES

Mr. Tomasz KOZLUK **Observer**
Economist
Economics Department
Country Studies 6
OECD
2 rue André-Pascal
75775 PARIS Cédex 16
FRANCE

Dr. Adrian MOORE
Vice President
Reason Foundation
3415 S. Sepulveda Blvd Suite 400
LOS ANGELES, CA 661/477-3107
UNITED STATES

Prof. James ODECK
Professor of Transport Economics
Department of Civil and Transport Engineering
Norwegian University of Science and Technology
Po Box 7491
N-TRONDHEIM
NORWAY

Mr. Jan PERSSON
Economics Department
OECD
2 rue André-Pascal
75775 PARIS CÉDEX 16
FRANCE

Mr. Dae Ho SONG
Economics Department
OECD
2 rue André-Pascal
75775 PARIS CÉDEX 16
FRANCE

Dr. Charles RAUX
Directeur
Laboratoire d'Économie des Transports (LET)
Université de Lyon
14 Avenue Berthelot
F-69363 LYON CEDEX 07
FRANCE

Dr. Sergi SAURI-MARCHAN
Project Manager
Center for Innovation in Transportation (CENIT-UPC)
c/ Jordi Girona, 29, 2A (Nexus II Building)
08034 BARCELONA
SPAIN

Dr. Jens SCHADE
Technische Universität Dresden
Faculty of Transportation Science
Traffic Psychology Unit
D-01062 DRESDEN
GERMANY

Dr. Simon SHEPHERD
Principal Research Fellow
Institute for Transport Studies (ITS)
University of Leeds
LEEDS LS2 9JT
Yorkshire
UNITED KINGDOM

Prof. Donald SHOUP
Department of Urban Planning
University of California, Los Angeles
LOS ANGELES, CA 90095-1656
UNITED STATES

Mrs. Maria Teresa Da Costa SOUSA
Expert
International Relations Department
Institute for Mobility and Land Transport (IMTT)
Av. das Forças Armadas, 40
P- LISBOA 1649-022
PORTUGAL

Mr. Tuomo SUVANTO
Senior Adviser
Ministry of Transport and Communications
FINLAND

Dr. Terje TRETVIK
Senior Scientist
SINTEF Technology and Society
Transport Research
NO-7465 TRONDHEIM
NORWAY

Prof. José M VIEGAS
Professor of Transportation
Instituto Superior Técnico – Tech. University of Lisbon
Av. Rovisco Pais
1049 – 001 LISBOA
PORTUGAL

Mrs. Monique van WORTEL
Senior Policy Advisor
Ministry of Transport, Public Works and Water Management
DG Mobility
P.O. Box 20901
2500 EX THE HAGUE
NETHERLANDS

INTERNATIONAL TRANSPORT FORRUM SECRETARIAT

Mr. Jack SHORT
Secretary General

JOINT TRANSPORT RESEARCH CENTRE:

Mr. Stephen PERKINS
Head of Centre

Dr. Kurt VAN DENDER
Chief Economist

Dr. Michel VIOLLAND
Administrator

Dr. Jari KAUPPILA
Administrator

Ms. Françoise ROULLET
Assistant

Mrs. Julie PAILLIEZ
Assistant

ALSO AVAILABLE

Oil Dependence: Is Transport Running out at Affordable Fuel? Series ITF – Round Table 139 (2008)
(74 2008 03 1 P) ISBN 978-92-82-10121-6

The wider Economic Benefits of Transport: Macro-, Meso- and Micro-Economic Transport Planning and Investment Tools. Series ITF – Round Table 140 (2008)
(74 2008 04 1 P) ISBN 978-92-821-0160-5

17th International Symposium on Transport Economics and Policy: Benefiting from Globalisation – Transport Sector Contribution and Policy Challenges (2008)
(74 2008 01 1 P) ISBN 978-92-821-0168-1

Privatisation and Regulation of Urban Transit Systems. Series ITF – Round Table 141 (2008)
(74 2008 06 1 P) ISBN 978-92-821-0199-5

The Cost and Effectiveness of Policies to Reduce Vehicle Emissions. Series ITF – Round Table 142 (2008)
(74 2009 01 1 P) ISBN 978-92-821-0212-1

Port Competition and Hinterland Connections. Series ITF – Round Table 143 (2009)
(74 2009 02 1 P) ISBN 978-92-821-0224-4

Terrorism and International Transport: Towards Risk-based Security Policy. Series ITF – Round Table 144 (2009)
(74 2009 03 1 P) ISBN 978-92-821-0231-2

Competitive Interaction between Airports, Airlines and High-Speed Rail. Series ITF – Round Table 145 (2009)
(74 2009 05 1 P) ISBN 978-92-821-0245-9

Integration and Competition between Transport and Logistics Businesses. Series ITF – Round Table 146 (2010)
(74 2010 01 1 P) ISBN 978-92-821-0259-6

18th International Symposium on Transport Economics and Policy – The Future for Interurban Passenger Transport – Bringing Citizens Closer Together (2010)
(74 2010 02 1 P) ISBN 978-92-821-0265-7

To register for information by email about new OECD publications: www.oecd.org/OECDdirect
For orders on line: www.oecd.org/bookshop
For further information about ITF: www.internationaltransportforum.org

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(74 2010 04 1 P) ISSN 2074-3378 – No. 57461 2010

147

IMPLEMENTING CONGESTION CHARGES

Recent advances in the scientific understanding of urban traffic congestion have only strengthened the already solid case for congestion charges as an element of a successful urban transport policy. But examples of real-world congestion charging systems remain few and far between. What can be done to improve the chances of their more widespread adoption in practice? This report draws lessons from attempts to introduce congestion charges.

Technology is not an obstacle, and technologies should serve policy purposes instead of define them. Charging systems are not cheap and thus should only be used where congestion is severe. Public acceptance is seen to be the key to successful implementation. Although environmental benefits and careful deployment of toll revenues may improve acceptance, a charging system should never lose sight of its principal aim, which is to reduce congestion.

www.internationaltransportforum.org

OECD*publishing*
www.oecd.org/publishing

(74 2010 04 1 P1) ISBN 978-92-821-0284-8

9 789282 102848