OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ECMT

EUROPEAN CONFERENCE OF MINISTERS OF TRANSPORT

TRANSPORT RESEARCH CENTRE

TRANSPORT INFRASTRUCTURE CHARGES AND CAPACITY CHOICE

SELF-FINANCING ROAD MAINTENANCE AND CONSTRUCTION

ROUND TABLE

135

TRANSPORT RESEARCH CENTRE

REPORT OF THE
ONE HUNDRED AND THIRTY FIFTH ROUND TABLE
ON TRANSPORT ECONOMICS

on the following topic:

TRANSPORT INFRASTRUCTURE CHARGES AND CAPACITY CHOICE

SELF-FINANCING ROAD MAINTENANCE AND CONSTRUCTION

OECD

ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT

ECMT

EUROPEAN CONFERENCE
OF MINISTERS OF TRANSPORT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies 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, 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, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities 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.

EUROPEAN CONFERENCE OF MINISTERS OF TRANSPORT (ECMT)

The European Conference of Ministers of Transport (ECMT) is an inter-governmental organisation established by a Protocol signed in Brussels on 17 October 1953. It comprises the Ministers of Transport of 44 full Member countries: Albania, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FRY Macedonia, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and the United Kingdom. There are seven Associate member countries (Australia, Canada, Japan, Korea, Mexico, New Zealand and the United States) and one Observer country (Morocco).

The ECMT is a forum in which Ministers responsible for transport, and more specifically the inland transport, can co-operate on policy. Within this forum, Ministers can openly discuss current problems and agree upon joint approaches aimed at improving the use and ensuring the rational development of European transport systems.

At present, ECMT has a dual role. On one hand it helps to create an integrated transport system throughout the enlarged Europe that is economically efficient and meets environmental and safety standards. In order to achieve this, ECMT assists in building bridges between the European Union and the rest of the European continent at a political level. On the other hand, ECMT also develops reflections on long-term trends in the transport sector and, more specifically, studies the implications of globalisation on transport.

In January 2004, the ECMT and the Organisation for Economic Co-operation and Development (OECD) brought together their transport research capabilities by establishing the Joint Transport Research Centre. The Centre conducts co-operative research programmes that address all modes of inland transport and their intermodal linkages to support policy-making throughout Member countries.

Ministers at their Dublin Council in May 2006 agreed a major reform of ECMT designed to transform the organisation into a more global body covering all modes of transport. This new international transport Forum will aim to attract greater attention to transport policy issues, and will hold one major annual event involving Ministers and key sectoral actors on themes of strategic importance. 2007 is a transitional year for the setting up of the Forum. The new structure will be fully operational as of 2008.

Also available in French under the title:

Tarification des infrastructures de transport et dimensionnement de la capacité

L'AUTOFINANCEMENT DE L'ENTRETIEN ET DE LA CONSTRUCTION DES ROUTES

Further information about the ECMT is available on Internet at the following address:

www.cemt.org

2, rue André-Pascal, 75775 PARIS CEDEX 16, France

TABLE OF CONTENTS

THE ROLE OF ROAD FUNDS IN IMPROVING MAINTENANCE

Kenneth GWILLIAM
Economic Adviser, Transport
Florida
UNITED STATES

SUMMARY

Florida, June 2005

1. INTRODUCTION

1.1. Objectives of the report

This report is presented as a contribution to a broader discussion of infrastructure charges and infrastructure capacity choice, particularly concentrating on charging policy in a fee-for-service perspective. The terms of reference for this contribution were to present some views on charging and investment in the context of the experience of the international financial institutions in trying to set up infrastructure funds. In interpreting this brief some attention is paid to the theoretical structure and incentive effects of various tax and quasi price revenue sources, but the main concerns of the paper are with the institutional arrangements for revenue collection and disbursement.

1.2. Structure of the report

Chapter 2 identifies the perceived problems of developing countries in road finance and discusses the essential institutional and policy characteristics which have contributed to those problems. Chapter 3 summarises the evidence of the attempts of early road funds (hereafter RFs) to ensure secure funding for roads, and presents the macroeconomic arguments against the traditional tax hypothecation approach. Chapter 4 describes the development and nature of the "second-generation road funds" (hereafter SGRFs), presents the theoretical responses of the primary international macroeconomic agencies to these new developments and identifies the criteria which have been suggested for their appraisal. Chapter 5 discusses some available evidence on the development and performance of the SGRFs and presents some preliminary conclusions about the institutional requirements for effective performance. Finally, Chapter 6 summarizes the conclusions on road funds in developing countries and attempts to assess the relevance of that experience for the industrialised countries.

2. THE PERCEIVED PROBLEM FOR DEVELOPING COUNTRIES

2.1. The scale of the maintenance problem

Road maintenance is seen as the problem of highest priority in many developing countries. In Africa it was estimated that during the two decades of the 70s and 80s, road stock of a value of $45 billion was lost due to inadequate maintenance, which could have been averted by expenditures on preventive maintenance of only $12 billion (Brushett, 2005). Similarly, in Latin America and the Caribbean, with road networks valued at over US$350 billion, it is reported that more than US$30 billion are being wasted annually due to the absence of adequate road maintenance, and individual countries in the region are losing between 1% and 3% of their annual GNP due to an unnecessary increase in vehicle operating costs and loss of road asset value alone (Zietlow, 2004). The extent of underspending is exemplified by World Bank comparisons of actual and recommended road maintenance expenditures under existing arrangements in a number of countries (See Table 1).

Table 1. **Road maintenance expenditure for selected countries**

million $US/year, current

Year	Country	Network	Spent (S)	Recommended (R)	S/R Ratio	"Recommended" means:
1990	Bangladesh	R&H Dept	24.6	42.4	58%	Maintain existing network on a sustainable basis (constant flow of services) in quantity terms
1988-91	Honduras	All	11.0	45.0	24%	Id.
1992	Nepal	All	2.6	18.1	15%	Id.
1988	Nigeria	Fed. Highways	112.3	248.5	45%	Eliminate backlog in five years
1992	Zambia	Roads Dept	6.1	32.7	19%	Maintain existing network on a sustainable basis

Sources:
Bangladesh, Second Road Rehabilitation and Maintenance Project, SAR, 1993.
Honduras, Transport Rehabilitation Project, SAR, 1992.
Nepal, Road Maintenance and Rehabilitation Project, SAR, 1994.
Nigeria, Road Sector Strategy Paper, The World Bank, 1991.
Zambia, Financial Performance of the Government-Owned Transport Sector, The World Bank, 1992.
Road User Charges Model, Version 1.0, Ian Heggie and Rodrigo Archondo Callao, 1996.

2.2. The elements of the maintenance problem

In discussing road expenditures, a distinction is usually made between four categories, namely:

a) *Routine maintenance*, which includes ditch and culvert cleaning and filling of potholes;
b) *Periodic maintenance and rehabilitation*, defined as surface treatment or surface renewal, including asphalt overlays and re-gravelling;
c) *Reconstruction,* defined as selective repair and re-strengthening of the pavement or shoulder after demolition of the existing structure;
d) *New construction,* involving new routes or increases in capacity of existing routes.

The maintenance problem is concentrated particularly on routine and periodic maintenance, for which between 2% and 3% of the replacement value of the road stock is generally required in order to maintain the condition (and hence the value) of the road stock. The reason for this is that funds for rehabilitation and reconstruction are still more readily available from the international financial institutions and donors, while funds for routine and periodic maintenance have to compete for limited domestic current budget funds.

2.3. The institutional incentives issue

The cost of restoring roads which are allowed to deteriorate to the point at which reconstruction is required is substantially greater than that associated with a policy of timely and effective maintenance. The ratio has been variously estimated at 2.5 to 1 for Costa Rica and Chile (Gyamfi, 1992) and even up to ratios of between 3 and 5 to 1 (Harral and Faiz, 1988), the differences being explained by differences in the initial conditions and the traffic volumes of the roads considered. In those circumstances it might appear sensible for governments to maintain in the current period in order to avoid much greater expenditures later. However, most developing countries have a pervasive and perennial shortage of public funds, so that expenditures in many other sectors, as well as roads, would yield high social returns. If the future expenditures on road rehabilitation are discounted at the marginal rate of return on funds currently invested in the economy, then it may well be the case that a perfectly standard economic calculation would suggest that deferral of expenditure in the roads sector is a correct policy. This conclusion is accentuated if, as is commonly the case for the poorest countries, funds for rehabilitation can be obtained from the international institutions for reconstruction, and not for routine and periodic maintenance, on very favourable terms.

On the other hand, it has also been estimated that it costs about three dollars in present worth in additional vehicle operating costs for every dollar economised by road agencies in under-funding maintenance. Because these figures relate to the same period of time, the ratio between them depends solely on the technical conditions of the road system and on traffic volumes, and does not depend on the rate of discount. For the economy as a whole, it is thus not economically rational to defer maintenance where even the current year costs of deferral exceed its benefits.

The problem is that the benefits of improved current maintenance accrue primarily to road users, while the costs fall on the public agencies. The myopia of governments in regard to road maintenance is thus not so much an inability to rationally calculate the effects on the public budget but an inability to recognise, or mobilise, the great savings to road users accruing from more timely maintenance. The reasons for substantial under-funding of maintenance are thus essentially institutional.

2.4. The role of road funds in improving road maintenance

The "second generation" road funds discussed in the next chapter are seen essentially as a mechanism to compensate for political or administrative myopia and to ensure the allocation of resources to a low-profile economic activity with particularly high rates of social return (Heggie and Vickers, 1998). In addition, significant operational efficiency gains can be reaped by overcoming uncertainty in the level and timing of funding, thereby creating a stable and skilled subcontracting industry with appropriate incentives for efficient execution. However, these advantages do not arise solely from the creation of a road fund. In looking at efficiency improvements a distinction needs to be made between:

- *Funding*, which can be through government budget finance or through an independent or quasi independent fund; and
- *Implementation*, which can be through a public sector highway agency, on force account, or by private contractors on a system or project contract basis.

Figure 1: Implementation through a roads agency

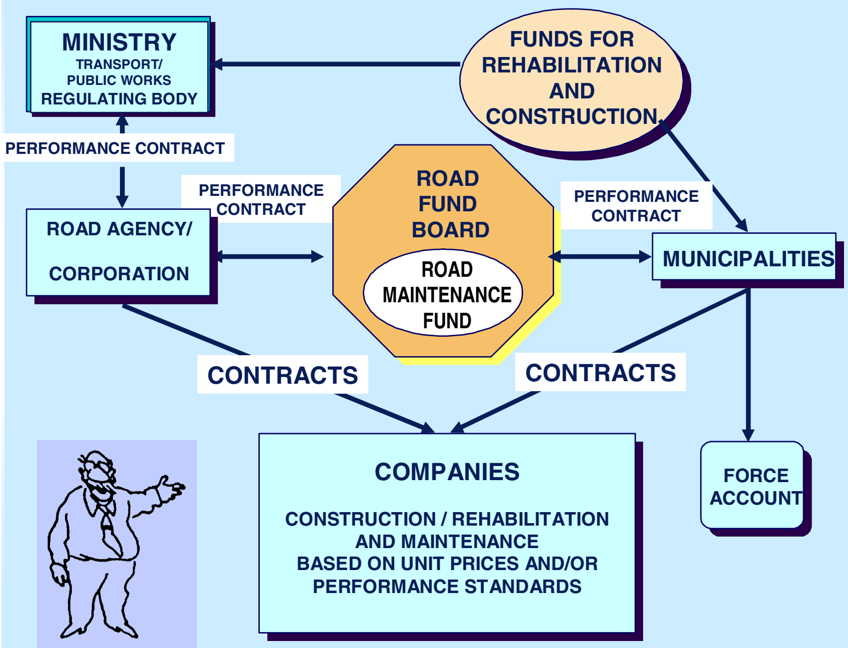
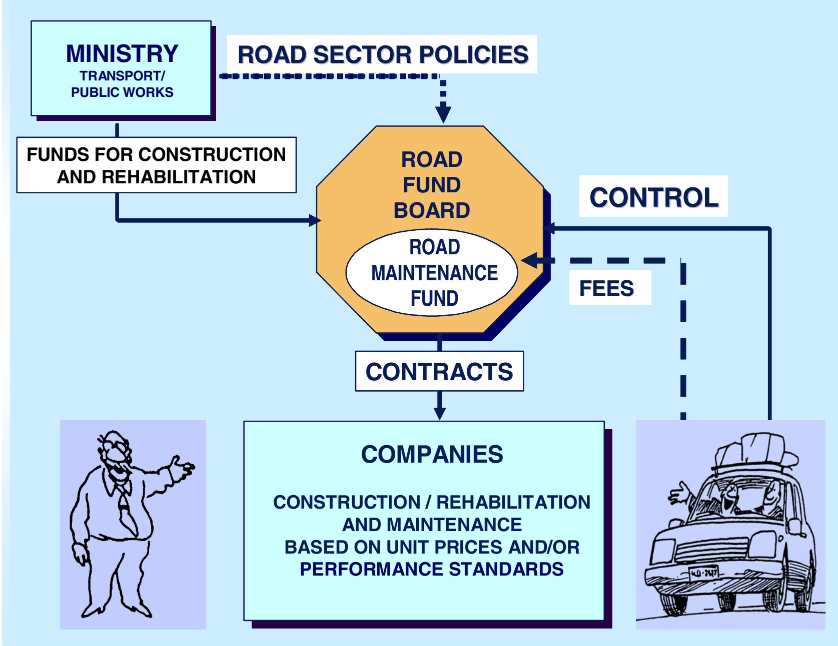


Figure 2: The fund as implementing agency



Source: Zietlow, 2004.

The relationship between the funding institutions and the implementation institutions may take different forms. If an efficient road agency or corporation exists, a performance contract between the agency and the board might be the best choice (see Figure 1). Alternatively, the road could itself act as the implementing agency, contracting all road maintenance to road maintenance management companies (see Figure 2).

While reform of the funding arrangements may facilitate reform of the implementation arrangements, it does not necessarily do so. For example, it has recently been estimated that contracting out works on a performance specified basis can yield cost savings of up to 40% (Frost, 2001). A comprehensive policy for road maintenance reform would need to address both. In this paper we only consider the former.

3. THE HISTORY OF ROAD FUNDS

3.1. Types of road fund

A dedicated road fund is a separate fund, outside the central government's general budgetary framework. It may be financed partly from user charges, and may be responsible for maintenance and/or capital expenditures. It is essentially a funding agency, and not an implementation agency. It may be combined with a traditional public sector implementation or with an independent or private-sector implementation function.

Within this broad definition, Potter (1997) identifies five different models of road fund, namely:

1. Post-war trust funds, often for capital expenditure on highway construction, in developed countries like the US and Japan;
2. RFs established in the 1960s and 1970s in Africa to protect resources for road maintenance;
3. RFs set up in the 1990s in the FSU in response to tight budget conditions;
4. Arrangements which seek to pursue an agency model of service, as in the UK or Finland (though these may not include road funds as defined in this paper);
5. Arrangements which aim to commercialise road maintenance, and are financed partly through road user charges, as in New Zealand, Sweden and the "second-generation road funds" in some African countries.

Within this taxonomy, Potter distinguishes between categories 1-3, which are pure tax hypothecation devices, unrelated to any reform of the traditional public sector delivery arrangements, category 4, which attempts to gain benefits from an independent supply agency arrangement without hypothecation of taxes, and category 5 (later described as "second-generation road funds") which attempts to commercialise road maintenance financing by combining the redefinition of what has traditionally been viewed as a pure tax (particularly the fuel tax) as a user charge with more commercial disbursement and procurement mechanisms.

3.2. The performance of the early road funds

The evidence of the effects of early road funds (and indeed of more recently created funds) on the actual allocations to road maintenance is relatively weak. An early study of 37 developing countries, over the period 1955-65, showed a positive correlation between earmarked taxes as a proportion of gross investment, and road expenditure as a proportion of gross investment, supporting a tentative conclusion that there is a positive relationship and a likely causality between the amount earmarked and the amount spent on roads. However, the differences in the cross-section analysis were not statistically significant (Eklund, 1967).

Wherever the government retains control over the level of the user charges or over the allocation of complementary funds, the total level of funding may be just as vulnerable with a road fund as without one. In Colombia, for example, although the "earmarked funds" for the national road fund grew at the same rate as GDP between 1979 and 1987, total funding for roads grew more slowly than GDP and the road network continued to deteriorate (Dick, 1989).

3.3. The macroeconomic objections

The longstanding macroeconomic objections to earmarking of taxes are based on the weakening effect of earmarking on the degree of flexibility with which governments are able to control both the total level of public expenditure and the distribution of that expenditure between sectors and objectives (Deran, 1965). That view has been most strongly expressed by the International Monetary Fund (IMF) (Premchand, 1983) but also strongly supported by many of the macroeconomists within the World Bank. Any special or dedicated fund which is created outside the normal budgetary procedures is treated as a threat to good governance (Teja, 1988). On that basis, Potter (1997) unambiguously rejects the old-style hypothecation arrangements (categories 2 and 3 in the list in section 3.1.), though curiously not criticising the trust fund arrangements. The arrangements which involve independent agency management but no hypothecation (category 4 above) are viewed as benign. Only the newer style arrangements (category 5) are judged with any degree of ambiguity. It is to the characteristics of these "second-generation road funds" that we now turn.

4. SECOND-GENERATION ROAD FUNDS

4.1. Origins of the second-generation road fund (SGRF)

The SGRF was developed primarily as an instrument for the achievement of the objectives of the Road Management Initiative (RMI) of the Sub-Saharan Africa Transport Programme (SSATP). This initiative was launched by the World Bank in 1988, and subsequently joined by other donors and funding agencies. From its original nine African member countries, RMI has now expanded to 19. The SGRF was based on the observation that inadequate maintenance had resulted in losses of road stock over a period of two decades of a value equal to the total of WB investment in roads in the same period. The SGRFs were thus explicitly and virtually exclusively devoted to the finance of maintenance and not capital investment. Of 28 sub-Saharan countries covered in a recent RMI review, 24 have such funds (see Annex 1).

The trend soon spread to Latin America. In 1993, the International Road Federation, the UN Economic Commission for Latin America and the Caribbean, the World Bank and the Pan American Institute of Highways joined forces in promoting regional and national seminars on improving the highway system in Latin America and the Caribbean, under the name PROVIAL. The financial and institutional reform of road maintenance played and still plays a major role in all of these seminars. After several countries expressed their interest in embarking on such a reform, in 1994 the International Road Federation and the German Agency for Technical Cooperation initiated a project to assist these countries in creating roads maintenance funds and contracting out road maintenance by performance standards (Zietlow, 2004).

4.2. The distinguishing features

SGRFs usually involve the following features which distinguish them from traditional road funds:

(i) Establishment of charges in addition to and entirely independent of the determination of levels of taxes on road users for general revenue purposes;

(ii) Direct transfer of such charges into the fund, outside of the procedures for allocations from the consolidated fund;

(iii) The management of the road fund by a board representing road users, who can simultaneously determine the level of charge and service which they prefer;

(iv) Establishment of efficient internal allocation procedures by the road board to determine day-to-day allocation decisions.

The concept of the second-generation road fund and board was thus that of an autonomous agency controlling the funding of road maintenance, directed predominantly by road users, having the power to raise revenue and control funding allocations, and having a strong incentive to insist on commercially and professionally efficient management. It was not part of the concept that it should execute works, that task being normally exercised by separate national and regional road agencies which would submit plans and programmes for consideration by the roads board for financing from the road fund.

4.3. The IMF view of SGRFs

During the early nineties, IMF technical missions in African, Middle Eastern and FSU countries consistently advised countries to avoid the creation of off-budget institutions which could undermine expenditure control and distort resource allocation. The development of SGRFs under World Bank guidance, ostensibly designed to avoid these consequences, led to discussions between the Bank and the Fund and eventually to the formulation of a written response to the challenges posed.

That response (Potter, 1997) began by expounding three major principles of the Fund approach to budget preparation and public expenditure management:

- A macroeconomic aggregate needs to be set, within which the overall total of public expenditure should be constrained;

– Budget priorities should be settled within a transparent and predictable institutional arrangement;

– Good financial management, including flexible use of balances, is crucial.

The response continued, however, by recognising a strong argument for public sector reforms to simulate market disciplines to secure efficient and effective public service delivery. Potter referred specifically to the “Anglo-Saxon reforms”, which he took to include the separation of policy determination; privatisation of activities where possible and the creation of arms-length agencies for public service delivery when full commercialisation was not possible; and giving agencies specific objectives and managerial authority by disengaging them from the restrictions of civil service wages and responsibility structures. He gladly conceded that even where the conventional budgeting approach delivers an adequate volume of road maintenance, there could be a case for a dedicated fund to secure the agency benefits.

Despite this, Potter was still concerned that RFs might undermine good budget governance. In particular, he questioned whether there could truly be a hard budget constraint on a road fund. He argued that if there is an inelastic demand for the service then any relevant tax, fee or charge may not be independent from the government’s capacity to raise other revenues. This is particularly likely to be the case in developing countries where fuel tax is usually the source of a sizeable proportion of total tax revenues (Gupta and Mahler, 1995). Moreover, with an implicit monopoly for the fund, he feared the emergence of rent-seeking behaviour on the part of its managers. He therefore concluded that RFs would be acceptable only:

a) “where there are adequate budgetary procedures to maintain macroeconomic control as well as capturing the agency benefits; or

b) “where there are inadequate budgetary procedures but the right expertise and political and financial circumstances to make the RF a success.”

In the latter part of his paper, Potter attempted to define what these required organisational and financial circumstances were. Organisationally, they included transparency in accounting; accountability to board, truly representing users as well as government; freedom from political interference (except on the determination of the budget envelope); and freedom from corruption. Financially, they included a charges system, well reflecting the sources of costs; strict and clear principles for the determination of the share of fuel taxation going to the fund; and no automatic “topping up” of user charges from budget funds.

The IMF position may thus be summarised as follows. A road fund is not considered appropriate where it represents the earmarking of general revenues for one purpose, with the service delivery still pursuing a traditional model. But where a RF is a genuine attempt to pursue a purchasing agency approach then, in principle, it might be acceptable, subject to satisfying some stringent institutional and financial conditions. Hence the desirability of RFs in future needs to be assessed on a case-by-case basis.

4.4. The World Bank view

The highway sector within the Bank, as the originators of the SGRF, and particularly the staff associated with the Road Maintenance Initiative, unsurprisingly retain their general support for it.

That rosy view is not unambiguously supported, however. In an initial review of the scope and role of SGRFs, Gwilliam and Shalizi (1999) highlighted the tension between the conventional macroeconomists' approach, which tended to presume that good governance existed, and the sector specialists' approach, which was based on their perception that it did not. They therefore argued that road funds should be seen as a step towards *either* the re-establishment of good governance in the allocation of public revenues *or* the more general commercialisation of road use, and that road funds should not be assessed on general principles but on a case-by-case basis in terms of their macro- and micro-economic effects. In particular, they suggested that the establishment or continuation of a second-generation road fund should be judged in terms of objective indicators of its effects (or likely effects) on efficiency of resource allocation, operational efficiency and rent-seeking behaviour. The implication of their argument was that the funds being established should only have a limited life.

Gwilliam and Shalizi were clearly driven to that conclusion by their recognition of the inherent validity of *both* the macroeconomic desirability of fiscal flexibility *and* the micro-economic argument for reformed processes to improve the efficiency of performance within the sector. But the history of the argument appeared to offer no compromise. In order to finesse the earmarking arguments, the proponents of road funds had constructed an edifice in which part of a fuel tax was renamed as a user charge, and in support of that device had given its determination the status of an internal business decision of what might be construed as a road users' co-operative. While macroeconomists might not be fooled by the finesse, their fears could perhaps be appeased by the imposition of a "sunset clause" on new road funds.

In a later review, based on the evidence of actual outcomes discussed in the next chapter of this paper, Gwilliam and Kumar suggested a revision of that position. They argued that much of the observed benefit of SGRFs came from the changes in process which they initiated rather than through an increase of funding arising from tax earmarking/user charging. They thus concluded that, even without any loss of ultimate control by ministers of finance over total tax allocations to road maintenance, the SGRFs had improved performance and should not thus necessarily be subject to a sunset provision.

4.5. The current state of the debate

The current state of the debate can be gauged from a discussion which took place at the World Bank's March 2005 Annual Transport Forum in Washington, which can be accessed on the World Bank website. Contributors from the Bank highway sector and the Sub-Saharan Africa Transport Program emphasized the importance of road maintenance and the economic losses arising from inadequate maintenance, and asserted that the fuel levies on which the SGRFs mainly depend did not constitute earmarked taxation but were genuine user charges. Speakers from the macroeconomic Poverty Reduction and Economic Management Vice-Presidency of the Bank (PREM) and from the IMF restated the standard arguments mentioned in the previous sections, particularly pointing out the impacts on fiscal flexibility and control. In effect, the two parties talked past each other rather than to each other.

However, some element of a resolution did suggest itself. Insofar as the highway engineers argue that the cost savings to users exceeded the extra costs that might be imposed on them through increased user charges directed to road maintenance, they implicitly argue that there was no adverse effect on general taxable capacity. This is, in principle, capable of verification in two stages, namely:

- Demonstration on a case-by-case basis of the actual balance between increased maintenance expenditures and operating cost reductions;

- Demonstration on a case-by-case basis that the users actually recognise this.

That may be the evidential basis on which the tension between the micro and macro arguments will ultimately be resolved, consistent with the positions adopted by macroeconomists at the Bank and the Fund, as discussed above. On that basis, the next chapter explores the empirical evidence that is available on the effects of SGRFs.

5. THE EVIDENCE ON SGRFS

5.1. The African road funds

In pursuit of its earlier recommendation that road funds should be considered on a case-by-case basis in terms of a number of well-defined indicators of performance, the World Bank undertook a review of the evidence on the performance of a number of African RFs, based on data from 1999-2000 (Gwilliam and Kumar, 2003). The analysis was undertaken in terms of the performance of SGRFs in satisfaction of the structural, process and outcome objectives of the original architects of SGRFs. A broader and more recent compilation of data, though containing less detail, taken from the "RMI Matrix" produced by SSATP, is contained in Annexes 1 and 2.

5.1.1. Structure

Indicators of structure

The indicators of structure are logically derived from the declared objectives of second-generation road funds, including the strength of the legal basis of the road fund; the size, composition and chairmanship of its managing board; the duties and powers attributed to the boards; and the provisions to assure the technical competence of its permanent staff. The results are set out in Table 2 below.

The institutional form

The legal basis of the road funds and boards is generally secure as established by statute but the composition of the RF boards differs significantly among countries. In Ghana, Kenya, Malawi, Tanzania and Zambia (Chipewo, 2003) the boards have a dominant private sector representation, while in Ethiopia and Benin the boards are dominated by the public sector. The more recent review (Annex 1) shows exactly half of the boards to have a private sector majority. The chairmanship arrangements also vary. In Ethiopia it is legally reserved for a public official; in other cases the chairman is appointed by the minister, but not necessarily from the public sector (Malawi, Kenya) or chosen by the board from its members (Zambia). In Benin, Ethiopia and Ghana, the Minister of Works is the board chairman. Only in Tanzania is the chairperson directly appointed by the President.

Table 2. **The new structure of road maintenance administration**
(as of 2001)

	Benin	Ethiopia	Ghana	Kenya	Malawi	Tanzania	Zambia	Uganda
Year of establishment	1996	1997	1997	2000	1998	1999	1994	**
Statutory basis	Law	Law	Law	Law	Law	Law	Order	**.
Private sector share on Board	5/9	4/15	8/13	8/13	9/12	5/9	7/11	**.
Board Chairman	Minister of Works	Minister of Works	Minister of Works	Private Sector	Private Sector	Ex-PS, Finance Ministry	Private Sector	**
Planning separate from execution	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Board sets fuel tax levy	No	No	No	No	No	No	No	**
Board proposes fuel tax levy	Yes	No	No	Yes	Yes	No	Yes	**n.a.
Professional management	Partly	Yes	Yes	In process	Yes	Yes	Yes	Yes
Board has additional functions	No	No	No	No	Yes	No	Yes	**
Independent executing agencies established	No	No	No	No	Partly	Yes	No	Yes

Notes: * See text: in practice the NRA always seeks ministerial approval.

** See text: Uganda chose not to set up the road fund and instead committed to meet road maintenance requirements through the normal budgetary process.

Source: Gwilliam and Kumar, 2003.

One of the weaknesses of the SGRF concept concerns its ability to respond to the range of policy objectives and stakeholder interests that exists in practice. In particular, the danger is that the main commercial interests to be represented on the board would not be concerned about rural roads, non-user interests, environmental impacts, etc. That has been partly addressed by superimposing appraisal processes on the allocation process -- to be discussed later. But it does also suggest a wide-ranging public sector presence. The Road Board in Ethiopia has high-level public, especially regional, representation on the Board, which has given it strong support from the Government as laid out in the Road Sector Development Program (RSDP).

The professional base for the RF board has been weak in most countries, even though in most of the sample countries the secretariat salary is competitive with the private sector. Where they are not, as in Kenya, there is great difficulty in attracting qualified staff. The Ethiopia RF secretariat took about three years to fill its designated professional positions.

The role of the road boards

The declared role of the road boards varies from country to country. All are in principle responsible for resource generation, evaluation and allocation. Most RF boards serve as financer of services rather than as provider of services, with the programming, tendering, evaluating, negotiating, awarding, supervising and managing of contracts generally being the responsibility of the road agencies to which the funds are allocated. But there are variants. In Malawi, the National Road Administration was also, at 2000, a “de facto” road implementation agency, having taken over most of the professional road staff of the Ministry of Works and Services, though it is intended ultimately to separate out implementation staff in an independent central roads authority. Several boards have responsibility for advising on general road policy matters, such as vehicle overloading. In Zambia, the Roads Board provides advice to agencies on procurement and monitors performance under the national road sector investment plan, ROADSIP, and the Roads Board is responsible for co-ordination of donor financed programmes, including management of the ten-year road investment programme.

The powers of the boards

The actual resource generation and distribution powers of the roads boards are much more restricted than originally envisaged. In no case are they able to finally determine the level of the tax levy which is the core of their income. Even where they have the power to recommend levy rates, the evidence so far is that the ultimate decisions of government continue to be motivated more by general fiscal considerations than by perceived sector needs. In other respects, the autonomy of the boards is legally restricted. For example, the Benin Board is required to implement policy “as laid down by the Review Council” (a body consisting of one representative of government and two of donors). In most cases the programmes of expenditure have also to be approved by one or more government ministries. Only in Malawi does the board formally have the power to determine the levy, though in practice it always seeks the approval of the Minister of Transport and Public Works and the Minister of Finance.

Implementation agencies

Lack of executing capability is the most serious element in poor road maintenance, particularly for regional and rural roads. While road works are increasingly contracted out to the private construction industry on a competitive basis, SSA governments are reluctant to delegate road management to autonomous agencies operating according to sound business practices (Nyangaga, 2001). Government departments (ministries) are mostly cumbersome and largely ineffective, with little or no commercial approach to the task at hand. In most of the countries, main roads are managed by the roads department (Ministry of Works), which is suffering all the ills of the civil service and little has been done to implement a decentralised management. Because of the focus on the road fund and road board, institutional reform of the executing agencies is not necessarily provided for in most of the founding statutes.

In contrast, the creation in Uganda of an independent engineering organisation to procure resources from the market by competitive tendering is a core part of a reform, not associated with the establishment of a SGRF (Kumar, 2001). In that country the Government introduced a Road Sector Development Plan (RSDP) in 1995. The objective of the plan was to restructure government involvement from the direct provision of services to the provision of policy guidelines and a clear legal framework for implementation. The institutional reforms within RSDP included commercialising and contracting of technical services on a performance basis under hard budgetary conditions; separating the planning and financing functions from procurement and implementation; decentralisation of maintenance delivery on district, urban and community roads. But it did not provide for an autonomous road board. Policy and regulation continued to be the responsibility of the

Ministry of Works, Housing and Communications, and although a commitment was made in the plan to stable and secure funding for road maintenance it was left on budget. The weakness of MoWHC in implementation was addressed by the provision for an autonomous road agency, preceded during the formation period by a Road Agency Formation Unit (RAFU), responsible to MoWHC. This reform is not yet complete. In practice RAFU is acting now as a professional procurement management agency with heavy external financial and technical support. Moreover, it has a focus on rehabilitation and reconstruction rather than maintenance -- it remains to be seen how well it will transform into a sustainable maintenance-focused roads agency.

5.1.2. Process

Indicators of process

It was the explicit intention in creating SGRFs to develop a businesslike, rather than bureaucratic, process in order to provide adequate and stable sources of financing for road maintenance. The indicators of *process* derived from this include the way in which revenue and expenditure totals are set, funds transferred and secured, expenditures allocated and implementation (both by the board and the executive agencies) monitored and audited. Their outcomes are shown in Table 3.

Table 3. **Process changes in selected road funds**

	Benin	Ethiopia	Ghana	Kenya	Malawi	Tanzania	Zambia	Uganda
Fuel tax levies paid direct to RF	No	Yes	No	Yes	Yes	No	No	
% of funds from fuel levy	75%	25%	85%	>90%	95%	95%	95%	
% of revenue from own resources[1]	10%	75%	0%	10%		0%	0%	
Regular monthly transfer of funds	Yes	Yes	Yes	Yes	Yes	No	No	
Independent technical audit		In process	Yes	No[2]	No	Yes	No	In process
Independent financial audit		In process	Yes	No[2]	Yes	Yes	Yes	In process
Performance audit of implementation/management		Yes	Yes	No	No	No	No	Yes
Program to empower SME contractors	Yes	Yes	Yes			Yes	Partly	In process
Road fund financing objectives	Maint.	Maint.	Maint./ rehab.	Maint./ repair	Maint.	Maint.	Maint./ rehab.	

Notes: 1. Other duties and fees earmarked to the road fund.
2. Audits have not yet been completed, as the secretariat is newly in place, but are required by law.

Source: Gwilliam and Kumar, 2003.

Sources of revenue

The process for determining income is the starting point. As of 2000, the fuel levy was the dominant source of contribution to these RFs in all the countries examined (more than 90%). The nominal basis for the levy varies from country to country. In Zambia and Benin, the fuel levy is a fixed percentage of the wholesale price, set at 15% in Zambia (4 US cents per litre in 2000) and 10% in Benin (2.5 US cents in 1999). In Ghana, Ethiopia and Tanzania, the fuel levy is a fixed charge per litre, set at 3.5 US cents, 4.4 US cents and 9.0 US cents, respectively, in 2001. The 2004 data shows the levy usually falling in the range between 2 and 9 cents per litre. The danger of setting a fuel levy as a percentage of the fuel price is that collections may have no relation to the maintenance needs and the RF revenues may fluctuate with changes in the macroeconomic environment.

Arrangements to diversify road user charges will need to be explored. As Annex 1 shows, the fuel tax levy is the dominant, if not the sole source of revenue in most countries. Direct tolls are one possible source; Benin already gets 10% of the total road fund collection from tolls and weight controls deposited directly into its account by the toll concessionaire. But with traffic volumes below 1 500 vehicles per day, the high level of toll collection is likely to make this impractical. Zambia's National Roads Board has solicited additional funds from international transit tolls, weigh bridge fines, motor vehicle licence fees, etc., although legislative reform leading to broadening of the road fund resource base has not been effective so far. In Ethiopia, the dominant contribution to the road fund originally came from almost 75 per cent of total collections from the sales tax and municipality tax, which was replaced by VAT in 2004, without contributions to the RF being affected. Most importantly, if charge levels are to reflect the costs imposed it will be important to raise standing charges, or mileage-related charges based on axle loads for heavier vehicles.

Setting charge levels

In principle, the fuel levy, and other charges, should be set on the basis of assessed maintenance "needs" (as practiced in Ghana and Ethiopia). In practice, the initial level is more usually fixed on the principle that the "user charge" element was in addition to what would have been levied for general tax purposes in its absence. This appears to have been important in getting ministers of finance to agree to an initial level of a fuel tax surcharge, and to be the principle driving the continued insistence of ministers of finance that they should have the final say in setting the levy. To do otherwise would threaten their fiscal viability, especially in countries with already high levels of fuel tax, or for those neighbouring countries such as Nigeria with very low tax levels, where the price elasticity of demand might be sufficiently high to ensure that any increase in fuel surcharge did effectively reduce the yield of the basic fuel tax, even if its nominal level was unchanged. The onus is very much on road interests to demonstrate that general tax yield is independent of the levels of surcharge imposed. So, in the event, governments have generally insisted on retaining effective control over the level of fuel tax surcharges.

Securing revenues

Arrangements for securing the revenue vary. In Ethiopia, the fuel levy is credited directly from the petroleum enterprise into the road fund account, on a regular monthly basis, creating a stable foundation for the RF. In Kenya, the payments are made from customs to the Ministry of Public Works and Highways and then on to the Road Fund. In Benin and Zambia, the fuel levy is channelled into the road fund account from the Petroleum Commission through the Ministry of Finance, creating delays of four months and two months respectively. Even in Ethiopia, where the payment is direct from the petroleum enterprise to the Road Fund, a delay of four months is involved in the transfer.

Even worse, however, is the interposition of either a ministry of works or a ministry of finance in the channel, which is highly likely to generate both leakages and irregularities in the flow. For example, in Zambia in 1999, only about two-thirds of the fuel levy delivered by Zambia National Oil to the Zambia National Revenue Authority, which should then be credited to the Road Fund through the Ministry of Transport and Communications, actually found its way on to the Road Fund. The Zambia Roads Board is working to reduce delays and streamline procedures, beginning FY01. In Tanzania, the process is even more cumbersome. In the Dar es Salaam region, a fuel levy is collected by the Tanzania Revenue Authority from the fuel importers and deposited in the Ministry of Finance account, from where the amounts are passed on to the Ministry of Works (for main roads) and Local Government (for district roads) in the agreed proportion. The amounts, in turn, are deposited into the road fund account, which is responsible for disbursements to the executing agencies. In addition, there are significantly more steps in the collection process from the up-country regions. The overall transit time between the payments of the fuel levy by the importers and the time when this credit reaches the road fund account exceeds 60 days, during which the funds remain on the TRA accounts. Most countries do not allow the board any redress against non-payment or delayed payment of the established "user fees" to the fund account. The 2004 data in Annex 1 shows that tax levies are being channelled directly into the fund in 14 out of 24 of the fund countries.

It is of course too simplistic to assume that road funds will always be inviolate. It is essential to design road funds to maximize the probability that they will not be abused, rather than simply legislating a road fund into existence. Raiding of the road fund may still be possible because of legal or bureaucratic holes in the system. One of the key advantages of setting up road funds is that user charges provide a link between the level and quality of service and price to be paid. However, in practice, governments are reluctant to relinquish control of cash flows and of opportunities to "borrow" funds for other purposes when the need arises (Nyangaga, 2001). MoFs usually find a way for using funds for national emergencies (which may even be justified) as well as for non-emergencies. The structural safeguards necessary to provide protection to the road funds require a strong political will. In addition, adequate user representation and transparent dissemination of the board's activities are required to establish a check-and-balance system to improve accountability in the day-to-day activities. It is important for the road funds to be supported by well-established enforcement tools to recover money owed to the road funds.

Prioritising expenditures

Most SGRFs have found it difficult to prioritise expenditures. In terms of types of maintenance the problem arises partly because, as a result of inadequate funding over the past decade, a "maintenance backlog" has resulted and most of the road networks require substantial amounts of rehabilitation works to be carried out before they are in a maintainable condition. Moreover, in some countries adverse weather conditions enforce emergency works, which pre-empt the largest part of the maintenance budget. In those circumstances, the development of long-term stability in road maintenance expenditure and road conditions is difficult to achieve.

In principle, second-generation road funds, with their commercial orientation and strong constituency, should be much better set up than the traditional first-generation road funds to allocate resources efficiently within the sector. In practice, however, resource allocations for road maintenance continue to be mainly dictated by "standard formulae" rather than a planned review of programmes put forward by various road administrations. This is most apparent in Zambia, Kenya, Ethiopia and Benin. For example, in Kenya substantial contractual commitments have been made on the non-prioritised core network, while demands are increasing on the priority core network.

The allocations are shown in Annex 2 to be generally biased towards main roads to the detriment of the rural/feeder road network (probably reflecting the dominance of central government ministries and national suppliers' organisations in the composition of boards). However, in Kenya, 40% of the road fund revenues are allocated by rule to constituencies and districts for lower category road maintenance, while 57% goes to the Roads Department for the main road network. In addition, in some countries even the planned expenditures are particularly poorly disbursed for the rural road network, primarily because of lack of capacity at the regional level. In Ethiopia, for example, only about 20% of the planned allocations for the rural road network were disbursed during FY99 because of the lack of absorptive capacity. In Tanzania, less than 40% of the programmed works for district roads were undertaken during FY01.

Monitoring performance

The available evidence on the effectiveness of monitoring performance of the flow of funds and the quantity, quality and cost of the road works is mixed. Financial auditing is the easiest to introduce and is now common. Ghana's Roads Fund Board has established proper planning and programming of road works with well-defined disbursement and accounting procedures, which have facilitated timely contracting arrangements. In Ethiopia, procedures have been introduced to monitor the performance of the national roads agency (Ethiopian Road Authority, ERA) based on the preparation and approval of payment certificates and monthly progress reports. Currently, they are required to report on funds utilisation. Until recently disbursements were made to the agencies in Ethiopia on a monthly cash flow basis: that has now been changed to a payment system based on verified payments certification for work performed.

Technical audit is much more difficult. Ideally this should involve continuous auditing of projects-in-progress for improving performance, involving: (i) establishment of a credible and independent external auditing process; (ii) establishing appropriate responsibilities for reporting and follow-up of the audit recommendations to ensure its effectiveness; (iii) developing an updated and rationalised inventory and condition survey of the classified road network. Together this would eliminate projects being technically audited after rather than during the event. But systematic arrangements to carry out independent technical audits are still lacking in most countries, with the remarkable exception of the Ghana Road Fund, which has a firm basis, with detailed internal and external monitoring procedures to ensure efficient use of money, and accompanied by monthly progress reports and external financial and technical audits. In Zambia, the National Roads Board was operated for three years on the basis of an Annual Works Programme, based on the identified needs of the road agencies, as in the case of the IDA-financed strategic investment plan.

The road funds in Zambia and Kenya have not been immune to the kinds of political interference that has impacted the classical first generation road funds. Resources have been diverted to rehabilitate roads in the capital city over the past two years to the exclusion of road maintenance needs of the country's network. This is not surprising as, after years of resource constraints and for the first time since establishment of the road fund, executing agencies have resources at their disposal. The first beneficiaries have been residents of large capital cities, in view of their high political profile, and the fact that the most ardent supporters of the road fund are urban residents, who also account for a dominant share of car ownership. However, to ensure sustainability of the road fund and adequately address the broader issues related to the quality of the road network, it is critical that in the coming years, attention is given not only to high-volume urban roads but also to rural/feeder road networks, to ensure the equitable distribution of resources.

Development of the implementation agencies

While setting up adequate and sustainable financing arrangements is a necessary condition in the road reform programme, it is equally necessary to ensure that the resources allocated are used in the most effective and efficient manner by the road executing agencies. The available evidence presents mixed results. In the Gwilliam and Kumar set, some countries had already set up semi-autonomous road agencies (Uganda, Tanzania), siphoning off the executing and service delivery functions from the Ministry of Works, while others (Kenya, Ghana, Zambia) had attempted to reform the existing government departments. Still others (Malawi) had juxtaposed road financing and road executing functions within the same agency. In the larger RMI set (Annex 2), twelve of the 24 countries with a road fund also had independent road agencies. In general, the agencies lacked a commercial environment and have continued to work as government departments (except Uganda), without a clear understanding of the mission objectives, clearly stated functions or a work programme. The reporting arrangements between multiple regional offices and headquarters remain weak.

The capacity of road agencies at the district level raises even greater concern. The quality of works carried out is affected by: (a) a lack of uniform standards followed by different districts/councils; (b) a lack of donor coordination and use of different mechanisms to channel funding for district roads; (c) different interests influencing the decision on the priority of the road maintenance works, not always justified on social, engineering or economic grounds; (d) insufficient capacity. It is critical to develop institutional and implementation arrangements to deliver an efficient road maintenance programme.

5.1.3. Performance outcome

Indicators of performance

Whatever the structure and process of the SGRFs, the ultimate test of their ability to improve efficiency lies in their actual performance. While data is far from good, some indicators of performance are available. These include trends in allocation and disbursement; efficiency of implementation; and the ultimate outcomes of road conditions. These are contained in Table 4.

Adequacy of funding

In all the countries reviewed by Gwilliam and Kumar, funding for maintenance showed consistent increases after the establishment of the SGRFs. However, though the question of whether the SGRFs caused the increase or were a condition of the Multilateral Financial Institutions (MFIs) for contributing to a road reconstruction programme is perhaps a matter of interpretation rather than determinable fact.

The extent to which they have reached an adequate level is more difficult to assess. The most recent RMI review (Annex 2) showed that the road funds were only able to fully meet the requirements for *routine* maintenance in nine of the 24 countries, and nowhere able to cover full maintenance requirements.

In contrast to these road fund countries, it should be noted that the Government of Uganda, which has not adopted a road fund, has consistently met most of its maintenance budget requirements as well as counterpart funding needs through a normal budget process. It has committed greater resources for road maintenance per kilometre of road than any of the "securely funded" neighbouring countries.

Table 4. **Performance impacts of road maintenance administrations**

	Benin	Ethiopia	Ghana	Kenya	Malawi	Tanzania	Zambia	Uganda
RF administrative expenses as % of income	2%	>1%		3%	<3%	1%	<3%	
Annual change in % roads in good condition	+9.4	+3.0	+4.5				+4	
Annual change in % roads in bad condition		-10.0	>-10.0		>-10.0		-5.5	
Annual change in % maintenance works contracted out		+10	+30		+30	>+30	>+30	+10
% of estimated needs financed		75%	50%		30%	30%	30%	80%
% of funds allocated (trunk/urban/rural)			52/27/20		48/42/10	70/30[3]	40/20/40	
% of allocated funds disbursed (trunk/urban/rural)	55 (ave)	65/55/20		50/40/2		100/50[3]	24/69/8	

Notes:

1. The rates of change shown in the table are not all for the same period, but are selected to reflect the average annual rate of change since the introduction of the new arrangements.
2. In the absence of a detailed road inventory in most of the countries, it is difficult to comment on percentage change in quality of the road network, which explains the blank cells. In addition, heavy investments by donors in rehabilitation and reconstruction of the paved roads have significantly improved road quality and it is difficult to isolate the impact of improved maintenance funding on road quality.
3. Allocation available only between two agencies, the first responsible for the trunk network and the second responsible for urban and rural roads.

Source: Gwilliam and Kumar, 2003.

It has also given explicit undertakings to sustain maintenance funding, as well as the salary levels and conditions of employment offered by the executing agencies after the main donor financing ends. Available evidence thus suggests that setting up dedicated financing arrangements, even under second-generation principles, while useful, is not a sufficient condition to ensure that a sustainable and stable basis for road maintenance is established which ultimately translates to improved service delivery and operational efficiency. Government commitment is also required.

Disbursement experience

Even where funds are allocated to road maintenance, they may not be disbursed if the budgetary process is delayed, particularly where climatic conditions also limit the period during which effective maintenance work can be performed. Creation of an SGRF should improve this by permitting the carry forward of funding. Immediate improvement was achieved in some countries. For example, Ghana has been the most successful in disbursing over 90% of the allocated funds in 1998 and 1999. Benin improved its disbursement proportion from 55% in 1998 to 95% in 1999. In other countries the improvement has been less dramatic. In Ethiopia, less than 40% of the road fund was disbursed over the first two years of the SGRF. Zambia showed a similar result. To some extent, that poor outcome has been the result of overoptimistic planning by the road boards, whose capability to plan far outstripped the capability of the executing agencies to perform. However, it might be expected that performance would continue to improve as implementation capability expands.

Disbursement performance also differs significantly by type of road, with rural roads having very low disbursement rates, usually less than half of that for main highways and urban roads. Several factors contribute to this, including:

- Poor executing capability in rural areas;
- Political pressure to maintain roads in the capital cities. For example, in Zambia the Government put much emphasis on the rehabilitation of roads in Lusaka in the first years of the road fund, and 50% of the allocation in Malawi for 1999 was to the cities of Lilongwe and Blantyre - half of which was for rehabilitation;
- Low representation of rural road interests in most road boards.

Implementation capability

SGRFs appear to have helped insulate road maintenance contracting and payment arrangements from financial uncertainties. This has facilitated improvements in work programming and the emergence of a domestic contracting industry, which has brought efficiency gains in resource use. In Zambia, the number of local contractors has increased from four in 1994 to 450 in 1999 and local consultancy from six to 20 over this period. The share of maintenance works contracted out has increased to almost 90% in Zambia and Ghana.

It must be noted, however, that the effect of SGRFs on private participation is indirect as they are not themselves the implementation agency. Change depends also on reform of the agency function, which may be slower to occur. For example, though road maintenance expenditure in Ethiopia more than doubled over the first five years of the SGRF, and the Ethiopian Roads Authority (ERA) sought to improve its effectiveness by establishing commercial operations in maintenance districts and jointly implementing performance contract agreements with the RFA, road maintenance continued to be carried out using force account. Similarly in Benin, the share of maintenance works carried out using force account has declined from 47% in 1997 to 40% in 1998.

Operational efficiency

In terms of operational efficiency, improved contract management and disbursement arrangements have resulted in a reduction in road maintenance costs per kilometre by 10% to 20% in Zambia, Ethiopia and Ghana. However, SGRFs did not immediately eliminate all operational problems. Evaluation of maintenance works carried out in 2000 in Zambia and Kenya revealed limited local capacity, technical constraints and inability to manage contractual arrangements. In Kenya, the absence of a fully functional maintenance management system makes it difficult to ensure that the maintenance budget is correctly allocated, despite what appears to be an adequate supply of small contractors to undertake minor works, with the consequence that funds have been allocated to roads with little economic priority.

Roads network condition

Absence of reliable time-series data on road conditions limits the precision with which changes in the quality of the network can be estimated, and the statistics should be treated cautiously. Nevertheless, the general picture after the introduction of the SGRFs is promising. In Zambia, the impact of reforms in the roads sector in terms of improved quality of the road network is substantial and there appears to be in place a sound strategic framework to reverse the deteriorating trend and address the neglect of past decades. In Ethiopia, the proportion of main roads in "good" condition has increased from 15% in 1996 to 25% in 1999. In Ghana, the proportion of "good" roads has increased from 21% in 1997 to 30% in 1999. Nevertheless, by 2004 the proportion of roads in bad condition was still 35% in Ghana and 38% in Ethiopia (see Annex 2). Progress remains very slow.

Also on the down side, the quality of feeder/rural road network continues to deteriorate in some countries. This is partly a reflection of an inadequate planning and programming framework and partly a lack of capacity in the regional administrations. Years of neglect has limited the capacity of the road agencies to carry out maintenance works, a deficiency most apparent in rural and feeder road agencies. Gains in productivity have been registered only when the road funds were instrumental in fostering the outsourcing of works and services with private suppliers. (It should be noted, however, that this has been achieved without the establishment of RFs in some countries such as Albania and Serbia.) Revenue-raising through road funds should thus match absorptive capacity rather than identified maintenance expenditure needs. Although the ability of the contracting sector to absorb the allocated funds has improved, it is still not adequate in several countries. Despite these limitations, costs of maintenance have been reduced and the long-term trend of decline in road quality has been arrested and in some cases significantly reversed.

5.1.4. Conclusions on SGRFs in Africa

It may seem too early to jump to a final judgment on the "second generation" road funds. But some interesting patterns can already be discerned which are of relevance to the traditional arguments on the topic.

First, it is clear that, whether the boards have a majority of private sector members and a private sector chairman or not, the freedom that they have to redistribute national resources from other sectors to road maintenance is severely constrained. Most countries are still not able to fully fund their desired levels of road maintenance because of residual controls by their ministry of finance over the level of the fuel tax levy. In that sense the worst fears of the macroeconomists have not been realised. The use of those funds has improved. Although delays still occur, transfer of approved revenues to the roads account is now more secure. The funding arrangements are more transparent, with financial audits working well in most cases. Though formal allocation procedures are still not well developed and technical auditing is limited, the arrangements are much better than before the establishment of SGRFs in most cases. Despite a continuation of limitations on overall funding, there is already evidence of increased efficiency in implementation, resulting not from the availability of full funding for desired programmes but from the increased security of a sufficiently high proportion of finance on a multi-annual basis to maintain continuous work programmes and to effect payments in a timely fashion. There is no strong and systematic link between the form of the fund (user majority on boards, private sector chair, etc.) and their performance (reduction in costs, improvement in road condition). Even continued reliance on the budget for a substantial part of funding has not been a particular impediment.

The impediments to improvement are increasingly seen in implementation. Many countries remain unable to disburse funds that are available because of the low absorptive capacity of the maintenance contracting sector. The capacity and effectiveness of implementing agencies has developed least rapidly, though there are signs of progress, particularly in respect of private contracting capability, to which the MFIs are increasingly turning their attention. As reform of works execution is to some extent a consequence of changes in their financing, planning and monitoring, improvements in execution are likely to be lagged. But improved execution also depends on the willingness of governments to take reforms further than the establishment of a road fund and board. As the experiences of Uganda, Serbia and Albania show, it is possible to improve works execution without the creation of a second-generation road fund. But that can only be achieved by the adoption within the traditional structures of many of the procedures and processes (separation of implementation from financial control, improved managerial procedures, extended private sector contracting, etc.) associated with the new road funds. Improvement in performance is thus largely independent of either structure or resource allocation but highly dependent on some critical aspects of process.

In the event, governments have been willing to recognise the advantages of most of the process change brought by the second-generation road fund argument, without yielding the principle of their ultimate control over either taxation or expenditure responsibilities for an important public good. The outcome has often been surprisingly favourable in terms of the efficiency of maintenance expenditures, quite irrespective of whether those expenditures met the needs as assessed by the sector. But it must be repeated that it is still too early to judge conclusively; the evidence on road network conditions presented in Annex 2 shows that road conditions remain generally very poor in the sub-Saharan African countries, whether with road funds or without them.

5.2. The Latin American road funds

In 2001, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) undertook a review of experience of RFs in Latin America (Metschies and Zietlow, 2001). Detailed studies were undertaken of Honduras, Guatemala, Costa Rica, Nicaragua, El Salvador and four states of Brazil. Some general characteristics of these funds are set out in Table 5.

Table 5. **Characteristics of road funds in Latin America, 2001**

Country/ Date of creation	Main sources of income	Fuel levy in US cents per litre	Fuel levy % of total income	Composition of board public-parliament-private
Costa Rica 4/1998	Fuel levy	7.5 - 4.3 25% for local road fund	95	4 - 0 - 3
Guatemala 2/1997	Fuel levy	3.1 - 3.1	100	3 – 0 - 3
Honduras 1/1999	Fuel levy	8.2 – 2.6	100	4 – 0 - 3
Estade de Mato Grosso do Sul 8/1999	Fuel levy and taxes on agricultural goods	0.4 – 0.8	50	5 – 1 - 2
Estade de Mato Grosso 3/2000	Fuel levy and taxes on agricultural goods	1.5 – 1.5	17	7 – 0 - 0
Nicaragua 6/2000	Federal budget	---	---	3 – 0 - 3
El Salvador 11/2000	Fuel levy	5.3 – 5.3	100	2 – 0 - 5
Estado de Paraná 12/2000	Fuel levy	0.4 – 0.8	100	6 – 1 - 9
Estado de Goiás 1/2001	Vehicle licensing fee	---	---	3 – 1 - 6

Source: Zietlow, 2004.

5.2.1. Structure

In structural terms, the Latin American SGRFs all do involve the private sector, sometimes as a majority, but usually retain the chairmanship and even the formal power to appoint the private sector representatives for the public sector, usually in the body of the minister responsible for road construction. Some examples are shown in Box 1.

Box 1. Membership and Characteristics of Some Road Fund Boards in Latin America

Road Maintenance Fund of Honduras
The Board has executive functions and consists of 7 members: three ministers or vice ministers (Transport and Public Works, Finance, and Economy), the Director of Roads of the Ministry of Transport and Public Works, one representative of the Association of Municipalities, and three from the private sector (Chamber of Commerce, Association of Transport Enterprises, and College of Engineers). The Minister of Transport and Public Works is chairman of the Board and appoints the representatives of the private sector upon nomination by the organizations they represent. The Board contracts out all execution and supervision of road maintenance to the private sector and uses the Ministry of Transport and Works for planning purposes.

Road Maintenance Fund of Guatemala
The Board has some executive functions and consists of 6 members: two ministers or vice ministers (Transport/Public Works and Finance), the Director of Roads of the Ministry of Transport and Public Works, one representative from the Road Transport Association, one representative of the Chamber of the Construction Industry, and one representative of the Chamber of Agriculture. The Minister of Transport and Public Works is chairman of the Board and appoints the representatives of the private sector upon nomination by the organizations they represent. The Board contracts out all execution and supervision of road maintenance to the private sector.

Road Maintenance Fund of the State of Paraná (Brazil)
The Board has executive functions and consists of 16 members: three Secretaries of State (Transport/Infrastructure, Industry/Commerce, and Agriculture), Director General of Public Works, one representative from Parliament, one representative from the municipalities, and ten representatives from the private sector (Agricultural Federation, Federation of Industries, Federation of Freight Transport Enterprises, Federation of Passenger Transport Enterprises, Chamber of Commerce, Transport Syndicate, Syndicate of Transport Related Services, Federation of Agricultural Workers, Syndicate of Freight Transport Enterprises, and a road user selected by the Consumer Protection Agency). The Chairman of the Board is the Secretary of Transport. The Board contracts out all execution and supervision of road maintenance to the private sector

Source: Zietlow, *op. cit.*

5.2.2. Process

The process impacts of Latin America vary from country to country, but show some general characteristics which in some respects differ from those of Africa. The general characteristics include:

- The spatial scope of the fund varies from country to country. In Honduras, the board was initially responsible only for the main network, but will take over responsibility for the whole network by 2009. In Costa Rica there are two road funds, with 75% of the earmarked proportion of fuel taxes (30%) going to the national road fund and 25% to municipal roads.

- The funds have varying responsibility for rehabilitation and network development as well as maintenance. For example, in Costa Rica the fund has full responsibility for all rehabilitation and improvement of the national road network, while in Nicaragua only up to 10% of the fund's annual budget can be spent on minor rehabilitation work. The Brazilian RFs in Paraná and Goiás are exclusively concerned with maintenance, but those in the Mato Grosso states have broad responsibility for road network development.

– Most funds are predominantly dependant on fuel tax for their revenue. The main exceptions to this are the agricultural states of Brazil (Mato Grosso and Mato Grosso do Sul) where the RF is responsible for road construction and rehabilitation as well as maintenance and is funded by levies on agricultural goods as well as on fuel.

– In most cases, all works are contracted out to the private sector, usually by obligation of the founding statute of the fund. Attached to this requirement is often a limit on the proportion of revenues which can be used for administration (2% in Guatemala).

– Funding is now going directly into the accounts of the road fund, except in the State of Goiás, where funds are still channelled through the Secretary of Finance.

5.2.3. Outcome

The conclusions on outcome in Latin America were very varied. In *Honduras* the introduction of a road fund stabilised maintenance finance and promised to increase funding by 30% over a period of eight years and to increase the number of kilometres of road maintained by the fund from 5 743 to 14 602 over the same period. In *Guatemala* the introduction of a fund in 1996 led to an increase of maintenance expenditure from $30 million to between $70-80 million over the next three years. In both countries the costs of administration were limited at a low level (2-2.5% of turnover), and works were executed by the private sector. The proportion of roads in good condition in Guatemala increased from 30 to 85% over a period of four years. In *Costa Rica* experience was less successful. Only half of the funds which should have been transferred from the Ministry of Finance as required by the National Road Fund Law were in practice transferred, and though some improvements were achieved, roads in general remained in a poor state. The introduction of funds in Nicaragua, El Salvador and the three Brazilian states of Mato Grosso, Mato Grosso do Sul and Paraná, were too recent for changes in outcome to be reliably assessed, but certainly in Nicaragua and El Salvador current fiscal conditions limited the extent to which government was willing or able (by persuading parliament) to increase funding significantly.

5.2.4. Conclusions on SGRFs in Latin America

The general conclusions reached in the GTZ review were that the establishment of road funds was important, but only part of the general issue of re-establishing financial discipline in the countries. In periods of intense fiscal pressure, road maintenance charges can only be raised gradually and, unless the road fund initiatives are spear-headed by a very high-ranking minister and associated with a financial stabilisation strategy approved by the ministry of finance, they were likely to fail. Here again, as in the World Bank review, the conclusions appear to be that it is unrealistic and unproductive to try to force the complete independence of the RFs from government tutelage, and that it is attention to process -- including the introduction of more sophisticated means of assessment of expenditures and competitive private sector implementation -- which matters most.

6. CONCLUSIONS

6.1. On road funds in developing countries

The foregoing analysis supports a number of important conclusions:

a) Quasi independent road funds in developing countries have been established primarily as a response to the perceived under-funding of maintenance and their activities have therefore usually concentrated on the funding of routine and periodic maintenance, and sometimes extended to reconstruction and new construction.

b) SGRFs constitute an intermediate arrangement between the re-establishment of good governance of a public sector agency and the complete commercialisation of road funding and financing.

c) Increased budget allocations for maintenance following the establishment may reflect the recognition of the problem of funding which would have implemented with or without the establishment of the fund and are not an inevitable consequence of the establishment of an SGRF. There appears already to be some evidence of a shift back to emphasis on rehabilitation and new construction due to its easier availability of finance from the international institutions.

d) Ministers of finance and often ministers of public works have usually maintained their influence and power in overall levels of road funding, whatever the formal composition and powers of the road board which manages the SGRF.

e) As reform of works execution is to some extent a consequence of changes in financing, planning and monitoring, improvements in execution efficiency may be lagged, and depend on the willingness of governments to undertake even broader reforms of public agencies.

f) Improved disbursement of funds allocated and improved effectiveness of funds disbursed have arisen as the consequences of the possibility of multi-annual budgeting of the SGRF by the board and improved selection, supervision and auditing arrangements.

g) The traditional macroeconomic objections to RFs can be addressed by (i) retention of a significant MoF influence on overall funding decisions, and (ii) demonstration at the specific nation level of the impacts of increased road user charges on changes in taxable capacity.

h) The objective of improving the reflection of costs in charges/taxes implies that both fuel tax and vehicle-specific annual taxes should be included in the revenue sources of the SGRF, and the balance between them analysed and set in the light of their effects on efficient incentives for operators.

6.2. On the relevance of developing countries' institutions for industrialised countries

SGRFs in developing countries have been established in recent years as a response to a perceived failure in governance, namely the very myopic under-funding of maintenance. For that reason they have tended to be strictly limited to the funding of periodic and routine maintenance, and have usually been put at arms from the implementation institutions. They have usually existed in countries where traffic volumes were not sufficient to yield enough in toll revenues to justify the complete commercialisation of primary road networks. They have also tended to exist in countries which still rely heavily on funding for capital investment in new roads from the various international financial institutions, and the latter have required evidence of willingness to face the problem of maintenance financing as a condition of further lending for rehabilitation and new construction.

These conditions differ significantly from those in most ECMT countries, which would be characterised by:

- Well-managed budgetary processes, with mechanisms for intergovernmental transfers ensuring that fiscal allocations match the distribution of responsibilities between jurisdictions;
- Co-location of responsibility for construction and maintenance at the same jurisdictional level;
- Reasonably well-maintained infrastructure;
- Well-established private sector contracting capability.

Because of this, the weakness of the fiscal mechanisms for ensuring an adequate revenue flow for basic road maintenance is less likely to be the core problem in ECMT countries than those of adequately integrating construction and maintenance mechanisms and of mobilising private sector business skill in system maintenance. Hence it is arguable that the concern in ECMT countries should be more directed to the establishment of business-oriented road agencies, combining implementation with financing responsibilities. While road tolls and other revenues may be paid directly into the fund, this will tend to be not for revenue security reasons as much as for administrative convenience. The main source of funding could still be directly budget controlled.

ANNEX 1

Formal status of road funding administration in 28 countries, August 2004

26-Aug-04

Countries	Road Fund									
	Established?	Creation Date	Has a Board?	Board with Private majority?	% Share of Road Fund resources from user charges	% Coverage of routine maintenance needs from all sources(3)? (4)	% Coverage of Total Maintenance needs (3)	Fuel Levy US cents/liter Gas	Diesel	Direct channeling of Road user charges?
Angola	yes	-	no	-	-	yes				no
Benin	yes	1997	yes	no	90	yes				no
Burkina Faso	no	(4)	-	-	-	-	-	-	-	-
Burundi	yes	2001	yes	no	95	no	no	2	2	yes
Cameroun	yes	1998	yes	yes	90	no	no	4	8	no
Cape Verde	yes	1999	yes	no	100	no	40	0% of the taxe		no
CAR	yes	2000	yes	no	80	no	no	9	9	yes
Chad	yes	2000	yes	yes	90	70	no			yes
Cote d'Ivoire	yes	2001	yes	no	30	60	no	7	2	yes
Ethiopia	yes	1997	yes	no	40	100	80	1.50	1.03	yes
Gabon	yes	1997	yes	yes	46	no				no
Ghana	yes	1997	yes	yes	100	80	40	4	4	yes
Guinea	yes	2000	yes	yes	100	no	no	2.6	2.6	yes
Kenya	yes	2000	yes	yes	100	yes		7	7	yes
Lesotho	yes	1996	yes	no	100	yes	30			no
Madagascar	yes	1997	yes	no	50	100	50	5	5	yes
Malawi	yes	1997	yes	yes	80	50		6	5	yes
Mali	yes	2000	yes	yes	19	75	52	0.6	0.6	yes
Mozambique	yes	1999	yes	no	>95	yes	85	9	8	no
Niger	yes	1999	yes	yes	20	no	no	6	6	no
Nigeria	no	-		no	-	-				-
Rwanda	yes	2000	yes	no	100	no	30	3	3	yes
Senegal	no	-	-	-	-	yes	no			
Tanzania	yes	1998	yes	no	90	90	50	9	9	no
Togo	yes	1997	yes	yes	80	60	50	7	7	yes
Uganda	no	-				100	60			-
Zambia	yes	1994	yes	yes	95	35		8	8	no
Zimbabwe	yes	2001	yes	yes				2.5	2.5	yes

Source: SSATP-RMP, The RMI Matrix.

ANNEX 2

Process and performance in 28 African countries, August, 2004

Countries	Allocation of RMF resources			Road Agency (2)				Road Network Condition (3)	
	Main network	Rural network	Urban network	Established ?	Since When?	Has a board?	Private Majority?	% in good / fair	%in bad
Angola	70%	20%	10%	yes	1990			30%	70%
Benin				no		-	-	75%	25%
Burkina Faso	-	-	-	no		-	-	57%	43%
Burundi	80%	20%	-	no				5%	95%
Cameroun	75%	12%	13%	no	-	-	-	60%	40%
Cape Verde				yes	2003	yes	no	58%	42%
CAR	90%		10%	no				75%	5%
Chad	90%	0%	10%	no		no		30%	70%
Cote d'Ivoire	100%			yes	2001	yes	no	60%	40%
Ethiopia	70%	20%	10%	yes	1997	yes	no	62%	38%
Gabon	-	-	-	no					
Ghana	47%	23%	25%	yes	1974	-	-	65%	35%
Guinea	68%	25%	7%	no		-	-	66%	34%
Kenya	57%	40%	-	no		-	-	57%	43%
Lesotho	(5)	80%	20%	no (7)		-	-	96%	4%
Madagascar	80%	5%	15%	no		-	-	30%	70%
Malawi	60%	20%	20%	yes	1997	yes	yes	63%	37%
Mali	80%	20%	-	yes	2005 (8)	yes	yes	44%	56%
Mozambique	70%	20%	10%	yes	2001	yes	no	70%	30%
Niger	90%		10%	no				57%	35%
Nigeria				no		-	-		
Rwanda	100%	0%	0%	no		-	-		
Senegal				yes	2002	yes	no		
Tanzania	70%	20%	10%	yes	1999	no	no	50%	50%
Togo	70%	10%	20%	no	-	-	-		
Uganda				yes	1998	no	no	81%	19%
Zambia	33%	55%	12%	yes	2004 (8)	yes	yes		
Zimbabwe	39%	45%	16%	no				80%	20%

Source: SSATP-RMP, The RMI Matrix.

BIBLIOGRAPHY

Brushett, S. (2005), "Management and Finance of Roads in Africa", *SSATP Discussion Paper No. 4, Road Management and Financing – RMF Series*, World Bank, Washington DC.

Chipewo, Henry (2003), "National Roads Board of Zambia – Origin of the Board and involvement of Road Users in the Decision Making Process", Presentation to PIARC World Congress, Durban, South Africa, October.

Deran, E.Y. (1965), "Earmarking and Expenditures: a Survey and a New Test", *National Tax Journal,* No. 18, pp. 354-61.

Dick, M. (1989), "Earmarking of Transport Funds in Colombia", in: *Earmarking, Road Funds and Toll Roads. A World Bank Symposium*. Report INU 45, World Bank, Washington DC.

Eklund, P. (1967), "Earmarking of Taxes for Highways in Developing Countries", *Economics Department Working Paper No. 1*, IBRD.

Frost, Michael (2001), "Imperatives in Future Road System Management: The Australian Case", Paper presented to 14th IRF World Road Congress, Paris.

Gupta, S. and W. Mahler (1995), "Taxation of petroleum products: theory and empirical evidence", *Energy Economy 17* (2):101-116.

Gwilliam, K.M. and Z. Shalizi (1999), "Road Funds, User Charges and Taxes", *World Bank Research Observer 14,* August, pp. 159-185.

Gwilliam, K.M. and A.J. Kumar (2003), "How effective are Second Generation Road Funds?: A preliminary appraisal", *World Bank Research Observer 18 (1)* Spring, pp. 113-128.

Gyamfi, P. (1992), *Infrastructure Maintenance in LAC: The Costs of Neglect and Options for Improvement, LACTD Regional Studies Program Report No. 17,* World Bank, Washington, DC.

Harral, C. and Asif Faiz (1988), *Road Deterioration in Developing Countries: Causes and Remedies,* World Bank, Washington, DC, June.

Heggie, I. (1995), *Management and Financing of Roads: An Agenda for Reform,* World Bank Technical Paper 275, World Bank, Washington DC.

Heggie, I. and P. Vickers (1998), *Commercial Management and Financing of Roads*, World Bank Technical Paper 275, World Bank, Washington DC.

Kumar, Ajay (2000), Assessment of Selected Road Funds in Africa. Case Study of Benin, Ethiopia, Ghana, Kenya and Zambia, *SSATP Working Paper No. 51*, World Bank, Washington DC.

Metschies, G.P. and G. Zietlow (2001), "Introducing Road Funds in Latin America -- the GTZ experience". Paper presented to the ADB Regional Workshop on Road Funds Strategy, Manila, March.

Potter, B.H. (1997), "Dedicated Road Funds: A Preliminary View on a World Bank Initiative", *Paper on Policy Assessment of the International Monetary Fund, PPAA/97/7,* International Monetary Fund, Washington DC, June.

Premchand, A. (1983), *Government Budgeting and Expenditure Controls: Theory and Practice,* IMF, pp. 158-160, Washington, DC.

Teja, R.S. (1988), *The Case for Earmarked Taxes: Theory and an Example,* IMF, Washington, DC, February.

World Bank (1986), *Road Funds from Earmarked Sources: Interim Guidelines.* World Bank Transportation Note 1, Washington, DC, .May.

World Bank (1989), *Earmarking, Road Funds and Toll Roads.* A World Bank Symposium Report INU 45.

Zietlow, Günter (1998), *Cutting Costs and Improving Quality through Performance Specified Road Maintenance Contracts. Pilot Experiences in Latin America offers lessons for Africa,* SSATP Note No. 14, Road Management Initiative.

A more detailed bibliography of individual country experience of road funds is contained in the Asian Development Bank document, *Road Funds and Road Maintenance,* ADB, Manila, 2003.

BUDGETING FOR ROAD MAINTENANCE

Barry POTTER
International Monetary Fund - IMF
Washington D.C.
UNITED STATES

SUMMARY

Washington, July 2005

1. INTRODUCTION AND BACKGROUND

The maintenance of highways and main roads is not a subject with universal appeal. For the politician, it is unlikely to be an area where flights of rhetoric or political commitment will change the outcome of elections. For most bureaucrats, even in these days of tighter measurement of or targets for individual performance, it is perhaps not the most promising route for advancement. For most consumers of road services, it is widely taken for granted – except on those rare occasions, where some belatedly seen pothole damages a tyre, wheel or axle.

Against such a background, it is hardly surprising that, in the past, those with responsibility for maintaining roads have found the budgetary arrangements lacked imagination or initiative. The budget regime was the subject of a familiar litany of the standard complaints about public sector provision: too restrictive because road maintenance was too low a priority, and just too easy to postpone for one more year. Those responsible for delivery complained, as public servants often do, that they found themselves caught between too many regulations and too little resources. Many have searched for imaginative ways of ensuring that the roads are well maintained. For, underneath this unpromising political and institutional backdrop, there is a general (if theoretically inaccurate) acceptance that the maintenance of high-quality roads is in the nature of a public good, with all of society benefiting from facilitating the movement of people, goods and services, etc. Nowhere is this more important than in less developed economies. Just maintaining the major roads is not an avenue to a breakout into economic prosperity; failing to maintain them, however, is likely to thwart progress, even when the other preconditions have been met.

The purpose of this paper is to examine the evolution of budgetary arrangements for road maintenance. In particular, the paper describes the evolution of so-called road funds that provide for road maintenance in an increasing number of developing countries. While such ideas have been taken even further in the approach to road maintenance in New Zealand, the concept has attracted little attention in other industrialised countries. The paper considers whether there are any lessons to be learned – both for developing and industrialised countries – from the evolution of present institutional frameworks governing road maintenance, that may have some wider implications.

From the adoption of the main roads as public roads in Europe, dating back to the 18th and 19th centuries, road maintenance – at least for major highways – was provided by government and financed from general taxation resources. The paper thus first reviews the theoretical case against the use of earmarked, rather than general, taxes for the provision of road maintenance and then briefly traces the emergence of the first generation earmarked road funds. The reforms that led to the introduction of second generation road funds under the World Bank's Road Management Initiative are then set out. A brief review is undertaken of the relatively more successful outcomes on road maintenance, which seem to have emerged following the introduction of the second generation of road funds. But their success has not been replicated, either in former Soviet Union or East European countries or (generally) among OECD members. The exception – the innovative approach in New Zealand – is briefly described. Finally, the paper asks whether, drawing on both the experience of these second generation road funds and that of New Zealand, there are reasons why a similar approach has not been adopted in industrialised countries; speculates on why the approach remains more traditional; and asks what the future may therefore hold for the road funds now established in the developing world, and for more innovative financing of roads in industrialised countries.

2. EARMARKING AND THE TRADITIONAL APPROACH

Prior to the emergence of the first road funds in the 1950s (though some earlier earmarked arrangements date back to the 1930s), the maintenance of main roads was a responsibility of central government and funded from consolidated general government receipts. Arrangements varied across countries, depending on how far national roads and highways were the responsibility of central versus either state governments under a federal system or local governments. But the model was largely the same: public funding and public sector provision.

In theory, there is an argument to be made against regarding the provision and maintenance of major roads as a public good. The classic definition of public goods is that where, if the service is provided by the State, no individual can be excluded, the best-known examples being national defence and perhaps foreign relations. Clearly, use of the roads can be charged for – indeed they already are in the form of vehicle registrations, etc., and in countries like the United Kingdom there were extensive private toll roads prior to their adoption as public roads. But the wider view of public goods has been generally invoked in the case of roads – that the externalities of maintaining good roads for the community as a whole exceed the private benefits to the road user and thus justify public sector provision.

Once the case was accepted that roads were a public service, up until the early 1960s, for economists, the arguments for funding all public services from general revenues and avoiding earmarking of particular revenues for particular public services were well-established and largely went unquestioned. An article 40 years ago (Deran, 1965) summed them up as follows:

Earmarking:

(a) can hamper effective budgetary control;

(b) can lead to misallocation of funds, giving excess revenues to some public services, while others are underfinanced;

(c) makes the revenue side of the government accounts inflexible, denying legislatures the chance to reshape the tax regime to make it better in the sense of more optimal;

(d) tends to ossify therefore both the structure of public spending and the tax regime; and

(e) infringes on the policymaking powers of central government, and the legislature by removing a part of fiscal resources from their purview.

A quick inspection reveals the negative nature of many of these criticisms: they lay out what damage earmarking can do, and several of the arguments are essentially practical rather than economic in nature. The corollary, however, was that funding of individual public services from general taxation was a superior model, which lacked any of the theoretical or practical weaknesses of earmarking. Underlying such a view was a particularly, and then wholly orthodox, benevolent view of public expenditure. It was very much that of the social planner attempting to maximize a well-defined social welfare function – indeed a welfare function on which government, legislature, the electorate and the bureaucrat were universally agreed. By the late 1960s, that model was coming under increasing attack.

The first element of that attack came from the emergence of the public choice school. This was founded on the realisation that voters are presented with a package of choices at elections, in effect a whole bundle of public goods, by the different political parties. The voter must choose (to characterize the situation a little too simply) between one particular package and one level (and mix) of taxation and one or more other packages and associated sets of taxes. To give a recent example, the author does not believe that specific provisions for major road maintenance featured strongly, if at all, in any of the election manifestos of the three major parties in the May 2005 UK elections, though he is told it rated an honorable mention by the party that came in third. While subsequent literature suggested that economic outcomes were not necessarily any less optimal from this "packaged" approach, at the very least, it cast doubt on whether the right decisions, in economic welfare terms, were necessarily being made on each and every individual service within the overall package.

Secondly, more attention began to be paid to the primary countervailing economic argument: that the use of earmarked taxation could be justified by invoking the benefits principle of taxation. This argues that taxes should be borne, wherever possible and practical, by those who most benefit from the associated expenditure. Thus, in such situations where the beneficiary group can be clearly identified but the implementation of a user charge is administratively difficult, an earmarked tax can be justified as a second-best instrument of finance. In short, the benefits principle argued that, where there is an earmarked tax for a discrete function with identifiable beneficiaries, it can lead to a better outcome than one in which the resources available for a particular function, in relation to actual electoral preferences, cannot be clarified. Two important points to note are that earmarked taxes are themselves in such a situation only second best to a direct user charge – an issue to be considered in more depth later; and secondly, that even though the main beneficiaries can be identified (who incur lower private costs as a result of the service) there still may be wider social benefits that do not in themselves argue for funding from public revenues. A corollary is that any such good or service can, by definition, not be a "pure" public good, although fiscal economists have long established that "pure" public goods are rare.

Furthermore, there is another externality-related argument for an earmarking approach that is somewhat specific to road maintenance and often overlooked. One dimension of externality is that when an individual uses the road, it causes damage which is suffered by others, and is not paid for by that individual. The idea that we can then bring about a Pareto improvement by taxing road use – the standard Pigovian tax argument – requires that those who lose out as a result of the tax should be compensated from the proceeds of the corrective tax itself. Even with something less strict than Pareto optimality, political economy also suggests that compensation will be important – especially when the losing group is politically powerful/numerous. In theory, the proceeds should also be returned to those who pay it as a lump sum amount, i.e. one they perceive to be independent of their own actions. In practice, that cannot be done: but one way to finesse the compensation problem is to use the proceeds instead to provide some form of in-kind benefit, that is valued only/mainly by those who use the tax; which would be a form of earmarking.

Third, doubts grew by the early 1970s about whether the bureaucracy was capable of delivering the preferences (leaving aside whether the bundling problem allows these to be properly articulated), of not just the electorate, but also of the executive and legislative branches of government. The executive itself, not least in the everyday political arguments about priorities within its budget, may for political reasons give lower priority to road maintenance than it should. Then the legislature, for the same reasons and, further, often handicapped by less knowledge than the executive branch, may also make skewed decisions. Finally came recognition of the role of the bureaucrat – perhaps in terms of personal advancement or defending entrenched models of delivery in making choices amongst competing claims on resources. This further weakened faith in traditional methods of public sector delivery and thus strengthened not only the case for earmarking but arguably also for taking provision, as distinct from financing, out of the hands of the public sector.

To summarize, the normative case for earmarked taxes takes into account that groups and individuals in society are likely to have different preferences and that, where the principal beneficiaries of a particular public service are identifiable, earmarking may provide the electorate with a better opportunity to reveal their preferences than when such services are bundled into the wider package. This argument is further strengthened where there may be both imperfections in translating electoral intentions into programmes and one or more conflicts of interest between the executive and legislative branches and the bureaucracy.

While the case for financing public services wholly from general tax revenues, even for functions like road maintenance, generally tended to be accepted in the international financial institutions, some cracks were appearing by the 1980s as explained below – see Teja (1988). Even, however, as late as the mid-1990s, the conventional wisdom in the IMF at least was against the use of earmarked taxes.

3. THE EMERGENCE OF ROAD FUNDS

The starting point for the development of earmarked road funds in the late 1980s and early 1990s was the general observation that the volume (and also the quality) of maintenance for major roads was inadequate in developing countries; that the condition of roads in such countries was deteriorating as a result; that budgetary problems and pressures had contributed to this, both through under-allocation relative to need in the budget process, and through the then subsequent diversion of monies away from road maintenance during execution of the budget. These arguments were particularly forceful in Sub-Saharan Africa, where in many countries road conditions were visibly deteriorating as a result of inadequate maintenance, with more widely damaging effects on the economies. It was often easier to attract capital for financing the building of new roads (for example, from donor institutions) than to find the resources to maintain the existing ones. The first generation road funds were thus essentially attempts to insulate road maintenance from the impact of wider budgetary pressures.

However, the first generation road funds were not so much genuine and separate funds as an earmarking arrangement for the financing of a public-sector-provided function. Essentially, they were an earmarking of particular revenues, typically, some or all of fuel taxes only, for the purposes of providing for road maintenance expenditures: thus, they were designed by, as it were, sectioning off one part of generalised tax revenues, to get around the inadequate allocation of monies to road maintenance under general budgetary procedures. But there was no attempt, under these first generation road funds, to reform how road maintenance was delivered (that is, whether by direct labour organisations within the public sector or contracted out to the private sector). Nor was there much attempt to make explicit a connection between the rates of taxation on fuel and the levels of road maintenance provision. Indeed, some road funds were topped up with general tax revenues, and some received less than (and a changing proportion of) the relevant tax base. Moreover, the intended insulation of the monies for road funds from wider budgetary pressures often proved more theoretical than real. As before, road funds, i.e. the resources set aside for road maintenance, were "raided for other purposes". Sometimes this was done explicitly, other times monies seemed to get stuck in some consolidated fund or treasury arrangement. In some of the worst cases, the taxes that were part of the tax base – largely but not exclusively – were often diverted on a sustained basis to other uses altogether. Some, particularly African economies, found that fuel duty was one of the more reliable and buoyant sources of tax revenue and simply could not afford to have it fully diverted to one function.

On the basis of the practices and experiences of these first road funds, it would be difficult to avoid concluding that the opponents of earmarking were proved correct. Certainly, the practical experience of the first generation of road funds played to the traditional arguments of the IMF against the earmarking of general taxation revenues. The IMF had always been suspicious of such arrangements: there is a natural tendency in an organisation that deals with macroeconomic issues, to look for a consolidated set of fiscal accounts and to look at earmarked funds as a potential complication at best and as a means of downright evasion of fiscal discipline at worst. When one is seeking to assess the overall fiscal situation and to define the public sector on the broadest basis, it becomes essential to consolidate central, regional and local governments' fiscal positions, as well as those of every relevant public fund. Relevance here is defined on the basis of whether ultimately the government is seen as being liable for the debts of the relevant institution.

But the situation was worse to the extent that country authorities argued that such earmarked funds also took the resources out of the public sector altogether, such that the arrangement was defined as extra-budgetary – i.e. not relevant to the overall fiscal position. (There is a whole art/science in the particular definitions of the public sector for the purposes of IMF programmes which is skipped over here.) It was a short step from such treatment to the setting up of road funds as a way of avoiding the fiscal discipline which the IMF wanted to put in place as a part of its macroeconomic adjustment programme. Indeed it became clear that, in a very few cases, this had been the real motivation for establishing the funds.

4. SECOND GENERATION ROAD FUNDS

The second generation of road funds emerged in the 1990s, following the World Bank's roads management initiative (RMI). They sought, through the design, to avoid the many problems associated with the first generation model. While Gwilliam and Shalizi (1999) suggest that "the process was akin to (runs parallel with and principally for the same reason) the privatization of state enterprises," it seems more plausible that the new-style road funds were initially more driven by a desire to overcome the manifest problems associated with the first generation, than to privatise the road network. That said, there can be little dispute that the practical consequence was to go quite far in privatising the road maintenance function – even though no change in the ownership status of the roads themselves was involved.

The key features of these new second-generation road funds – which owe much to the vision of then World Bank official, Ian Heggie (Heggie, 1995) – were that the road fund would be financed wholly from earmarked taxes as a surrogate user charge. The fund would be responsible for financing the maintenance of major (interurban or so-called trunk) routes, not for the construction of new roads. Moreover, it would be a public-sector body – the roads would still belong to central government – but it would not deliver the service itself; rather it would contract out to the private sector the actual work on the maintenance of the roads. Other key details included the following:

(i) Independent boards would be created to be in charge of road funds; members would be appointed by government and a careful balance would be struck such that they would represent consumer, not just producer, interests; direct government representation on the board was allowed but not mandated and would not constitute a majority interest.

(ii) They would be proper separated funds, not just an earmarked budgetary account within the government sector; they would not be so much extrabudgetary funds as having a status which rendered that question irrelevant.

(iii) They would be financed wholly by user charges – typically by surrogates for use of road space and so far as possible for the damage done to road surfaces by different types of road user: the surrogates included fuel taxes, vehicle registration duties (differentiated by size of vehicle and sometimes linked to axle weights), cross-border tolls for heavy goods vehicles, again where possible linked to axle weights, etc.

(iv) There would thus be the classic separation of the purchaser and provider roles. The road fund would commission road maintenance, i.e. contract it out, rather than provide road maintenance itself.

(v) There would be an independent audit of the accounts. The board of the road fund would be responsible to the parliament – the roads were still owned by government; and capital expenditure would in most cases continue to be financed from general government revenues, borrowing, etc.

It is important to unbundle all the elements of the second generation funds and understand the motives behind them. First and foremost, in the mind of the Bank, was a need to overcome a budgetary problem. The earlier model had not succeeded in cracking the vagaries of the old budget system: the tendency to underallocate resources, or shift them during budget execution, had not been prevented. Thus, one element of the package aimed to find a way that would genuinely both define a source of financing that provided adequate resources and a status for the fund that prevented subsequent raiding during budget execution. To achieve the former required a board for the fund that would represent consumer and not just producer (i.e. government) interests. Second, however, was the recognition that, in order not just to get value for money from the resources made available for road maintenance but also to try and limit the scope for corruption, actual delivery needed to be taken out of the hands of government and, in many countries, the (notorious) direct labour forces employed. Thus, the insistence on contracting out the delivery: while that did not, by any means, eliminate the scope for corruption, it did make it more challenging. With the benefit of hindsight, therefore, the road maintenance initiative seemed less a bold decision to privatise the road network than a series of careful steps to try and overcome the budgetary and (related) corruption problems that bedeviled road maintenance at the time.

But the road management initiative was not, internally within the Bretton Woods institutions, an easy sell. Indeed, the early stages of the initiative in the mid-1990s were, in part, held up by the IMF's continuing concerns over earmarking in general and the potential use of extrabudgetary funds to evade fiscal discipline. As the Bank proceeded with the reforms on a pilot basis, there was an almost comical mix of the Bank creating or opening such funds, with Fund staff arriving sometime later and suggesting they be closed down. In short, the prevailing view remained within the IMF that earmarking was in general not to be supported and the (partly justified) prevailing suspicion was that this was another extrabudgetary route being exploited to avoid fiscal discipline. One would have thought that, even separated by the 100 feet of 19th Street in Washington, DC, the two institutions could have resolved the matter more effectively.

As already noted, the IMF's objections were rooted in long-standing concerns about extrabudgetary funds of any kind. But already there had been some internal research, notably the article by Teja (1988), which cast doubt on the traditional economic arguments against earmarking; though in all fairness, as demonstrated earlier, most of those arguments rested less on high principles of economics than on practical experiences. In a paper, subsequently published as a book by Jack Diamond and the author (1997), the IMF view had also moved to be more grudgingly tolerant of extrabudgetary funds by defining conditions under which they might be acceptable. However, our wording was cautious and our approach frankly sceptical, as can be seen from the two boxes extracted. The burden of proof was very much placed on the proposer of the fund, and the hurdles put in place by the second box, in particular, were somewhat challenging.

Pros and Cons of Extrabudgetary Funds

Pros

- Can increase efficiency by simulating private market conditions where levels and standards of service are linked directly to fees or charges.
- Can provide more consistent source of funds for expenditures that yield high benefits yet do not get much recognition (road maintenance expenditures are a primary example).

Cons

- Can result in a loss of aggregate expenditure control; such expenditure may be outside the control of the ministry of finance.
- Can distort allocation of resources by circumventing the budget process and review of priorities.
- Earmarked revenues can become entrenched so funding is no longer based on priority needs.
- Less transparency may lead to inefficiency and/or misuse of funds.
- Can facilitate rent-seeking and abuse of monopoly power.
- Leads to less flexibility at the margin to reallocate when budget is under stress.
- Is incompatible with good cash management practices.

Key Questions Concerning Extrabudgetary Funds

What is the purpose of the extrabudgetary fund? What is the rationale for keeping such a fund off-budget?

Financing Issues

What is the source of funding? Does the source of funding make sense; does it help to relate marginal benefits to marginal costs – for example, user fees? How are user fees determined; are there limits to prevent abuse of monopoly power (especially if demand is inelastic)? Are there general benefits (positive or negative externalities, public goods arguments) in addition to user benefits that justify support from general budget revenues? If there is a split, how is the share of financing determined? Is the source of financing an important government revenue, and can the government afford to lose the associated degree of flexibility in prioritising expenditures? Do earmarked revenues detract from the government's capacity to collect traditional revenues?

Expenditure Decisions

How are expenditure decisions made by the extrabudgetary fund? What use is made of cost effectiveness or cost-benefit analysis? Does the management of the extrabudgetary fund promote efficiency, for example, through quasi-market mechanisms or through mission statements, objectives, performance measures? How are consumer interests represented and taken into account in expenditure decisions? If governed by a board, is membership of the board biased toward certain needs – for example, regional needs?

Management Issues

Does the management of the extrabudgetary fund meet good governance requirements? Is it free of political interference or unduly influenced by suppliers or trade unions? Is it possible for funds to be diverted to other uses? Can these accounts be "raided" for other uses? Is the extrabudgetary fund independently audited?

How are the cash resources of the extrabudgetary fund handled? Does the government have access to these funds for overnight borrowing to minimize government borrowing needs? Does the treasury or ministry of finance have the legal right to reduce funds available for expenditure in extrabudgetary funds if the budget is under severe pressure?

The impasse on the second generation road funds was resolved, at least temporarily, by the author's own article of 1997. Predictably, in the cautious style of the IMF, this recommended a case-by-case approach – the classic approach to conceding the principle, while remaining vigilant, if not downright skeptical, on the practice. The article set out criteria for deciding when a road fund might be appropriate. These were as follows:

First, do the existing budgetary arrangements require an RF solution?

(a) For developed countries following the simple agency approach (i.e. the UK or Finland models), there should be no objection. Such countries will usually be seeking to capture the microeconomic efficiency gains from the agency model: the necessary macroeconomic, budgetary and financial management requirements are likely to be already in place. There should also be scope for such agencies to take on road capital, as well as maintenance expenditures.

(b) Where there is to be a user charge also, it would be useful to ascertain whether the user charge would compromise the scope for effective fiscal and budgetary management. In most developed country cases this is unlikely. Petroleum duties will typically form a relatively small part of the tax base. Any marginal loss of macroeconomic manoeuvre may well be less important than the gains from a clearer link between the level and quality of road services and petrol prices, particularly if the tax burden and tax system are relatively stable. Moreover, to the extent that full commercialisation can be envisaged as the long-term policy goal, then with road maintenance outside the public sector, the arguments about any interaction between the user charge and general revenue-raising capacity will ultimately not apply.

(c) For countries where existing budgetary arrangements are inadequate, there is a choice between (1) changing the budgetary arrangements (which may offer a quicker and more effective solution); and (2) putting a road fund (RF) in place, where there are adequate, minimum organisational and financing requirements. The requirements are discussed below. But for many countries, this may be quite a hard test to pass. Correspondingly, in many situations it may be better to concentrate efforts on improving budgetary procedures.

Can the necessary organisational requirements be defined? The minimum requirements would seem to be:

(a) that the RF is dedicated 100 per cent to the task in question – and is not simply a means of avoiding budget discipline, whether for roads or wider public expenditures;

(b) that it is in the form of an agency principally as a purchaser, not as a provider of services. Thus it should have, as a minimum, a mission statement, objectives, physical output indicators and total input cost envelopes, etc. Ideally, the service provision should come from the private sector: for many countries this will be hard to achieve;

(c) that the financial management system of the RF can handle the more complex tasks involved in managing total cost envelopes and associated accounting requirements. The government should have access to the cash balances for cash management reasons; and

(d) that there is a management board with a significant private-sector presence but genuinely free from a producer (whether supplier or trade union) interest; the criteria must be objectivity and impartiality.

The article concluded, in a predictably cautious IMF fashion:

> "To sum up, the desirability of dedicated road funds in the future needs to be assessed on a case-by-case basis. Provided the right conditions are met, road funds can be endorsed in practice, as well as in principle. The question is just how often the right conditions can be met."

That position was then endorsed by the rather more comprehensive study by Gwilliam and Shalizi (1999). This also favoured a case-by-case approach. But, interestingly, it also noted that "many

European countries with good governance maintain road systems through traditional public sector budgeting arrangements." This is an issue to be returned to later in this paper. But it was also the basis for a lingering difference of view both within the Bank and within the IMF. Were road funds an appropriate solution to the provision of road maintenance only when traditional budgetary arrangements were inadequate? If so, was the right approach to introduce the road fund – or would it be better to reform budgeting arrangements so that something closer to OECD practice might be introduced? Both the Bank and the IMF had raised this question – and left it hanging, where, in a sense, it remains to this day: and as will be suggested later, it will have to be addressed as developing countries mature further.

One final piece of recent history is also relevant. As the second generation dedicated road funds expanded through Africa, into Latin America (particularly Central America) and Asia, while there was some variation in practice, they tended very much to the same model. Typically, governments did split road funds from normal government operations. These road funds were financed principally from fuel duties, but also sometimes from vehicle excise, that is, motor registration fees, and sometimes linked to charges for heavy road vehicles, tolls, etc. While it would not be true to say that they faced no diversion of funds to other purposes, or that their monies never became stuck in the system, the scale of such abuses was limited and tended to decrease over time. With one or two halts and interruptions, there was also wider use of contracting out – even in African countries where the private sector is often less well developed. Problems of corruption did not disappear but they do seem to have moderated.

A case study by Kumar (2000) showed that, while there were differences in the experiences of the five African countries studied, the creation of second generation road funds had brought about a major improvement in road conditions and efficiency gains in its provision. This tended to confirm the earlier and understandably more tentative findings of Gwilliam and Shalizi. A GTZ project in Asia produced similar positive results, as did some work by the AsDB which looked at both Asian and Latin American experiences. However, much of the research is in effect a time series look at the impact of road funds on the provision of maintenance in each of a number of sample countries; more convincing, and more illuminating, might be a comparison of the experiences on road maintenance of those countries which did and did not set up second generation road funds. Moreover, it is not easy to discern from the research where the gains came from: was it through an improvement in allocative efficiency such that adequate budgetary funds were at least made available? Or was it that – again an allocative gain – less of the money was hived off through corruption? Or was it a productive efficiency gain: more was got out of the money by capturing the efficiency gains from private versus public sector provision? Whatever the reality, some of what was observed in Africa probably reflected general improvements in the provision of public services. As noted, it would be interesting to test whether there are cases of just as significant improvements on road maintenance delivered under the more traditional public service framework... .

Yet in some countries in central Europe and the former Soviet Union, the progress observed with second generation road funds was not quite fast enough. In the early 2000s, some final skirmishing took place between the IMF and the World Bank over road funds in Poland, Georgia and Russia. On the one hand, the IMF was content for a road fund to open up in Bulgaria in 1996, but, partly in response to pressures from the IMF, road funds in Russia, Lithuania and Georgia were closed as part of wider budgetary reforms. This was, however, not really a re-run of the earlier Fund/Bank differences of view. Many such extrabudgetary arrangements had been found to be particularly prone to corruption arrangements, particularly where they stretched back to the former regimes. Moreover, these particular road funds were in nature much more like first generation than second generation funds; that is, arrangements to escape from general fiscal budgetary discipline, including that which the IMF was seeking to apply as a condition for its loans, often, as noted, allied with providing opportunities for graft and corruption. The closure may have been precipitate, perhaps reform in one or two cases might have been better: but the reasons for closure did not stem from some fundamental rethink by the IMF on the desirability of earmarked road funds.

5. NEW ZEALAND

At the other end of the world (and indeed, as it were, the spectrum), is the experience of New Zealand. Starting from the 1983 reforms to its public sector, New Zealand has applied its skeptical and imaginative approach to the problems of public provision in general to that of road maintenance. From the abolition of the Ministry of Works and Development in 1988, road maintenance moved initially and temporarily to the Ministry of Transport, before becoming the responsibility of an independent agency, Transit New Zealand. In 1996, that was transformed into Transfund, a separate funding body for all land transportation.

New Zealand now has a system, under the national land transport programme, in which the road maintenance function is carefully fitted within the framework of an overall mission statement, a wider set of transport objectives and specific programmes to deliver those objectives. All financing is from road user charges, principally at present fuel excise taxes, motor vehicle registration and licensing fees. With a small amount subtracted and directed to road safety organisations, the remainder of these revenues flow into a single national land transport account. Nearly half of that account is in turn spent on road maintenance. This is a classic fund role, with the allocation based on each authority's strategic objectives, asset management plan, road pavement modelling and performance measures on its road network. The roads themselves are now the responsibility of Transfund, including ownership of the assets.

Thus New Zealand, perhaps more than any other country, has already moved effectively to fully privatise road maintenance, indeed road transport as a whole. Government agencies, to the extent that they are involved at all, function only as relatively minor intermediaries between a fund financed from its own hypothecated revenues, through to the private sector contractors that undertake the actual work on road maintenance. Moreover, New Zealand has already begun to look beyond the reliance on taxation as a source of revenue and to move directly to road user charges, even (for the future) involving imaginative satellite tracking systems. The small population relative to the land area may make New Zealand particularly well suited technically to such an approach, whatever the economic and political merits. But this may, of course, be a particularly attractive approach for wider economic reasons: as noted earlier, the theoretical case for regarding roads as a pure public good is not strong. Most of the costs borne arise from individual user-damage to vehicles, of course, but also from lost time and the higher gas consumption caused by congestion: and congestion also imposes an external cost through pollution, etc. Thus, reliance on fuel excise duty as the principal form of user charge penalises rural road users relative to urban and is negatively not positively related to environmental damage caused by pollution. To date, no effective way has yet been found through, as it were, an unbundling of the separate components of taxes and user charges to relate them to the different private and social costs. But the recognition of the different elements is clear in New Zealand thinking (Dunlop, 1996) and the website of Transfund shows sustained interest in finding better solutions.

6. NEXT STEPS IN REFORM

But this evolution in thinking in New Zealand on road maintenance, and more generally on privatising road transportation, does not seem to have been widely matched elsewhere among industrialised countries. While there are many other experiments on the economics of road transport, the financing of road maintenance – where this is a responsibility of central government – remains with either a government department or in many instances with a government agency. A quick desk review of European practices indicated no direct adoption of the second generation road fund model, although some of the key features are there, particularly where the function is carried out by an independent agency. (The reality is perhaps more of second generation road funds initiating best practice in the government agency model than the other way around.)

In many industrialised countries, road maintenance for highways and other major roads has remained as a public service provided by government and financed through the budget – this includes some outsourcing of the work, however, to the private sector. Such arrangements seem to apply in France, Spain, Norway and the Netherlands, for example. One or two have gone in the direction of the separate but still government-owned agency – the United States (through the Federal Highways Authority), the United Kingdom, Sweden and Finland, for example – but the agencies are largely funded from general taxation and not from an earmarked tax or other form of surrogate user charge.

Thus, we seem to be in a position where there are – admittedly, to oversimplify – three models in existence. First, New Zealand, with essentially full privatisation of the roads now in place. Second, many developing countries have adopted the second generation road funds model, and in that sense, moved further in privatising their road maintenance operations than industrialised countries. Third many European countries, the broader group of OECD members and other industrialising countries, still rely on traditional budget financing of their road maintenance programmes, which are carried out either as a government programme or by a government agency, albeit with some or total outsourcing of the actual work to the private sector.

Why have the industrialised countries not followed this particular avenue of "privatisation" when they have done so more generally – not just on utilities but also on some other forms of transport, such as buses and airlines? Before suggesting some possible answers and indicating areas for potential study, there are two possibilities that can be more or less readily dismissed. First, just because there seem to be very few road funds around among industrialised countries should not be taken as evidence of lack of interest in user charges or other innovations in road financing. There is no shortage of innovative ideas (though much less implementation of them) for tackling some of the other economic problems associated with road usage – safety, environmental congestion, etc. For example, a number of European cities are experimenting with, or have made operational, city congestion taxes, park-and-ride schemes, private sector road provision paid for through tolls, etc. While sheer bureaucratic inertia should never be too readily dismissed in looking at patterns of public sector provision, given the willingness to seek radical solutions elsewhere for road transport, innate conservatism seems more likely to be rooted in a particular view of the merits of earmarked funds. Second, there is no general outcry that would suggest the budgeting for road maintenance is insufficient or the level of provision is inadequate. (This is not to deny that matters may be far from optimal: but there would not seem to be a perceived problem specifically in providing for road maintenance, that is not seen in other public sector areas.) That may put reform some way down the agenda of the politicians.

While it is always difficult to be well briefed on planned reforms in this or other public policy areas, there do not seem to be any plans for the adoption of the road fund model or for the general adoption of the earmarked tax route. A number of issues/questions arise – both about the practices in OECD and developing countries – and even more, about what the likely future pattern of provision might be:

- Does this mean that the budgetary arrangements for highway road maintenance in the OECD and other industrialised countries are therefore adequate and no further innovations are necessary?

- If so, should we conclude that most of the benefits in terms of securing better road maintenance that seem to have emerged from the second generation road funds in at least some developing countries, probably came from contracting out rather than from the hypothecated fund aspect of road funds?

- Are there reasons why – particularly in certain otherwise reform-minded countries – the problems of using an earmarked tax are so great that, whatever the other benefits available, the authorities are unwilling to establish a road fund?

- If so, does this suggest that the developing countries that are at present working through road funds can safely revert to traditional road maintenance arrangements after a certain level of development has been reached? And, if so, what is that level of development?

There are no easy answers to these questions: but it is worth starting with the basic hypothesis that budget systems are sufficient to generate adequate resources for road maintenance (bearing in mind that the focus here is on highways and major roads)[12]. At first sight, and at the risk of considerable generalisation across OECD major road systems, road conditions on the major highways seem to be generally good. There is no public outcry about adequate road surfaces even if complaints of hold-ups, bad scheduling or occasionally poor quality road maintenance are still there. Moreover, it is very difficult to get away from the problem that led to the creation of road funds in the first place: that road maintenance is likely to be a lower political than economic priority. Despite widespread budgetary reforms in public sectors, and the greater introduction of performance information and performance measurement, it may be quite difficult to find evidence of much improvement in the allocative efficiency of budget systems (Diamond, 2003). Indeed, as is well recognized, performance information may present to officials and politicians more vexing questions – whether poor performance is indicative of the need to cut back or to augment resources.

Moreover, despite all the budgetary reforms, it is difficult to believe that government budget systems, where incremental budgeting remains the norm, have found a way around the perennial problem: that the final allocation of resources to individual public sector programmes is influenced not just by economic issues, but rather by political perceptions. And, as noted at the very beginning of this paper, road maintenance seems likely to be a plausible victim in such a game, not a winner. As one ex-Permanent Secretary of the UK Ministry of Transport put it: “In a regime in which education and health are funded by general tax revenues, transport is always likely to fare badly.” Within transport as a whole, road maintenance seems particularly well suited to be a victim of budgetary procedures.

However, while admittedly the evidence is more anecdotal than substantive that the budget systems of today may not be any better (in the sense of allocating to economic priorities) than in the past, there is more persuasive evidence that reforms to budget systems have led to greater productive efficiency in the public sector. In short, it is fairly widely accepted that the greater use of contracting out to the private sector, the revised tougher accounting regimes, the general introduction of performance information and performance measurement, etc., have overall contributed to an improvement in performance. If this is correct, it might mean that we are securing more and/or better quality road maintenance from given budgets. This would tie in with the lack of any evidence that greater resources are being allocated to road maintenance, in a general environment of restricted budgets. In short, a plausible hypothesis is that improvements in the productive efficiency of the resources available for road maintenance have been enough to satisfy most road users that road maintenance is broadly adequate.

On this basis, there may thus be something in the argument: "If it ain't broke, don't mend it." Governments may have been dissuaded from radical changes in policy in this area simply on that basis. But, could there be other reasons why road funds have not taken off in major OECD countries; and might those reasons be related less to the delivery of road maintenance than to issues of paying for it through surrogate user charges?

First, as many others have pointed out, the treatment of road user charges is bedeviled by misperceptions and misunderstanding of history. Why should people pay for the maintenance of roads through user charges that are already there? (their transfer from private sector to government is too far in the past to be deemed relevant). In short, the starting point is probably an in-built resistance to the idea that road users should face any charges, whether directly or on a surrogate basis through the tax regime. (The fact that they already do so through vehicle registration fees to use the network is conveniently forgotten, or seen as a road safety, not road usage issue.)

Second, there may be differences in the sheer scale of the economic costs experienced from poor road maintenance. As is well established, those costs fall in part on the private user (delays as a result of increased congestion, longer driving times, greater fuel usage) and the wider impact on the economy (the impact of longer travel times, greater fuel usage on the costs of all goods and services, damage to the environment, etc.). In reality, in industrialised countries the volume and quality of road maintenance provided is associated more with delays and associated congestion costs and to a lesser extent with environmental pollution that afflicts the individual. There is less of an observable or perceived wider impact on economic efficiency. The problems associated with poor road maintenance (and differentiating those which arise from a lack of road capacity) are probably seen as more in the nuisance or inconvenience category than as a genuine economic cost.

But in many developing countries, the basic condition of the roads is often relatively poor, so that it is not just a question of nuisance or inconvenience but of long-standing, higher in-built costs associated with very long journey times for the transit of goods, services and persons. In short, the scale of the economic damage is likely to be much wider and, just as importantly, is understood by the electorate and the government to be important. Another way of putting this argument is that, to the extent that there is a perception of the wider economic damage caused, it may lead to road maintenance being given a higher political priority in the budget systems of developing countries. That, in turn, may make governments in such circumstances more open to the idea of introducing a road-fund-type solution. Frustration with the inefficiency of their own public sectors probably reinforces the attractiveness of the road fund solution.

But could there be further arguments why the road fund solution has not proved attractive in the industrialised countries? In the longer-established democratic societies that typify industrialised countries, there is much greater sensitivity to the structure and level of taxes, which may also make a road fund solution unattractive.

First, no matter how well the idea may be presented, any changes in the tax structure to introduce a road fund financed from surrogate user charges (in the form of fuel duties, vehicle excise duties, etc.) will likely be presented by opponents as an additional tax. That has bedeviled tax reform in many countries, notably the United States. No matter what offsets are offered through other forms of taxation, the politics of adding any new tax to the armory are becoming perennially more difficult.

Second, the main tax that is being used to finance road funds – fuel duties – is perhaps one of the most politically sensitive areas. There has been substantial disquiet about rising fuel prices, particularly in recent years: within those times, civil action has been seen by groups varying from road haulers to farmers in countries like the Netherlands, France, Germany and the United Kingdom. For that reason alone, it would take a courageous political decision to enter into this area. Some have raised technical objections on this score also: to the extent that it is a buoyant tax, it may lead to overprovision for road maintenance (while denying central government resources needed for other public services – one of the familiar arguments against earmarked taxes). This can, of course, be got around by setting it at a flat amount, and perhaps indexed thereafter to a consumer price index, not oil prices.

Third, however, in setting a rate of taxation on petroleum, governments may be moving to a position where one item of expenditure, and often one associated tax, is being asked to serve multiple objectives: revenue raising; actual (or potential) hypothecation to a fund for road maintenance; congestion tax; pollution tax; even accident prevention. One possible way forward may be to unbundle some of the elements: a start has been made on this in New Zealand by taking out a discrete element for road safety. But technology could offer much better ways to expand and differentiate charging instruments and link them to particular dimensions – notably congestion costs (time-based charges for road usage, etc.) and environmental taxes linking usage to location (rural *vs*. urban). All this may be to come.

Finally, any consideration of a revised tax structure needs to recognise that overall levels of taxation tend to be higher in industrialised than in developing countries – though not universally true, this is the broad pattern. Put another way, it may be easier to sell a change in the tax structure for what is seen as an economic priority in developing countries, but which is of much less significance in the industrialised countries.

CONCLUSION

Much of the above is admittedly somewhat speculative. But it might appear that some combination of inertia and an adequate improvement in productive efficiency led by contracting out of the service (rather than the creation of independent road funds), has provided sufficient resources for road maintenance and perhaps discouraged any real search for alternatives. Where embracing more radical solutions for road maintenance would involve an explicit move to a real user charge which can be politically challenged, it is unsurprising that few have moved in that direction.

But this is not to say that this situation will last: it is clear that growing concern about the environment, about congestion costs and indeed the various new approaches being pursued, whether through private toll roads or city congestion schemes, means that more radical use of taxes and user charging is quite conceivable. Technology may come to our aid in the search for a more appropriate pricing and charging regime. But for the moment, one might tentatively conclude that, while road funds are a rational solution to a more intense, widely perceived weakness of budget systems and clearer economic need in developing countries, they are not needed in the industrialised countries.

That, in turn, raises the interesting question of whether, and if so when, developing countries themselves might safely move to reintegrate road maintenance into the budget system – an issue for the future. It is perhaps feasible that some developing countries, perhaps having graduated to the point where the quality of main roads matches that of OECD members, may conclude that the road fund is no longer necessary and reabsorb the funding into general taxation.

But it seems improbable.

One must surely, however, conclude that the movement would more likely be in the other direction: that, under the pressure to take on the challenge of congestion costs, pollution and the rationing of increasingly scarce road space, it is the industrialised countries which will make a move toward the New Zealand practice. Thus, it would be the industrialised world which would make the change in the direction of the developing. That would indeed be unique.

NOTE

1. An interesting dimension not explored here is whether some of these questions can be posed at a more general level. For, as pointed out in Potter (1997), while social security and very rarely health taxes are levied under earmarked arrangements, in principle, many other public services could be open to such an approach. Why therefore are there not more education or social funds? One answer may be pure coincidence: relatively steady growth and inelastic demand, combined with the use of specific taxes, tend to mean that revenue from petrol taxes is relatively stable – as is the expenditure year to year on road maintenance. Such relative stability in expenditure need and appropriate revenue resource may be difficult to match elsewhere.

BIBLIOGRAPHY

Asian Development Bank (2003), "Road Funds and Road Maintenance: An Asian Perspective", Manila, July.

Deran, E. (1965), "Earmarking and Expenditures: A Survey and a New Test", *National Tax Journal*, December.

Diamond, Jack (2003), "Performance Budgeting: Managing the Reform Process,", IMF Working Paper 03/33, Washington: International Monetary Fund.

Dunlop, Robin J. (1999), "The New Zealand Experience in Restructuring Road Administration: New Zealand Road Reform", *Transportation*, Vol. 26, No. 1, February.

Gwilliam, Ken and Zmarak Shalizi (1999), "Road Funds, User Charges, and Taxes", *The World Bank Research Observer,* Vol. 14 , No. 2, pp. 159-85, August.

Heggie, Ian G. (1995), "Management and Financing of Roads: An Agenda for Reform", World Bank Technical Paper 275, Africa Technical Series, Washington: World Bank.

Kumar, Ajay (2000), "Assessment of Selected Road Funds in Africa: Case Study of Benin, Ethiopia, Ghana, Kenya and Zambia", Sub-Saharan Africa Transport Policy Program Working Paper No. 51, Washington: World Bank, December.

Metschies, Gerhard P. and Gunter Zietlow (2001), "Introducing Road Funds in Latin America: the GTZ Experience", Manila: Asian Development Bank, March.

Potter, Barry H. (1997), "Dedicated Road Funds: A Preliminary View on a World Bank Initiative", IMF Paper on Policy Analysis and Assessment 97/7, Washington: International Monetary Fund, June.

Potter, Barry H. and Jack Diamond (1999), "Guidelines for Public Expenditure Management", Washington: International Monetary Fund.

Teja, Ranjit S. (1988), "The Case for Earmarked Taxes: Theory and an Example", IMF Working Paper 88/18, Washington: International Monetary Fund.

SELF-FINANCING OF INFRASTRUCTURE SERVICES

Erik T. VERHOEF
Free University Amsterdam*
Amsterdam
NETHERLANDS

*Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam.

SUMMARY

Amsterdam, July 2005

ABSTRACT

This paper considers the relation between transport infrastructure pricing and capacity choice and financing from the economic perspective. We present the self-financing result of Mohring and Harwitz (1962) and review the literature that generally shows how the result remains valid for various important extensions, as long as tolls are equal to marginal external congestion costs and capacity is continuous. But even with lumpy capacity, pooling of surpluses and deficits across roads or over time would make aggregate self-financing less unlikely than for every individual road segment at every instant. The empirical evidence, insofar as available, suggests that the extent to which the constant economies condition is fulfilled, necessary for exact self-financing, may vary over applications, but on the whole it appears to be a reasonable approximation.

Keywords: *Traffic congestion, Road pricing, Road capacity choice*.

1. INTRODUCTION

A well-functioning and adequate transport infrastructure is generally seen as an essential precondition for economic growth and prosperity. Transport is necessary to allow spatial specialisation, clustering and trade, which are among the major sources for productivity growth. But infrastructure also facilitates the movement of people, thus allowing consumption and productive activities to be carried out in a larger area. This may contribute to individual wellbeing via different channels, including a greater access to variety in consumption goods, better matches on labour markets, efficiency gains and cost reductions through more competitive markets, etc.

Traditionally, the provision of transport infrastructure has been regarded as a public responsibility, motivated by a mix of practical, political and economic considerations. For example, the building of new infrastructure involves the use of land owned by a possibly very large number of individuals, each of whom might convey strategic behaviour when a private infrastructure firm would attempt to acquire the land necessary for construction. The resulting multilateral negotiations are usually easier to undertake and strategic behaviour by landowners easier to counter, when the constructor of the new infrastructure has public and legal authority. Furthermore, the design of new infrastructure (e.g. route, capacity) is best optimised from a network perspective, in accordance with longer-run spatial development plans, while taking into account negative effects on local (and global) environments – all suggesting that the operator should take a rather broad, social perspective when considering its investments. From the welfare economics viewpoint, an important motivation for public intervention has always been that, at least until congestion sets in, transport infrastructure has the characteristics of a public good, with consumption of its services being at least non-rival and, to a lesser extent, non-excludable. As Dupuit (1844) pointed out, the optimal price for an uncongested piece of infrastructure (a bridge, in his example) is zero, which of course leaves little hope for private provision under optimal pricing.

This traditional economic view on infrastructure as a pure public good has become somewhat obsolete for significant parts of the contemporary instances of new infrastructure supply. When a basic transport network already exists, most new supply will be motivated by capacity shortages, implying that the existing infrastructure cannot have been uncongested. The consumption of infrastructure space then becomes rival, and hence the optimal price is not equal to zero. Technologies for automated vehicle identification have caused the consumption of specific infrastructure services to become potentially perfectly excludable. Because the availability of public funds has been under pressure in many countries, the implied possibilities to raise revenues from the users of new infrastructure often invoke political interest in user charging as a potential source of funding for new facilities.

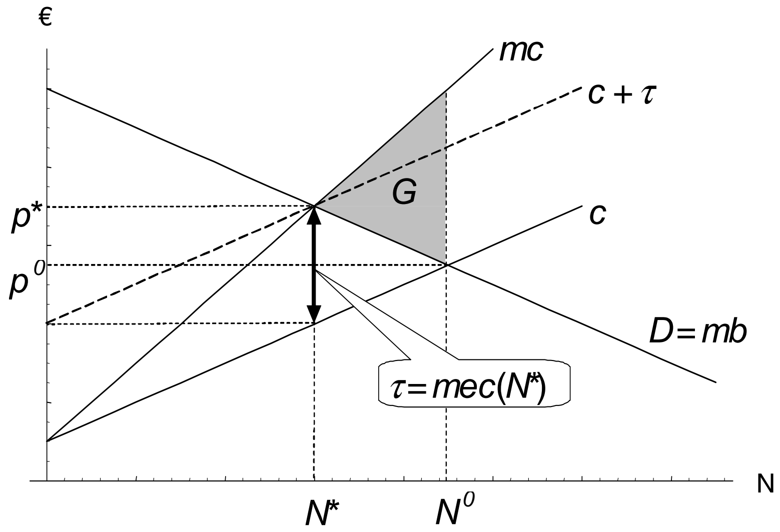
This paper studies user charging for the financing of infrastructure capacity supply from the economic perspective. We will begin, in Chapter 2, by deriving and presenting one of the landmark results in transport economics, known as the Mohring and Harwitz (1962) self-financing theorem. This theorem states that, under certain technical conditions, the revenues from optimally priced infrastructure are exactly equal to the cost of supplying the optimal capacity for that infrastructure. This, of course, creates a close link between infrastructure charging and the revenues it generates on the one hand, and capacity choice and the costs involved in it on the other. Chapter 2 will discuss these matters in the context of a simple but transparent example, with the aim to provide the intuition behind the result as clearly as possible. Chapter 3 will discuss a number of practical complications, as extensions to the basic exposition in Chapter 2 and, in doing so, addresses the question of whether the self-financing result "survives" the hitches one is bound to encounter in reality when determining infrastructure capacities and charges for real facilities. The extensions considered include lumpy capacity, time-of-day dynamics, network issues, user heterogeneity, maintenance, external effects and modes other than car traffic. Chapter 4 will address a specific type of complication, namely, second-best issues. In Chapter 5, the empirical validity of the technical assumptions underlying the self-financing result will be briefly discussed. Chapter 6 provides some conclusions, including directions for further research.

2. THE FINANCING OF OPTIMALLY DESIGNED ROADS: A BASIC EXPOSITION

2.1. Optimal road pricing

Before turning to the economics of capacity choice for infrastructure in relation to the application of user charging, it is instructive to first consider the workings of the market for the use of that infrastructure in the short run, defined as the term over which capacity is to be considered fixed. We will do so by considering what is probably the simplest possible example; namely, a single road used by identical individuals. For capacity expansion to become relevant to consider, congestion should initially occur on this road, and Figure 1 presents the conventional textbook analysis of the economics of traffic congestion and road pricing for such a setting.

Figure 1. **The economic basics of traffic congestion and road pricing**



The horizontal axis measures the volume of road use, *N*, while the vertical axis measures costs and benefits. The function *D* represents the (inverse) demand for trips and therefore coincides with the marginal benefits *mb* from road use. The function *c* gives the average cost for a trip, encompassing both financial costs and (the value of) travel times. For that reason, *c* rises with *N*: congestion causes travel times and hence *c* to rise. *Absent tolling:* users will enter the road up to N^0, beyond which the benefits attached to a next trip (*mb*) will fall short of the cost incurred by the user (*c*), so that the trip will not be made. N^0 exceeds the efficient level of road use, N^*, which occurs at the intersection of marginal cost *mc* and *mb*, implying that the addition of another vehicle would cause an equal amount of additional social costs (*mc*) as benefits (*mb*). Figure 1 shows that *mc* is above *c*, so that N^* is to the left of N^0. The economic reason why *mc* exceeds *c* is that *mc* not only includes the cost incurred by an additional traveller himself (*c*), but also the costs imposed on all other travellers because the travel time goes up – the marginal external cost, *mec*. The mathematic reason why *mc* is above *c* is that *c* is rising, and an average (*c*) can of course only go up if the value for the last unit added (*mc*) exceeds that average. Also note that the marginal cost can be expressed analytically in terms of *c* as the derivative of total cost $N{\cdot}c(N)$ with respect to *N*, or (by the product rule of differentiation): $c + N{\cdot}\partial c/\partial N$. The latter term is positive when *c* is rising, and represents the marginal external cost: *N* users, multiplied by the term $\partial c/\partial N$ which gives the increase in costs for each of these *N* users when an extra traveller enters the road.

By avoiding the use of the road by all traffic between N^* and N^0, a net welfare gain, given by the shaded triangle *G,* can be realised, which is the difference between the social cost saved (the area under *mc* between N^* and N^0) and the social benefits foregone (the area under *mb* between N^* and N^0). With an optimal toll τ in place, this optimum is "decentralised" as a market equilibrium, because the perceived generalised price *p* for a trip then becomes equal to $c+\tau$. Note that the optimal τ is equal to the marginal external cost in the optimum, $mec(N^*)$. This optimal congestion toll is often referred to as the "Pigouvian" tax, after its spiritual father, Arthur Pigou (1920).

For future reference, it is convenient to also derive this optimal toll analytically. To that end, we denote the social objective of social surplus as S, and define it as the difference between total benefits B and total cost C. The latter can be decomposed into user cost C_u and capacity cost C_{cap}, but because the latter is constant in the short run, it can be ignored in the formal short-run optimisation problem. Noting that B can be written as the integral of mb and C_u as the product of use and average cost, the optimisation problem, subject to the equilibrium constraint, can be written as:

$$\begin{aligned} &\underset{N}{\text{Max}}\, S = B - C = \int_0^N D(n)\mathrm{d}n - N \cdot c(N) \\ &\text{s.t. } c(N) + \tau - D(N) = 0 \end{aligned} \tag{1}$$

The optimality condition for this problem, in equation (2) below, states that marginal benefit mb=D be equal to marginal cost, which implies by the constraint that the optimal toll τ should be set equal to the marginal external cost:

$$D - c(N) - N \cdot c'(N) = 0 \quad \Leftrightarrow \quad \tau = N \cdot c'(N) \tag{2}$$

This result is fully consistent with the graphical exposition of Figure 1.

2.2. Optimal road pricing and optimal capacity choice

Using the same conceptual approach as above, it is now possible to determine the optimal capacity K for a road, given that optimal road pricing will be implemented. To make the problem meaningful and to avoid an optimum with an infinitely wide road, there should of course be a cost of supplying road capacity. We denote it as $C_{cap}(K)$. A larger road capacity will affect the relation between average user cost c and traffic volume N; presumably, $c(N,K)$ generally falls if K rises, and in any case does not increase in K.

The resulting social optimisation problem becomes too complex for an insightful diagrammatic exposition, because there are now two variables to be optimised: N and K. We therefore only present the analytical derivation, which is a straightforward extension of problem (1) above:

$$\begin{aligned} &\underset{N,K}{\text{Max}}\, S = B - C_u - C_{cap} = \int_0^N D(n)\mathrm{d}n - N \cdot c(N,K) - C_{cap}(K) \\ &\text{s.t. } c(\cdot) + \tau - D(\cdot) = 0 \end{aligned} \tag{3}$$

The optimality conditions are:

$$\frac{\partial S}{\partial N} = D(\cdot) - c(\cdot) - N \cdot \frac{\partial c(\cdot)}{\partial N} = 0 \quad \Leftrightarrow \quad \tau = N \cdot \frac{\partial c(\cdot)}{\partial N} \tag{4a}$$

$$\frac{\partial S}{\partial K} = -N \cdot \frac{\partial c(\cdot)}{\partial K} - \frac{\partial C_{cap}(\cdot)}{\partial K} = 0 \tag{4b}$$

The first of these shows that the Pigouvian tax from the short-run model remains intact also in the long run, when considering the possibility of capacity expansion. The optimal charge remains equal to the marginal external cost. The second condition states that the optimal capacity is reached when the marginal benefits of further expansion (the first term in the middle expression) has become equal to the marginal cost (the negative of the second term). The equality of marginal cost and marginal benefits is of course a familiar economic optimality condition, that we indeed already encountered in a different form in equations (2) and (4a).

Unless we are prepared to impose somewhat more structure on the two cost functions C_u and C_{cap}, it is not possible to derive any further general results on the relations between optimal capacity and road prices, and between capacity cost and toll revenues (for given cost functions, numerical results would of course be possible to obtain). As it turns out, it is instructive to explore these relations in the absence of economies and diseconomies of scale for both functions. For C_u, this assumption means that there are *constant returns to scale in congestion technology* (Small, 1992), meaning that the travel time remains unaffected when both the use level N and the capacity of the road K change in the same proportion. As a result, average user cost can be written $c(N/K)$. For C_{cap}, this means that it can be written as $K \cdot c_{cap}$, where c_{cap} gives the fixed unit cost of capacity, so that there are *constant scale economies in road construction*.

Note that these two assumptions together do not require K to be measured in constant units such as traffic lanes. Instead, if, for example, both conditions are fulfilled when K is expressed as the square of road width, economies of scale in congestion technology when expressing K in road width would be exactly compensated for by diseconomies in road construction when expressing K in road width, and there is a way of redefining K (namely, as K^2) for which both conditions are still fulfilled. More generally, because we are free to choose the units in which capacity is measured, we can always choose a definition of K that makes the constant returns to scale in congestion technology hold. The question then becomes whether, given this particular definition of units of capacity, the assumption of constant scale economies in road construction still holds. This means that what looks like two constant-returns-to-scale assumptions can be brought back to a single assumption after an appropriate choice of units. The empirical plausibility of the constant-economies-of-scale assumption will be addressed in Chapter 5 below.

2.3. Self-financing under optimal road pricing and optimal capacity choice

Under the constant-economies-of-scale assumption just introduced, we can derive the self-financing theorem, either analytically or graphically. The quickest derivation is analytical, and rephrases problem (3) as follows:

$$\begin{aligned} &\underset{N,cap}{\text{Max}} \int_0^N D(n)\mathrm{d}n - N \cdot c(\frac{N}{K}) - K \cdot c_{cap} \\ &s.t.\ c(\cdot) + \tau - D(\cdot) = 0 \end{aligned} \tag{5}$$

The optimality conditions are:

$$\tau = N \cdot \frac{\partial c(\cdot)}{\partial N} \tag{6a}$$

$$-N \cdot \frac{\partial c(\cdot)}{\partial K} = c_{cap} \tag{6b}$$

These conditions mean that total toll revenues R and capacity costs C_{cap} will amount to:

$$R = N \cdot \tau = N^2 \cdot \frac{\partial c(\cdot)}{\partial N} \tag{7a}$$

$$C_{cap} = K \cdot c_{cap} = -N \cdot K \cdot \frac{\partial c(\cdot)}{\partial K} \tag{7b}$$

Note that the following relation exists between the two partial derivatives of the average user cost function c:

$$\left.\begin{array}{l} \dfrac{\partial c(N/K)}{\partial N} \equiv \dfrac{\partial c(N/K)}{\partial (N/K)} \cdot \dfrac{\partial (N/K)}{\partial N} = \dfrac{\partial c(N/K)}{\partial (N/K)} \cdot \dfrac{1}{K} \\ \dfrac{\partial c(N/K)}{\partial K} \equiv \dfrac{\partial c(N/K)}{\partial (N/K)} \cdot \dfrac{\partial (N/K)}{\partial K} = \dfrac{\partial c(N/K)}{\partial (N/K)} \cdot \dfrac{-N}{K^2} \end{array}\right\} \dfrac{\partial c(N/K)}{\partial K} = \dfrac{\partial c(N/K)}{\partial N} \cdot \dfrac{-N}{K} \tag{8}$$

Substituting the right-hand side of (8) into (7b) gives:

$$C_{cap} = N^2 \cdot \frac{\partial c(\cdot)}{\partial N} \tag{7b$'$}$$

which is identical to the optimal revenues in (7a).

This self-financing result can also be illustrated graphically, which requires derivation of the long-run average cost functions for a congested road – similar to what we do when seeking a firm's optimal production capacity. A long-run average cost function in the present context gives the average cost as a function of the level of road use, *with capacity optimized for each level of road use*. Three such functions are relevant in our analysis. The first is the long-run average user cost function ($lrac_u$), which is the long-run counterpart of the c function. This function gives c as a function of N when the capacity is optimized for each N. The second is the long-run average capacity cost function ($lrac_{cap}$), which gives the capacity cost per user (C_{cap}/N) when capacity is optimized for each N. The sum of these two gives the long-run average total cost ($lratc$); the third function of interest.

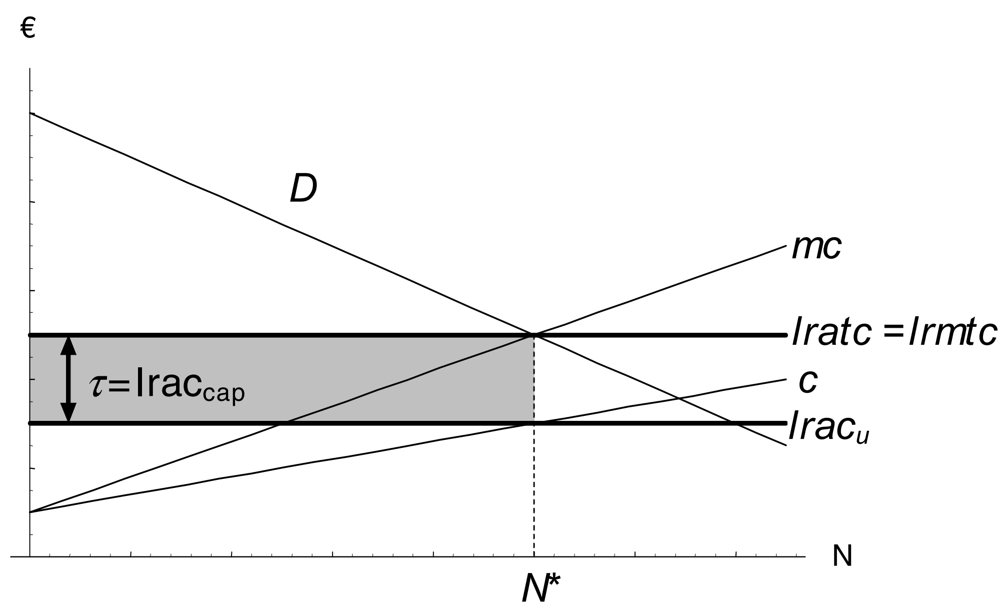
To derive these long-run average cost functions, a first task is to determine the optimal capacity for each N. For any given N, the optimal capacity is of course the one that maximizes social surplus S as given in equation (5). Because total benefits B only depend on N, which we now take as given, we thus in fact seek the capacity that minimizes the sum of total user costs (C_u) and total capacity costs (C_{cap}), given N. And, again because we treat N as given, we may also seek the capacity that minimizes the average total cost (*atc*): the sum of average user costs (*c*) and average (per user) capacity costs (C_{cap}/N). Under the constant economies of scale assumption, this average total cost can be written as:

$$atc = c(N/K) + c_{cap} \cdot K/N \tag{9}$$

Because *atc* depends only on the ratio between *N* and *K*, the solution to the problem of minimizing *atc* for a given *N* will, independent of the value of *N*, always lead to the same ratio between *N* and *K*. The fact that the optimal *N*/*K* ratio is independent of the value of *N*, in turn, implies that the optimal capacity will be proportional to *N*. This means that *lratc*, $lrac_u$, and $lrac_{cap}$ will all be independent of *N*, and will therefore be horizontal functions. These three facts are sufficient to secure that the self-financing theorem applies.

Figure 2 illustrates why. The long-run optimum is found at the intersection of the inverse demand function with the (bold) long-run average total cost function *lratc*. The reason is that *lratc* can be interpreted as the long-run marginal total cost function *lrmtc*: it gives the additional total cost (user cost plus capacity cost), for an extra user on the road, if capacity is optimised to account for this extra user. With capacity optimised, the average cost *c* at N^* is by definition equal to the (bold) long-run average user cost $lrac_u$. Next, because an optimal road price τ is in place, *D* will be equal to the short-run *mc*. We therefore have the following equalities: $D = mc = lratc$, and $c = lrac_u$. These equalities mean that $mc - c = lratc - lrac_u$, or: $\tau = lrac_{cap}$. In words: the optimal toll equals the average (per user) capacity cost. Multiplying both by N^* yields the result that total toll revenues are equal to the total capacity cost. Both are given by the shaded rectangle in Figure 2. This confirms, graphically, the equality between total revenues in (7a) and capacity cost in (7b′).

Figure 2. **A graphical illustration of the self-financing theorem**



At least one issue is worth emphasizing. The long-run average total cost function *lratc* is horizontal, and therefore equal to the long-run marginal total cost function *lrmtc*. This may lead one to conclude, erroneously, that in the long run the marginal external cost is zero, and hence no congestion tax should be levied in the long run. This, however, is not the correct interpretation of long-run cost functions. That is, after the capacity has been set at its optimal level, short-run functions *c* and *mc*, as shown in Figure 2, of course still apply, although their slopes depend on the capacity chosen. Given the capacity chosen, therefore, congestion still exists and the optimal toll should be based on the interplay between the resulting short-run cost functions and the inverse demand function, as shown in Figure 1. This reflects that for a given capacity, an additional user of course will not bring along additional road capacity when entering the road. Indeed, at N^* and with an optimised capacity, the marginal cost for a next user added is given by *mc*, not by *lrmtc*.

2.4. Interpretation and implications of self-financing

We have just reproduced one of the most famous results in transport economics, due to Mohring and Harwitz (1962), who were the first to establish that, under appropriate technical (constant-economies and continuous capacity) conditions, the revenues from optimal congestion pricing will be just sufficient for financing the cost of optimal capacity supply. As we will see in Chapter 3, the theorem has been shown to "survive" various extensions of the admittedly simplistic set-up in which we derived it. Before turning to these extensions, it is useful to emphasize some of the theorem's more important policy implications.

Indeed, the theorem may be highly relevant for practical policymaking. Its application in practice would in the first place help in achieving an efficient road system, in terms of optimal capacities and optimal pricing. It furthermore firmly reduces the need of using tax revenues from other sources for the financing of roads. This may improve efficiency further, because these other taxes are often distortionary. It may in addition help in overcoming problems of public acceptability of road pricing. The resulting scheme may be perceived as "fair" (only the users of a road pay for its capacity) and "transparent" (there are no "hidden" transfers surrounding the financing of roads). Finally, it may lead to improved transparency in political decisions on infrastructure expansion. It is easily demonstrated that if the technical economies-of-scale assumptions are fulfilled, and other external costs are optimally priced, road capacity should be expanded when short-run optimal congestion pricing yields revenues per unit of capacity that exceed the unit (capital) cost of capacity[1]. Comparable rules can be formulated for discrete units of capacity – say, lanes. These rules, however, will be more complex due to the inherent "integer problem" (see also section 3.2 below). The market would thus indicate whether or not expansion is socially warranted, which will generally help improve the transparency and credibility of cost benefit analyses.

The static setup chosen for explaining the theorem in its simplest form clouds one possible, potentially important, misinterpretation of the theorem, i.e. it implies that all revenues from optimal congestion tolling should be used to finance further capacity expansions. The capacity cost C_{cap}, however, should not be misinterpreted as investment costs: instead, it represents the capital costs. Just as the total toll revenues R should be interpreted as a flow variable (e.g. revenues per year), also the capacity costs C_{cap} would represent capacity costs per year. In the simplest case, in the absence of maintenance and depreciation and with a constant interest rate ρ, this would be ρ times the investment costs. In a steady state, the yearly toll revenues would then be equal to yearly opportunity costs of the invested capital, and there should be no grounds for believing that efficiency would be served by reinvesting toll revenues in additional road capacity.

3. OPTIMAL CHARGING AND CAPACITY CHOICE, AND SELF-FINANCING: SOME EXTENSIONS

The self-financing result from our basic model suggests a very simple and clear relation between infrastructure charging and capacity choice: when the charge is set optimally, the optimal capacity to choose would be the one that implies a flow of capacity costs that equals the flow of toll revenues. An important question is to what extent this result is a fluke, resulting from specific simplifying assumptions in the basic model, and to what extent it carries over to more elaborate settings. This chapter will consider a number of complications that were ignored above, but that will be relevant in practical applications. Our discussion will follow and sometimes draw from earlier reviews, as given by Small (1992), Lindsey and Verhoef (2000) and De Palma and Lindsey (2005).

3.1. (Dis-)economies of scale

Given the repeated reference to the importance of constant-economies conditions in the exposition of Chapter 2, it should be no surprise that if these are violated, the self-financing theorem will break down. As for any market, optimal pricing under economies of scale will lead to a deficit for the road operator, and under diseconomies of scale to a profit. This was in fact already derived by Mohring and Harwitz (1962) themselves; their original result is more general than predicting exact self-financing under neutral scale economies. Instead, their result states that the ratio between toll revenues and capacity cost equals the elasticity of total capacity costs with respect to capacity, which we may denote as ε_K. Only with neutral scale economies would this elasticity be unitary, and would exact self-financing therefore occur. An above-unitary elasticity reflects diseconomies of scale and hence a surplus, and a below-unitary elasticity reflects economies of scale and hence a deficit.

To derive this result formally, we follow Small (1992) and define a measure for economies of scale $s_K = 1/\varepsilon_K = (C_{cap}/K)/(\partial C_{cap}/\partial K)$. An above-unitary value of s_K now denotes economies of scale, and a below-unitary value diseconomies of scale. There are neutral scale economies when $s_K = 1$. Next, observe that equation (8) implies that the following equality is true as long as c is a function of N/K (so, irrespective of whether there are constant scale economies in road construction), which can always be secured through an appropriate choice of units for capacity:

$$-K \cdot \frac{\partial c(N/K)}{\partial K} = N \cdot \frac{\partial c(N/K)}{\partial N} \tag{10a}$$

Total revenue from optimal pricing can therefore be written as:

$$R = N \cdot -K \cdot \frac{\partial c(\cdot)}{\partial K} \tag{10b}$$

When economies of scale in road construction need not be neutral, (6b) is no longer valid, but instead the more general optimality condition for optimal investments (4b) should be reconsidered. This condition (4b) can be rewritten as:

$$-\frac{\partial c(\cdot)}{\partial K} = \frac{1}{N} \cdot \frac{\partial C_{cap}(\cdot)}{\partial K} \tag{10c}$$

Using (10c), total revenue in (10b) can next be written as:

$$R = K \cdot \frac{\partial C_{cap}(\cdot)}{\partial K} \tag{10d}$$

Finally, using the definitions of $s_K = (C_{cap}/K) / (\partial C_{cap}/\partial K)$ and $\varepsilon_K = 1/ s_K$, we finally obtain:

$$R = \frac{C_{cap}(\cdot)}{s_K} = \varepsilon_K \cdot C_{cap}(\cdot) \tag{10e}$$

Note that (10e) confirms our earlier statements: exact self-financing occurs with $s_K = 1$ (i.e. with neutral scale economies), a deficit arises with $s_K > 1$ (i.e. when economies of scale are present) and a surplus occurs when $s_K < 1$ (i.e. with diseconomies of scale).

A violation of the constant-economies-of-scale assumption therefore leads to a breakdown of the exact self-financing result. This is, however, nothing peculiar to road infrastructure, but is true for any market on which marginal cost pricing applies. In what follows, we will refer to (10e) as "the self-financing result" (which thus stands for the degree of self-financing being equal to ε_K); the special case of $\varepsilon_K = s_K = 1$ will be referred to as "the exact self-financing result".

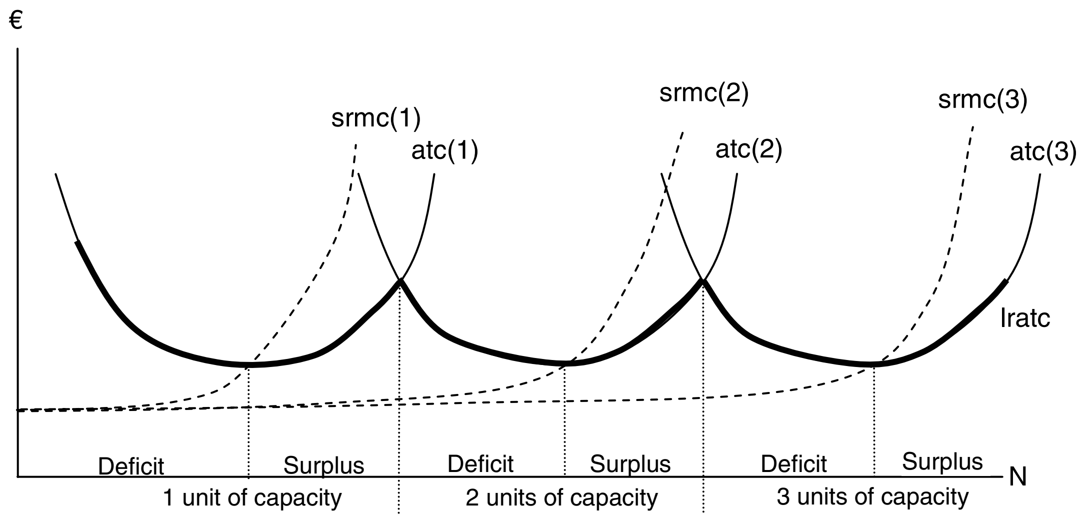
3.2. Discrete capacity

Our exposition so far assumes that capacity is perfectly divisible, which is generally not true in reality because the number of lanes is discrete. Certainly, capacity can still be varied by widening lanes, by resurfacing, or by re-grading or straightening a stretch of road, which probably makes it "less discrete" than it would seem from the discreteness of the number of lanes. But especially for smaller roads, discreteness of capacity will remain an issue of potentially high significance. For one thing, a road that is less than one lane wide makes little sense to construct: and the next capacity beyond a single lane (in a certain direction) would be double that capacity (although direction-switching on highway lanes, e.g. between morning and evening peaks, actually allows the use of "half-a-lane" in some cases). The discreteness of capacity generally causes the self-financing result to break down.

This is most easily seen for a road for which demand is so small that it will be uncongested once constructed at minimum feasible capacity, such as a road only meant to serve a small, newly built neighbourhood. The optimal road price on such a road, if never congested, will be zero, and hence the road cannot be self-financing.

Furthermore, for a road that would be congested at smaller capacities, discreteness of capacity would render the self-financing theorem no longer (exactly) valid, not even if the constant-economies assumption holds on average over the relevant range of capacities. The *lratc* curve from Figure 2 would then become a scalloped lower envelope of U-shaped *atc*(*l*) curves, valid for different discrete numbers of lanes, as shown in Figure 3, where *l* indicates the number of lanes. The relevant short-run marginal cost function, *srmc*(*l*), for the determination of optimal road use, in every segment of that scalloped *lratc* curve, cuts the *lratc* curve through its local minimum. Because the optimal toll will be equal to the difference between *srmc* and *c* (with *c* not drawn in the figure), and the per-user capacity costs will be equal to the difference between *atc* and *c*, a surplus occurs in any optimum with *srmc* > *atc* (and *atc* rising) and a deficit in any optimum with *srmc* < *atc* (and *atc* falling).

Figure 3. **Surpluses and deficits with discrete capacity**



Note that the long-run marginal cost function *lrmc* in Figure 3, also not drawn, would be given by the saw-toothed function that is defined by the *srmc* curves over their valid domains, connected by vertical lines at the segment boundaries. The reader may verify that as the discrete units become smaller, the *lratc* in Figure 3 becomes flatter and the relative differences between *lratc* and *lrmtc* diminishes. When units of capacity become infinitesimally small, both functions become horizontal and equal to each other, and we are back in the world of Figure 2. But with discrete units of capacity, the optimum occurs at an intersection between the saw-toothed *lrmc* function and the inverse demand function (also not shown). The peculiar shape of the *lrmc* function implies that multiple local optima would be possible; discrete optimisation should then reveal which of these is globally optimal. Depending on exactly where in a particular interval the global optimum occurs, a deficit or surplus will arise, as indicated in Figure 3.

Discreteness of capacity will thus generally cause the exact self-financing theorem to break down, even when there are constant economies of scale in the aggregate (the minima of the *atc* functions depict no upward or downward trend). Whether this is a serious problem for the self-financing of a full network remains to be seen. One consideration is that the operation of a larger number of roads of course allows “deficit and surplus pooling”, making the problem less important, relatively speaking, for a full network than for individual roads. A second consideration is that, when demand grows steadily over time, even for each individual road, periods of surpluses and deficits may alternate, causing the discounted net deficit or surplus to be smaller, relatively speaking, than the result in a particular year might suggest (see also section 3.4 below). These considerations would make the issue seem less important in the aggregate than what might appear from Figure 3 above. On the other hand, if a substantial share of the network’s roads has the minimum possible capacity and no congestion, a gross deficit would become more likely to arise. A second bias, but now in favour of a gross surplus, would be the existence of and pricing for externalities other than congestion (see section 3.9). All in all, it seems hard to predict what will be the end result for a full network. If anything, our discussion suggests that there may be compensating forces towards deficits and surpluses at work.

3.3. Time-of-day dynamics

One of the more disturbing simplifications of the models so far is that congestion is assumed to be a static, stationary-state phenomenon. This is helpful in keeping our discussion transparent, but rather unrealistic when looking at actual instances of traffic congestion. It is therefore important to face the question of whether the self-financing result remains intact when taking the time patterns of congestion and optimal congestion tolls into account. Arnott and Kraus (1998a) have shown that this is indeed the case, so that equation (10e) remains valid, under rather general conditions, provided tolls can be varied optimally over time.

A specific example of this result can be given for the conventional bottleneck model, first introduced by Vickrey (1969), and later analysed in greater depth by Arnott, De Palma and Lindsey (1993). In its basic form, this model considers dynamic queuing at a simple bottleneck, for which the outflow is equal to the inflow whenever there is no queue and attempted inflow is not above the bottleneck's capacity, and the outflow is equal to the bottleneck's capacity in all other cases. The dynamic equilibrium condition states that no user should be able to reduce travel costs by unilaterally changing the departure time. When all users have the same most-desired arrival time t^* and the same travel cost parameters α for the value of time, β for the shadow price of arriving early, and γ for the shadow price of arriving late, the travel cost for an individual arriving at time t can be written as the sum of *travel time cost* ($\alpha \cdot T(t)$, where $T(t)$ gives the travel time for someone arriving at time t) and *schedule delay cost* ($\beta \cdot (t^* - t)$ for an early arrival, or $\gamma(t - t^*)$ for a late arrival). Without tolling, the equilibrium will be such that $T(t)$ grows at a rate β/α for early arrivals and declines at a rate γ/α for late arrivals, so that the sum of travel time cost and schedule delay cost is constant throughout the peak. The queue therefore grows during the period of early arrivals and shrinks during the period of late arrivals.

With such pure bottleneck congestion, all time spent queuing is a pure loss: a longer queue does not reduce aggregate schedule delay cost (because the outflow of the bottleneck is constrained at the capacity) but does raise aggregate travel time cost. It is therefore always optimal to shorten queues. However, traffic should not be spread over time so strongly that the bottleneck is used below capacity during the peak: increasing the flow then reduces aggregate schedule delay cost without increasing aggregate travel time cost. The first-best optimum therefore entails a constant inflow, equal to capacity, throughout the peak: the bottleneck is used at its capacity without creating a queue. This optimum can be decentralised by a toll that mimics the equilibrium pattern of $\alpha \cdot T(t)$ from the untolled equilibrium. For further details and explanation, see Small (1992) and Arnott, De Palma and Lindsey (1993).

From the perspective of our discussion, what is important is that the model allows us to derive reduced-form cost functions, which relate total travel cost (travel time cost and schedule delay cost together) to the total number of travellers N during the peak. Because optimal time-varying tolling eliminates all queuing, a different reduced cost function applies with it than without it. The reduced cost functions thus incorporate equilibrium departure time choices. The reduced-form average user cost functions without and with optimal time-varying tolling turn out to be as follows (Small, 1992; Arnott, De Palma and Lindsey, 1993):

$$\overline{c}^{\,no\ t.v.toll} = \delta \cdot \frac{N}{K}$$
$$\overline{c}^{\,opt\ t.v.toll} = \tfrac{1}{2} \cdot \delta \cdot \frac{N}{K}$$
$$with: \delta = \frac{\beta \cdot \gamma}{\beta + \gamma} \qquad (11)$$

These are the cost functions that can be used to assess the optimal choice of capacity. Both without time-varying tolling and with optimal time-varying tolling, the average user cost depends on the ratio *N*/*K* only, which means that the analysis of section 2.3 and the more general equation (10e) are directly applicable. Note that the upper cost-function applies in absence of *time-varying* tolling, and remains valid with "flat", *time-independent* tolling. In other words, both without and with optimal time variation of the toll, the optimal bottleneck will be exactly self-financing, provided optimal tolls are used (that is, either the optimal time-varying toll or the second-best optimal flat toll), and capacity is optimised given the prevailing toll regime.

This brief exposition of the basic bottleneck model thus serves as an illustration of the more general finding from Arnott and Kraus (1998), and is furthermore helpful in illustrating that a time-independent, reduced-form cost function can be derived from dynamic equilibrium models and can be used to determine optimal capacity.

3.4. Dynamics over the years

The basic model of chapter 2 is not only static in its neglect of time-of-day dynamics, which we just addressed, but also by ignoring that investments in roads usually are supposed to remain productive over a large number of years, during which it is not inconceivable that demands may change. Arnott and Kraus (1998a) address a number of complications that will arise when allowing the associated longer-run dynamics to enter the analysis. They find that, whether or not capacity is added continuously or in discrete portions, and whether or not the timing of investments is optimal, the self-financing theorem remains valid in present value terms, provided the size of capacity additions is optimised, conditional on the timing of investments and, of course, user charges are optimised and the economies-of-scale assumption is satisfied.

3.5. Networks

The self-financing result of (10e) also continues to hold when extending the analysis from a single road or bottleneck to a full network. Yang and Meng (2002) demonstrate that self-financing will hold for every individual link in an optimally priced network, and therefore for the network at large also. The straightforward extension from a single link to a full network is an example of the so-called *envelope theorem* at work, which implies that indirect effects of policy variables (link tolls, link capacities) upon other markets (links, in this case) can be ignored in welfare analysis, provided that optimality conditions on these other markets are fulfilled. In other words, one can then analyse each link individually, and derive (10e) for it, without worrying about network effects.

As we shall see in chapter 4 below, network effects do lead to a breakdown of the self-financing result if other parts of the network are *not* optimally priced. In other words, the extension of the result to full networks is to be understood under the condition that optimal pricing also extends to the full network. De Palma and Lindsey emphasize that the capacities of other links need not be optimised for exact self-financing to hold for an individual link, as long as tolls on these other links are optimised, conditional on their non-optimal capacities. However, self-financing would then break down for these other links, and hence for the network as a whole: and the capacity set for the link considered will then turn out to be non-optimal once these other links' capacities have been optimised.

3.6. Heterogeneity of users

Arnott and Kraus (1998b) address heterogeneity across users and find that, as long as every user faces an optimal charge, this does not undermine the self-financing theorem. As for networks and time-of-day dynamics above, therefore, the important pre-condition is that marginal cost pricing applies to all users. An inability to differentiate tolls perfectly across users is then harmless as long as the marginal external costs caused by an individual only vary by observable characteristics, and the toll can be adjusted according to those characteristics. For example, suppose that congestion is more severe in the peak than outside it, that higher-income and higher-value-of-time drivers drive inside the peak and lower-income and lower-value-of-time drivers outside the peak, and that the marginal external cost is independent of one's own income and/or value of time. What matters then is whether tolls can be varied by time of driving. One does not need to differentiate by income or value-of-time to apply first-best tolls – even though, through self-selection, it will be the higher-income groups who pay the higher toll. The optimal toll is then said to be "anonymous", because it can be set without knowing the identity or characteristics of the user.

3.7. Road damage and maintenance

A further complication that has received attention in this literature concerns the inclusion of road maintenance costs, in combination with the choice of durability (e.g. thickness of the pavement) in the construction phase and optimal charging for road damage. A rather detailed technical explanation, for example, is provided in De Palma and Lindsey (2005); we will only mention some main conclusions here.

Newbery (1988, 1989), for example, considered this issue and concluded that "*if there are constant returns to scale in roads construction (for roads of given strength), and if there are strictly constant returns to road use (in the sense that heavy vehicles distribute themselves uniformly over road width), then the optimal road user charge (congestion charge plus road damage charge) will recover all road costs (maintenance and interest on capital) even if there are substantial economies of scale in road construction* (Newbery, 1989, p. 167)." The result requires that there are constant returns to scale in constructing roads with a given strength at different capacities. If this is the case, economies of scale in strengthening roads are unimportant. The optimal policy response to road damage externalities is a flat charge per "equivalent standard axle load" and Newbery (1988) established a self-financing property of this tax that is analogous to the Mohring-Harwitz result with respect to the congestion externality. Small, Winston and Evans (1989) considered issues of congestion and road damage in a multi-product framework. Even though they find substantial economies of scale in providing pavement, they conclude that the road system as a whole comes close to self-financing because of the diseconomies of scope involved in the simultaneous production of road capacity for passenger cars and trucks. The presence of these diseconomies means that the joint supply of infrastructure for cars and trucks causes additional cost, in comparison to the provision of separate infrastructure for both types of traffic. Overall, neutral scale economies approximately hold.

3.8. Distortions in land markets

A next issue that has received attention concerns the market for land. Contrary to the standard assumptions, the supply function of land for road capacity expansion may be rising when the road is sufficiently large to drive up the local price of land. In this case, the distinction between "returns to scale" (a property of production functions) and "economies of scale" (a property of cost functions) becomes important (as explained by Small, 1999). A general equilibrium analysis by Berechman and Pines (1991) shows how in this case the so-called "imputed profits" will have the same sign as the degree of returns to scale of the production function – where imputed profits are based on a cost measure for land that is obtained by multiplying the amount of land used by its shadow price. However, as Small (1999) observes, only for a factor-price taker would these imputed profits correspond to actual profits. A road authority may often have market power in the land market. Small (1999) shows that the sign of actual profits from highway operation under first-best marginal cost pricing will then still be determined by the degree of scale of the actual cost function (which differs from the degree of returns to scale of production with a rising supply curve for land). This holds both for the case with and without price discrimination by a monopsonistic road authority in the land market. The critical condition for exact self-financing under marginal cost pricing thus involves the degree of economies of scale of the cost function, and not the degree of returns to scale of the production function.

3.9. Other externalities

Of course, traffic congestion is not the only motivation for applying road pricing. Other externalities that warrant corrective Pigouvian charges include traffic accidents, noise annoyance and environmental pollution (see ECMT, 1998, among many others). The associated revenues from such Pigouvian taxes should analytically be kept separate from the capacity funds from congestion pricing as considered above – just like the payments that road users make for fuel, vehicle maintenance, etc. Should we include them in a model of optimal pricing and capacity supply, in which otherwise the default constant-economies assumptions apply, we would find a surplus from optimal pricing. But when comparing the pure congestion toll revenues alone to capacity costs, the familiar, exact self-financing result should re-emerge.

3.10. Other modes

Most of the economic literature on the relation between infrastructure charging and capacity choice has focussed on road transport – which is not fully unjustified when looking at empirical modal split and infrastructure budget figures. Nevertheless, it is important to ask whether the conceptual apparatus developed above is directly applicable to modes other than private road transport – notably rail, bus and airlines. In principle, the answer is "yes"[2], but various specific complications would hamper a straightforward practical application.

One of these is that, because of traffic control and discreteness rolling or flying stock, it is far from straightforward to determine marginal external congestion in modes such as rail and aviation. A second complication is that congestion effects are often, at least partly, compensated for by economies of density in transport markets with scheduled services, which may also be hard to quantify when attempting to determine the appropriate user cost functions.

Thirdly, scheduled services are often offered by a few (sometimes only one) operators per (congested) node. As pointed out by Brueckner (2002), such operators would internalise the part of the congestion costs that is imposed upon own services, which calls for a downward adjustment of congestion tolls and hence for a deviation from the Pigouvian congestion tolls underlying the self-financing theorem. Moreover, as pointed out by Pels and Verhoef (2004), when these operators compete under oligopolistic conditions – or act like a monopolist – a further downward adjustment of tolls may be called for, to counter overpricing tendencies due to market power. A negative second-best optimal net toll is not inconceivable.

This all leads to potentially strong deviations from the idealised world for which the self-financing result was originally derived. The relation between optimal infrastructure charging and capacity choice for modes other than private road transport seems to be fertile ground for further research, both theoretical and empirical.

3.11. Conclusion

This chapter has demonstrated that the self-financing result of Mohring and Harwitz, and in particular their more general result, that the ratio of optimal toll revenues over optimal capacity costs is equal to the local capacity cost elasticity as shown in equation (10e), is rather robust with respect to various extensions of the simplest possible model. It is therefore not surprising that the result has been so central in the economic literature on the relation between transport infrastructure charging and capacity choice. Many potential complications were seen to be harmless at closer inspection, as long as the principle of marginal cost pricing remains adhered to. This was, for example, the case for flat congestion pricing with dynamic congestion, and with heterogeneity of users.

One important reason why the result might break down involves discreteness of capacity choice but, as we saw, the law of large numbers may nevertheless make this problem less important over a full network and over a range of years than it may appear for a single road at instantaneous traffic conditions. Another important reason involves a departure from the marginal cost pricing principle. This brings us to the issue of second-best circumstances, to which the next chapter will be devoted.

4. SECOND-BEST ISSUES

The marginal-cost-based tax rules applied above are often referred to as “first-best” taxes, because there are no constraints on the pricing instrument itself (when marginal external cost differs according to users, so does the toll, and all users can be charged), and because there are no market distortions other than the congestion externality itself (transport markets and the wider economy otherwise operate under first-best conditions). Although useful as a theoretical benchmark, such first-best pricing is increasingly recognised to be of limited practical relevance. Various constraints may prevent the setting of Pigouvian tolls from being feasible (i.e. the charging instrument may be imperfect) or, in fact, economically desirable (there may be indirect effects, stemming from market failures “elsewhere”, to be taken into account). Even if feasible and desirable from an efficiency viewpoint, considerations of public acceptability, transparency or practicality may still cause regulators to prefer to deviate from strict marginal cost pricing.

The economic literature on second-best pricing is growing rapidly, and is too large to be reviewed here (Lindsey and Verhoef, 2001, provide a review). Instead, we choose to present two examples of second-best issues and to illustrate the possible implications for the self-financing theorem. Section 4.1 considers two instances of imperfect charging and section 4.2, two examples of distortions through mis-pricing elsewhere in the economy – or, more precisely, at other links of the network.

4.1. Non-optimal pricing

Suppose that, for some reason, it is not the optimal toll of equation (4a) that applies, but an arbitrary toll τ_A instead. This creates a second-best capacity choice problem of the type that was considered by Wheaton (1978), Wilson (1983) and d'Ouville and McDonald (1990). It can be represented by the following Lagrangian[3]:

$$\Lambda = \int_0^N D(n)\mathrm{d}n - N \cdot c(N,K) - C_{cap}(K) + \lambda \cdot \left(c(\cdot) + \tau_A - D(\cdot)\right) \tag{12}$$

where λ is the Lagrangian multiplier associated with the constraint that τ_A is given, and with it the equilibrium wedge between marginal benefit and average cost. The optimality conditions are:

$$\frac{\partial \Lambda}{\partial N} = D(\cdot) - c(\cdot) - N \cdot \frac{\partial c(\cdot)}{\partial N} + \lambda \cdot \left(\frac{\partial c(\cdot)}{\partial N} - \frac{\partial D(\cdot)}{\partial N}\right) = 0 \quad \Leftrightarrow \quad \lambda = \frac{N \cdot \frac{\partial c(\cdot)}{\partial N} - \tau_A}{\frac{\partial c(\cdot)}{\partial N} - \frac{\partial D(\cdot)}{\partial N}} \tag{13a}$$

$$\frac{\partial S}{\partial K} = -N \cdot \frac{\partial c(\cdot)}{\partial K} - \frac{\partial C_{cap}(\cdot)}{\partial K} + \lambda \cdot \frac{\partial c(\cdot)}{\partial K} = 0 \quad \Leftrightarrow \quad \frac{\partial C_{cap}(\cdot)}{\partial K} = \left(N - \lambda\right) \cdot -\frac{\partial c(\cdot)}{\partial K} \tag{13b}$$

Note that (13a) and (13b) become identical to the first-best conditions (4a) and (4b) when τ_A happens to be set optimally and, as a consequence, the multiplier λ becomes equal to zero. In all other cases, a second-best capacity choice problem arises.

With undercharging, λ is positive and the marginal benefits of capacity expansion are calculated as if fewer than the actual number of road users N are present (with overcharging, the opposite occurs). This reflects that with undercharging, the social net benefit of an additional user is smaller than the private benefit because of the undercharged congestion externality (and reversely with overcharging). Limiting capacity is one way of discouraging this additional user to enter the road. However, whether the equilibrium second-best capacity will be smaller or larger than the first-best capacity cannot be said with certainty, but will depend on demand and cost elasticities and their impact on equilibrium use levels (d'Ouville and McDonald, 1990; Small, 1992): also the N's will be different between first-best and second-best optima. For example, with undercharging, N on the right-hand side of (13b) will be larger than with optimal charging, but λ will be positive, leaving the net effect ambiguous. Still, Wheaton (1978) showed that reducing the toll infinitesimally below the first-best level results in an increase in optimal road width. In accordance with this result, Wilson (1983) derives global conditions, which he judged reasonable, under which second-best road capacity with under-pricing exceeds the first-best level.

Even if all other conditions would favour exact self-financing, the forced departure from marginal cost pricing as described above would generally cause self-financing under second-best optimal capital choice to break down. An extreme but most straightforward example would be $\tau_A = 0$. This yields zero toll revenues, but certainly not necessarily a zero second-best optimal capacity.

Non-optimal pricing thus potentially leads to a breakdown of the self-financing result in (10e). But it is important to emphasize that it is not, in fact, the difference between first-best and second-best tolling that matters here, but rather the difference between marginal cost pricing and deviations from it. This somewhat cryptic statement can be clarified with reference to the example discussed in section 3.3, where second-best "flat" tolling of a bottleneck was seen to lead to a relation between toll revenues and capacity costs satisfying (10e). The reason was that the second-best flat toll in that example is equal to marginal external cost, given that no time differentiation in tolling is applied and hence no departure rescheduling is initiated. It is an example of a case where the self-financing result continues to hold under second-best pricing, because second-best pricing entails marginal external cost pricing for all users.

4.2. Second-best distortion through network spillover[4]

As a second application of second-best analysis, let us consider the case where not all roads in a network are priced. How does this affect the degree of self-financing for an individual tolled road? Verhoef (2005) presents an analysis that produces some answers as a by-product. There appears to be a major difference between the case of unpriced parallel capacity, and unpriced serial capacity. We consider both cases in what follows, and treat both in the context of a simple two-link network in which for one link (T) the toll and capacity can be optimised, while an unpriced link (U) of given capacity exists in parallel or in series to the link T under consideration.

4.2.1. Unpriced substitute

The case of second-best congestion pricing with an unpriced substitute has been considered by various authors, including Lévy-Lambert (1968). The inclusion of capacity as a second policy instrument has been less common. Using superscripts U and T to denote the untolled and the tolled alternative respectively, and assuming that K^U is to be treated as given, the surplus-maximising second-best toll τ^T and capacity K^T can be found by solving the following Lagrangian:

$$\Lambda = \int_0^{N^U+N^T} D(n)\mathrm{d}n - N^U \cdot c^U(N^U, K^U) - N^T \cdot c^T(N^T, K^T) - C^U_{cap}(K^U) - C^T_{cap}(K^T)$$
$$+\lambda^U \cdot \left(c^U(N^U, K^U) - D(N^U + N^T)\right) + \lambda^T \cdot \left(c^T(N^T, K^T) + \tau^T - D(N^U + N^T)\right) \quad (14)$$

The set of first-order conditions (w.r.t. N^U, N^T, K^T, τ^T, λ^U and λ^T) can be solved to yield:

$$\tau^T = N^T \cdot \frac{\partial c^T}{\partial N^T} - N^U \cdot \frac{\partial c^U}{\partial N^U} \cdot \left(\frac{-\dfrac{\partial D}{\partial N}}{\dfrac{\partial c^U}{\partial N^U} - \dfrac{\partial D}{\partial N}} \right) \qquad (15a)$$

$$-N^T \cdot \frac{\partial c^T}{\partial K^T} = \frac{\partial C^T_{cap}}{\partial K^T} \qquad (15b)$$

The second-best optimal toll (15a) is the same as the one reported by Lévy-Lambert (1968) and is therefore unaffected by the possibility of setting a second-best optimal capacity for route *T*. It consists of a first term that is equal to the marginal external cost on route *T* itself, and adds to this a second term, which is a second second-best correction term. A more detailed interpretation of this toll expression can be found in Verhoef *et al.* (1996), but note that it falls short of the marginal external congestion on route *T* [the first term in (15a)], in order to optimise the congestion spillover onto route *U*. The second-best optimal capacity rule (15b) is similar to the first-best rule (4b). Given the equilibrium level of use of route *T*, N^T and given the generalised equilibrium price $c^T+\tau^T$, it is optimal to set capacity K^T such that the flow N^T is served at a combination of K^T and τ^T that minimises the social cost of carrying N^T. The corresponding condition is therefore the same as for a road without a substitute.

Given that we find the familiar rule for optimising capacity K^T, it will be no surprise that the same calculations as for the case of a single road [equation (7)-(10) in chapters 2 and 3 above] could subsequently be performed to derive that, without the second-best correction term, the net surplus or deficit on the road would be as given by (10e). The second-best distortion, leading to a downward adjustment in pricing, therefore causes the self-financing result of (10e) to break down. This is consistent with earlier findings in that the second-best toll (15a) is unequal to marginal external costs on route *T*.

In summary, under the neutral-scale-economies conditions that would produce a balanced budget for a road in isolation, we would find a deficit when an untolled, congested substitute is present.

4.2.2. Unpriced complements

Second-best pricing with an unpriced complement (such as a serial link) has received considerably less attention in the literature than the unpriced substitute. Maintaining the assumptions of a single origin-destination pair, and control over instruments at one of the two links only, the second-best optimum can be found by solving the following Lagrangian:

$$\Lambda = \int_0^N D(n)\mathrm{d}n - N \cdot c^U(N, K^U) - N \cdot c^T(N, K^T) - C^U_{cap}(K^U) - C^T_{cap}(K^T) + \lambda \cdot \left(c^U(N, K^U) + c^T(N, K^T) + \tau^T - D(N)\right) \qquad (16)$$

The set of first-order conditions (w.r.t. N, K^T, τ^T, and λ) can be solved to yield:

$$\tau^T = N \cdot \left(c_N^T + c_N^U\right) \tag{17a}$$

$$-N^T \cdot \frac{\partial c^T}{\partial K^T} = \frac{\partial C_{cap}^T}{\partial K^T} \tag{17b}$$

Intuitively, the second-best optimal toll of (17a) perfectly internalises the marginal external congestion costs for both links jointly. Compared to the first-best tax rule, there is now a positive correction term. Again, the rule that defines optimal capacity in (17b) has the by now familiar form. No effects on link U are present in this rule, which reflects that indirect effects of changes in K^T upon congestion on link U cancel because the toll in (17a) already perfectly internalises this congestion.

Using the same logic as for the unpriced substitute case of equations (14)-(15), it is now easy to see that under the neutral-scale-economies conditions that would produce a balanced budget for a road in isolation, we would find a surplus when an untolled congested complement is present. Again, this is of course consistent with the fact that the second-best toll in (17a) is unequal to marginal external costs on route T.

4.3. Conclusion

This section has illustrated that for the self-financing result of (10e) to remain valid, the toll(s) should be set equal to marginal external congestion cost(s). Second-best pricing schemes often (but not always) deviate from this principle, which would cause the theorem to break down.

5. EMPIRICAL ESTIMATES OF SCALE ECONOMIES IN ROAD PROVISION

A number of studies have looked into the question of scale effects in the supply of infrastructure. We briefly review this literature here, drawing heavily from Verhoef and Rouwendal (2004).

Keeler and Small (1977) considered the implementation of the Mohring-Harwitz (1962) analysis to urban expressways from an empirical perspective. The homogeneity of degree zero of the transport cost function was not assessed empirically, but assumed on the basis of “considerable empirical evidence” (p. 3), and they measured the cost of capacity as the sum of construction cost, land acquisition cost and maintenance cost. They found no evidence of (dis-)economies of scale in construction cost. This study therefore confirms Mohring and Harwitz’s conjecture that the conditions under which the self-financing result holds are plausible.

Kraus (1981) used engineering data in a simple network setting and concluded that there are probably increasing results to scale. According to his best estimates, optimal tolling would cover 84 per cent of capital cost. He judges that his findings “are not that different from those of Keeler and Small” (p. 20) who, as we saw, found constant returns to scale.

Small, Winston and Evans (1989) point out that there are probably increasing returns to scale in the construction of single roads of a given type since the costs of shoulders and median strips are independent of capacity, whereas integration of such roads in networks (by means of intersections, etc.) would probably cause decreasing returns to scale. In their base-case numerical model there are practically constant returns to scale.

More recently, Levinson and Gillen (1998) report a point estimate for the ratio between long-run average and marginal cost of 0.92 for auto, but 1.45 for single trucks and 1.96 for combination trucks, suggesting mild diseconomies for passenger cars and considerable economies for trucks.

These results taken together suggest that economies of scale in road investment are, on average, approximately neutral. Further study, however, appears warranted.

6. CONCLUSION

This paper considered the relation between transport infrastructure pricing and capacity choice and financing from the economic perspective. We presented the self-financing result of Mohring and Harwitz (1962) and reviewed the literature that generally shows how the result remains valid for various important extensions, as long as tolls are equal to marginal external congestion costs and capacity is continuous. But even with lumpy capacity, pooling of surpluses and deficits across roads or over time would make aggregate self-financing less unlikely than for every individual road segment at every instant. The empirical evidence, insofar as available, suggests that the extent to which the constant-economies condition is fulfilled, necessary for exact self-financing, may vary over applications, but on the whole it appears to be a reasonable approximation.

Further research into this matter would, however, be a welcome addition to the literature. Other issues that await further treatment include the operationalisation of the theorem to other modes, and a more empirically-based consideration of second-best constraints in transport pricing and capacity choice.

In terms of policy implications, the insights from the self-financing theorem may have far-reaching consequences. Its application in practice would, in the first place, help in achieving an efficient road system in terms of optimal capacities and optimal pricing. It furthermore firmly reduces the need of using tax revenues from other sources for the financing of roads. This may improve efficiency further, because these other taxes are often distortionary. It may in addition help in overcoming problems of public acceptability of road pricing. The resulting scheme may be perceived as “fair” (only the users of a road pay for its capacity) and “transparent” (there are no “hidden” transfers surrounding the financing of roads). Finally, it may lead to improved transparency in political decisions on infrastructure expansion. If the technical economies-of-scale assumptions are fulfilled and other external costs are optimally priced, road capacity should be expanded when short-run optimal congestion pricing yields revenues per unit of capacity that exceed the unit (capital) cost of capacity. The market would thus indicate whether or not expansion is socially warranted, which will generally help in improving the transparency and credibility of cost-benefit analyses.

NOTES

1. To see why, observe that for a given demand function, both the short-run optimal congestion price (i.e. for a given capacity) and the road use per unit of capacity are decreasing in capacity. Short-run optimal toll revenues per unit of capacity therefore exceed the unit cost of capacity with a below-optimal total capacity, and fall short of it with an above-optimal total capacity.

2. For example, Morrison (1983) shows that if capacity is divisible and costs are homogeneous in volume/capacity ratio, then social-marginal-cost pricing leads to exact cost recovery for airports. And Oum and Zhang (1990) show that with lumpy capacity, social marginal cost pricing need not guarantee cost recovery for the airport – as in section 3.2 above.

3. A "Lagrangian" function represents a constrained optimisation problem, and consists of the objective function plus the (set of) constraint(s), (each) multiplied by an auxiliary variable called a Lagrangian multiplier, which will obtain a value representing the shadow price(s) of the constraint(s) in the second-best optimum.

4. This section draws heavily from Verhoef (2005).

BIBLIOGRAPHY

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

Arnott, R. and M. Kraus (1998a), *Self-financing of Congestible Facilities in a Growing Economy,* in: D. Pines, E. Sadka and I. Zilcha (eds.) (1998), *Topics in Public Economics: Theoretical and Applied Analysis*, Cambridge University Press, Cambridge, UK, pp. 161-184.

Arnott, R. and M. Kraus (1998b), "When are anonymous congestion charges consistent with marginal cost pricing?", *Journal of Public Economics*, 67 (1), 45-64.

Arnott, R. and A. Yan (2000), "The two-mode problem: second-best pricing and capacity", *Review of Urban and Regional Development Studies*, 12, 170-199.

Berechman, J. and D. Pines (1991), "Financing road capacity and returns to scale under marginal cost pricing", *Journal of Transport Economics and Policy*, **25,** 177-181.

Bichsel, R. (2001), "Should Road Users Pay the Full Cost of Road Provision?", *Journal of Urban Economics*, 50**,** 367-383.

Brueckner, J.K. (2002), "Airport congestion pricing when carriers have market power", *American Economic Review*, 92 (5), 1357-1375.

De Palma, A. and C.R. Lindsey (2005), "Relation between pricing, toll revenues and investment", Task Report 2.1, Project REVENUE, DG-TREN Fifth Framework Programme.

Dupuit, J. (1844), "On the measurement of the utility of public works", in: D. Murphy (ed.) (1968), *Transport*, Penguin, London.

ECMT (1998), *Efficient Transport for Europe: Policies for Internalisation of External Costs,* OECD/ECMT, Paris.

Keeler, T.E. and K.A. Small (1977), "Optimal peak-load pricing, investment and service levels on urban expressways", *Journal of Political Economy*, 85**,** 1-25.

Kraus, M. (1981), "Scale economies analysis for urban highway networks", *Journal of Urban Economics*, 9**,** 1-22.

Levinson, D.M. and D. Gillen (1998), "The full cost of intercity highway transportation", *Transportation Research,* 3D**,** 207-223.

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

Lindsey, C.R. and E.T. Verhoef (2000), "Congestion modelling", in: D.A. Hensher and K.J. Button (eds.) (2000), *Handbook of Transport Modelling, Handbooks in Transport 1*, Elsevier/Pergamon, Amsterdam, pp. 353-373.

Lindsey, C.R. and E.T. Verhoef (2001), "Traffic congestion and congestion pricing", in: D.A. Hensher and K.J. Button (eds.) (2001), *Handbook of Transport Systems and Traffic Control, Handbooks in Transport 3*, Elsevier/Pergamon, Amsterdam, pp. 77-105.

Mohring, H. and M. Harwitz (1962), *Highway Benefits: An Analytical Framework,* Northwestern University Press, Evanston Il.

Mohring H. (1970), "The peak load problem with increasing returns and pricing constraints", *American Economic Review*, 60**,** 693-705.

Morrison, S.A. (1987), "The equity and efficiency of runway pricing", *Journal of Public Economics*, 34**,** 45 -60.

Newbery, D.M. (1988), "Road damage externalities and road user charges", *Econometrica*, 56**,** 295-316.

Newbery, D.M. (1989), "Cost recovery from optimally designed roads", *Economica*, 56, 165-185.

d'Ouville, E.L. and J.F. McDonald (1990), "Optimal road capacity with a suboptimal congestion toll", *Journal of Urban Economics*, 28, 34-49.

Oum, T.H. and Y. Zhang (1990), "Airport pricing: congestion tolls, lumpy investment and cost recovery", *Journal of Public Economics*, 43**,** 353-374.

Pels, E. and E. Verhoef (2004), "The economics of airport congestion pricing", *Journal of Urban Economics,* 55**,** 257-277.

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

Small, K.A. (1992), *Urban Transportation Economics. Fundamentals of Pure and Applied Economics,* Harwood, United Kingdom.

Small, K.A., C. Winston and C.A. Evans (1989), *Road Work,* Brookings Institution.

Verhoef, E.T., P. Nijkamp and P. Rietveld (1996), "Second-best congestion pricing: the case of an untolled alternative", *Journal of Urban Economics*, 40**,** (3), 279-302.

Verhoef E.T. (2005), "Second-best road pricing through highway franchising", Discussion paper, Free University, Amsterdam.

Verhoef, E.T. and J. Rouwendal (2004), "Pricing, capacity choice and financing in transportation networks", *Journal of Regional Science*, 44 (3) 405-435.

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

Yang, H. and Q. Meng (2002), "A note on 'Highway pricing and capacity choice in a road network under a build-operate-transfer scheme'", *Transportation Research*, 36A**,** 659-663.

Wheaton, W.C. (1978), "Price-induced distortions in urban highway investment", *Bell Journal of Economics*, 9**,** 622-632.

Wilson, J.D. (1983), "Optimal road capacity in the presence of unpriced congestion", *Journal of Urban Economics*, 13**,** 337-357.

ROAD USER CHARGES AND INFRASTRUCTURE

Georgina SANTOS
University of Oxford
Oxford
UNITED KINGDOM

SUMMARY

Oxford, July 2005

1. INTRODUCTION

Applied microeconomics prescribes that in order to achieve an efficient equilibrium, externalities need to be internalised, and those causing them should pay a corrective charge equal to the marginal external cost they impose on others[1]. Internalising externalities, however, does not guarantee an efficient outcome when there are externalities in other (related) sectors in the economy which are not internalised. Another obstacle is the calculation of the precise marginal external cost, which will typically vary with the conditions prevailing at the time. For example, pollution will be lower on windy days when particulate matter may be more rapidly dispersed. Even if the exact calculation of the marginal external cost were possible perhaps in real time, the cost effectiveness of these corrective charges would be dubious.

The transport sector constitutes a text-book example of a sector that generates a number of negative externalities that, if not dealt with, will typically result in an excessive level of activity. There is general agreement on the idea of the polluter-pays principle and even though it may not be practical to charge for the marginal external cost, there is at least consensus that those causing externalities with their travel should pay some kind of corrective charge.

A different but related problem is the provision and maintenance of infrastructure. Under-provision will cause excess demand and congestion, which will in turn increase the external cost. Over-provision will cause a waste of resources. Once the efficient level is known, the resources to provide and maintain it need to be determined.

Transport infrastructure has traditionally been funded by governments, either from general funds or with the revenues from charges or taxes levied on the users. This paper concentrates on the relationship between the social costs of road transport and provision and maintenance of road infrastructure. It is concluded that a well-designed system of road user charges will internalise most external costs and will cover the costs of maintenance of the road network.

1.1. The external costs of road transport

Vehicles impose four main costs on the rest of society: accidents, environmental costs, congestion and road damage (Newbery, 1990).

Accident externalities arise when extra vehicles on the road increase the probability that other road users will be involved in an accident. Newbery (1990) argues that accident externalities could be as large as all other externality costs taken together.

Environmental externalities of road transport comprise global warming, which affects the whole of the Earth, and air pollution, with regional effects such as acid rain and local effects such as health problems linked to high concentrations of particulate matter. Other local environmental costs usually associated with road transport include noise and visual obstruction.

Congestion arises when additional vehicles reduce the speed of other vehicles, and hence increase their journey time. The problem lies in the fact that drivers are not faced with the true social costs of their trips and only pay for their private cost. The difference between their private cost and their social cost constitutes the congestion externality.

Finally, the passage of vehicles causes damage to the road. The most costly combination is of course a heavy vehicle on a thin road. The road also gets damaged by the weather. Damage costs are borne by the highway authority, who repairs the road, and by other drivers, who bear extra vehicle operating costs caused by this damage (Newbery, 1990).

The way in which these externalities can be internalised to achieve an efficient equilibrium depends on the nature of the externality. The problem is discussed below.

1.2. Internalisation of the external costs of road transport

Transport accidents impose a range of impacts on people and organisations. The UK Department for Transport (2004) describes them as follows:

- Medical and healthcare costs[a]: in the UK these are borne by the National Health Service, and hence, by tax-payers, or more generally, society;
- Lost economic output[a]: this is a cost imposed on society;
- Pain, grief and suffering[a]: this is borne by the individual involved and friends and family;
- Material damage[b]: this is borne by the insurance companies;
- Police and fire service costs[b]: these are borne by tax-payers, or more generally, society;
- Insurance administration[b]: this is borne by the insurance company and ultimately by the individual who paid the insurance policy;
- Legal and court costs[b]: these are borne by the individual and by society.

Those impacts marked [b] are closely related to the number of accidents, while those marked [a] are related to the number of casualties (Department for Transport, 2004).

Accident externalities are already being internalised at least up to a point. The questions are "what fraction of external costs are not reflected in insurance costs?", and "how much should be related to distance driven?" (Newbery, 2005). Edlin (2002), DeCorla-Souza (2002) and Greenberg (2003) argue that insurance could take the form of charges per mile, rather than per year. Other factors such as driving record, type of car, etc., could also be considered. Discussing the different ways in which road traffic accident externalities can be internalised, however, falls beyond the scope of this paper.

Since the main road transport environmental externalities (i.e. global warming and pollution) are related to fuel emissions, which in turn are closely linked to fuel consumption, it seems practical to tax fuel. Emissions will vary slightly according to weather and traffic conditions, but these variations can be ignored. Keeler and Small (1977), for example, find that over a range of plausible optimal rush-hour speeds, fuel consumption is constant. The alternative for the environmental externality would be to charge emissions directly. This would slightly improve the linkage between the externality and the charge at a, probably, very high administrative cost. Fuel duties are a reasonably effective way of dealing with the environmental externality of road transport. In the UK, these are complemented with differentiated vehicle excise duties to reflect the different emissions per unit of fuel consumed by different vehicle types. Thus, diesel vehicles pay a higher vehicle excise duty than petrol vehicles because they are more polluting.

Almost nowhere are congestion costs internalised, with the exception of London and Singapore, which have urban congestion charging schemes in place. Some motorways in the US also have some kind of congestion charging. In any case, it is clear that it is possible to implement congestion charging when there is a political will and revenues are shown to improve the transport system. Congestion charging has been studied and debated by economists, physicists, engineers and policymakers since at least the early 1960s, and arguing for it is beyond the scope of this paper.

Finally, road damage costs in Europe are typically borne by the highway authority, which will repair the damage caused by the passage of vehicles (Newbery, 1990). In the UK, the highway authority follows a condition-responsive maintenance strategy. Thus, each road is repaired when its condition reaches a pre-determined state. If maintenance is condition-responsive then it is not necessary to charge vehicles for the damage they create for vehicles coming after them, which have to drive on a damaged tarmac as, on average, the condition of the road remains constant (Newbery, 1990). The resources to repair highways in the UK ultimately come from the Treasury, which in turn receives £31.5 billion (€45.7 billion) in fuel duties, VAT on fuel duties and vehicle excise duties (National Statistics Online, 2004). Heavier vehicles pay higher vehicle excise duty. For example, as of 2005, a goods vehicle no heavier than 7 500 kg pays an annual vehicle excise duty of £165 (€240), whereas a goods vehicle between 27 000 and 44 000 kg pays £650 (€944), unless it is 4 or more axled rigid, in which case it pays £1 200 (€1 743) (Driver and Vehicle Licensing Agency, 2005).

Since accidents and road damage can be treated separately, the two candidates for corrective charges are the environmental and congestion externalities. The environmental externality can be easily dealt with by taxing fuel. The congestion externality needs a finer system, which will at least differentiate between peak and off-peak times, and hence, peak and off-peak traffic conditions. Before discussing the problem of road transport externalities and their internalisation any further, let us now discuss the problem of road provision.

2. PROVISION OF ROADS

When there is excess capacity all the time, which of course implies a wasteful use of resources, a road is a public good. As such, it has the characteristics of non-rival and non-excludable consumption. Roads, however, do get congested, and when they do, the non-rivalrous feature disappears. Equally, it is not impossible to exclude certain vehicles from using the road. Roads are therefore impure public goods. Both pure and impure public goods constitute an example of market failure and, as such, they tend to be provided by governments with revenues from taxes. This is indeed the case of roads.

Having said that, if we consider a congested toll road, we are in the presence of a private good. It is excludable because only those paying the toll are allowed to use it, and consumption is rival, as in congested conditions each additional vehicle delays the others. An uncongested toll road, on the other hand, is excludable but non-rival, and thus it has been defined as a club good (Kopp, 2002).

Although traditionally roads have been publicly provided, governments in both developed and developing countries are facing dramatic growth in highway needs, both for new investment and for maintenance. This is a fairly serious problem in developing countries, many of which have heavily relied on private road toll financing since the 1980s (Fisher and Babbar, 1999). The World Bank estimates that private toll road development accounts for 8 per cent of the annual market for private infrastructure projects worldwide (Fisher and Babbar, 1999).

When, like in most countries in Europe, public expenditure is subject to the regular, yearly resource allocation procedures, including the share for road infrastructure, investment plans become uncertain (Ratouis *et al.,* 1999).

An alternative way of providing road infrastructure would be that of a regulated public utility, which is discussed in the next section.

2.1. Private provision of roads with government regulation

Newbery and Santos (1999) offer an alternative to road provision that would encourage private sector investment and, at the same time, provide the right price signals and protect road users from being overcharged and/or underprovided.

The idea they proposed consists of three key institutional changes:

- The creation of an agency responsible for managing highways;
- A regulatory agency to oversee its activities;
- The restructure of road charges to make them more transparent, thus reflecting external costs better. Part of these road charges could go into a road fund or be earmarked to new construction and maintenance of highways.

It is very likely that road user charges set according to external costs would produce a surplus for payment as interest or dividends to the Treasury (representing the public owner) which might prove adequate to finance investment, though the Treasury might wish additions to capital stock to be financed by borrowing, to ensure that its dividend stream was predictable.

The key for the system to work would be for the regulatory agency to ensure that the highways manager does not abuse its monopoly position, to set minimum standards of provision and to regulate the charges paid by road users by means of a price cap which could be reset periodically.

Newbery (2005) points out that the regulated road user charges could be set to recover the operating and capital costs of an efficiently run network. The capital cost has two components: the interest (or return) on the opening value of the Regulatory Asset Base (RAB) and the depreciation of this value. Determining the RAB, its interest and its depreciation seem therefore to be of crucial importance. Depreciation can typically be approximated by the maintenance costs needed to keep the equivalent asset value.

Newbery (2005) computes all these numbers for the case of the UK. He estimates the RAB at £200 billion (€ 290 billion). The interest rate to be used has recently been the topic of much debate in the UK. For many years, the Government used an interest rate of 6 per cent until 2002, when a revision took place. There was a period when the two rates, the old and the new were used, and both results were reported in appraisals. However, as of 2005, the interest rate adopted by the Government is 3.5%. Needless to say, the annual capital costs are highly sensitive to the choice of interest rate. Using the £200 as the RAB and the two interest rates, the range of capital costs would be £7-12 billion (€10-17 billion), excluding depreciations, which can be measured by maintenance costs. Maintenance is about £3-4 billion (€4-6 billion) per year. Total road costs are thus £10 to 16 billion (€14-23 billion) per year (Newbery, 2005), whereas road tax revenues, as will be pointed out below, are around

£26 billion (€38 billion) (National Statistics Online, 2004). The expenditure on roads is roughly £5.5 billion (Department for Transport, 2003, Table 7.13), which fully covers maintenance but not capital costs.

If a plan of this sort were to be actually implemented, the question that would need to be answered is how should these charges be designed. This is the topic of discussion in the next chapter.

3. WHICH SOCIAL COSTS SHOULD BE COVERED BY USER CHARGES?

The White Paper (European Commission, 2001) suggests that "transport users are entitled to know what they are paying for and why". Ironically, it then goes on to say that the "*fundamental principle of infrastructure charging is that the charge for using infrastructure must cover not only infrastructure costs… but also external costs, that is, costs connected with accidents, air pollution, noise and congestion* (p. 70)."

There seems to be some confusion between user charges and infrastructure charges. User charges may include external costs and even a pure tax element, whereas infrastructure charges are, as the name indicates, charges made on users for the use of the infrastructure. Further down the text, however, some difference between the two is made, implying that infrastructure costs and external costs are two different problems.

Road user charges should be transparent and consist of two elements. The first would be a set of road user charges based on a careful assessment of the relevant costs to be charged out. As stated above, these would include congestion charges and "green" taxes on environmental impacts such as pollution and greenhouse gas emissions[2]. The second would be a pure revenue raising tax on road transport.

User charges are price-like instruments that can be used for club goods, where it is possible to exclude those that have not paid. Although in the case of roads, exclusion of offending vehicles with physical barriers would not be practical as it would cause chaos and congestion, fining of offending vehicles[3] would offer an equally effective alternative.

An efficient road user charge system in the EU would increase investment on roads and internalise external costs, with the consequent effects of optimising the level of traffic. The only distortive element would be the (unavoidable) pure tax component (Newbery and Santos, 1999), which would need to stay in place, as fuel is an attractive tax base for any government in need of fiscal revenues (and they all are!). Let us now discuss the external costs that should be charged for.

3.1. Green tax

The idea that whoever causes pollution and global warming is responsible for the cost of repairing any damage or compensating those affected (polluter-pays principle) has become generalised and is even taught in schools. Since environmental damage from road transport is closely linked to quantity of fuel consumed, fuel duties constitute an effective way of charging, without prohibitive administration costs. As was argued under the section on the internalisation of externalities, they are not a perfect way but they are good enough, as the proxy they use, fuel consumption, approximates the quantity of emissions quite well. The more polluting the fuel the heavier the tax, and as long as the right incentives are provided vehicle users will make efficient decisions. As Newbery and Santos (1999) point out, these are politically attractive taxes as they are likely to command public support, since they will awaken some sense of stewardship or responsibility for preserving the environment.

Newbery and Santos (1999) argue that green taxes should be distinguished from other taxes or charges and set at levels that will at least approximate the environmental damage caused by emissions. In the UK, for example, although the high fuel duties are meant to internalise the environmental damage, "Mr Smith" does not know which part of the duty he pays is for the environmental cost he imposes when driving and which part is pure tax that will go to the Treasury to fund schools and hospitals. One idea then, which would increase transparency, would be a clear description of the components of road taxes and charges.

If green taxes are to be defended as corrective taxes that internalise environmental externalities, these will have to be quantified (Newbery, 2005). In theory, the green tax should be equal to the marginal environmental cost. This is easier said than done, as computing the marginal environmental costs of road transport is not an easy task. Besides, as Newbery (2005) points out, the damage per litre of fuel varies across vehicle types and ages, and although the vehicle excise duty can be fine-tuned to allow for those differences, fuel duties cannot. Emission standards such as the ones in place in the EU are a good complement of fuel and vehicle excise duties.

Santos *et al.* (2000) and Santos (2004) examine the effects of cordon tolls on pollution for a number of urban areas in England and find a very close relationship between congestion and the environmental benefits of cordons. Given the very wide range of environmental cost estimates of road transport emissions available in the literature, the high estimates of the total environmental costs they find are about 15 times as high as the low estimates, showing the considerable uncertainty attached to the figures. The pollutants they consider are carbon dioxide, carbon monoxide, volatile organic compounds, nitrogen oxide, particulate matter, methane, nitrous oxide and ammonia. The only costs considered in the analysis are global warming and health effects from pollution. They find that in all cases the optimal congestion toll was the same toll that would both maximise the increase in social surplus and the reduction in environmental costs.

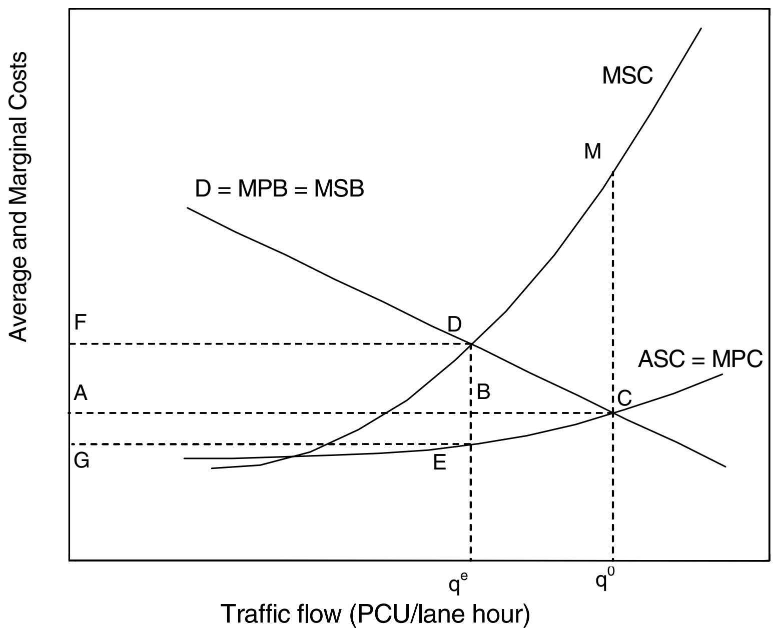
3.2. Congestion charge

Higher traffic flows lead to lower average speeds and higher travel times and costs per km. Additional traffic imposes an external cost on all other road users. Under congested conditions, particularly in urban areas, and in the absence of efficient road pricing, traffic will be undercharged and hence excessive.

In Figure 1, a simple road is assumed with uniform lane width and no junctions. Road users are identical apart from their marginal willingness to pay for a trip, given by the inverse demand curve D, which represents marginal private and social benefits. Marginal private (MPB) and marginal social benefit (MSB) are assumed identical (i.e. distributional considerations are ignored). The efficient equilibrium is at point D, where the marginal social cost (MSC) is equal to the marginal willingness to pay. The actual equilibrium however is at point C, where the average social cost (ASC)[4] is equal to the marginal willingness to pay. At this point, drivers are paying for their ASC but not for their MSC, leaving thus an externality or marginal congestion cost (MCC) that can be measured by the segment

MC. The inefficiency of incorrect pricing of scarce road resources or deadweight loss is then measured by the area DMC.

Figure 1. **The economics of congestion**



Unfortunately, these costs are not charged for, so motorists only value their own costs of making a trip, failing to consider the impact of their decision to add an extra vehicle to the route on other motorists.

The theoretical foundations of road congestion pricing were first proposed by Pigou (1920) as a way of increasing allocative efficiency. The efficient level of traffic is point q^e in the figure and one solution that has been suggested to achieve this optimal allocation is to internalise the congestion externality by raising the average cost for road users via a congestion charge, equal to DE. As well as optimising the flow of traffic, this solution would have the additional benefit of raising revenue equivalent to DEFG in Figure 1, which could then be used to improve other transport services.

It is difficult to implement this very well-known policy recommendation. The reasons for this difficulty are mainly the following:

– Marginal cost pricing in the road sector poses public and political acceptability issues, linked sometimes to concerns about equity;
– Introducing marginal cost pricing in the transport sector does not guarantee an efficient outcome when there are externalities or distortions in other (related) sectors in the economy, which are not priced according to marginal cost. Examples of these are unpriced externalities (such as environmental pollution), or distortionary taxes (for instance on the labour market in the case of commuter traffic). A congestion toll, for example, raises the overall costs of commuting to work and discourages labour force participation, which might be already quite discouraged by the presence of labour taxes[5]. Inefficiencies in related sectors are therefore likely to be the rule rather than the exception;

– Marginal cost pricing has proved difficult to implement. In today's technologically advanced world the calculation of instant marginal cost pricing may not be very difficult to envisage. Its cost effectiveness, however, would be dubious. How much would it cost to charge every road user for his or her MCC? How transparent would such a system be? Drivers would not know the congestion charge they would be required to pay before starting their journey[6]. Marginal cost pricing would require highly differentiated pricing systems in time and space, which would be expensive to provide and confusing to users (Nash and Sansom, 2001).

Since it is difficult to charge for the marginal congestion cost, an approximation can be used. A number of practical alternatives have been suggested over the years. These include an area licensing scheme, like the one in place in Central London, and a cordon toll, like the one which operated in Singapore between 1975 and 1998[7].

Even if a simpler system like an area licence or a cordon toll is chosen, there is still the problem of the charge level. The level can depend on the time of day and perhaps, vehicle congestive effects (e.g. lorries could pay a higher charge than cars). Designing an efficient congestion charging cordon or area licence is complex. There have been some theoretical attempts to do so but they are not applicable to real world situations. These include Verhoef (2002), Shepherd and Sumalee (2004), May *et al.* (2002), Mun *et al.* (2003) and Zhang and Yang (2004).

One alternative would be to have a target average speed in the problematic area or road where most of the congestion takes place, and to adjust the congestion charge until the target speed is achieved. This is the way the Singaporean system operates. The revenues from a congestion charge could be earmarked for investment in infrastructure and/or public transport. As of 2005 there are no congestion charges in Europe, except for the one in London. In London, the charge was £5 (€7.3) between February 2003 and June 2005. In July 2005 it was raised to £8 (€11.6). The charge level was not decided on the basis of any MCC estimate but was rather a political (almost arbitrary) decision. Yet, it achieved an important traffic reduction and an increase in average speed.

If road user charges included a green tax and a congestion charge, then they would give the right signals to both users and providers. To give the right signals, however, they would need to approximate the external costs they claim to internalise.

Assuming the environmental and congestion externalities were internalised, there would still be the problem of financing infrastructure. How would the revenues from those corrective charges manage to cover the maintenance and operating costs of roads? Luckily, the costs of congestion and infrastructure are not divorced. In the chapter that follows we discuss the link between congestion costs and highway costs.

4. CONGESTION CHARGES AND HIGHWAY COSTS

Keeler and Small (1977) build on previous work by Mohring and Harwitz (1962), Vickrey (1963), Strotz (1964) and Mohring (1970) and show that revenues from an efficient congestion charge (equal to the marginal congestion cost) are equal to the rental cost of the highway under the assumptions of constant returns to scale, indivisibility of road construction and independence of demand in each period from prices in other periods. Although at first sight we may see no link between efficient congestion charges and the rental cost of the road, there is one, and this is given by a function of net benefits that includes both benefits and costs to the users and the costs of the highway. It is both interesting and fascinating that congestion charges can cover highway costs. Congestion and road damage are two different externalities with different causes. They are not linked or dependent on each other and yet the internalisation of one may pay for the other. The model they propose can be summarised as follows.

The rental cost of the road where trips take place has three components: construction costs (on which interest and amortization are computed), maintenance costs and land acquisition costs. They call the sum of these three components $\rho(w)$. They then define the average variable cost function as $C_t = C_t(Q_t, w)$, where C_t is the average cost which includes all the costs borne by the user, such as vehicle operating costs and time costs, and the costs borne by the government, such as police costs. C_t is assumed to increase with the flow of vehicle trips on that road per unit of time, Q_t, and decrease with the width of the road, w.

They then define the objective function to be maximised as the net benefits of all trips on that road over the life of the road, computed as the difference between the net benefits to road users and the rental cost of the road. Here is where the link between congestion and rental cost of the road is to be found. By differentiating the objective function with respect to C_t and to w and setting each derivative to zero they arrive at two fairly standard results: (a) that the price paid by the user should equal the average cost plus the congestion externality; and (b) that lane capacity (or width) should be expanded to the point where the marginal cost of an extra unit of capacity is equal to the marginal value of user cost savings (typically, vehicle operating and time cost savings) brought about by that expansion. The beauty of their model is that with some simple algebraic manipulation these two results can be combined to show that the revenues from this efficient congestion charge are exactly equal to the rental cost of the road, assuming there are constant returns to scale in road construction, which makes $\rho(w) = aw$.

They then briefly discuss the consequences of the relaxation of their assumptions of constant returns to scale, indivisibilities and inter-temporal demand independency.

If there are increasing returns to scale, the road will need to be subsidized and if there are decreasing returns to scale the road will earn a surplus. Hau (1998) argues that all three cases are possible in the real world: decreasing returns to scale, constant returns to scale and increasing returns to scale. Newbery (2005) explains that for inter-urban roads in not very populated areas, there is some evidence of mildly increasing returns to scale. Hau (1998) gives the example of doubling a two-lane road to a four-lane road. This, he says, would almost double the capacity per lane and hence would almost quadruple total capacity. The reason for this is that an important proportion of road provision is dead space. Also, in rural areas, there are often earth-moving costs involved, such as filling valleys or levelling hills which, once incurred, will not be increased further by building an extra two lanes. On the other hand, both Hau (1998) and Newbery (2005) note that intersections will tend to be costly, regardless of whether the road expansion takes place in an urban or a rural area.

In densely populated urban areas, on the other hand, the cost of land acquisition is high and so road expansion will probably have decreasing returns to scale (Hau, 1998; Newbery, 2005).

If there are inter-temporal demand dependencies, demand in one period will be affected by prices in another. Thus, the efficient congestion charge will be the same in the short-run but will be different in the long run.

If there are indivisibilities, as they assume, then some roads will be either too large or too small. However, they argue, Neutze (1966) shows that when there are constant returns to scale in road construction and maintenance, although some roads will make money and some will lose money, a large group of roads will tend to break even overall. With increasing returns to scale, the system overall will lose money (and will need a government subsidy), and with decreasing returns to scale, the system overall will make money.

Their model is simple and elegant and goes to show that congestion charging revenues would cover highways costs under the assumptions mentioned above. The main problem with their model is that they assume the rental cost of the highway, $\rho(w)$, to be exclusively dependent on its width, and not on traffic. Newbery (1989) improves the model by assuming that maintenance to repair road damage does depend on traffic. Rather than having just a congestion charge, he proposes that vehicles are also charged for the damage they cause on the road. A road user charge would then include a congestion charge element and a road damage charge element. The congestion charge would be computed on the basis of passenger car units[8], and the damage charge would be computed on the basis of equivalent standard axles[9].

4.1. Road funds and earmarking

Two possible ways in which road infrastructure finance can be guaranteed with a stable flow of revenues are the setting up of a road fund and earmarking of taxes or charges.

A road fund is a fund that will receive revenues and provide them for the purpose of financing new road infrastructure and maintenance, including lighting and safety.

Earmarking, or hypothecation of taxes, will obligate the revenues from those taxes to some pre-defined allocation. For example, road taxes could be hypothecated to investment and maintenance of roads. A road fund would indeed need some hypothecation in order to survive.

Newbery and Santos (1999) report that the British Treasury has always resisted hypothecating or earmarking taxes to particular purposes, with only a few exceptions. The Treasury's view is that hypothecation of revenues should be limited to a few specific instances for which there is a very good case.

Table 1 presents the advantages and disadvantages of earmarking taxes.

Table 1. **Advantages and disadvantages of earmarking taxes**

Advantages	Disadvantages
Road users would benefit from paying road taxes related to road use, provided the revenue was actually spent on roads.	The government should have freedom to decide how tax revenues should be spent each year, and is also the best judge with the best information.
There would not be underinvestment on roads, or at least this would be reduced.	Road funds may discourage financial discipline, rather than increasing the efficiency of the road sector.
It may be a successful way to increase public acceptability of higher taxes[10].	It may prevent the funds from being spent somewhere else where there might be greater urgency or need.

Source: Queiroz (2003, p. 7) and author's own opinion.

Queiroz (2003) argues that many governments depend on fuel and other taxes for general revenues. Fuel taxes are relatively easy to administer and collect and so it is difficult for governments to let go. The US, however, does allocate most of the road tax revenues to expenditure on transport. Fuel taxes there are lower than in other developed countries, and about three-quarters goes to the highway account, a small portion is allocated to public transport and only the residual goes to the general budget.

4.2. A brief history of the Road Fund in the UK

In 1920 the *Roads Act* established the Road Fund, which received all the resources previously held in its predecessor, the Road Improvement Fund. The Road Fund's revenues came from the licence duties on "mechanically propelled vehicles", imposed from January 1921 (Newbery and Santos, 1999). Between 1930 and 1934, fees for drivers' licences, goods vehicle licences and driving tests were added to the revenues of the Road Fund. The Fund was also originally supplemented by a grant from the Treasury but the last such grant was made in 1923. Before too long, the Fund started to pay the Treasury. The main outlays were grants from the Fund to local authorities for road construction and maintenance. However, these grants were reduced by annual payments to the Treasury under the *Local Government Acts 1929* and the *Finance Acts* of 1935 and 1936. In 1936-37, payments to the Treasury amounted to over 40 per cent of all disbursements.

In 1937 the hypothecation of motor vehicle licence duties to the Road Fund was ended. The Road Fund itself was eventually abolished in 1955. Bracewell-Milnes (1991) quotes the relevant paragraph of the Second Report from the Select Committee on Estimates for 1953-54:

> *"…it would lead to greater clarity in the Estimates and more information being available to members if the Road Fund were abolished, and expenditure on roads provided for in a normal departmental Vote."*

Between 1937 and 1955, the Road Fund had no revenues of its own and became merely an administrator of the grant in aid.

Newbery and Santos (1999) summarize the lessons from the history of the British Road Fund. First, while the Fund existed, the only revenues were from motor licences, driving tests and driving licences. Fuel duties were too clearly tax-like and went to the Treasury. Second, although this did not really work in practice, the original aim of the Road Fund was to allocate revenues to *additional* expenditures required for motorised traffic. Third, and this survived until 1995, the accounts presented to Parliament were cash flow and not proper resource accounts. They only identified the amount spent by national and local governments on building, maintaining and policing the roads network, together with expenditure on roads administration, research and road publicity. Newbery and Santos (1999) point out that proper resource accounts would include the interest on the value of the capital and depreciation, which might be approximated by maintenance expenditure, and would treat investment in new facilities as a capital, not a current item. As a result, the accounts were ill-suited to the proper management of, and charging for, a capital-intensive activity in need of considerable early investment.

Finally, the problem with the major road investment programme that Britain clearly needed was that it would require large borrowing in the early stages to build up a capital stock. This would generate revenues to pay off the loans -- initial deficits would be matched by subsequent surpluses. It seems reasonably easy to finance activities with earmarked taxes where there is a fairly constant and predictable stream of expenditure closely related to the services supplied. Revenues can be matched to expenditures without major fluctuations from year to year, and hence without requiring detailed government scrutiny of the reasons for changing taxes.

The model of hypothecated revenue, where income and expenditure are in close balance, thus fails for roads, at least in the early transitional stage. Of course, the whole purpose of setting up a fund is that it will enable some smoothing of revenues and expenditures, but the scale of the gaps between road taxes and expenditure on roads could make some governments, like the British Government for example, uncomfortable (Newbery and Santos, 1999).

In the UK, road taxes[11], exclusive of VAT receipts, collected almost £26 billion (€38 billion) in 2002-03, or 6.7 per cent of total tax revenue. From that, only £5.5 billion (€8 billion) were spent on roads[12] (National Statistics Online, 2004, Table 7.13). Newbery and Santos (1999) report similar gaps since the 1950s. Interestingly, the allocation of road tax revenues to other sectors in the economy is not taking place because there is no need for investment in roads. In fact, investment in the road network in the UK is failing to keep up with the growth in traffic levels. Between 1991 and 2001, road traffic in the UK increased by 15 per cent, yet road length only grew by 9 per cent (RAC, 2003).

4.3. Toll roads

Although the provision of roads can be done through general taxation or through earmarking of taxes with or without a road fund, there is a third option, which has gained support in many countries, both developed and developing. This is the entirely private financing of roads, from the moment of design and construction all the way through to maintenance.

When a financially viable project with positive social benefits cannot be undertaken with government funds, be it because there are severe constraints limiting the borrowing capacity of the public sector or because of other priorities, the private sector may offer a solution by providing the necessary funds. When the private sector finances roads completely, the source of payment to the private sector will be the revenue generated by a given road infrastructure, i.e. the toll paid by the users (Ratouis *et al.*, 1999). Needless to say, there would need to be a regulator making sure prices are not monopolistic and that the quality of the service exceeds a minimum standard.

There are a number of possible contractual arrangements. Table 2 presents the ones which Ratouis *et al.* (1999) list as the most common.

Table 2. **Contractual arrangements**

Acronym	Name	Description
BOO	Build Own Operate	A private entity finances and builds an infrastructure project, which is owned, tolled and operated by that company for an unlimited time.
BOT	Build Operate Transfer	A concession is awarded to a private entity to finance, build and operate a tolled infrastructure during a limited period (usually of 20 to 40 years), at the end of which the infrastructure is transferred free of charge to the government.
DBFO	Design Build Finance Operate	A private entity is selected through competitive tender to build, own and operate infrastructure for an unlimited time. Payment is made to the private owner/operator by the public sector in the form of shadow tolls.
BTO	Build Transfer Operate	A private entity finances and builds the infrastructure, but upon completion transfers its ownership to the State. The infrastructure is then leased from the State, operated and tolled by the same, or another private company during a limited period (usually 20 to 40 years), at the end of which all rights have to be transferred to the State. Although the State can "own" the infrastructure from the first day of operation, the private company often keeps the full financial responsibility, which is not transferred to the State.
LIO	Lease Improve Operate	A private entity leases existing infrastructure, upgrades or repairs it, then operates and collects the revenues generated (usually tolls) over the duration of the lease.

Source: Ratouis *et al.* (1999).

Although there are many examples of toll roads around the world, two that seem worth mentioning, as they are both linked to congestion in some way, are the M6 Toll and the toll rings in Norway.

4.4. The M6 Toll in the UK

In the UK there is only one toll motorway: the M6 Toll. It is a 27-mile, three-lane motorway, which opened in the West Midlands in December 2003. The toll system places vehicles into different categories and charges separate rates for each category depending on the size of the vehicle. Rates also alter according to the time of day, with day and night tariffs.

As of the road's opening, tolls were £1 for motorcycles, £2 for cars, £5 for vans and £10 for lorries, each to rise by £1 after the first ten million vehicles. This figure was achieved in August 2004. A lower price is available during off-peak hours (23:00 - 06:00) as well as at the Langley Mill tolls for northbound exit and southbound entry to the motorway. On 23 July 2004, the toll for heavy goods vehicles was reduced from £10 to £6 due to the low numbers of lorries using the new motorway. Although the originally predicted daily numbers of vehicles that would use the M6 Toll was between 70 000 and 100 000, the average never exceeded 58 300.

Prices rose on 14 June 2005 by 50 pence for cars and motorbikes and £1 for larger vehicles and a 5 per cent discount was introduced for existing tag customers and anyone signing up to a tag account. Table 3 shows the charges as of July 2005.

Table 3. **M6 charges, July 2005**

	6:00 to 23:00	**23:00 to 6:00**
Motorcycles	£2.50	£1.50
Cars	£3.50	£2.50
Car & Trailer	£7	£6
Van/Coach	£7	£6
Heavy Goods Vehicle	£7	£6

Source : http://www.m6toll.co.uk/pricing/
Note: Wide and unusually heavy goods vehicles have higher rates.
Customers who have a tag get a 5 per cent discount per trip.

The original publicly-operated M6, which was built to carry 72 000 vehicles a day, copes with up to 145 000. As a result, the motorway is one of the most congested in the UK. The early suggestion was to widen it, but since part of the M6 is built on a viaduct, widening was not possible. After an extensive public consultation process, the decision was made to build a new motorway (M6 Toll, 2005) in order to relieve congestion.

In 1991, the UK Government decided the M6 Toll would be a privately funded venture. Midland Expressway was awarded the concession to run the M6 Toll for 53 years to 2054. The M6 Toll runs from Junction 3a on the M6 and arcs around the north-east of the West Midlands conurbation, rejoining the M6 at 11a. It links the M6 and M42 to the south of Birmingham and the M6 north of Birmingham, while bypassing the most congested section of the M6 which runs past Birmingham (M6 Toll, 2005).

There are various modes of payment. For each journey on the M6 Toll, vehicles pass through a toll station, either on the M6 Toll, or at a junction. As they approach the toll station, they see a number of signs indicating how and where to pay. They can choose an attended or an automatic lane and pay by cash or credit/debit card, although they must have the right change in all automatic lanes. They can also pay in advance of their journey using the M6 Toll tag, which is an electronic system that consists of a tag attached to the windscreen of the vehicle. The tag has a microchip that is automatically read as the vehicle enters the toll lane. If there is credit on the account, the vehicle will be able to go through without having to stop. If there is not enough credit the barrier will not rise.

4.5. The Norwegian Tolls

In Norway, road toll payment as a way to finance the building of bridges, tunnels, roads and fjord crossings, was established more than sixty years ago. The urban toll systems are much more recent: the first one was put into operation at Bergen at the beginning of 1986. It was followed by Oslo in 1990 and Trondheim in 1991. These are the three largest cities in Norway, with 250 000, 500 000 and 150 000 inhabitants, respectively.

The Norwegian Public Roads Administration has the responsibility for the planning, construction and maintenance of the toll projects as well as the toll collection system (Ramjerdi *et al.*, 2004). A private company operates the toll system. The shareholders in these private companies are usually public bodies such as the affected municipalities. The Road Acts were amended in the mid-1990s to allow the allocation of some of the revenues from tolls to investment in public transport (as opposed as investment on infrastructure only). Then, in June 2002, a new amendment was approved, allowing the modification of the tolls to manage traffic demand, and hence congestion. Thus, tolls, which were constant throughout the day, are now allowed to change with the time of day (Ramjerdi *et al.,* 2004).

The original aim of the Bergen scheme was to provide supplementary funding for a package of road projects. In 2003, about 45% of toll revenues were allocated to road construction and 55% to projects that improve environmental quality and safety. Because tolls were set at a low level, the impact on demand was never important. The objective of the scheme was to raise revenues and it succeeded in so doing (Ramjerdi *et al.*, 2004).

In Oslo, the toll was introduced in February 1990. The electronic payment system became operational in December 1990. Revenues are allocated to an investment package on public transport projects, referred to as “Oslo Package 2”. Originally, the toll revenue, supplemented by about equal funds from the central government, was to finance the “Oslo Package” (now referred to as “Oslo Package 1”), comprising some 50 new road projects to increase road capacity. About 30 of these road projects are tunnels that divert traffic from city streets. It is estimated that by 2007 the total contribution of the scheme to “Oslo Package 1” will amount to NOK 9.1 billion at 2002 prices (€1.2 billion at 2005 exchange rates), approximately 15-20% above the initial estimate.

The toll ring in Trondheim started in October 1991 with the full use of the electronic toll payment system. The revenue from the toll scheme funds a package of road infrastructure projects, mainly to increase road capacity, with some earmarking to public transport and facilities for pedestrians and cyclists. The toll revenue will contribute to about 40% of the total cost of the package, estimated at NOK 2.52 billion at 2002 prices (€328 million at 2005 exchange rates).

The tolls in all three cases are very low. For light goods vehicles, for example, the tolls are between NOK 10 and NOK 15 (€1.3 and €2).

5. CONCLUSIONS

This paper has discussed the external costs of road transport, the internalisation of externalities and the finance of road infrastructure. Although under certain (restrictive) conditions charging for congestion and road damage would recover the capital costs of highways, including interest and depreciation, this may not necessarily be so if the congestion charge is set at a level too different from the efficient corrective charge or if the assumptions of that theorem do not hold.

If road taxes were to be restructured in order to make them more transparent and earmark all or part for infrastructure finance, there would be a challenge: how to phase the changes.

The idea of charging for road use, reforming road taxation and internalising the externalities is attractive. However, there are two difficulties that would need to be dealt with before policymakers embark on changes of that sort.

To begin with, there is the problem of local authorities *vs*. central government. Congestion charges would be collected at a local level, whereas fuel and vehicle excise duties, or rather, the reduced fuel and vehicle excise duties, would be collected by the central government. This would present some tension, as there would be essentially a transfer of revenues from the central government to local authorities. An additional problem would be that of people coming from neighbouring towns and villages to buy cheaper fuel, although this might be dealt with by some kind of number plate registration for residents in that urban area, who would pay lower fuel duties.

If policymakers managed to surmount these difficulties, then restructured road user charges would internalise externalities and finance capital and infrastructure maintenance costs. The lag in investment not meeting demand for road trips would disappear as there would be a secure source of revenue to plan well ahead.

NOTES

1. The principle also applies to external benefits, where those generating them should be compensated.

2. All, as opposed to some, polluting modes of transport ought to be treated in the same way and charged according to the marginal environmental cost or best possible and practical approximation.

3. Offending vehicles could be tracked down via the Driver and Vehicle Licensing Agency or equivalent.

4. Average social cost is equal to marginal private cost. The assumption consistent with this is that road users are atomistic price-takers. As the price they consider includes time costs, one could rephrase this as "price and speed-takers". The marginal private cost they consider (the cost considered by the marginal user when contemplating whether or not to use the road) thus includes time costs and vehicle operating costs for their trip, given the level of road use (including themselves, but as they are atomistic, this of course does not matter). This is equal to the costs incurred by every user, and hence the average (social) cost.

5. Parry and Bento (2001) find that the resulting welfare loss in the labour market can easily exceed the Pigouvian welfare gain from internalising the congestion externality. However, if the revenues from the toll are used to reduce labour taxes, the net impact on labour supply is positive, and the overall welfare gain from the congestion tax can increase by up to 100 per cent.

6. This could be in part overcome with screens at all priced roads, showing the charge drivers would be required to pay further down. Thus, drivers could decide between paying the charge or using an alternative toll-free road or lane. This solution may work on motorways or in big cities (indeed it is used in the US) but it would be unlikely to work in English towns and cities where it is already difficult to find alternative routes, given the high number of one-way streets and the narrowness of some of them.

7. The cordon toll in Singapore was erroneously named as an area licensing scheme.

8. Passenger car units are a measure of the relative disruption that different vehicle types impose on the network. A car, for example, has a rating of 1, whereas a light goods vehicle has a rating of 1.5, etc. In the US, passenger car equivalents are used instead. The meaning, however, is the same.

9. One equivalent standard axle is the damaging power of an 18 000 pound (8.2 tonnes) single axle.

10. For example, RAC (2003) reports the results of a survey, where they find that motorists would be more supportive of congestion charges if the money raised was spent on improving existing roads (RAC, 2003).

11. Road taxes in the UK are fuel duties and vehicle excise duties.

12. This expenditure included new construction, improvement, structural maintenance, bridge maintenance and strengthening, safety and public lighting.

BIBLIOGRAPHY

Bracewell-Milnes, B. (1991), "Earmarking in Britain: Theory and Practice", Part 2 of R. Teja and B. Bracewell-Milnes (eds.), *The Case for Earmarked Taxes,* Institute of Economic Affairs, London.

DeCorla-Souza, P. (2002), "Estimating the benefits from mileage-based vehicle insurance, taxes and fees", *Transportation Research Record,* 1812, 171-178.

Department for Transport (2003), *Transport Statistics Great Britain,* 2003.

Department for Transport (2004), *Transport Analysis Guidance,* The Accidents Sub-objective, Webtag Unit 3.4.1, December. http://www.webtag.org.uk/webdocuments/3_Expert/4_Safety_Objective/3.4.1.htm

Driver and Vehicle Licensing Agency (2005), *Rates of Vehicle Excise Duty.* http://www.dvla.gov.uk

Edlin, A.S. (2002), "Per-mile premiums for auto insurance", Department of Economics, University of California, Berkeley, http://ideas.repec.org/p/wpa/wuwple/0303001.html

European Commission (2001), *White Paper: European Transport Policy for 2010 - time to decide*, Office for Official Publications of the European Communities, Luxemburg.

Fisher, G. and S. Babbar (1999), Private Financing of Toll Roads, RMC Discussion Paper Series 117, World Bank. http://rru.worldbank.org/Documents/Toolkits/Highways/pdf/05.pdf

Greenberg, A. (2003), "Comparing the benefits of mileage and usage pricing incentives with other government transportation expenditures", published in the CD-Rom of the 82nd Annual Meeting of the Transportation Research Board.

Hau, T. (1998), "Congestion Pricing and Road Investment", in: K.J. Button and E.T. Verhoef (eds.) (1998), *Pricing, Traffic Congestion and the Environment*, Edward Elgar, Lincolnshire.

Keeler, T. and K. Small (1977), "Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways", *Journal of Political Economy*, 85(1), pp. 1-25.

Kopp, A. (2002), "Compressed Analytics of Infrastructure Pricing and Capacity Choice", ECMT Mimeo.

Lindberg, G. (2001), "Traffic Insurance and Accident Externality Charges", *Journal of Transport Economics and Policy,* 35(3), pp. 399-416.

M6 Toll (2005). Homepage. www.m6toll.co.uk/

May, A., R. Liu, S. Shepherd and A. Sumalee (2002), "The Impact of Cordon Design on the Performance of Road Pricing Schemes", *Transport Policy,* 9(3), pp. 209-20.

Mohring, H. (1970), "The Peak Load Problem with Increasing Returns and Pricing Constraints", *American Economic Review*, 60(4), pp. 693-705. Cited in Keeler and Small (1977).

Mohring, H. and M. Harwitz (1962), *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston, Ill. Cited in Keeler and Small (1977).

Mun, S., K. Konishi and K. Yoshikawa (2003), "Optimal Cordon Pricing", *Journal of Urban Economics,* 54(1), pp. 21-38.

National Statistics Online (2004), Tax revenues, taxes and excise duties: Environmental Accounts: Government revenues from environmental taxes (£ million). http://www.statistics.gov.uk/STATBASE/Expodata/Spreadsheets/D5688.xls

Nash, C. and T. Sansom (2001), "Pricing European Transport Systems", *Journal of Transport Economics and Policy*, 35(3), pp. 363-380.

Neutze, G. (1966), "Investment Criteria and Road Pricing", *Manchester School of Economic and Social Studies*, 34 (January), pp. 63-73. Cited in Keeler and Small (1977).

Newbery, D.M. (1989), "Cost Recovery from Optimally Designed Roads", *Economica*, 56 (May), pp. 165-85.

Newbery, D.M. (1990), "Pricing and Congestion: Economic Principles Relevant to Pricing Roads", *Oxford Review of Economic Policy*, 6 (2), pp. 22-38.

Newbery, D. (2005), "Road User and Congestion Charges", in: S. Cnossen (ed.), *Theory and Practice of Excise Taxation: Smoking*, *Drinking, Gambling, Polluting and Driving,* Oxford University Press, Oxford.

Newbery, D. and G. Santos (1999), "Road Taxes, Road User Charges and Earmarking", *Fiscal Studies*, 20(3), pp. 103-132.

Parry, I. and A. Bento (2001), "Revenue Recycling and the Welfare Effects of Road Pricing", *Scandinavian Journal of Economics,* 103(4), pp. 645-671.

Pigou, A. (1920), *The Economics of Welfare*, McMillan and Co., London.

Queiroz, C. (2003), "A Review of Alternative Road Financing Methods", paper presented at the Transport Infrastructure Development for a Wider Europe Seminar, Paris, November 27-28. http://unpan1.un.org/intradoc/groups/public/documents/UNTC/UNPAN013148.pdf

RAC Foundation for Motoring (2003), RAC Report on Motoring: *Making the Most of Britain's Roads,* London.

Ramjerdi, F., H. Minken and K. Østmoe (2004), "Norwegian Urban Tolls", in: G. Santos (ed.), *Road Pricing: Theory and Evidence,* Oxford: Elsevier.

Ratouis, D., J. Charoud and O. Georges-Picot (1999), "Financing of road infrastructure: Guide for New Methods of Financing and Public/Private Partnership", http://rru.worldbank.org/Documents/Toolkits/Highways/pdf/03.pdf

Santos, G. (2004), "Urban Congestion Charging: A Second Best Alternative", *Journal of Transport Economics and Policy,* 39 (3), pp. 345-369.

Santos, G., L. Rojey and D. Newbery (2000), "The Environmental Benefits from Road Pricing", DAE Working Paper 0020, Department of Applied Economics, University of Cambridge, October. http://www.econ.cam.ac.uk/dae/repec/cam/pdf/wp0020.pdf

Shepherd, S.P. and A. Sumalee (2004), "A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problem", *Networks and Spatial Economics,* 4, pp. 161-179.

Strotz, R. (1964), "Urban Transportation Parables", in: J. Margolis (ed.) (1964), *The Public Economy of Urban Communities*, Johns Hopkins Press, Baltimore. Cited in Keeler and Small (1977).

Verhoef, E. (2002), "Second-best Congestion Pricing in General Networks: Heuristic Algorithms for Finding Second-best Optimal Toll Levels and Toll Points", *Transportation Research B,* 36 (8), pp. 707-729.

Vickrey, W. (1963), "Pricing in Urban and Suburban Transport", *American Economic Review*, 53 (May), pp. 452-65. Cited in Keeler and Small (1977).

Zhang, X. and H. Yang (2004), "The Optimal Cordon-based Network Congestion Pricing Problem", *Transportation Research B,* 38 (6), pp. 517-537.

SUMMARY OF DISCUSSIONS

Andreas KOPP
Chief Economist
OECD/ECMT Transport Research Centre

SUMMARY

1. INTRODUCTION

The White Paper of the European Commission (EC, 2001) initiated a broad and extended discussion in Europe on charges for infrastructure use. A similar, but narrower discussion continues in the USA on “value pricing” of road use, i.e. extending the existing road system by the provision of fast lanes which will be charged for, while other lanes can be used for free (Small, 2001). These discussions had focused on charges for infrastructure use as a means to contain transport demand, and on the internalisation of external costs, mainly environmental and safety, by relating the level of charges to these costs at the margin. Despite extensive discussion, many stakeholders and policy analysts are disappointed by the level of implementation. Where infrastructure charges have been introduced -- the London toll ring and the German motorway charges for heavy duty trucks -- the pricing rules employed differ markedly from the concepts recommended by policy planning (Prud’homme and Bocarejo, 2005). The level of political resistance to a more general introduction remains high. In many cases, fiscal motivations for introducing charging schemes were as important as objectives to increase the efficiency of the transport sector.

The fiscal reasoning for the introduction of infrastructure charging is based on the claim that only revenues from infrastructure charging could ensure that sufficient resources are available to finance the transport infrastructure required to achieve economic development objectives. This claim is, sometimes implicitly, associated with the argument that these resources cannot be generated from general taxes. This, in turn, is due either to the impossibility to increase taxes or the fact that transport policy allocations fall short of what an optimal fiscal plan would indicate.

Against the backdrop of this situation, the Round Table addressed the following questions, which have been neglected in much of the earlier discussions:

-- Is it possible to design a quasi-market of infrastructure services? Such a market would have to be based on prices for infrastructure services which are not only aimed at target demand levels, but which guide supply decisions at the same time. The supply of infrastructure services would be determined by the decision on the capacity of the infrastructure stock.

-- Would the pricing rules that have been discussed in recent debates have to be modified? This would mainly concern the question of to what extent external costs would and should be taken account of in defining a pricing rule that would help to provide infrastructure services at the least possible cost.

-- Would charging and the provision of infrastructure services at least cost be possible by relying exclusively on distance-related charges? If not, how should the difference between full costs and the prices required for optimal use be covered?

-- Is political resistance to road pricing inevitable? The introduction of infrastructure charging has been argued for by presenting it as an efficiency enhancing step of fiscal reform. Infrastructure users perceive the introduction of infrastructure charges as just another tax.

The gap between the implementation of charging systems and its justification -- e.g. pricing according to broad vehicle classes which are only loosely connected to environmental costs -- has led to the perception, by at least some user groups, that infrastructure "pricing" would in fact mean the introduction of just another tax with a justification lacking plausibility and credibility.

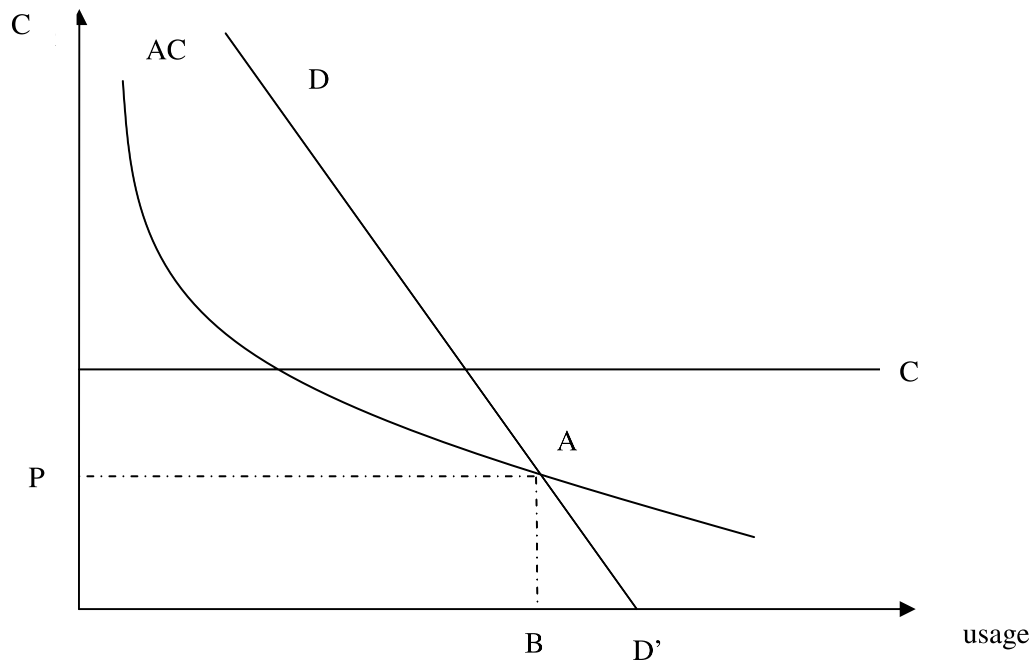
-- What mechanisms are required to ensure that the revenues of infrastructure charging are indeed used as payment for infrastructure services and not diverted to other purposes? This question is closely related to the debate on the desirability of infrastructure funds, most prominently the one between the World Bank and the International Monetary Fund. That discussion focused on how to ensure an appropriate funding of road maintenance. While the World Bank sought a remedy to cure the under-funding of road maintenance in some client countries, the IMF was concerned that off-budget cost recovery mechanisms would increasingly interfere with rational fiscal programming. There were also doubts that the abuse of infrastructure funds for other purposes could be avoided. The Round Table discussion showed that beliefs about the likely outcomes of budgetary processes did not differ as much as earlier discussion had suggested and that there was agreement on the need for careful institutional design of the management of road funds to respond to the concerns of the IMF.

2. PRICING PRINCIPLES AND COST RECOVERY

It might be instructive to recall why transport infrastructure services are normally not provided on private markets, and why local or national governments play a major role in taking infrastructure investment decisions to build up capital stock which determine the capacity of individual facilities and networks overall. The salient feature of transport infrastructure services is the fact that they require a durable input good -- transport infrastructure capital stock -- which can only be installed in large, discrete units. Once established, the investment is sunk, i.e. there is no possibility to turn infrastructure capital stock into another physical good, nor does a resale market normally exist for these capital goods. These characteristics imply that, at least at low levels of demand, the costs per user of the infrastructure decrease with an increase in the number of users and the level of service demanded by the individual user.

What this means for pricing may be clarified by an extreme example. It is possible that, in sparsely populated geographical areas, all of the infrastructure costs are independent of the actual low levels of usage. The initial construction costs would imply that a certain annual opportunity cost of finance (debt service or other public service forgone by having built the infrastructure facility) and administrative as well as maintenance costs are determined by the fixed annual costs of a minimum administrative and maintenance unit. Different use levels of, let's say, a road system by the local population do not make a difference to the routine activities of the administration and maintenance unit. In this case, user charges per kilometre of road use would not improve the provision of infrastructure services but worsen it, as may be illustrated by Figure 1.

Figure 1. **Only fixed costs: optimal prices are zero**

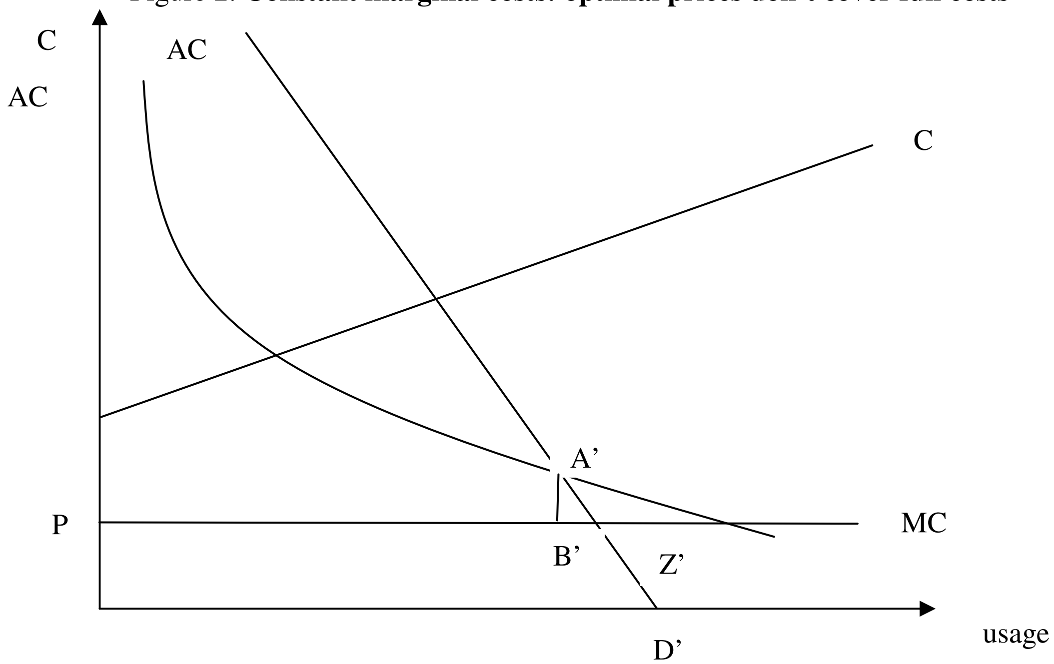


The average cost curve (AC) indicates the decreasing costs which accrue annually per unit of service, depending on the total demand for services. As all annual costs are fixed (C), they decrease continuously with an increasing use level. DD' indicates demand for infrastructure services depending on charges per unit of usage. The area below the demand curve indicates the benefits resulting from the use of the infrastructure facility. Any price charged would lead to a loss of benefits, due to the reduction in road use that is associated with a price of zero, without leading to a reduction in costs (being all fixed due to low levels of usage). Should a provider be required to supply the service at prices per unit of service that just cover costs, the price would be P. At this price level, the users would lose benefits equivalent to the triangle (ABD') without any associated reduction in costs. The optimal per-km user charge in such a case would be zero. While this might explain why there are normally no private markets to provide infrastructure services, the above argument does not imply that a provider cannot or should not recover the (fixed) costs of providing the service. The optimal way of recovering costs is to levy a lump-sum access charge for all users of the transport infrastructure facility, i.e. a charge that is unrelated to individual usage[1]. The decision on capacity choice would be a trivial one. The road administrators would invest in a facility of minimum size (e.g. two lanes) and a desirable quality.

A more important reason for government involvement in the provision of infrastructure is the fact that the unregulated private provision of infrastructure would lead to an underutilisation of the resources that have been invested in transport infrastructure stock. In fact, even if all infrastructure costs were fixed costs, an unregulated provider would increase the price up to a point where the effects of the price increase on revenues and the induced decrease of demand cancel out. The price would be much higher than P in Figure 1. In other words, a (monopolistic) provider will seek to turn consumer benefits into rents and thereby reduce the benefits to society overall[2].

This argument remains valid if the administration and maintenance costs increase with an increase in the number of users or an increase in infrastructure use per user, if the increases are constant. A linear relationship between infrastructure costs and usage, i.e. constant marginal costs, still leads to decreasing average costs due to the fixed construction and maintenance costs, as shown in Figure 2. The slope of rays from the origin of the co-ordinate system to the cost function, indicating the average costs, decreases with an increase in usage.

Figure 2. **Constant marginal costs: optimal prices don't cover full costs**



With positive marginal costs, additional use of the facility will directly lead to additional costs. If users are not charged for these additional costs, they will tend to overuse the facility. If a cost-recovery mechanism like a lump-sum access charge were in place, users with very high individual demand would cause additional, use-dependent costs that could partly be shared by users with low demand, above the level to which they have contributed to variable costs. To avoid incentives to overuse the infrastructure in this sense, and to avoid the negative distributional consequences, the price per unit of infrastructure use should be equivalent to the (constant) marginal costs.

The revenue from pricing would still not cover the full costs: the fact that the average costs decrease indicates that the extra costs following from additional usage fall short of the unit costs. A price higher than the marginal costs would, however, indicate an under-utilisation of the accumulated stock of (sunk) investment. The collective of users would lose benefits equivalent to the triangle A'B'Z'. To cover the total costs, a fixed access charge, for example a vehicle tax, is required that covers the difference between total costs and the revenues from marginal cost pricing. As long as congestion continues to be unimportant, a linear cost function for infrastructure use does not alter the above argument on capacity choice: infrastructure managers will choose the minimum capacity, as expected demand remains below the level where time losses, due to crowding, occur. As in the case with only fixed costs, an unregulated provider will try to implement charging policies that do not only cover fixed costs but that maximise monopoly rents.

3. EXTERNAL COSTS OF INFRASTRUCTURE USE

3.1. External road damage costs

As mentioned in the introduction, external costs, i.e. costs generated by one group of infrastructure users that accrue to other users or to non-users, and their inclusion in the calculation of prices, have played a major role in pricing discussions. The environmental costs of most political concern are those that affect non-users. The external part of road damage and congestion costs concerns other road users. As we will see, pricing schemes to recover costs should include some of these costs, but should exclude others if the prices and revenues in equilibrium are supposed to give guidance to investment policies.

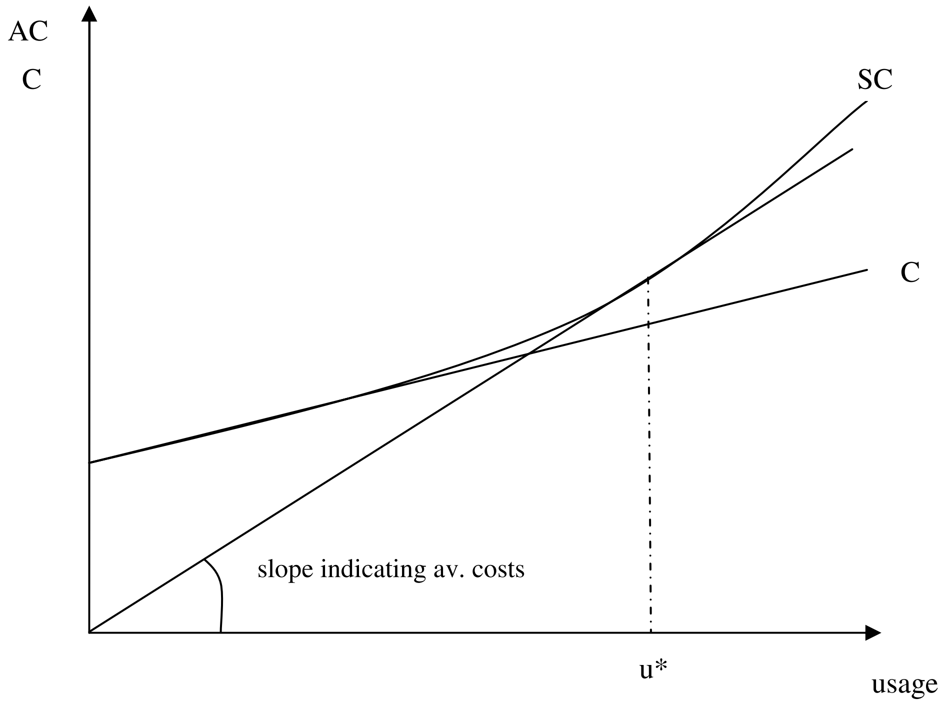
A first category of "external" cost, closest to the core administrative and maintenance costs, arises from road damage (Newbery, 1988). Similar externalities may hold for other types of transport infrastructure. When a vehicle damages the road surface, the increased roughness of the surface leads to increased *vehicle costs* for subsequent traffic. The increased vehicle operating costs constitute a road damage externality. On well-trafficked inter-urban roads these vehicle operating costs are estimated to be between 10 and 100 times as high as maintenance costs (Newbery, 1988, p. 298). This has not generally been considered in debates on the internalisation of external costs in road pricing to date.

Road damage caused by vehicles advances the date at which the road needs to be repaired. The most important type of damage is best expressed by the increased roughness of the surface, which can be quantified by instruments like the Bump Integrator. For a well-designed road, its initial roughness will be low and will continuously increase with the passage of traffic. Depending on pavement type, the roughness reaches, after 10-20 years, a level at which major maintenance, such as an asphalt overlay, is required to restore the surface to its initial low level of roughness.

The damage vehicles cause depends on the type of vehicle and the type of road. The damage of vehicle types can be measured as some fraction or multiple of a standard damaging unit as, for example, the equivalent standard axle load for paved roads.

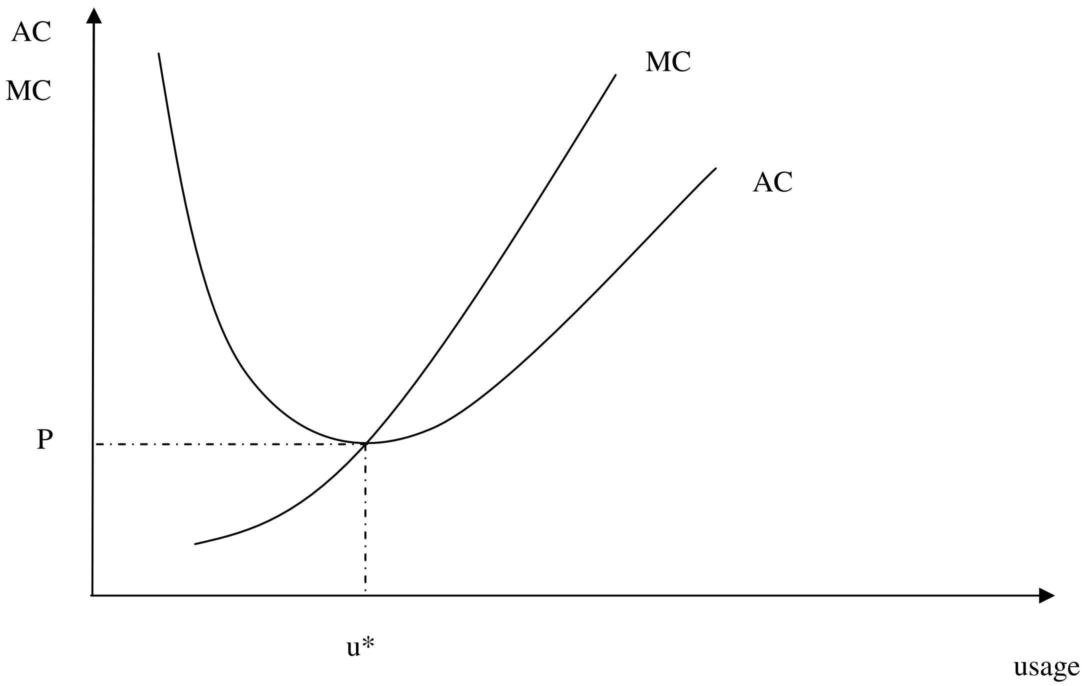
What is important for the discussion of the link between pricing and investment decisions is the fact that adding the external road damage costs to the cost function of road infrastructure services makes the total cost function convex to the origin. In other words, the additional (internal and external) costs of transport infrastructure services increase with increasing demand. This has the important consequence that the total infrastructure costs per user or per unit of service no longer -- as in the cases of only fixed costs or fixed costs and constant marginal costs -- decrease continuously with increasing demand. Due to the fact that the increase in marginal costs is weak for low levels of traffic, an increase implies that the average costs might decrease for low levels of demand, reach a minimum and increase with strongly increasing external road damage costs, as shown in Figure 3.

Figure 3. **Infrastructure costs and road damage externality**



That is, even if the cost function, which comprises only administrative and maintenance costs, should be linear, the external road damage costs in the form of increased vehicle operation costs, for example, due to an increased roughness of road surfaces, might lead to an exponentially, i.e. more than proportionately, increasing total cost function. The minimum of the associated average cost function is associated with the equality of average and additional costs caused by the last extra unit of service, indicated by u* in Figure 3. Figure 4 shows the marginal cost curve, MC, indicating additional maintenance and administrative costs as well as vehicle damage, resulting from increased roughness of road surfaces due to an additional unit of road service, and the average cost curve AC, indicating total costs per unit of service. It illustrates that, for low levels of usage, average costs are higher than marginal costs. That is, for levels of usage below u* its increase reduces the amount of resources required per unit of service. In other words, the efficiency of the infrastructure sector is increased as the resources bound by past investment and spent for maintenance allow an increase in transport services.

Figure 4. **Average costs and optimal capacity**



Pricing according to marginal cost at u* would just cover the average costs. At this usage level, charging only for the additional internal and external costs covers the full costs of the facility, despite the fact that transport infrastructure facilities are usually associated with relatively high, fixed construction costs. For demand that is at least as large as the usage at minimum average costs (u*), per-unit-of-service charges cover the full costs.

That demand at a minimum average cost price exactly happens to be equal to the minimum cost level can only happen by chance.

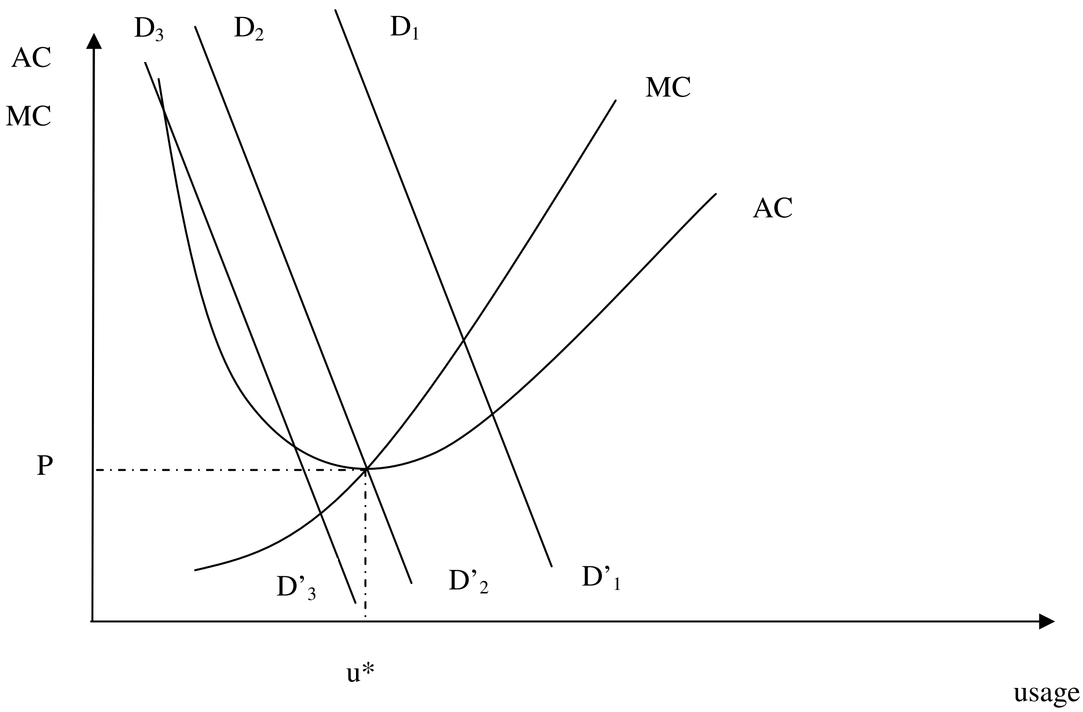
For the decision on the capacity of the facility, this can have three possible consequences:

1. The demand level is smaller than the minimum cost use level. In that case, the existence of a minimum average cost is irrelevant for infrastructure managers. An increase in usage will lead to a decrease in average costs, and therefore the above arguments for the cases of only fixed costs and proportional additional costs apply. The price should not cover the full costs, as it would lead to an under-utilisation of the infrastructure facility. A fixed charge (e.g. an access charge, a vignette or a vehicle tax) should be used to cover the difference between total costs and revenues from minimum AC. Knowing that demand is and will remain low, the pricing infrastructure managers will choose the minimum capacity.

2. Demand just equals the use level. A price which just covers the additional costs resulting from an extra unit of demand will cover the full costs. The capacity of the facility equals the minimum average costs.

3. Demand at the minimum average cost price is greater than the optimal use level. In that case, a marginal cost price will lead to revenues which are greater than the total costs of the facility. At the same time, infrastructure users will suffer from high levels of vehicle damage, due to the high level of wear and tear of infrastructure facilities. With increasing usage, it will become cheaper for society overall to add lanes or establish a second facility -- leading to lower costs per service unit despite the need to finance a second block of fixed costs -- because it avoids the high external costs of vehicle damage[3]. With an increase in the number of facilities (road lanes, railway tracks), the demand per facility might fall short of optimal use levels u*. The per-unit-of-service price should always be equivalent to minimum average costs, and possible deficits should be covered by a fixed fee.

Figure 5**. Optimal capacity and equilibrium demand**



3.2. Congestion costs

Congestion costs have been more important than the vehicle costs resulting from overuse of infrastructure for the development of pricing rules as well as for practical policy discussions. The quantity dimension of congestion costs is measured in time units as the delay caused by a number of simultaneous users of a facility which impedes a free flow of vehicles (defined by adequate safety standards). A long and protracted discussion exists on how to value these delays (cf. Round Table 127). The present Round Table discussed the role of congestion costs for rules of infrastructure pricing and decisions on the capacity of infrastructure facilities.

Congestion costs are "external" because individual decisions to use a road or any other crowded transport facility give rise to time delays for other users. This has led to the view that congestion is to be considered like other external costs. Behaviour leading to external costs should be contained by taxing their negative outcomes.

The conceptual literature following this view paid little attention to the informational function the price for infrastructure use has for decentralised infrastructure investment decisions, but focused strongly on demand management aspects (e.g. Proost *et al*., 2003b).

The early literature on optimal pricing and self-financing of infrastructure, taking account of congestion costs, made rather specific assumptions on the relationship between congestion costs and the increase of transport infrastructure capacity. Disregarding non-linear pricing or two-part tariffs, it assumed that there are "constant returns to scale in the congestion technology" (Small, 2005). That is, it was assumed that travel times remained unaffected when both the infrastructure use and the road capacity expanded continuously and in the same proportion. This, in turn implies that there are constant economies in road construction. The fact that most transport infrastructure can only be expanded in major, discrete steps is given little importance in these considerations (cf. the review in Verhoef, forthcoming).

Another element of the pricing discussion, of more relevance for the link between pricing and the capacity of infrastructure facilities, is the literature explaining to what extent the production of public goods can be decentralised (cf., for example, Starrett, 1988, ch. 4). Public goods are facilities which can be utilised by more than one user, without one curtailing the usage of the other. Infrastructure goods belong to the particular class of these goods where "non-rivalry" in consumption is due to the physical indivisibility of the good. This implies that, at low levels of demand, more users can enjoy the services without any reduction of others' consumption. Congestion is another term for the crowding of such public goods. What makes crowding or congestion interesting when thinking about public finance is the fact that it allows for a decentralised, market-like supply of this type of public good. The collection of users is sometimes called a "club", to raise the connotation that supply should be organised as if the group of users would collectively decide how this should be done.

The substance of this argument is identical to the one in the preceding section on external costs of road damage. The "indivisibility" means that infrastructure facilities are associated with high fixed costs. The additional administrative and maintenance costs which can be attributed to an additional user are low and often considered to be constant. That is, without taking external costs into account, average costs are decreasing. Adding congestion costs to the picture, the social costs will at some usage level become higher than the core infrastructure costs and increase more than proportionately, as in Figure 3. When this is the case, there will be a minimum average cost level, which indicates how the unit of service should be priced and what size the facility should have. The level of congestion costs is positive at the point of minimal social costs. That is, infrastructure charging does not aim at the complete removal of congestion. Rather, beyond a use level of u*, the users favour the expansion of infrastructure capacity over reducing infrastructure use through increased prices. Taking both road damage and congestion into account will have the consequence of making the exponential growth of the social costs of infrastructure services stronger.

3.3. Other external costs

In recent pricing discussions, in particular in the European Union, other external costs, for example for environmental damage from transport and accident risks, have been included in the calculations for pricing prescriptions. This has to some extent to do with the focus on the demand management dimension of pricing; environmental damage due to pollution is seen as analogous to congestion costs. The environmental damage associated with using transport infrastructure should be priced to contain behaviour which incurs costs on others. Including all kinds of external costs in calculating prices for infrastructure use might have the advantage of saving costs for the fiscal administration. It has, however, important disadvantages for establishing a quasi-market for infrastructure service, linking a "fee for service" to supply decisions.

In general, external costs should be corrected for as close to their causes as possible. In the case of air pollution, pollutants should be taxed as a matter of principle. This would give the most direct and strongest incentives to avoid pollutants by changing transport behaviour or the technologies used. Such specific measures to correct for external costs might be expensive to implement and require less specific proxy instruments.

However, when infrastructure charges are designed to be used for infrastructure investment and/or maintenance, the attempt to internalise external costs by charging per unit of service leads to dysfunctional cost recovery mechanisms: if, for example, environmental costs are included in the calculation base, the funding of infrastructure is more generous the lower the environmental standards of the vehicle fleet.

In the process of policy decisionmaking, road managers, the construction industry, etc., might thus have reason to oppose legislative measures to reduce the environmental damage caused by transport. The exclusion of environmental or other costs from the infrastructure charging scheme requires other, more specific measures to reduce them. The need for specific corrective measures for specific external costs was discussed on the basis of the background paper by Santos (forthcoming), which also looks into the costs and benefits of earmarking corrective taxes. Measures to correct external costs, for example fuel taxes to reduce CO_2 emissions, will shift the demand curve in the above diagrams to the left. The above arguments on optimal pricing and capacity choice remain unaltered if "demand" is the demand that results after appropriate corrective measures (other than charges for infrastructure use) have been put in place.

4. INFRASTRUCTURE FUNDS AS COST RECOVERY MECHANISMS

To establish an argument that a quasi-market for infrastructure services is conceivable and feasible does not necessarily imply that it is desirable. One section of the Round Table discussion tried to clarify whether off-budget cost recovery mechanisms are needed, and what the advantages of such a mechanism would be compared to a traditional fiscal system of infrastructure investment and maintenance funding.

The Round Table discussion referred to the debate between the International Monetary Fund and the World Bank on the usefulness of road funds (mainly in developing countries). The IMF saw the establishment of infrastructure funds as a threat to an ordinary budget process, limiting the opportunities for rational fiscal programming, particularly in times of changing tax revenues. There was also the concern that infrastructure funds could and would be abused for other than infrastructure funding purposes or for ostentatious projects. In fact, the application of (some version of) optimal taxation concepts would result in the same outcome as the quasi-market sketched above (e.g. Diamond, 2003a): public goods should be provided by raising finance through fixed charges and linear taxes on net trades in goods and services. The question is whether the outcome of the political process to provide transport infrastructure will be equivalent to a planning result that exclusively takes the infrastructure users' interests into account (Potter, forthcoming).

The World Bank's experience was in stark contrast to such an expectation (Gwilliam, forthcoming). Inadequate road maintenance is seen as a problem of the highest priority in many client countries. In Africa, it was estimated that during the two decades of the 70s and 80s, road stock to a value of US$45 billion was lost due to inadequate maintenance, which could have been avoided by expenditure on preventive maintenance of only US$12 billion (Brushett, 2002). A systematic comparison between planned and realised expenditures for transport investment and maintenance only exists in rare cases. A World Bank assessment of such information revealed that in all countries in the sample, actual expenditures were far below the planned expenditure levels. The highest value was 58 per cent and the lowest 15 per cent.

While the problems of funding transport infrastructure investment and maintenance may be less severe in countries with a highly developed fiscal administration, there is an almost universal effort to increase the number of public-private partnerships in transport infrastructure in order to mobilise funding for planned transport infrastructure projects. This suggests that there is the perception that the budget process favours other portfolios relative to transport policy. In these cases, off-budget cost-recovery mechanisms, following the self-financing concept presented above, would improve the contribution of the transport sector to the overall economic development.

In part as a response to the concern that infrastructure funds could be abused, to earn hidden profits or bureaucratic rents, the concept of "second generation road funds" has been developed. This concept gives an example of how to contain the risks of the diversion of funds and to ensure that they work to the advantage of infrastructure users. The infrastructure funds should be organised according to the following features:

i) Charges should be established in addition to and entirely independent of the determination of levels of taxes on road users for general revenue purposes.

ii) Charges should be directly transferred into the fund, outside allocations from the general budget.

iii) The infrastructure fund should be managed by a board representing infrastructure users who would simultaneously determine the level of charge and the service.

iv) The users' board should decide efficient internal allocation procedures to determine day-to-day allocation decisions.

That is, the "second generation infrastructure funds" should have the status of an autonomous agency, controlling the funding of maintenance, and possibly investment. They should be directed mainly by the users to give a strong incentive to insist on commercially and professionally efficient management.

A critical discussion of infrastructure funds continues, however. This is mainly due to the fact that the discussion of institutional issues and good fund governance is detached from a discussion of where the resources to feed the fund should come from. In many countries, in particular those where the fiscal administration is weak, the parties supporting infrastructure funds propose to fund them by fuel taxes. As fuel taxes are levied for a variety of reasons, some of which are largely unrelated to a national infrastructure policy, this would violate the principle of the second generation funds. As fuel taxes reduce CO_2 emissions, feeding infrastructure funds by fuel taxes would lead to the undesirable association of the infrastructure quality with damage done to the environment. The often high level of fuel taxes would also raise doubts about the credibility of an infrastructure fund policy supporting concerns about bureaucratic or monopoly rent seeking.

The implementation of infrastructure funds should be strictly associated with the efficient rules to endow them set out above. This is also likely to increase the acceptability of a quasi-market for infrastructure services: all revenues from charging will be returned to supply. Capacity choices are made to provide the service at minimum average costs, including the external costs for road damage and congestion. Per-unit-of-service charges will be based on the extra costs from small increases in the level of usage. With such a model, users can more easily perceive the charge as a payment for present or future infrastructure services. It would avoid the impression existing in public discussions that infrastructure charges are tax increases by another name.

5. CONCLUSIONS

The current debate on infrastructure charging is dominated by demand management considerations, i.e. the question how transport can be contained or the modal split changed to better take account of, for example, environmental or accident costs associated with transport. In this discussion, "prices" appear as a form of tax. The Round Table had the objective to take this discussion a step further and find answers to the following questions:

-- Is it possible to identify a charging scheme that would not only convey cost signals to the users of the transport system but also provide information for infrastructure managers on where to invest and which capacity to choose?

-- Would the pricing rules, discussed under the restriction to a demand management perspective, remain in place when charged with the broader function of guiding infrastructure investment?

-- Should a charging and cost recovery system entirely rely on a per-unit of service (in many cases a per-km) charge?

-- Should the charges for infrastructure services be considered a tax?

-- How can a cost recovery mechanism be designed to work to the benefit of the user? How can risks of abuse of a cost recovery mechanism like an infrastructure fund be minimised?

The Round Table arrived at the following answers:

Charges for infrastructure services can provide information on the location and scale of infrastructure investment. To do so they have to follow strict pricing rules.

The charges per unit of infrastructure service should be based on the additional costs caused by the last additional unit of service for road administration, maintenance and external costs which are directly associated with the provision of the services, i.e. vehicle operation costs that derive from a road damage externality and from congestion. The inclusion of other external costs at the margin, like environmental costs and accident costs, may lead to an under-utilisation of infrastructure facilities. Fiscal corrections of these external costs of transport have to use other instruments than per-unit-of-service infrastructure charges.

Where the external costs of road damage and congestion costs are not high, i.e. in cases of relatively low levels of usage, revenues from pricing according to the additional costs for the last unit of service will not cover full costs. In such cases a fixed charge has to complement a per-unit charge to install a mechanism to recover the full costs of infrastructure services.

A cost recovery mechanism that is separated from the budgetary process could help to correct for a disadvantageous position which transport policy sometimes seems to have in the budgetary process. It would also avoid the impression to the taxpayer and the public at large that the introduction of an infrastructure charging scheme is just another tax increase.

The organisational design of a cost recovery mechanism like an infrastructure fund has to make sure that the resources made available to the fund are those identified by the above-mentioned concepts of optimal pricing and investment. It should make sure that infrastructure users have control rights over the fund management.

NOTES

1. If there are major differences between users in vehicle types used, a distributional problem can arise. On how to respond to this problem by differentiating fixed charges, cf. Kopp (2005).

2. Depending on the provider's opportunities to price discriminate, different prices for different users could be implemented to turn consumer benefit into profits for the providing firm.

3. The fact that, without charges and without a proper accounting of external costs, the perceived demand by planners has often been too high relative to the social optimum, has certainly contributed to the restriction to demand management discussions in some contexts like dense urban areas or environmentally sensitive areas (cf. ECMT, 2003b).

REFERENCES

Brushett, S. (2005), Management and Finance of Roads in Africa, SSATP Discussion Paper No. 4, Road Management and Financing, RMF Series, World Bank, Washington.

Commission of the European Communities (2001), White Paper: European Transport Policy for 2010. Time to Decide, Brussels.

Diamond, P. (2003), *Taxation, Incomplete Markets, and Social Security,* Cambridge, Mass.

ECMT (2003), *Reforming Transport Taxes,* Paris.

Gwilliam, K. (forthcoming), The role of road funds in improving maintenance, in: OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice Round Table 135,,*OECD, Paris.

Kopp, A. (2005), Fairness, efficiency and the simultaneity of pricing and infrastructure capacity choice, *European Transport,* Vol. 31, pp. 15-27.

Newbery, D.M.G. (1988), Road damage externalities and road user charges, *Econometrica*, 56: 295-316.

Potter, B. (forthcoming), Budgeting for road maintenance, in: OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice, Round Table 135,* OECD, Paris.

Proost, S., K. Van Dender, C. Courcelle, B.L. De Borger and J. Peirson (2002), How large is the gap between present and efficient prices in Europe?, *Transport Policy*, 9: 41-57.

Prudhomme, R. and J.P. Bocarejo (2005), The London congestion charge: a tentative economic appraisal, *Transport Policy*, 12: 279-87.

Santos, G. (forthcoming), Road user charges and infrastructure, in: OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice. Round Table 135*, OECD, Paris.

Small, K. and Y. Jan (2001), The value of ”value pricing”: second-best pricing and product differentiation, *Journal of Urban Economics*, 49: 310-336.

Starrett, D.A. (1988), *Foundations of Public Economics,* Cambridge, Mass.

Verhoef, E. (forthcoming), Self-financing of infrastructure services, in: OECD/ECMT Transport Research Centre (ed.), *Transport Infrastructure Charges and Capacity Choice, Round Table 135*, OECD, Paris.

LIST OF PARTICIPANTS

Professor Werner ROTHENGATTER **Chairman**
Universität Karlsruhe
Institut für Wirtschaftspolitik und
Wirtschaftsforschung (IWW)
Postfach 69 80
D-76128 KARLSRUHE
Germany

Prof. Kenneth GWILLIAM **Rapporteur**
Economic Adviser, Transport
12720 Grand Traverse Drive
DADE CITY, Florida 33525
USA

Mr Barry POTTER **Rapporteur**
Director, Office of Budget and Planning
IMF - International Monetary Fund
700 19th St. NW
WASHINGTON DC 20431
USA

Dr. Georgina SANTOS **Rapporteur**
Departmental Lecturer
University of Oxford
University Offices
Wellington Square
GB-OXFORD OX1 2JD
United Kingdom

Prof. Dr. Erik T. VERHOEF **Rapporteur**
Free University Amsterdam
Faculty of Economics and Business Admin
Department of Spatial Economics
De Boelelaan 1105
NL-1081 HV AMSTERDAM
Pays-Bas

Mrs. Nina BARYSHNIKOVA
Head of Protocol Unit
Foreign and Economic Relations Department
Ministry of Transport and Communications
14 Peremohy Avenue
UKR-KIEV 01135
Ukraine

Professeur Andrea BOITANI
Catholic University
Department of Economics
1, Largo A. Gemelli
I-20123 MILANO
Italie

Professeur Alain BONNAFOUS
Directeur de l'ISH
Laboratoire d'Économie des Transports (LET)
ISH
14 avenue Berthelot
F-69363 LYON Cedex 07
France

Mr Francis CHEUNG
Senior Economic Adviser
AVV Transport Research Centre
Ministry of Transport, Public Works and Water Management
Boompjes 200
P.O. Box 1031
NL-3000 ROTTERDAM BA
Pays-Bas

Monsieur Alain FAYARD
Conseil Général des Ponts et Chaussées (CGPC)
Direction des Routes
Ministère de l'équipement, des transports,
de l'aménagement du territoire, du tourisme et de la mer
F-92055 LA DEFENSE Cedex 04
France

Prof. Charles Robin LINDSEY
Department of Economics
Office 9-10 Tory
University of Alberta
EDMONTON, Alberta
Canada T6G 2H1

Ass. Prof. Edoardo MARCUCCI
Professore Associato in Economia Applicata
Istituto di Scienze Economiche
Università degli Studi di Urbino "Carlo Bo"
Via Saffi 22, 61029 URBINO (PU)
Italie

Monsieur David MEUNIER
Département d'Économie et des Sciences
Ecole Nationale des Ponts et Chaussées
28 rue des Saints-Pères
F-75007 PARIS

Mrs Eva MOLNAR
Expert
Muhar u. 50
H-1028 BUDAPEST
Hongrie

Mrs. Kateryna MYHAYLICHENKO
Deputy Director,
Department for Financial Regulation and Social and Economic Policy
Ministry of Transport and Communications
14 Peremohy Avenue
UKR-KIEV 01135
Ukraine

Mr. John NELLTHORP
University of Leeds
Institute for Transport Studies
36 University Road
GB-LEEDS, LS2 9JT
Royaume-Uni

Dr. Esko NISKANEN
Strafica Ltd and STA Research
Temppelikatu 7A9
00100 Helsinki
Finland

Professeur André de PALMA
Université de Cergy-Pontoise
Département d'Économie
33 Boulevard du Port
F-95011 CERGY-PONTOISE
France

Prof. Stef PROOST
Katholieke Universiteit Leuven
Faculty of Economics
Centre for Economic Studies
Naamse Straat 69
B-3000 LEUVEN
Belgique

Professeur Rémy PRUD'HOMME
Université Paris XII Val de Marne
6, rue des Haudriettes
F-75003 PARIS
France

M. le Professeur Émile QUINET
Chef du Département
École Nationale des Ponts et Chaussées
Département d'Économie et des Sciences
28 rue des Saints-Pères
F-75007 PARIS

Mrs Catharina SIKOW-MAGNY
National Expert
European Commission
Directorate General for Energy and Transport
rue de la Loi 200
B-1049 BRUXELLES
Belgique

Mr Mateu TURRO
Associate Director
Banque Européenne d'Investissement
100 Bld Konrad Adenauer
L-2950 LUXEMBOURG
Luxembourg

Prof. Roger VICKERMAN
University of Kent
Dept. of Economics
Keynes College
GB-CANTERBURY, CT2 7NP
Royaume-Uni

ECMT SECRETARIAT

Mr. Jack SHORT
Secretary General

JOINT TRANSPORT RESEARCH CENTRE

Dr. Andreas KOPP
Chief Economist

Dr. Michel VIOLLAND
Administrator

Mrs. Julie PAILLIEZ
Assistant

Mlle Françoise ROULLET
Assistant

Mr. Colin STACEY
Administrator

TRANSPORT POLICY DIVISION

Mr. Stephen PERKINS
Principal Administrator

ALSO AVAILABLE

Vandalism, Terrorism and Security in Urban Passenger Transport. Series ECMT – Round Table 123 (2003)
(75 2003 07 1 P) ISBN 92-821-0301-3

Transport and Spatial Policies: The Role of Regulatory and Fiscal Incentives. Series ECMT – Round Table 124 (2004)
(75 2004 09 1 P) ISBN 92-821-2321-9

European Integration of Rail Freight Transport. Series ECMT – Round Table 125 (2004)
(75 2004 06 1 P) ISBN 92-821-1319-1

Airports as Multimodal Interchange Nodes. Series ECMT – Round Table 126 (2005)
(75 2005 03 1 P) ISBN 92-821-0339-0

16th International Symposium on Theory and Practice in Transport Economics. 50 Years of Transport Research (2005)
(75 2005 05 1 P) ISBN 92-821-2333-2

Time and Transport. Series ECMT – Round Table 127 (2005)
(75 2006 01 1 P) ISBN 92-821-2330-8

National Systems of Transport Infrastructure Planning. Series ECMT – Round Table 128 (2005)
(75 2005 10 1 P) ISBN 92-821-2341-3

Transport Services: The Limits of (De)Regulation. Series ECMT – Round Table 129 (2006)
(75 2006 02 1 P) ISBN 92-821-2345-6

Transport and International Trade. Series ECMT – Round Table 130 (2006)
(75 2006 13 1 P) ISBN 92-821-1338-8

Transport and Decentralisation. Series ECMT – Round Table 131 (2006)
(75 2006 12 1 P) ISBN 92-821-1342-6

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 ECMT: www.cemt.org

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(74 2007 01 1 P) ISBN 978-92-821-0108-7 – No. 55531 2007

135

TRANSPORT INFRASTRUCTURE CHARGES AND CAPACITY CHOICE

SELF-FINANCING ROAD MAINTENANCE AND CONSTRUCTION

Much of the current policy debate has considered infrastructure charges as a form of "fiscal" instrument for managing transport demand. The Round Table analysed the opportunities for setting infrastructure service prices so that they also provide guidance for the supply of services. It discussed the possibilities of increasing the finances available for transport infrastructure investment and maintenance by introducing a quasi-market for transport infrastructure services. The general public impression is that charges are a just another tax increase in disguise. Infrastructure charging that uses a mix of "fees for service" and capacity expansion criteria is expected to correct this impression, and thus to improve the political acceptability of infrastructure pricing.

The background papers were prepared and presented by Georgina Santos (Oxford University), Erik Verhoef (Free University of Amsterdam), Barry Potter (Director, Office of Budget and Planning, International Monetary Fund) and Kenneth Gwilliam (formerly World Bank, now advisor to the Department of Transportation, Florida).

VISIT THE ECMT WEB SITE
www.cemt.org
LA CEMT SUR INTERNET

(74 2007 01 1 P1) ISBN 978-92-821-0108-7

9 789282 101087