IMF WORKING PAPER

WP/87/5

INTERNATIONAL MONETARY FUND

Fiscal Affairs Department

Road User Charges and the Taxation of Road Transport

Prepared by David M. Newbery 1/

Authorized for Distribution by Vito Tanzi

February 11, 1987

Abstract

The paper discusses the measurement of road use costs and the design of a system of road user charges. The first theorem states that road damage externalities are zero, so that road damage costs are equal to the attributable fraction of maintenance expenditures. The second states that with constant returns and optimal capacity, congestion charges recover total overhead costs. Freight vehicles should pay their road use costs, but additional pure taxes on passenger transport should be guided by the principles of indirect taxation. Road use costs will typically fall short of road expenditures, but the additional pure taxation will more than cover the shortfall.

1/ The author is a Fellow of Churchill College, Cambridge. The paper was written when he was a visiting scholar in the Fiscal Affairs Department of the IMF. Any views expressed represent the opinions of the author.

Contents

Summary

Road taxes are an important source of revenue, but their relationship to road expenditures and their relative weight on different instruments and vehicles varies widely across countries. Recent empirical work has provided new evidence on the sources and levels of road use cost, and this work has stimulated new research on the design of road user charges. The paper reports these findings and asks what guidance economic theory can provide for the design of a system of road user charges, and whether additional taxation of road users is warranted.

Standard arguments suggest that freight transport should pay a road user charge equal to the marginal social cost of using the highway, with no pure tax element added for revenue purposes, while final consumers of transport, especially private automobile owners, might pay these additional charges. The main road use costs arise from road damage and congestion. Road damage costs comprise the extra maintenance costs borne by the highway authority and the road damage externality, that is, the increase in cost of operating subsequent vehicles on the damaged pavement.

Two theorems illuminate the measurement of these costs. The first is that if roads are restored when their condition reaches a predetermined level, then the average damage externality for roads of varying ages is zero. Consequently, the road damage cost is equal to the fraction of the maintenance expenditures attributable to traffic (as opposed to weather). The second theorem is that if there are constant returns to highway expansion and roads are optimally adjusted to traffic levels, then the efficient congestion charge will recover the highway overhead costs, including interest on infrastructure expenditures. With economies of scale, only a fraction of these overheads will be so recovered.

In order to test the relevance of the second result, congestion costs must be estimated. This study reviews the theory and empirical evidence on these costs. Urban congestion costs are overwhelmingly important, and are the most difficult to measure and charge for. The feasibility of designing electronic pricing of roads has been demonstrated for Hong Kong, but unless the revenues collected replace other road taxes, the transfer of income would exceed the distortion costs by a large factor, and would undoubtedly be resisted. If electronic pricing cannot be used, the instruments available to levy road user charges are very blunt, and the study reviews these, but it is not difficult to design a set of charges which implements the best feasible system. Without an effective system of income taxes, taxes on automobile ownerships might be set at high rates. If these tax revenues are included, road taxes will more than cover road expenditures, though efficient road user charges alone are likely to fall short of road expenditures.

I. Introduction

The industrialized market economies appear to collect about 6 percent of total government revenue from road taxes, or just over 2 percent of GNP (Table 1). 1/ World Bank data suggest that road taxes are even more important as a share of government revenue in developing countries, and the unweighted average of a sample of 24 countries gives the share as 11 percent (with a standard deviation of 3.6 percent). 2/ Road taxes are thus a very important component of government revenue, and in developing countries appear to be as important as taxes on individual incomes (Tanzi (1983), Table 4).

Average taxes on motor fuel appear to be responsible for just over half the total road tax revenue in the developed countries (with a range from 28 percent to 80 percent) and also for about half in the sample of developing countries (subject to all the provisos of footnotes 2 and 3). The obvious explanation for the importance of road taxes is that road expenditures are also important, and are most logically financed by taxes on road users. The main part of this paper is concerned with just this question.

Table 1 suggests that there is considerable divergence across countries in the extent to which road taxes recover road expenditures. The industrialized countries appear to fall into three main groups. The first group of countries (Netherlands, Great Britain, New Zealand, Sweden, Denmark, and possibly Germany) raise substantially more in road tax revenue than they spend on roads. The second group (Australia, Switzerland) spend about as much as they raise, and follow the pattern of operating a Road Fund, in which road expenditures are required to be financed by road taxes. The third group (Austria, Japan, the United States) collect significantly less in revenue than they spend on roads,

---

1/ These numbers should be treated with considerable suspicion, as they are compiled from local sources following a variety of different conventions. Road taxes are often underestimated as they exclude import duties on vehicles, but fuel and vehicle tax revenue may be overestimated if it includes value added tax (VAT). The source is International Road Federation (1985), which gives (some of) the necessary qualifications on the data.

2/ *The figures come from the Transport Department of the World Bank,* and are compiled from World Bank appraisal reports and the Fund's Government Finance Statistics Yearbook, mostly for single years in the period 1979-84. Many of the figures are incomplete, and in many countries the nominal taxes on fuel (which account for over half the revenue) often overstate the true taxes, as national ex-refinery prices have often been substantially below c.i.f. prices in many of these countries. At best, the figures can be taken as indicators of the intentions of governments which were often in conflict with the short-term goals during the period of the oil crisis.

Table 1. The Importance of Road Taxes in Selected Industrialized Countries, 1982-84

(Road taxes as percentage of:) 1/

Country	Road Expenditures 2/	Total Government Revenue	GNP
Netherlands	434	6.2	2.5
Great Britain	335	8.7	3.4
New Zealand	235	8.6	2.8
Sweden	230	6.3	1.9
Denmark	214	4.9	2.6
Germany	148	4.6	1.8
Australia	113	7.0	1.9
Switzerland	107	7.8	2.1
Austria	80	7.2	1.9
Japan	80	3.4	1.7
United States	63 3/	4.4	1.4
Unweighted average	185	6.3	2.2
Unweighted SD	113	1.7	0.6
Unweighted CV percent	61	27	25

Source: International Road Federation, World Road Statistics, 1985.

1/ Averages of percentages for years 1982-84.
2/ All figures of variable coverage and doubtful comparability.
3/ Source: U.S. Congress (1985).

Notes: SD: Standard deviation
CR: Coefficient of variation = SD/mean.

either by design, or because they have been unsuccessful in raising revenues in line with expenditures, despite their intention to operate a Road Fund in which expenditures are financed by revenues. The United States is a good example of this, for although it operates a Road Trust Fund, the fraction of disbursements covered by charges on road users fell fairly steadily from over 80 percent in 1960 to under 60 percent in 1982, though it has since recovered somewhat. The real value of highway receipts fell from over 6 ¢/vehicle mile in 1960 (at 1982 prices) to under 2 1/2 ¢/mile in 1980 (U.S. Congress (1985)). Nevertheless, the simple average of the whole sample of industrial countries suggests that road taxes are nearly twice as high as expenditures, and that highway users make substantial net contributions to government revenue.

The data for developing countries is very much weaker, but it also shows a considerable diversity across countries, with 18 countries collecting more revenue from road users (often substantially more) than they spend on roads, whilst four appear to spend less. On balance, highway users also benefit the exchequer in developing countries.

If there is considerable variation across countries in the level of road taxes relative to road expenditures, there is even greater variety in the structure of road taxes, defined as the relative level of taxes on different classes of road users. Table 2 shows that at one extreme Japan and Denmark charge only three to four times as much a year for a 16-ton truck as for a medium car, whilst Great Britain, Germany, Switzerland, and New Zealand all charge more than ten times as much. Heavier trucks are even more heavily charged in some countries, and the coefficient of variation in the level of taxes on heavy trucks at 105 percent is more than twice as high as for medium trucks and for private cars. On the other hand, most countries tax large private cars more than twice as much as small cars, and the coefficient of variation of this ratio is very small.

Comparable details for developing countries are harder to obtain, but certainly in some countries there is a substantial variation in taxes on different vehicles. Often gasoline is heavily taxed whilst diesel fuel is subsidized. Import and license fees for private cars are often very high, and for commercial vehicles much lower (although license fees for heavy trucks can be very high). The tax element in the price of diesel varies dramatically between countries and across time as the various forces of world oil prices, domestic inflation, pegged exchange rates, and domestic political pressures alternate in their relative importance and effect. Certainly the period since the 1973 oil price rise has been a turbulent one as far as fuel pricing is concerned. Tunisia, for example, faced with various objections to raising domestic diesel prices, has been able to recover the lost revenue by a set of heavy license fees which has substantially altered the balance of the system of taxing vehicles. Nor have the developed

Table 2. Structure of Road Taxes in Industrial Countries, 1984

Country 1/	Average Annual Taxation on: Medium Automobile	16-Ton Truck	32-Ton Truck	Large Auto / Small Auto	16-Ton Truck / Med. Auto	32-Ton Truck / Med. Auto
		(In SDRs)		(As a ratio of annual taxes on automobiles)		
Netherlands	721	4,167	7,902	2.7	5.8	11.0
Great Britain	533	5,306	11,561	1.9	10.0	21.6
New Zealand	329	4,981	27,389	1.5	15.1	83.2
Sweden 2/	399	3,968	6,818	2.5	7.4	17.1
Denmark	666	1,993	2,736	2.2	3.0	4.1
Germany	351	4,378	8,755	2.6	12.5	24.9
Australia	1,364	6,308	8,438	2.0	4.6	6.2
Switzerland	462	5,582	illegal	2.4	12.0	
Austria	1,137	NA	NA	11.4		
Japan	610	5,341	illegal	6.1	3.8	
Unweighted average 3/	657	4,223	10,511	2.3	8.2	24.0
SD	325	1,418	7,311	0.4	4.0	25.2
CV percent	49	34	70	18	49	105

Source: International Road Federation, World Road Statistics, 1985. Vehicle categories used are as follows:

small car 1,000 cc 15,000 km pa 1,200 litre gasoline;
medium car 1,500 cc 15,000 km pa 1,500 litre gasoline;
large car 4,500cc 15,000 km pa 2,700 litre gasoline;
16-ton truck laden weight, 75 percent capacity, 50,000 km, 40 litre diesel/100 km;
32-ton truck laden weight, 75 percent capacity, 80,000 km, 50 litre/100 km.

1/ In order of ratio of road taxes to road expenditure.
2/ Figures for 1982.
3/ Averages of ratios, not ratios of average.

countries been immune to dramatic variations in the real tax rates on gasoline over this period, as Tait and Morgan (1980) document for selected industrial countries.

The issue of the appropriate structure of road taxes is highly topical in a number of countries. The United States Federal Highway Authority published its Highway Cost Allocation Study in 1982, and since then several states have commissioned similar studies and reformed their systems of road taxes. New Zealand recently introduced a very sophisticated system of distance-weight charges (Starkie (1984)) and several states in the United States have introduced ton-mileage taxes. Singapore has had area licensing system in operation since 1979 (World Bank (1986)), and Hong Kong has recently successfully completed experimental trials of a system of electronic road pricing (Dawson and Catling (1986)). Thus, not only are road taxes quantitatively important for government revenues, but there are wide differences in the way they are collected, and in the structure of taxes across vehicle types. There is, therefore, considerable interest in first identifying the right structure of taxes, and then choosing the most appropriate instruments for levying these taxes. So far, however, most of the cost allocation studies have been undertaken by highway engineers, and have produced technical rather than economically efficient solutions, although there is a long and distinguished history of economic interest in road pricing and road user charges. It therefore seems timely to review the present state of theory and see what guidance economics can offer on road taxation.

1. Terminology

Vehicles impose a variety of costs--on other road users, on the highway authority, and on the rest of society. These costs may collectively be described as the road use cost--the social costs (excluding the private costs) arising from vehicles using the road system. Road user charges are levied on road users and treat the supply of road space as a publicly supplied service to be charged for, just as electricity is charged for by tariffs. The pure taxation of road users is the intentional excess of road taxes over the appropriate road user charge that is set to raise additional revenue for the exchequer. The pure tax element is the specific amount that transport or road use is taxed, and therefore excludes general taxes, which include most notably value-added taxes which apply to all (or most) goods and services. To take a concrete example, gasoline and motor diesel are subject both to excise and value-added taxes (VAT) in the United Kingdom, but the tax on fuel use is just the excise tax (commercial operations pay the excise tax but are rebated the VAT). In countries where the national ex-refinery price is controlled by the government, there may be an additional implicit tax or subsidy equal to the amount by which this ex-refinery price exceeds or falls short of the c.i.f. price. Similarly, government-controlled marketing margins may contain an element of tax or subsidy. Finally, some fraction of the tax on fuel will be the road user charge, with the balance being the pure fuel tax element.

The exact division between the road user charge and the pure tax element has a degree of arbitrariness, for two reasons. The first is reasonably simple and causes no conceptual problems. The efficient or economic road user charge is equal to the road use cost, just as the efficient price of electricity is the short-run marginal social cost of production. The actual road user charge may differ from this just as the pre-tax sales price of electricity may differ from the short-run marginal cost, for a variety of reasons mainly to do with balancing the budget or earning the required average rate of return on assets. It is often conceptually useful to work in terms of the efficient road user charge and treat the entire excess of the road tax as the pure tax element, but conventions vary among authors. The second problem is conceptually troubling at the level of individual elements of road taxation and arises because the exact allocation of the total (efficient) road user charge over different instruments (whether fuel taxes, tire taxes, vehicle purchase taxes, or license fees) is considerably arbitrary. It follows that one can only usefully discuss the pure tax element at the level of the vehicle, not at the level of the separate road taxes. Again, this is not a particularly serious problem once its nature is appreciated.

The main road use costs are congestion, pollution, accidents, and road damage. Pollution costs are not peculiar to road users and their calculation raises no conceptually new issues. The U.S. Highway Cost Allocation Study suggests that they will be modest, ranging from 4-6 percent of total road use costs for trucks in urban areas, to figures as high as 15 percent for suburban cars and pickups (whose other road use costs are smaller). The worst offenders may be responsible for costs of up to 12 ¢/mile, with urban cars and pickups less than 2 ¢/mile (U.S. Federal Highway Authority (1982), Table 12, Appendix E). They will therefore be ignored in this survey. Accident costs raise more interesting problems, for although single-person accidents are, in principle, internalized in insurance premia, multiperson accidents create externalities. Correctly charging for them would therefore recover more revenue than the cost of the damage done (Vickerey (1969)). However, to quote the Highway Cost Allocation Study: "Quantitative estimation of accident cost and vehicle volume relationships, however, has not yet proved to be satisfactory .... Attempting to combine these various effects into marginal cost figures leads to results that are small in magnitude and not especially plausible, so no tabulations have been incorporated into the user charge estimates" (U.S. Federal Highway Authority ((1982), p. E-37). Accident costs are therefore also ignored.

Congestion costs are the classic form of externality created by road users and study of these costs goes back to the early days of motorized road transport with the early dispute between Pigou (1912) and Knight (1924). They arise because additions to the traffic flow reduce the speed and increase the trip costs of other road users.

It was thought until fairly recently that congestion (together with road accidents) was the only interactive externality affecting road users. It was known that vehicle use damaged the road surface, and either hastened the date at which repairs or replacement were needed or raised the cost of these repairs. These pavement or road maintenance costs were borne by the highway authority and much research was devoted to determining the damaging power of different vehicles in order to allocate these road damage costs. The most systematic of these attempts, reported in Highway Research Board (1962), involved the construction of 169 test sections of pavement over which variously laden vehicles were driven millions of miles to determine the rate at which pavements were damaged by vehicle passage--an experiment which it is estimated would cost over $300 million at 1980 prices to replicate. 1/ The major finding of these tests was that the damaging power of a vehicle increased as approximately the fourth power of its axle load. The damaging power of a vehicle is thus measured by the number of Equivalent Standard Axle Loads (ESALs) where 1 ESAL has the damaging power of an 18,000 lb. (8.2 ton) single axle.

Subsequent research, most notably that reported in Paterson (1986), produced new evidence on the consequences of road damage, which suggest that vehicles, in damaging the road surface, create a new type of externality which is qualitatively quite different from congestion. The most convenient way to measure the damage done to the road by vehicle passage is by the increase in roughness (measured in International Roughness Index (IRI) units of m/km), which can be related to the U.S. system of measuring pavement quality in terms of its Present Serviceability Index (PSI) (Sayers, Gillespie, and Paterson (1986)). Since vehicle operating costs increase with roughness, the passage of an extra ESAL over a road has two effects--it advances the date at which maintenance will be required and hence raises the costs borne by the highway authority (the traditional pavement costs), and it raises the vehicle operating costs of subsequent vehicles, thus creating a road damage externality for subsequent road users. Over the life of a pavement (between major repairs such as overlays or reconstruction) vehicle operating costs are between 10 and 100 times as large as maintenance costs, and so potentially these road damage externalities to users are of the first importance. Rough calculations suggest that the

---

1/ Given the huge cost of the experiment and the enormous significance of the results, it is somewhat disturbing that the experimental data were analyzed in what appears to have been a statistically rather casual manner. Small and Winston (1986) report substantial differences to key parameters such as the returns to scale of strengthening roads when re-estimating the equations on the original data using best-practice techniques.

average increase in vehicle operating costs caused by roughness, though much less that the total vehicle operating cost, is comparable in importance to the pavement cost. 1/

II. Principles of Taxation Relevant to Road Taxation

The modern theory of public finance provides a powerful organizing principle for taxing and pricing. Under certain assumptions, policies should be designed to achieve production efficiency, with all distortionary taxes falling on final consumers. The conditions for this result, set out formally in Diamond and Mirrlees (1971), are roughly that production efficiency is feasible, and that any resulting private sector profits are either negligible or can be taxed away. The feasibility condition would be satisfied if the economy were competitive and externalities could be corrected or internalized. Activities which would run at a low if they set prices equal to marginal costs would not be viable as competitive private enterprises, but if they were located in the public sector, their overhead costs, or the difference between average and marginal costs, should be recovered from general tax revenue. The implication of this result is that there should be no pure taxation (as defined earlier) of intermediate goods such as commercial or freight transport, which in turn means that taxes on freight transport should be set equal to road use costs. Put another way, road user charges on freight vehicles should, if possible, be equal to road use costs (which are in turn equal to the marginal social cost of road use less the private cost borne by the road user). There is, however, no presumption that taxes on the final consumption of road services, that is, on passenger transport, should be equal to road use costs, and in general, unless the system of income taxes is more extensive and comprehensive than is typically the case, there is a strong presumption that additional taxes over and above the road use costs will be desirable. It is therefore logical to distinguish between the taxation of freight transport and passenger transport and to concentrate initially on setting road user charges for freight transport where efficiency, rather than social justice, is the appropriate objective (social justice being pursued by other, more suitable, means). The issue of whether the highway budget will or should contribute to general revenue will depend on the extent to which efficient road user charges cover total expenditures (including interest on the capital cost of the road network), which will depend on the importance of congestion and the extent of returns to scale in capacity expansion (discussed below in Section II.3) and the extent to which the pure taxation of passenger transport is a suitable method of redistributing income.

1/ On flexible rural interstate highways at strengths and traffic flows typical for the United States, the average increase in the vehicle operating cost over a repair cycle fixed at 15 years would be just over half the pavement cost, using figures presented in Newbery (1986b).

1. Road user charges and road use costs

If vehicles could be charged for the use of specific roads at different times of the day, then the efficient road user charge could be the road use cost. Provided such charges were visible, road users would then be encouraged to make socially efficient choices--whether to make the trip, and if so, by what means and when (these involve choices of transport mode, type of vehicle, time of day, etc.) and, in the long run, where to locate, what activities to undertake, and so on. In short, the road use charge would then have the same allocative effects as competitive or efficient prices, and might also serve the same function as a price in revealing to the highway authorities where to expand road capacity.

This suggests a logical sequence for the study of road user charges and road taxes. The first step is to identify the road use costs, the second is to see what methods are available for levying road user charges and how finely these can be adjusted to these costs. Since in many cases the only feasible methods are relatively crude, one can then ask how best to reflect road use costs via the available, imperfect, instruments. The next step is to examine how far these instruments have repercussions outside the transport sector, and, if so, how to take these repercussions into account in modifying the design of road user charges. The same sequence of steps is needed for setting road user charges on passenger vehicles, with the additional step of setting the pure tax rate on the final consumption of transport services.

2. Road damage costs

The major cost of maintaining roads is the cost of a periodic overlay (or, in extreme cases, of reconstruction), which may be required at intervals of 10-25 years, and which reduces the roughness of the road to an acceptable initial value (or raises the PSI).

Here the main theoretical result is due to Newbery (1986a), who provides the surprising result that if the road network has a uniform age distribution, and if maintenance policies are condition-responsive (that is, the road is overlaid or restored when its roughness reaches a predetermined value, or "minimum tolerable standard" in the terminology of U.S. highway engineers), then, averaging over roads of different ages, road damage externalities are identically zero in an important special case, and negligible in all reasonable cases. The special case has zero traffic growth and all road damage caused by vehicles. The general case allows for the effect of weather and time on the state of the road, as well as for growth in traffic volumes. In the special case, what might be called the fundamental theorem of road damage costs states that the efficient charge is the road damage cost exactly equal to the average cost of maintenance per ESAL km.

The argument goes as follows. The state of the road is measured by its roughness, R, and vehicle operating costs increase with R. In the simple case, R is a function of cumulative ESALs since last overlay, and the road will be overlaid when roughness reaches a predetermined level, $\bar{R}$, after which its roughness will fall to the initial value, R_0. The "age" of the pavement can be measured by cumulative ESALs. Imagine a road between two points, but of uniform age distribution. If its average lifetime is N ESALs before overlay, then a fraction m/N km will have an age of m or less, as shown in Figure 1. Initially, suppose that the youngest pavement is at the start of the road, and the oldest pavement, just requiring overlay, is at the end. Each year, if annual traffic is in ESALs, a fraction of n/N will be overlaid at a cost of C/km, or a total annual cost of Cn/N per km, or C/N per ESAL km. As time passes, the "age" of pavement at each distance will change as shown in Figure 1, but the "age" distribution (the portion of road of any age since overlay) will remain <u>unchanged</u>. Variations in the annual flow will alter the rate at which the "age" of a particular piece of road changes, but not the distribution. The cost of traversing the road will depend on the average roughness, which will depend on the age distribution of the road, but this will also be unaffected by traffic. Thus there is no damage externality, and the social cost of an extra vehicle is just C/N, the extra maintenance cost required. The marginal social cost of an extra vehicle will be equal to the average cost borne by the highway authority. 1/ It is important to realize that this result does not require an optimally set maintenance policy, only a consistent policy, in which (R_0, $\bar{R}$) are predetermined and consistently applied.

Another way to understand this surprising result is to examine the time path of vehicle operating costs (and roughness) shown in Figure 2. The effect of an extra ESAL is now to raise subsequent operating costs by the vertically shaded amount, to advance the date at which roughness reaches the critical level $\bar{R}$, and overlay occurs, and to lower subsequent vehicle operating costs as a result, by the amount of the horizontally shaded area. Averaging over roads of all ages, these two areas exactly balance in present discounted value.

The effect of weathering is to reduce the proportion of road damage attributable to traffic to between one half and three quarters of the whole for paved roads, depending on the life (in years) of the surface, the severity of the climate, and the stringency of the maintenance policy (in terms of R_0, $\bar{R}$). (The derivation of the exact formula is given in Newbery (1986a)). The effect of traffic growth is twofold. First, the road will require strengthening to withstand higher traffic

---

1/ Strictly speaking, the cost C should include the extra congestion costs caused by the disruption of maintaining the road. Whether this will cause the highway budget to make a surplus is discussed in Section II.3 below.

FIGURE 1.

AGE DISTRIBUTION OF ROAD

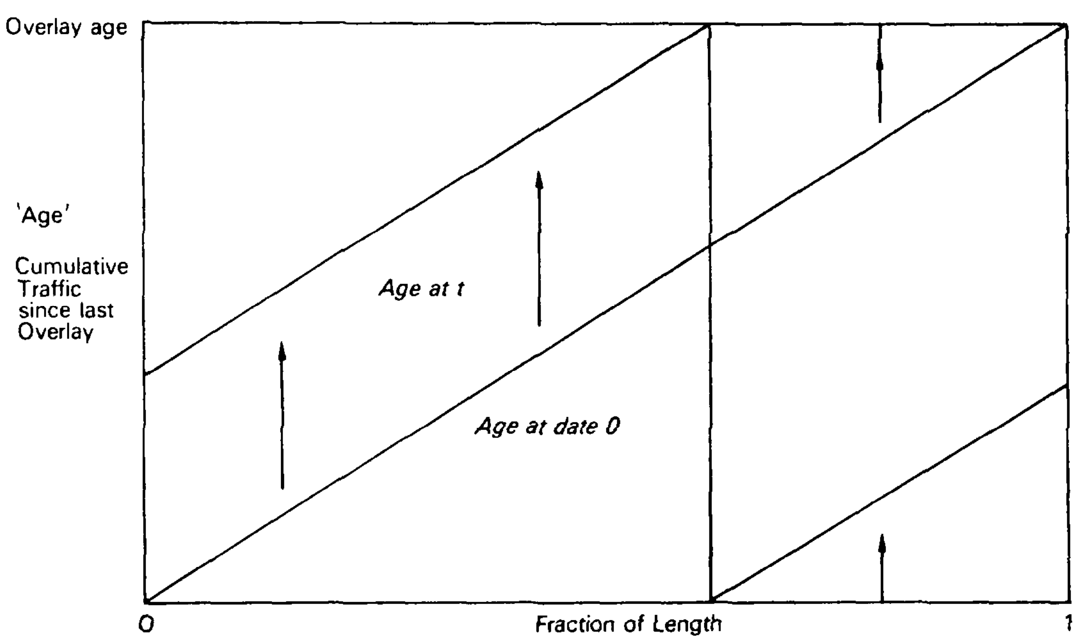
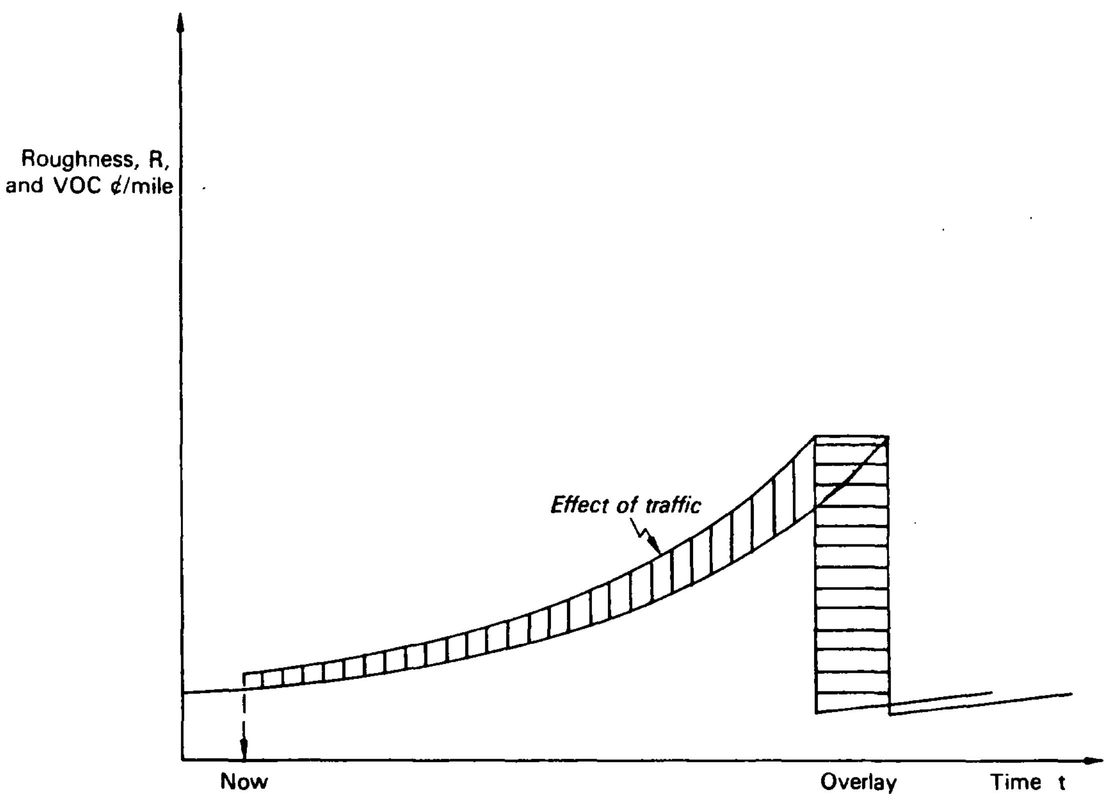


FIGURE 2.

TIME PATH OF VEHICLE OPERATING COSTS



volumes, and so, part of the maintenance is an expansion, or investment cost, more correctly attributable to future traffic levels. Second, the rest of the current maintenance cost is the result of past, different (lower) traffic volumes. Hence, there is a problem of relating the timing and fraction of maintenance expenditure to the damage done by current traffic. It is not difficult to calculate the appropriate level of road damage costs and to relate them to current maintenance levels, but there is no longer the simple formula of taking the correct level of maintenance cost and dividing it by lifetime ESALs. Calculations in Tunisia suggested that the combined effect of (high) traffic growth and weathering was that economic road user charges on lightly trafficked roads lasting 20 years would recover 55 percent of normal maintenance costs, and those for more heavily trafficked roads lasting 15 years would recover 65-75 percent. Allowing for the greater fraction of vehicle kilometers travelled on more heavily trafficked roads, an average figure of two thirds of maintenance costs recovered from road use charges seems reasonable (Newbery (1986c)).

The same arguments apply to paved and unpaved roads, but whereas it is normal for highway authorities to follow a condition-responsive maintenance strategy for paved roads, it is not uncommon for unpaved roads to be bladed or graded at predetermined intervals (in Tunisia, for instance, they are graded once a year before the harvest). In this case the earlier result that the road damage cost is some stable fraction of maintenance cost no longer applies, and road damage externalities may be appreciable. If, however, the period between maintenance is optimally chosen, then the earlier fundamental result reappears, this time as a consequence of an envelope theorem (Newbery (1986a)).

A special case of a noncondition-responsive maintenance policy of practical importance arises when paved roads have been allowed to deteriorate to the point of crisis, such that repair is urgently required. In this case extra traffic will not affect the date of overlay, and again the fundamental theorem does not apply. Sample calculations for Pakistan suggest that the road damage costs may be even greater in such cases as extra vehicles increase the costs of repair when the road is finally repaired (Newbery (1986e)).

The insights offered by the fundamental theorem are of considerable practical value, as they are robust to the exact form of the relationship between road damage, roughness, and ESALs and between vehicle operating costs and roughness--both of which are econometrically difficult to estimate with precision. They therefore allow a quick calculation of road damage costs given data on road and vehicle repair costs, traffic flows, and maintenance intervals. Thus, Newbery (1986b) was able to suggest that the estimates of road damage costs on U.S. rural interstate highways, given in the cost allocation study, appeared to be too high. On closer inspection the reason appears to be that calculations were done for a representative road which was assumed to be seven years away from next overlay. Extra traffic was assumed to raise vehicle operating costs over the next seven years, but there appears to be no credit item for the fall in these costs after overlay. Thus, in Figure 2, the extra vehicle operating cost shaded vertically has been counted, but not the benefits, showed horizontally.

The approach indicated can be extended to a variety of related problems. Thus, Newbery (1986d) investigates the time path of road damage costs and road use costs more generally on roads of different ages to see how the results are modified if a significant fraction of the highway system is nearing the end of its current overlay life (a worry sometimes expressed about the U.S. interstate system, and parts of the motorway system in the United Kingdom). Three factors are relevant to deciding whether road use cost will rise or fall as the road ages. The first is that the pavement cost (the present value of the costs borne by the highway authority) will rise as the date of repair approaches. This effect will be magnified if major expansion or reconstruction costs are required (and these may raise road damage costs by a factor of four). The second factor is that although on average road damage externalities are approximately zero, they start positive and finish negative, as the costs (vertical shading in Figure 2) remaining before overlay decrease, and the benefits (horizontal shading) increase in present value terms. Thus, as the road approaches overlay, the road damage externality falls and offsets the rise in pavement costs. On heavily trafficked roads which do not require reconstruction, this effect is likely to dominate the movement in pavement costs. Finally, as traffic increases, so congestion costs will increase, and these may overwhelm the other costs.

As mentioned earlier, road damage costs are proportional to the fourth power of axle load, and hence automobiles inflict essentially negligible damage compared to trucks; the ratio of damage done by heavier trucks will be 10,000 times as great as that done by medium automobiles (because demand increases as the fourth power of the axle load). The principal source of road use costs for automobiles are congestion costs, and to these we now turn.

3. Congestion costs

It has long been recognized that road users create congestion externalities, but study of these has fallen into four quite widely separate phases. The first phase was the theoretical exploration of optimal pricing and investment rules for congested and uncongested roads, which were developed by Ellet (1840), Dupuit (1844), Pigou (1912), and Knight (1924). The second stage really began with Walters' (1961) attempt to quantify the congestion externalities and the implied optimal tolls. This had to wait for the underlying congestion relationships to be empirically estimated, mainly by traffic engineers in the first instance (see the references in Walters (1961)). The magnitude of congestion costs appeared impressively high for urban streets, and the fact that the problem had been quantified and shown to be important stimulated a great deal of subsequent research, surveyed admirably by Winston (1985).

Walters had been primarily concerned with pricing issues, and hence with the short-run marginal social cost of extra traffic on a given road system. Other writers were quick to take up the related theme of the optimal investment rule, and Mohring and Harwitz (1962) and Mohring (1970) pointed out that if road capacity demonstrated constant returns and could be continuously adjusted, optimal congestion tolls would exactly recover the costs of providing the optimal degree of capacity. This qualitative result is attractive to economists who can argue that if traffic engineers have indeed chosen the correct capacity (perhaps on average), and if there are constant returns to scale, then the (average) congestion charge should be the average cost of capacity.

This potentially useful result was derived from a model in which roads were infinitely durable (and continuously adjustable to capacity). However, as has been shown, roads deteriorate under the influence of traffic, and although congestion is an increasing function of traffic (or the ratio of passenger car equivalents (PCE) to capacity), the wear on the roads, which gives rise to the need for repair, depends on cumulative ESALs. Thus, maintenance costs depend on cumulative ESALs and the road strength, while congestion costs depend on road capacity.

Economic theory suggests charging vehicles for the road damage they cause (proportional to the number of ESAL miles per vehicle), including the externalities, which involves calculating congestion costs. This raises a number of obvious questions. If the road network is to be optimally designed, how should the costs of strengthening the pavement to withstand higher axle loads per vehicle be apportioned between congestion costs and road damage costs? The discussion of road damage costs argued that if all damage was attributable to traffic, and there was no traffic growth, then charging the road damage cost would exactly recover the maintenance costs. It is tempting to conclude that optimizing the highway capacity will in turn lead to a set of additional congestion charges which will recover the capital costs, leaving the highway budget exactly balanced. But is it correct to allocate all the capital costs to congestion charges on a PCE basis when a large part of the capital cost is required to strengthen the road to a standard suitable for heavy trucks? It is common practice to allocate the minimal expenditure needed for a road suitable for automobiles on a PCE basis, and the balance, needed for heavier vehicles, on an ESAL basis. But is this correct, especially given the increasing returns to strengthening the pavement?

Newbery (1986f) argues that maintenance costs would depend on road capacity, or road width, which would in turn be adjusted to traffic volume, measured by PCEs. If heavy vehicles were effectively confined to the outermost lane, then the strength of the road and the period between overlays would be determined by the total number of heavy vehicles on the road, whilst the capacity would be adjusted to the volume of all vehicles. If there were constant returns to providing capacity, then the optimal congestion charge per PCE would recover the capital and maintenance costs, whilst road damage charges would also recover the maintenance costs. Mohring (1986) pointed out that whether the charges would more than recover costs depended upon the sense in which there were constant returns. His argument goes as follows. Suppose there are two types of vehicles: automobiles, N_a per annum on the road, with zero road damaging power, and heavy vehicles, N_h per annum. Vehicle operating costs depend on roughness, whose average value will not change for a condition-responsive maintenance strategy and an initially uniform age distribution (as in Figure 1), and travel time, t_a and t_h, respectively. Let v_i be the value of time of the two types of vehicle, then the total annual (variable) costs of providing the N_i trips can be written:

$$C = N_a v_a t_a(N_a, N_h, \omega) + N_h v_h t_h(N_a, N_h, \omega) + K(\omega, S) + M(N_h, \omega, S) \qquad (1)$$

where ω is road capacity (the number of lanes), K is the annualized capital cost of providing capacity ω, and strength S, whilst M is the annual maintenance cost of overlaying. The marginal social cost of a trip is given by:

$$\frac{\partial C}{\partial N_a} = v_a t_a + \sum_i N_i v_i \frac{\partial t_i}{\partial N_a}, \qquad (2)$$

$$(i = a, h)$$

$$\frac{\partial C}{\partial N_h} = v_h t_h + \sum_i N_i v_i \frac{\partial t_i}{\partial N_h} + \frac{\partial M}{\partial N_h}. \qquad (3)$$

The first term in each case is the private cost of travel, whilst the remaining terms are the road use costs, to be recovered by road use charges f_a, f_h, respectively. The optimal choice of capacity, ω, satisfies:

$$\frac{\partial C}{\partial \omega} = \sum_i N_i v_i \frac{\partial t_i}{\partial \omega} + \frac{\partial K}{\partial \omega} + \frac{\partial M}{\partial \omega} = 0. \qquad (4)$$

Mohring now argues that constant return to scale in capacity, ω, must involve t_i being homogenous of degree zero in its three arguments, and

$K(\omega,S)$ being of the form $\omega K(S)$. As far as maintenance costs go, constant returns would seem to imply that M is independent of ω, on the grounds that doubling the road width will halve the number of heavy axles on any lane, and hence halve the maintenance costs per annum per lane. Rather than prejudging this issue, suppose that the fraction of distance driven by heavy vehicles on the outer lane is $g(\omega)$, so that

$$M = \omega N_h \, g(h) \, m(S). \tag{5}$$

The special case considered by Mohring would make $g(\omega) = 1/\omega$, and M would then be independent of ω. Again, the more general case for the time cost functions would be to define road capacity as $h(\omega)$, and let t_i be homogenous of degree zero in N_a, N_n, and h. This would imply by Euler's theorem that:

$$N_a \frac{\partial t_i}{\partial N_a} + N_h \frac{\partial t_i}{\partial N_h} + h(\omega) \frac{\partial t_i}{\partial h} = 0 \qquad (i = a,h). \tag{6}$$

Total revenue from levying the road user charges f_a, f_h, will yield

$$R = \sum N_i f_i = N_a \sum_i N_i v_i \frac{\partial t_i}{\partial N_a} + N_h \sum_i N_i v_i \frac{\partial t_i}{\partial N_h} + N_h \frac{\partial M}{\partial N_h}$$

$$= \sum_i N_i v_i \left\{ N_a \frac{\partial t_i}{\partial N_a} + N_h \frac{\partial t_i}{\partial N_h} \right\} + N_h m(S) g(\omega) \omega \tag{7}$$

from (5). Substitute from (6) (noting that $\partial t_i/\partial\omega = h' \partial t_i/\partial h$) and (5) to give

$$R = - \frac{h(\omega)}{h'(\omega)} \sum N_i v_i \frac{\partial t_i}{\partial \omega} + M$$

and substitute from (4) to give

$$R = \frac{h}{\omega h'} \left(\omega \frac{\partial K}{\partial \omega} + \omega \frac{\partial M}{\partial \omega} \right) + M. \tag{8}$$

The last term in equation (8) confirms the fundamental theorem that the road damage costs are equal to the maintenance costs for the special case considered here. If $K = K_o + \omega K(S)$, where $K_o = 0$ for the case of constant returns, then from equation (5)

$$R = \frac{h}{h'\omega}\left\{K - K_o + \left(1 + \frac{\omega g'}{g}\right) M\right\} + M. \qquad (9)$$

In the case of pure constant returns, $h = \omega$, $K_o = 0$, $\omega g'/g = -1$, and

$$R = K + M \qquad (10)$$

or revenues exactly cover total cost. If, on the other hand, heavy vehicles are effectively confined to the outer lane, so that $g' = 0$, but otherwise there are no fixed costs, K_o, and capacity is proportional to width

$$R = K + 2M. \qquad (11)$$

(This case corresponds well to expansions above six lanes and fairly well to four-lane highways.) If, on the other hand, there are economies of scale in expanding capacity, then only a fraction of variable capacity costs, $K - K_o$, will be recovered, together with surplus over maintenance costs equal to the same fraction of the allocable share $(1+\omega g'/g)$.

The relative importance of capital and maintenance costs on flexible pavements can be found for a road of optimal strength by choosing S to minimize C:

$$\frac{\partial C}{\partial S} = \frac{\partial K}{\partial S} + \frac{\partial M}{\partial S} = 0, \qquad (12)$$

and noting that whilst $\partial K/\partial S$ is roughly constant, $m(S) = aS^{-\gamma}$, where $\gamma = 6.65$ (Paterson (1986)). If there are no fixed capacity costs, then from (12), $M = K/\gamma$, or maintenance costs will be about 15 percent of capital costs. Fixed costs will reduce this fraction. Finally, attributing some road damage to weather and time will further reduce the fraction of maintenance costs recoverable from heavy vehicles.

What light does this throw on the questions of cost allocation? First, the fraction of maintenance costs attributable to traffic damage (typically one half to two thirds) should be allocated to heavy vehicles in proportion to their damaging power, or number of ESALs. On wider roads these costs will be higher, and to the extent that heavy vehicles concentrate in the outer lane, these costs will increase with road width. Second, the congestion charge, to be allocated to PCEs, will, on roads of optimal capacity, recover a fraction of variable capacity costs ($K-K_0$) (or marginal expansion costs), together with the same fraction of the "congestive fraction" of maintenance costs, $1+\omega g'/g$. The fraction will be smaller, the greater the extent of increasing returns of capacity to road width, but it is unlikely to be less than one half and will be effectively more on wide roads (more than four lanes). The congestive fraction might be two thirds on two-lane to six-lane roads. As a practical matter, whereas the allocable fraction of maintenance costs can be charged to heavy vehicles regardless of the optimality of the maintenance or design policies, congestion costs will only cover their portion of the costs once traffic has reached design levels. The effect of returns to scale in road strengthening only affect charges by affecting the level of capital and maintenance costs, and the relationship of maintenance to capital costs. Thus the argument that the excess of construction costs above those needed to build a road suitable for automobiles should be allocated to heavy vehicles has no logical foundation (though the various envelope results may well make the optimal charges appear to be so determined).

How useful are these theoretical results for the practical question of determining congestion costs? On the face of it they appear very useful, as they require technical data (road expansion costs, extent of returns to scale, and so on) rather than more difficult-to-obtain observations on traffic flows on different roads. There are, however, several problems with this approach, though it remains a useful guide for rough orders of magnitude. The first problem is that highways cannot be smoothly adjusted to traffic (though steady small improvements can be made, and over the whole network the indivisibilities may not be too serious). The second is that there appear to be substantial economies of scale in expanding rural roads up to four lanes (but not much thereafter). On the other hand, expansion costs are often very high in urban areas, where most congestion occurs. The third, and most serious objection, is that there is no guarantee that highways have been optimally adjusted, short of measuring the congestion costs directly and comparing them with expansion costs.

4. The estimation of congestion costs

The theoretical approach for calculating the marginal congestion cost (MCC) of an extra vehicle in the traffic stream usually starts by postulating a relationship between speed (u kph) and flow (q PCE/hr) where PCE are passenger car equivalents, sometimes termed passenger car units or PCU. If the travel cost per km of a representative vehicle is

$$c = a + \frac{b}{u} \qquad (13)$$

when b/u is the cost per vehicle hour, which includes the opportunity cost of the vehicle occupants, then the total cost of a flow of q PCE/hr is cq. If an additional PCE is added to the flow, the total social cost is increased by

$$\frac{dW}{dq} = \frac{d}{dq}(cq) = c + q\frac{dc}{dq}. \tag{14}$$

The first term is the private cost borne by the vehicle, and the second term is the marginal externality cost borne by other road users. From equation (13), the MCC is given by

$$MCC = \frac{b}{u}\left(-\frac{q}{u}\frac{du}{dq}\right). \tag{15}$$

A more convenient measure of congestion is the marginal time cost

$$MTC = \frac{1}{u}\left(-\frac{q}{u}\frac{du}{dq}\right) \text{ veh hr/PCE km} \tag{16}$$

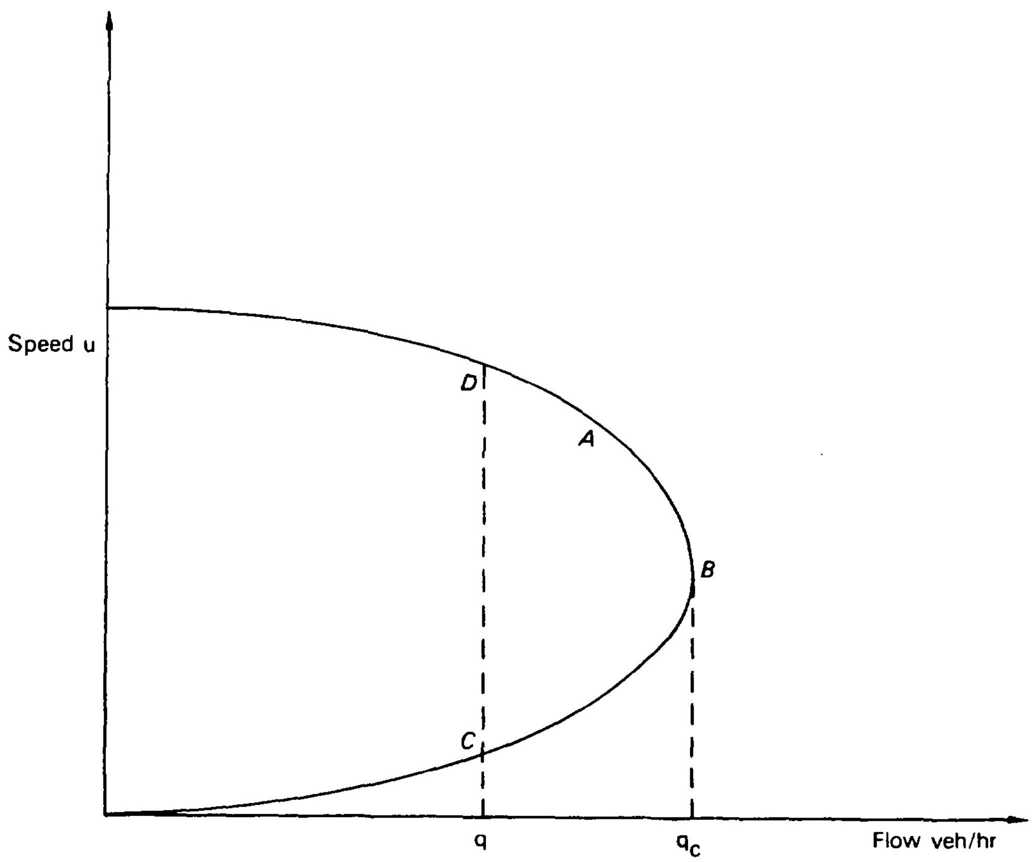
Since this is a technical, rather than an economic measure, it is more likely to be stable across countries and time than the MCC, which is typically calculated in local currency of the day, often making strong assumptions about the value of time which may themselves be disputed.

This simple approach has two serious limitations. The first is that the speed flow relationship is not single valued, and takes the typical form shown in Figure 3. This has the property that a road may have a relatively uncongested flow at a point such as A, but if traffic flow builds up to the capacity level, q_c at B then speed drops sharply, and so does flow, to a point such as C. On a freeway, this effect usually occurs at a point where additional traffic joins the stream, and if the flow builds up to capacity, then speed falls, and the boundary between uncongested and congested flow moves gradually upstream. Either the rate of entry to the freeway subsequently falls, or congestion spreads back and reduces entry (Hall, et al. (1986)).

The simple formulas (15) and (16) are clearly unsatisfactory, since at a point such as C, du/dq is positive, and the implied externality is negative, which is absurd. The problem arises because the implicit model behind (15) is one in which demand is for a traffic flow, q, whereas in practice demand is for completed vehicle trips, whose cost will be the costs of time, discomfort, etc., as well as the direct costs (vehicle costs, or transport fares). On this view, the time taken to

FIGURE 3.

REPRESENTATIVE SPEED-FLOW DIAGRAM



complete a trip will increase as extra vehicles precipitate congestion, perhaps very sharply, so that extra vehicles joining the flow will continue to exert positive externalities, which will mainly fall on subsequent vehicles. The simple flow-past-a-point model $u = u(q)$ fails to capture those dynamic or upstream effects on traffic flow and hence time (Else (1981); Carey and Else (1985)).

This dynamic effect is nicely described by flow density diagrams (Hall, et al. (1986), p. 203) shown in Figure 4. A possible sequence of events might be for initial traffic flows at A to be augmented by an inflow, increasing density $k = q/u$ and slightly reducing speed below u_f (the free flow speed). When volume reaches q_c, speed falls sharply and density rises to a point such as C. Eventually, speed will be able to increase from u', to u_f, carrying the same volume at a lower density, D, and traffic flow will revert to uncongested flow at D. (The same sequence is shown in Figure 3.)

If demand is ultimately defined over the time taken to complete trips, rather than the flow of vehicles, the model needed is not so much one of a single stretch of road, but of the time taken to transit points in a network, given demands arising elsewhere which generate traffic flows through the network. How this might be done will be discussed in the next section.

The second serious limitation of the static flow model $u = u(q)$ is that, by itself, it gives at least a partial equilibrium or short-run measure of the congestion costs caused by an additional vehicle unexpectedly incrementing the flow. In practice, the more important question to ask is what the new equilibrium would be if some changes took place (an increase in demands for trips, an increase in fuel taxes, the introduction of parking restrictions, higher parking charges, the introduction of electronic road charging, etc.) and all transport demanders have time to learn of the changes and adjust to them. This long-run, or general equilibrium, response may be quite different from the short-run impact, and will be discussed further below.

5. Modeling congestion in a network

At least three methods are available for determining the short-run congestion costs arising from increased demand for trips. Conceptually the simplest is to take the static flow model for a link in the network and assume that it continues to hold for average flows in the whole network or some part of it. A good recent example is provided by Harrison, et al. (1986) in their model of traffic flow in Hong Kong. Their approach was to divide the city into areas, identify the speed-flow relationships on links to and within areas, and then to generalize these relationships to cover the areas as a whole. These relationships were calibrated for each link at peak and interpeak flows, and it was then assumed "that within a range of approximately $\pm$ 20 percent of existing traffic flows, the average speed in a particular area would depend only on the total level of traffic flow in that area and would be independent of traffic pattern and distribution by link" (Harrison, et al. (1986), p. 141). The average speed and flow for the area was in turn found by averaging over the links in the area.

The second method is to simulate traffic flows through a network using a model of the junction delays at junctions controlled by signals. Such models are primarily intended to optimize the timing of traffic signals in a network, but can be used to simulate the effect of extra traffic on links in that network. Dewees (1979) reports the results of applying such a model to traffic flow in Toronto. Harrison, et al. (1986) tested their predictions against the results from a TRANSYT program run, where TRANSYT is a program for the detailed analysis of junction delays in fixed time-linked signal systems, described in Vincent, et al. (1980). Harrison, et al. concluded that the TRANSYT results were too sensitive to flow changes as they allowed no rerouting of the traffic, and this is their major problem (and the problem with using any single link-specific traffic flow model).

The final method is to estimate a city-wide traffic flow model. Bertrand (1978), draws on Zahavi's work (Zahavi (1976)) to derive estimates of congestion costs in Bangkok. Zahavi finds that traffic intensity (VKT/km2) N/A satisfies the following relationship (where N is VKT):

$$\frac{N}{A} = \alpha \frac{L/A}{\bar{u}} \qquad (17)$$

where L is road length, L/A is road density, and u is average speed. Thus, average speed can be written as

$$\bar{u} = \alpha\, L/N. \qquad (18)$$

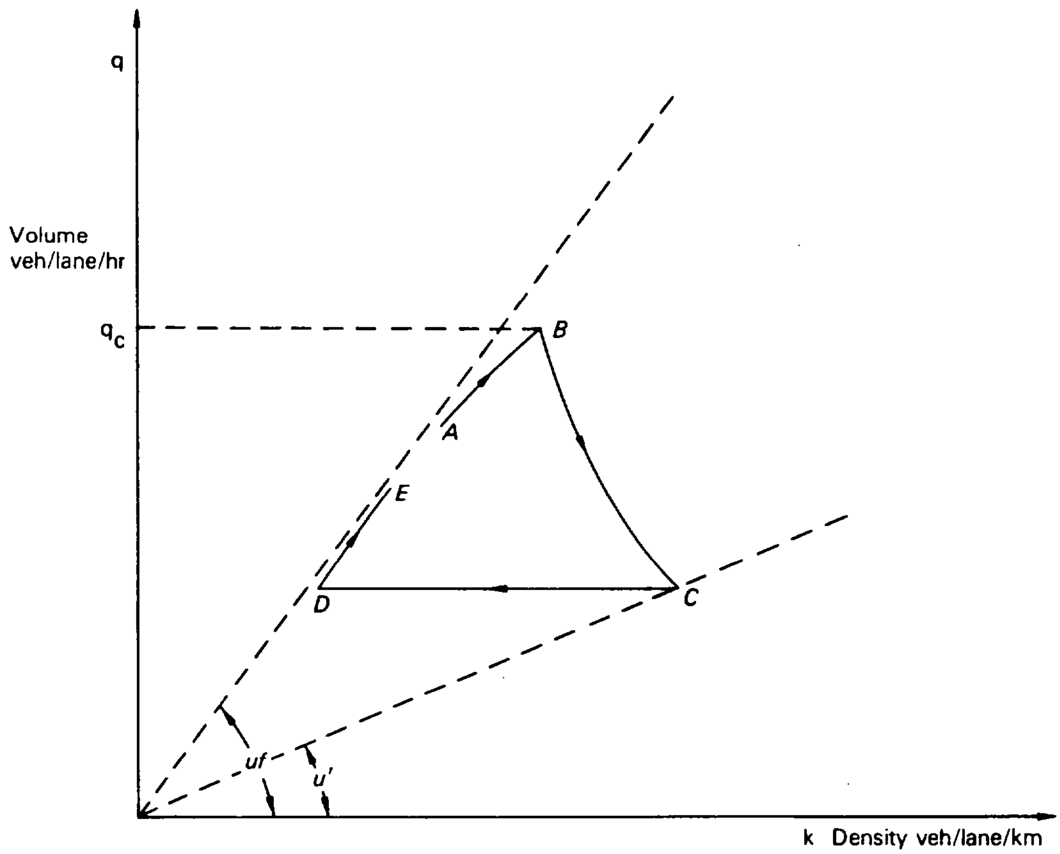
The marginal congestion cost of an extra VKT is then

$$MCC = \frac{2N}{\alpha\, L} = \frac{2}{u} \text{ veh hrs/veh km} \qquad (19)$$

which, since L is fixed and α is a parameter, appears to be independent of traffic flow. To avoid the apparent paradox that congestion externalities are no higher on congested than uncongested roads, Bertrand assumes that vehicle speeds vary with congestion. This procedure effectively assigns different values to the MCC in the heavy and lightly congested period, inversely proportional to (assumed) speeds.

FIGURE 4.

TIME-TRACED PLOTS OF FLOW-DENSITY DATA



The idea of modeling city-wide flow relationships based on cross-section data is attractive, but it is likely to produce long-run equilibrium relationships, rather than the short-run responses described by the preceding two methods. It also seems susceptible to rather arbitrary calibration methods.

6. Empirical estimates of short-run congestion costs

The best way of presenting different estimates of short-run congestion costs is as marginal time costs (MTC) which can then be valued appropriately, depending on local circumstances. Table 3 gives the area-wide estimates for Hong Kong deduced from the area speed-flow graphs in Harrison, et al. (1986). As such, they are subject to various errors of measurement (the slopes were the best fit with a ruler). Table 4 gives hypothetical values from Churchill (1972) derived from Greenberg's formula (Newbery (1985), Table 4 and below, Table 5). They are remarkably close to the value from Wanchai in Table 3 for corresponding mean speeds of 15 kph, but very sensitive to speeds at lower speeds. Greenberg's relationship is but one of many, and Table 5 summarizes four relationships discussed in Transport Research Board (1976), the second of which has been adopted in the U.S. Highway Cost Allocation Study (U.S. Federal Highway Authority (1982)).

The various formulas for the MTC all require calibration (of u_f or u_m) and in Table 4 this was done by setting the MTC equal to 0.1 at 15 kph, consistent with Churchill's calibration and the Hong Kong relationships. What emerges clearly from Table 4 is the great variability of the MTC across relationships away from the point of calibration (which, given the substantive differences between the relationships in Table 5, is not surprising).

Table 6 gives a comparison of MTC for a simple average of eight streets in Toronto the first estimated by Dewees using a traffic signal simulation model and, the remaining three estimated from standard speed flow relationships. The table confirms the sensitivity of the results to the choice of functional form and the method of calibration.

7. General equilibrium or long-run congestion costs

The short-run estimates presented above only measure the costs imposed on other traffic by an unexpected increase in the demand for trips, assuming no response from the existing traffic flow. Most policies that do deal with congestion remain in place for long enough so that commuters learn to adapt to the new equilibrium, and their responses may modify the impact in important ways. The long-run congestion cost of an extra vehicle allows for these responses, in particular for the displacement of existing vehicles using the road or network. The evidence in Thomson (1977) suggests that cities tend to reach an equilibrium in which the average speed settles down at about 15 kph. Extra vehicle journeys then cause existing traffic to adapt in

Table 3. Marginal Time Costs in Hong Kong

(In vehicle hours per PCE kilometer)

		Wanchai	Central	Mid-Levels	Ya Ma Tei	Simple Average
Area speed flow						
am peak	15 kph	0.1				
pm peak	14 kph	0.1				
inter peak	14 kph	0.1				
Normalized speed flow per lane						
speed	10 kph	0.17	0.24	0.22	0.14	0.19
	15 kph	0.13	0.13	0.20	0.11	0.14
	20 kph	0.06	0.06	0.08	0.06	0.07

Source: Harrison, et al. (1986), pp. 142-43, Figure 3, 3-4.

Table 4. Hypothetical Marginal Time Costs

(In vehicle hours per PCE kilometer)

Speed kph	Greenberg $u_m = 9.0$	Greenshields $u_f = 24$	Underwood $u_f = 27.33$	Bell Curve $u_f = 20.24$
10	0.90	n.a.	n.a.	n.a.
12	0.25	∞	0.39	n.a.
14	0.13	0.18	0.14	0.20
15	0.10	0.10	0.10	0.10
17	0.07	0.04	0.05	0.03
20	0.04	0.013	0.023	0.001
25	0.02	n.a.	0.004	n.a.

Sources: Churchill (1972), p. 125, and Table 5.

Note: Parameters u_m, u_f chosen to make MTC equal at 15 kph.

Table 5. Marginal Time Costs for Different Speed-Flow Models

Model	Relation Speed u	MTC	
Greenberg IV	$u_m \log (k_j/k)$	$\frac{1}{u(2x-1)}$	$x = u/2u_m > 1/2$
Greenshields I	$u_f (1-k/k_j)$	$\frac{1-x}{u(2x-1)}$	$1 > x=u/u_f > 1/2$
(HCAS)	$(u_f \{1 + \sqrt{1-q/q_c}\}/2)$		
Underwood V	$u_f e^{-k/k_j}$	$\frac{-\log x}{u(1 + \log x)}$	$\frac{1}{e} < x = u/u_f < 1$
Bell Curve VII	$u_f e^{-1/2(k/k_j)^2}$	$\frac{-\log x}{u(1/2 + \log x)}$	$0.6065 < x = u/u_f < 1$

Source: Newbery ((1985), Table 2) from TRB (1976), Figure 14.3. Roman numerals refer to numbers in Figure 14.3.

Note: k is density = q/u, u_m, u_f, k_j, q_c are parameters.

Table 6. Comparison of Congestion Costs for Toronto

	Marginal Time Cost vehicle hour/vehicle mile			
	Simulated (Dewees)	Smeed A	Smeed B	Mohring
Inbound (heavy traffic)	0.12	0.22	0.20	0.04
Outbound (lighter traffic)	0.02	0.03	0.03	0.02

Notes: Smeed's equation is $q = q_o(1 - u^2/u_f^2)$. Equation A used the street average volume-capacity ratio, q/q_o, to determine speed and cost. Equation B uses simulated average speed to infer the cost. Mohring's equation is $q = q_o \{-1.9 + 61.2/u - 324/u^2\}$ and is calibrated using the street average q/q_o.

various ways--perhaps seeking alternative routes, perhaps changing the timing of their departure, or even abandoning the journey. More congested cities have a higher proportion of the network which is congested, and a higher proportion of time occupied by the peak, which, in extreme cases, may be fairly uniform throughout the day. In other words, the effect of an extra journey does not just fall on the existing vehicles on the street, since they will react to the increased perceived private costs increase.

This effect can be illustrated as follows. Suppose that demand for vehicle trips, q^d depends on the private cost per km, c:

$$q^d = f(c). \tag{20}$$

The cost in turn depends on the flow, q, or c = c(q). If one more vehicle is added, then the new equilibrium has demand by the original road users as

$$q^d_1 = f\ \{c(q^d + 1)\}, \tag{21}$$

as shown in Figure 5. The right hand side can be expanded around the initial level of traffic, q_0, and will be accurate for the small increments of traffic which one extra vehicle represents:

$$q^d_1 = f\ \{c(q_0)\} + \left[q^d_1 + 1 - q_0\right] \frac{df}{dc}\frac{dc}{dq}. \tag{22}$$

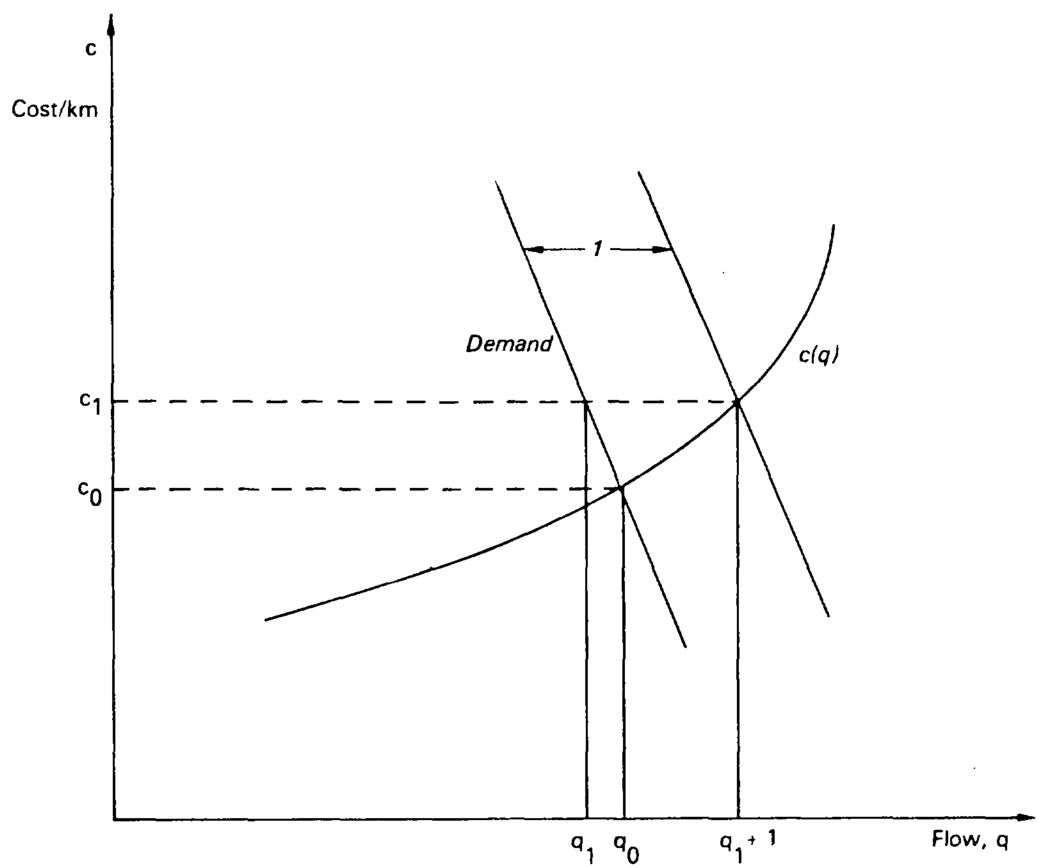
The effective increase in traffic flow is then

$$q^d_1 + 1 - q^d_0 = \frac{1}{1 - \frac{df}{dc}\frac{dc}{dq}},$$

$$= \frac{1}{1 + \varepsilon \frac{q}{c}\frac{dc}{dq}}, \tag{23}$$

FIGURE 5.

EFFECT OF AN EXTRA VEHICLE ON TRAFFIC



where ε is the elasticity of demand for vehicle trips (as a positive number). The long run MCC is equal to the short run or impact MCC multiplied by this factor to take account of the induced reduction in travel by other vehicles. In the limit if ε is very large, there will be no increase in congestion costs, for an extra vehicle joining the traffic flow will, in equilibrium, induce some other vehicles to leave, and the flow remains unchanged.

The effect of this adjustment can be quite large, for the corrected ratio of MCC to private cost now becomes

$$\frac{MCC}{c} = \frac{\mu}{1 + \varepsilon\mu}, \qquad \mu \equiv \frac{qdc}{cdq} = \frac{SRMCC}{c}. \tag{24}$$

The results given by Dewees ((1979), Table 1) suggest values of μ between 1.6 and 2.8 for the more heavily congested inbound streets and values between 0 and 0.5 for outbound streets. The effects of the elasticity of demand are then quite strong, as Table 7 suggests.

These calculations suggest that if the effect of extra traffic is to remove some current traffic from the road, the MCC will be reduced, possibly substantially and especially on congested routes. They confirm the importance of allowing for general equilibrium effects, and the wisdom of using area speed flow relationships, rather than simulated link behavior, as the Hong Kong study suggested.

The conclusion to emerge from this analysis is that the ability of the traffic flow as a whole to respond to extra vehicles on a particular route, either by withdrawing or reallocating to less congested routes, is to reduce the social congestion costs of the extra vehicle. If we assume a value of the trip elasticity of 1.0 then the long run MCC will be less than half the impact MCC on typically congested streets (u = 15 kph, MTC = 0.1) and less than one fifth if the elasticity is as high as 3.0.

This does not, however, imply that the social cost of congestion is low when the elasticity of demand is high, as Figure 6 demonstrates for the extreme case of perfectly elastic demand and hence zero long run MCC. Indeed, this social cost of congestion is highest when demand is perfectly elastic.

A quick calculation of the relative importance of this deadweight loss L is easy for the perfectly elastic case for

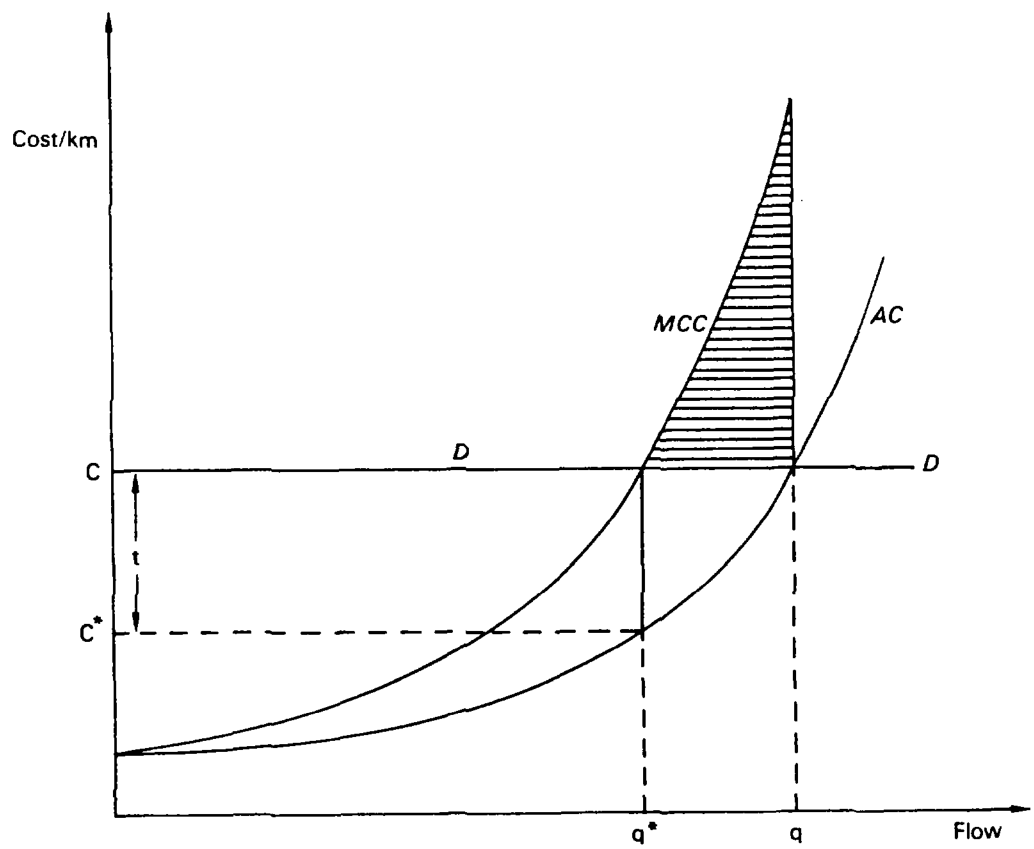
Table 7. Ratio of Long-Run Congestion Cost to Private Cost

Uncorrected Ratio of SRMCC to Private Cost	Ratio of LRMCC to Private Cost Value of Elasticity of Trip Demand ε				
	0.5	1.0	2.0	3.0	5.0
Inbound (congested)					
μ = 1.6	0.9	0.6	0.38	0.28	0.18
2.0	1.0	0.67	0.40	0.29	0.18
2.8	1.2	0.74	0.42	0.30	0.19
Outbound (less congested)					
0.2	0.18	0.17	0.14	0.13	0.10
0.5	0.4	0.33	0.17	0.20	0.14

Source: Equation (24).

FIGURE 6.

DEADWEIGHT LOSS OF CONGESTION



$$\frac{L}{qc} = \frac{1}{qc}\int_{q^*}^{q} q\,\frac{dc}{dq} - \frac{c(q)(q - q^*)}{qc} = \frac{\left[qc\right]_{q^*}^{q}}{qc} - (1-q^*/q) - \frac{1}{qc}\int_{q^*}^{q} cdq$$

$$= \frac{q^*}{q}\{1 - c^*/c\} - \frac{\int cdq}{qc} \qquad (25)$$

For the special case of the linear speed flow function $u = \alpha - \beta q$, and the case of pure time costs of travel, $c = b/u$, this expression is readily evaluated as

$$\frac{L}{qc} = \frac{q^*}{q}\left(1 - \frac{u}{u^*}\right) - \frac{u}{\beta q}\log(u^*/u). \qquad (26)$$

If MTC = 0.1 at u = 15 kph, then u^* = 17.37 kph, $q^* = 0.8944q$, t/c^* (the optimal charge) = 15.83 percent, and the proportional loss L/qc = 2.4 percent. The short-run ratio of MCC/c is 1.5, so the optimal tax is only 10 percent of the SRMCC, though since the LRMCC is zero, the SRMCC is potentially misleading in more ways than one. If the elasticity of demand is less than infinite, the congestion tax will be higher, the deadweight loss will be lower, and the LRMCC will be higher. While these figures may seem modest, it is important to remember that they are ratios of losses or taxes to the time inclusive cost of travel, which is here assumed to dominate the non-time costs.

8. Charging for road damage and congestion

The main reason for charging vehicles for the road damage they cause is, first, to encourage the right choice of vehicle (particularly the axle configuration, for spreading a given load over more axles dramatically lowers the damaging power of the vehicle). Second, the vehicle should not be overloaded, as this also increases the damage rapidly. Given the correct choice of vehicle and loading, both of which can be encouraged by legal restrictions on configuration and loading, as well as license fees, the final step is to charge vehicles in proportion to total damage, that is, in proportion to distance travelled. Distance-related taxes, such as taxes on ton miles, non-rebateable taxes on transport services, taxes on fuel and tires, and purchase taxes on vehicles and spare parts are more or less satisfactory for this purpose.

Since road damage costs vary across vehicles by such an enormous factor, there are potentially large benefits to be derived from adjusting road user charges to reflect variations in road damage costs. Small and Winston (1986b) calculate the welfare effects of imposing marginal cost pricing of road damage for heavy highway vehicles in the United States using 1982 data, removing existing (1982) federal and state mileage-related taxes, but leaving annual fees in place (on the rather doubtful argument that these are payments for urban congestion and such highway overheads as policing, signaling, and the like). They find that the welfare gain is roughly $1.2 billion per year, and an extra $10 billion is raised in tax revenue whilst annual highway maintenance and repair expenditures fall by nearly $3 billion. More than one third of the net benefits were attributed to an induced switch to rail, whilst the rest arises from a switch of freight from more to less damaging vehicles. To achieve this the road user charge has to be adjusted to encourage the right choice of vehicle and mode.

Although road damage costs for a given vehicle vary by a factor of ten or so across roads of different strengths, it is impractical to imagine charging vehicles different amounts for driving on different roads (especially as the highest road damage costs will occur on the least frequented roads), and it is difficult to imagine that there would be much advantage in so doing, even if it were feasible. Since road damage externalities are negligible, and since road damage costs are a small fraction of vehicle operating costs, the only purpose served by varying the charge across roads would be to discourage a few marginal trips on weak roads, with a small benefit to the highway authority and negligible benefits to other road users.

Exactly the opposite is true of congestion costs. Here there is little variation across vehicles on a given road under given traffic conditions, but a huge variation in congestion costs across different roads and different times. Moreover, although congestion substantially increases vehicle operating costs, the external costs inflicted on other road users can dwarf these private costs. In large part, the problem of road congestion arises because vehicles are not charged for the external costs they inflict on other road users. One logical solution therefore appears to be to charge road users for these costs. The correct level for these charges would be the marginal externality costs, which would depend on traffic levels and the road capacity. There are two obvious difficulties in implementing such a system. The first, which does not appear to create serious problems, is that the MCC varies from moment to moment. For any system of charging to work, the charges would have to be predictable and not too complex, so that a potential road user would be able to decide whether to make the trip, when to make it, and by which route. The solution is to have a limited number of charges, just as telephone charges are limited to three rates (peak, standard, and cheap rates), and perhaps a limited number of zones. The Hong Kong proposals provide an excellent example of this (Dawson and Catling (1986)).

The second difficulty is more serious, and that is physically how to meter road use and make the appropriate charge. The main options are to equip vehicles with an electronic device which either records road charges, or informs a central processor of road use, or to issue access licenses. Clearly, the first method is far better suited to confronting the road user with the right charge, whilst the second can only be a relatively crude way of charging for road use. The first method has been proposed for Hong Kong (Dawson and Catling (1986)) and looks very attractive for congested urban areas, especially in city states. The second approach has been tried in various forms and appears to be cost effective if carefully designed (May 1986). (The obvious problems are that the congestion is merely displaced from the licensed area to the immediate neighborhood.) Indirect alternatives are to charge very heavily for parking in certain areas, and hence reduce the demand for commuting by private car. Yet further alternatives are to improve public transport relative to private transport, and to be effective this must involve raising the private costs of private transport, either by increasing congestion, and/or reducing parking facilities, whilst protecting public transport from these costs (e.g., by segregating public transport).

What appears to emerge from a study of urban transport is that fuel taxes are a very indirect method for dealing with and charging for congestion, since they are so unselective, affecting congested and uncongested roads almost alike. Thomson (1977) implies that vehicle taxes (import duties and license fees) may be more effective, for in Colombia, where they were high, their effect was to reduce the demand for private car ownership and hence increased the demand for public transport in Bogota. The same appears to be true in Korea and Hong Kong (Armstrong-Wright (1986)). The benefits of this can be far reaching, for the finances of urban transport are precarious. If demand is high, finances improve, service improves, and this reinforces the demand for public transport. If its financial position deteriorates, then service quality may fall, and further lower demand. The argument for high license fees rather than high fuel taxes might then be made as follows. High fuel taxes may reduce vehicle traffic on uncongested roads proportionately more than on congested roads, since they will comprise a higher fraction of private costs on the uncongested roads. A tax on vehicle ownership which leads to an equal fall in vehicle km will presumably reduce vehicle km on both congested and uncongested roads in proportion to their previous use.

If this argument is accepted, the next question is to ask how best to reflect congestion charges in the charge for vehicle ownership. Taxes can be either on the value of the vehicle (purchase taxes), or on its annual ownership (license fees). The first may be desirable on equity grounds, but has little to recommend it as a congestion charge, as it will tend to encourage the use of vehicles which are small and old (and hence more prone to congestion-inducing breakdown). 1/ License

---

1/ Older vehicles typically have lower utilization rates, and purchase taxes have often been advocated as they will thereby vary somewhat with distance driven, but unless distance driven on congested roads increases with total distance driven (and the converse seems more likely) this argument will not apply.

fees have the obvious advantage that they can in principle be made locationally specific and set at a rate determined by place of residence. Insurance companies already vary their charges by residence and it allows a crude discrimination between congested urban areas and the rest of the country.

9. Efficient user charges for congestion

Congestion will have no implications for the taxation of fuel, vehicles, and other inputs such as tires and parts, if the government can selectively charge vehicles for the congestion they cause. If this is not feasible, then taxes which affect vehicle use will have an impact on congestion which cannot be achieved by other instruments, and this effect should be taken into account when setting these taxes. The question to be addressed now is at what level should the distance charge (such as the fuel tax) and the annual access charge (the vehicle license fee) be set to charge most efficiently for congestion externalities.

Distance taxes work by raising the private cost of trips, and hence discouraging vehicle use. The efficient charge is then a weighted average of the marginal congestion cost, where the weights depend on the distances driven on roads of differing congestion levels and the elasticities of demand for these trips.

Access charges increase the cost of vehicle ownership and discourage marginal owners from purchasing vehicles. To the extent that the marginal vehicle owner drives an above-average fraction of total distance in congested conditions, the access charge will have additional leverage over total congestion levels, though rough calculations suggest that if the license fee has to be countrywide, it should normally be set at a modest level. (City states like Hong Kong provide an obvious exception.) (Derivations of the formulas and an empirical illustration for Tunisia are given in Newbery (1986g).) If license fees can be made area-specific (by zip code of residence, possibly cross coded by zip code of place of work) then license fees will have substantially greater leverage than distance taxes, and very high rates might be suitable for car ownership in congested urban environments. One would expect this to have a significant impact on the demand for urban transport. The area licensing scheme in Singapore provides a rather crude model of this solution (World Bank (1986)).

10. The efficient level of road user charges

Road damage costs vary across different vehicles by a factor of between 1,000 and 10,000 to 1 (depending mainly on the proportion of vehicle kilometers travelled on unpaved roads, for which the variation is quite low). It is hard to think of a tax base with the same variation across vehicles. Congestion costs vary by a relatively small amount across different types of vehicle and are large relative to road damage costs (typically amounting to over half of all road use costs). Together they make up road use costs which vary across vehicle types by a factor of 5-20 to 1, comparable to the variation in fuel consumption and other input costs, suggesting that input taxes may provide a reasonable way of levying road user charges (at least, if electronic pricing is ruled out).

11. Road user charges on freight vehicles

For freight vehicles, the road use cost per ton mile was fairly uniform across different truck sizes in Tunisia, suggesting that either a ton-mile tax or a nonrebateable tax on transport services would be a good way of levying road user charges, were it to be feasible. Ton-mile taxes are levied by some states in the United States, but are evidently very demanding of tax administration skills. A tax on transport services may be easy to levy on professional hauliers, but not on the casual operators attracted into the haulage business following deregulation. Distance taxes at a vehicle-specific rate would be feasible for such vehicles if they were required to carry tamper-proof tachometers, and this could be encouraged by having low annual license fees for such vehicles. License fees and legal restrictions are required to encourage operators to choose vehicles with the correct configuration of axles designed to minimize road use costs by reducing road damage costs. (Doubling the number of axles reduces the damage by a factor of eight.)

Whilst these techniques would appear manageable for developed countries, they may be infeasible in developing countries, which will therefore have to fall back on taxes on inputs and license fees. On the face of it, fuel taxes appear to be a good choice, as they vary appropriately across size classes of trucks, increase with loading, and, set at U.K. rates, would appear to recover a major part of road use costs for legally laden vehicles. Tire taxes are similarly attractive, though the rate of tax is limited by the incentives they offer for dangerous misuse and retreading. At rates which are not excessive (20 percent), about one fifth of road use costs might be charged by tire taxes. Purchase taxes (nonrebateable) on vehicles and spares are also a good method of charging freight vehicles, as a large fraction of the tax would fall on distance (since trucks deteriorate primarily from use, not age), and overloading vehicles apparently accelerates the rate of wear. Purchase taxes at vehicle-specific rates would fall more heavily on less heavily utilized vehicles (since the interest costs on the tax would be higher for longer-lived vehicles) and might restrict entry somewhat by raising entry costs (though truck leasing would reduce the force of this objection). Together, taxes on these inputs could, in principle, recover a large fraction of road use costs, with vehicle-specific divergences being corrected by license fees. The key issue is whether it is desirable to levy a high rate of tax on diesel fuel, and this will be considered below.

12. Road user charges on passenger transport

Except for large buses, the overwhelming fraction of the road use cost consists of urban congestion costs, and, as mentioned above, it is hard to adjust charges finely to these highly variable costs. Even if charges can only be set at a weighted average of congestion costs, it appears that adjusting urban highway capacity may reduce the inefficiencies caused by congestion substantially (Mohring (1986)), though this remains controversial. If the main mechanism of adjustment to congestion levels is via location and the size and configuration of cities, then urban highway expansion may be largely self defeating.

The design of road user charges for passenger transport has two obvious dimensions--encouraging as far as possible the right level of road use (by time of day) and the right choice of mode (car, bus, train, taxi, bicycle, and foot). Bertrand (1977) offers two potentially attractive rules for setting the structure of congestion charges between different classes of passenger transport vehicles. The first is that if congestion charges do not affect the total number of passenger-kilometers traveled during a particular period of the day, then the "second best" (or the best feasible) set of charges should equalize the difference between the "tax" and the externality across modes (that is, should equalize distortions per passenger-kilometer across modes). ("Tax" includes all taxes and subsidies levied on the road user as a function of his choice of transport mode. If public transport is allowed to run at a loss, then this is equivalent to a subsidy.) For example, if private cars and taxis are taxed at a roughly uniform rate per kilometer for different levels of congestion, they will be undercharged on congested roads and overcharged on uncongested roads. If feasible, this rule would argue for subsidizing public transport on congested roads by the same amount per passenger kilometer and taxing them on uncongested (e.g., inter-urban) roads.

The second rule is that if raising charges on any mode reduces total passenger kilometer during that period, then any charge that can be raised during that period should be raised above the level that equates distortions across mode. The second rule substantially modifies the first, for if it is not feasible to vary charges on private transport by time of day, then only public transport fares can be raised. Since an increase in these fares will certainly reduce total passengers' kilometer during that period, there is a case for reducing the subsidy for rush-hour users of public transport.

In practice, these rules are mainly useful as rough tests to rule out obvious inefficiences. The exact level of charge and subsidy will depend not only on the level of congestion costs (which are hard to measure) but also on own- and cross-price elasticities of demand for trips by various modes (which are almost impossible to measure). Several conclusions can, however, be drawn. There appears to be a case for subsidizing rush-hour public transport if it is hard to charge for rush-hour private transport adequately. If bus subsidies lead to inadequate financial control and prevent otherwise desirable reforms such as privatizing bus services, or result in inferior and under-financed public transport services, then the case for raising charges on private vehicles greatly increases. Thus Japan, faced with very severe urban traffic congestion, has very high charges on private car ownership, and restrictions on car use (no car ownership without off-street parking, very limited on-street urban parking, stiff driver's exams, and so on. See McShane, Koshi, and Lundin (1984)). Second, it is clear that levying efficient congestion charges will by itself result in a very adverse increase in the cost of private passenger transport--the tax transfer will be very large (perhaps 5-20 times as large) compared with the reduction in inefficiency. (See, for example, Bertrand (1978), Tables 5 and 8.) Although traffic flows may improve and trip times fall, most, if not all, private motorists will be substantially worse off after these charges are levied, unless some way is found of returning the tax revenue to motorists. Here, Hong Kong offers a useful model, for without electronic pricing, annual road license fees were set at very high levels, but electronic pricing would allow the license fee to be reduced or eliminated. One logical sequence for introducing new methods of congestion pricing (including indirect methods such as increased taxes on parking) would be to introduce locationally-specific license fees as far as possible, and then link the introduction of the new measures to charge for congestion (daily access charges) with a reduction in the existing annual license fees.

13. The taxation of personal transport

Personal transport is a final consumption good, and hence a legitimate object for indirect taxation, over and above the collection of road use costs. Budget studies reveal that expenditures on gasoline, car purchase, and maintenance are amongst the most income-elastic expenditures in developing countries, and are thus an attractive tax base for countries where income tax collections are limited in coverage and effectiveness (Deaton (1986)). Vehicle and gasoline taxes are amongst the easiest to administer, and high rates of tax can readily be justified on distributional grounds. The only problem is to minimize tax avoidance and the inefficient use of vehicles. Heavy taxes on gasoline together with light taxes on diesel will encourage the purchase of diesel-engined alternatives. This can be avoided to some extent by imposing heavy annual license fees on diesel-powered private cars (at a level such that purchasers would make the same choice of vehicle as in the absence of all fuel taxes and extra license fees). The choice of overly sophisticated fuel-efficient vehicles can be offset by making purchase taxes and license fees a function of the value of the vehicle, rather than its power or cubic capacity. (Purchase taxes are an attractive way of achieving this and might be set at an escalating ad valorem rate.) The main problem will be that high automobile taxes encourage the use of pickups or utilities, which are also used for commercial purposes (and for which they should not be subject to pure taxation). This is probably a serious problem in several developing countries such as the Philippines, Tunisia, and Thailand. Whether it is possible to reduce this substitution by taxing pickups with more than, say, two seats, at automobile rates is not clear. If it proves difficult, the extent to which private cars can be heavily taxed without potentially costly distortions will be reduced. Nevertheless, European levels of taxation of gasoline are quite easy to justify.

14. The choice of fuel taxes for levying road user charges

Taxes on diesel fuel are a potentially attractive method of levying road user charges on freight vehicles. Gasoline taxes are a good method of charging private cars for congestion, though they discriminate inefficiently in favor of small cars, as Table 2 above demonstrates. Whether this variation across vehicles is warranted on distributional grounds seems doubtful in developed countries, though not in developing countries which have fewer effective methods of taxing income. Hence its rate can be set without regard for impacts elsewhere, except insofar as high gasoline taxes will encourage the use of diesel-fueled substitutes (and this can be offset by appropriate license fees on diesel-powered alternatives). Diesel and its very close substitutes, kerosene, gas oil, and other middle distillates, are extensively used outside the transport sector, both as an intermediate input and final consumption good. If diesel fuel for road use can be differentially taxed (as in the United Kingdom, Germany, and some other countries) and the illegal use of untaxed substitutes in road vehicles effectively prevented, then this need cause no concern. But in most developing countries this is not feasible, and before advocating a road user charge on diesel, one must explore the impact on the rest of the economy.

The impact depends on four factors--the extent of nontransport uses of middle distillates in production, the degree of substitutability in production and consumption, the extent of kerosene use by the poor, and the structure of the tax system. Hughes (1986) has studied the impact of raising the prices of diesel (and its close substitutes) in Tunisia, where 60 percent of such fuel is used outside the transport sector, using an input-output table with flexible coefficients, a proper model of tax shifting, and a consumer budget survey to examine the resulting welfare impact on consumers. The tax structure, together with the demand responses, determines the impact on government revenue and allows one to calculate approximate measures of dead-weight loss. The values for the substitution elasticities in production and consumption were taken from empirical studies elsewhere, as there are few studies available for developing countries. The results therefore give a feel for the importance of allowing for substitution responses, rather than a definitive answer for Tunisia, but are nevertheless of considerable interest.

He found that the derived demand elasticity for diesel was quite high, and that diesel taxes would therefore lead to quite high dead-weight losses. Thus, imposing a U.K. level of tax on the currently (1982) subsidized diesel leads to a dead-weight loss of over 50 percent of the revenue collected and would, therefore, be highly undesirable compared with other more broadly based taxes. Moreover, the distributional impact of the tax was also adverse, as kerosene is the good most preferentially consumed by the (predominantly rural) poor (as it is in many developing countries). (Subsidizing kerosene for distributional reasons and taxing diesel usually leads to massive adulteration of diesel, as the Indian evidence shows, since transport vehicles can run on a fuel mixture of 30 percent kerosene with no adverse effects and can tolerate even higher levels.)

The other result was that the structure of the rest of the tax system modifies the impact of raising the tax on diesel, as one would expect. Gasoline tax revenue increases, though there are a large number of less obvious revenue effects which depend sensitively on the structure of the economy and of the tax system.

The conclusion is that heavy diesel taxes are likely to be undesirable in developing countries and that therefore, vehicle purchase taxes will play an important role in levying road user charges on commercial vehicles, together with license fees. A corollary is that raising diesel prices to efficient (world price levels) is highly desirable, as subsidies on diesel will induce correspondingly high rates of inefficiency.

15. The taxation of other transport fuels

Although diesel and gasoline are the dominant transport fuels, some countries have, at various times, encouraged the use of methanol, ethanol, gasohol (i.e., a blend of gasoline and ethanol), LPG, and compressed natural gas (CNG), often at subsidized rates. The question arises as to how these fuels should be priced and taxed. Here the principles are easy. Most of the alternatives are gasoline substitutes rather than diesel substitutes (spark ignition rather than compression engines). Logically, provided the fuels have no nontransport uses, they should be priced at the (marginal) border (world) price and taxed at the same rate as gasoline. If, because of market power, and the low opportunity cost of sugar, the marginal export value of ethanol in Brazil is below the world price, but no lower than the marginal social cost of production, then this marginal value forms the base price on which the gasoline tax would be levied to produce the pump price. (Other arguments for subsidizing ethanol, mostly infant industry arguments, mainly apply to reduce the cost of production, not the export value.)

Of these fuels, the only problematic ones are LPG and CNG, for which there are many other uses. It may, nevertheless, be feasible to tax CNG but not ordinary natural gas, in which case there is no problem, but with LPG it is unlikely to be easy to discriminate between alternative uses. To that extent LPG is like diesel and should be treated in a similar way, with a relatively low tax on LPG and correspondingly higher license fees on LPG-powered private cars.

16. The impact of road user charges on the rest of the economy

Once a set of road user charges and road taxes on personal transport has been designed, taking into account these impacts elsewhere, the final questions any government will need to answer are: what impact will the switch to a new system of road taxation have on the price level and what impact will the change have on the distribution of income?

Again, the techniques used by Hughes (1986) can be employed to answer these questions, though the questions themselves have to be carefully formulated. As far as reforming road user charges is concerned, it only makes sense to consider the impact of a revenue-neutral tax reform. If road users are currently undercharged, and road taxes are to be raised, other taxes can be reduced. If government revenue is inadequate, it is a separate question to decide how best to raise further tax revenue, and not one that should be prejudged by raising road taxes.

The results are reassuring and appear to be robust, as they have been examined in three different developing countries (Thailand, Tunisia, and Indonesia). On balance, raising transport taxes and lowering sales taxes or import duties lowers the price level, as transport taxes are shifted backward onto factor incomes to some extent, whilst sales taxes are shifted forward onto final consumers. An increase in the cost of the freight transport of an exported good (such as rice in Thailand) will not raise the world price but will lower the farm gate or ex-factory price, and hence reduce farm wages or land rents (or both) or other factor prices.

The effect of an increase in road user charges on income distribution depends on the choice of road tax to be increased. The evidence is that gasoline taxes are quite progressive, diesel plus kerosene taxes (at the same rate) are somewhat regressive, and taxes on transport are almost exactly neutral. The effects are small (even for quite large tax changes) but very "noisy." That is, although the average impact may be small, the negative impact on some households may be quite large, while the impact on others may be beneficial, but these impacts are very poorly correlated with income. Politically, this is a drawback as the losers will be presumably more vociferous than the gainers, and the cost of the reform may appear higher than it really is. On the other hand, the fluctuations in the world price of fuels dwarfs the tax changes that are likely to be desirable (particularly given the arguments for rather modest levels of tax on diesel fuel) whilst recent falls in oil prices have opened the prospect of adjusting tax rates without adverse domestic price movements. The evidence from a wide variety of countries is that quite large domestic price adjustments of fuels are possible. It is also important to realize that road use costs are a modest fraction of vehicle operating costs and so the impact of the tax changes will not be as dramatic as the changes in the nominal tax rates might appear.

17. Conclusions

Recent theoretical advances have greatly clarified the nature and measurement of road damage costs and road damage externalities. It now appears that road damage externalities are negligible and that charging for road damage costs will recover between half and three quarters of the costs of road maintenance, almost entirely from heavy vehicles. The theoretical and empirical state of the art is less well advanced for the measurement of congestion costs, but if roads have constant expansion costs per unit of capacity, and are optimally adjusted, then congestion charges will recover marginal capital costs and a large fraction of maintenance costs. It is possible that efficient road user charges could cover all highway costs, though economies of scale in construction and indivisibilities in capacity make this rather less likely.

In the absence of electronic pricing, road user charges will have to cover both road damage costs and congestion costs simultaneously and imperfectly. Nevertheless, it is not difficult to devise a system of taxes on inputs, on vehicles, and on license fees, which is the best feasible, given the imperfect instruments available. If fuels for transport use cannot be differentially taxed, then diesel fuel is not a very suitable tax base for more than a fraction (perhaps one quarter) of road user charges. Taxes on ton miles or transport services, if feasible, are nearly ideal. Failing that, purchase taxes on vehicles and parts, together with modest rates of tire tax, and the balance recovered from license fees, appear to be a reasonable compromise for freight vehicles. High rates of gasoline tax also appear warranted, with compensating high license fees on diesel-engined private automobiles.

The impact on the economy of raising road user charges is mildly counter-inflationary and, with the exception of diesel taxes, has little effect on income distribution. Taxes on private cars are quite progressive.

Casual observation suggests that private cars are rather heavily taxed in European countries, unless congestion costs are very high. If congestion costs are indeed high, then highway capacity appears to be undersupplied. Heavy vehicles are often, but by no means always, undertaxed, though it appears that many highway cost allocation studies have overestimated their road damage costs.

References

Armstrong-Wright, A.T., "Road Pricing and User Restraint: Opportunities and Constraints in Developing Countries," Transportation Research, Vol. 20A (March 1986), pp. 123-28.

Bertrand, T.J., "Second Best Congestion Taxes in Transportation Systems," Econometrica (October 1977), pp. 1703-15.

_____, "Congestion Costs in a Transport System, with an Application to Bangkok," Journal of Transport Economics and Policy, Vol. 12 (September 1978), pp. 244-79.

Carey, M., and P.K. Else, "A Reformulation of the Theory of Optimal Congestion Taxes," Journal of Transport Economics and Policy, Vol. 19 (January 1985), pp. 91-94.

Churchill, A., Road User Charges in Central America (Baltimore: Johns Hopkins, 1972).

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

Deaton, A.S., "Determinants of Demand for Personal Transport in Developing Countries" (mimeographed, Washington: World Bank, TRP, 1986).

Dewees, D.N., "Estimating the Time Costs of Highway Congestion," Econometrica, Vol. 47 (November 1979), pp. 1499-512.

Diamond, P.A., and J.A. Mirrlees, "Optimal Taxation and Public Production, I: Productive Efficiency," American Economic Review, Vol. 61 (1971), pp. 8-27.

Dupuit, J., "On the Measurement of the Utility of Public Works," reprinted in Transport, edited by D. Munby (1844), pp. 19-57.

Ellet, C., Jr., "A Popular Exposition of the Incorrectness of the Tariffs on Tolls in Use on the Public Improvements of the United States," J. Franklin Institute, Vol. 29 (1840), pp. 225-32.

Else, P.K., "A Reformulation of the Theory of Optimal Congestion Taxes," Journal of Transport Economics and Policy, Vol. 15 (1981), pp. 217-32.

Gomez-Ibanez, Jose A., and Mary M. O'Keeffe, The Benefits from Improved Investment Rules: A Case Study of the Interstate Highway System, Research Report R85-2, John F. Kennedy School of Government, Harvard University (1985).

Hall, F.L., B.L. Allen, and M.A. Gunter, "Empirical Analysis of Freeway Flow-Density Relationships," Transportation Research, Vol. 20A (1986), pp. 197-210.

Harrison, W.J., C. Pell, P.M. Jones, and H. Ashton, "Some Advances in Model Design Developed for the Practical Assessment of Road Pricing in Hong Kong," Transportation Research, Vol. 20A (March 1986), pp. 135-44.

Highway Research Board, "The AASHTO Road Test: Pavement Research," HRB Special Report 61E (Washington, 1982).

Hughes, G.A., "Substitution and the Impact of Transport Taxation in Tunisia" (mimeographed, Washington: World Bank, Transportation Department, August 1986).

International Road Federation, World Road Statistics, 1985 edition.

Knight, F., "Some Fallacies in the Interpretation of Social Costs," Quarterly Journal Economics, Vol. 38 (August 1924), pp. 582-606.

May, A.D., "Traffic Restraint: A Review of the Alternatives, Transportation Research, Vol. 20A (March 1986), pp. 109-22.

McShane, M. P., M. Koshi, and O. Lundin, "Public Policy Toward the Automobile: A Comparative Look at Japan and Sweden," Transportation Research, Vol. 18A (March 1984), pp. 97-110.

Mohring, H., "The Peakload Problem with Increasing Returns and Pricing Constraints," American Economic Review, Vol. 60 (September 1970), pp. 693-705.

_____, "The Role of Fuel Taxes in Controlling Congestion" (mimeographed, Washington: World Bank, August 1986).

_____, and M. Harwitz, Highway Benefits: An Analytical Framework (Evanston, Il: North Western University Press, 1962).

Newbery, D.M., "Congestion Costs and Congestion Charges" (mimeographed, Cambridge, England: Churchill College, July 1985).

_____ (1986a), "Road Damage Externalities and Road User Charges" (mimeographed, Cambridge, England, revised May 1986).

_____ (1986b), "Efficient Pricing of U.S. Interstate Highways" (mimeographed, Washington: International Monetary Fund, Fiscal Affairs Department, July 21, 1986).

_____ (1986c), "Road Use Costs Attributable to Overlays in Tunisia" (mimeographed, Cambridge, England, redrafted July 3, 1986).

_____ (1986d), "The Time Path of Efficient Road User Charges" (mimeographed, Washington: International Monetary Fund, Fiscal Affairs Department, August 12, 1986).

_____ (1986e), "Road Damage Costs in Conditions of Road Crisis" (mimeographed, Washington: International Monetary Fund, Fiscal Affairs Department, July 28, 1986).

_____ (1986f), "On Cost Recovery from Optimally Designed Roads" (mimeographed, Washington: International Monetary Fund, Fiscal Affairs Department, July 29, 1986).

_____ (1986g), "Estimating Urban Congestion Costs" (mimeographed, Washington: International Monetary Fund, Fiscal Affairs Department, August 12, 1986).

_____, and N.H. Stern, The Theory of Taxation for Developing Countries, OUP for World Bank (1987).

Paterson, W.D.O., "Distress Mechanisms, Maintenance and Cost" (mimeographed, Washington: World Bank). Lecture presented at Denver, Colorado, March 31-April 4, 1986, National Highways Institute and Federal Highway Administration (1986).

Pigou, A.C., Wealth and Welfare (London: MacMillan, 1912).

Sayers, M., T.D. Gillespie, and W.D.O Paterson, "Guidelines for Conducting and Calibrating Road Roughness Measurements," Technical Paper No. 46 (Washington: World Bank, January 1986).

Small, K.A., and C. Winston, "Efficient Pricing and Investment Solutions to Highway Infrastructure Needs," American Economic Review (May 1986), pp. 165-69.

Smeed, R.J., "Traffic Studies and Urban Congestion," Journal of Transport Economics and Policy, Vol. 2 (1968), pp. 33-70.

Starkie, D.N.M., "Report to the National Freight Enquiry: The New Zealand Road Charging System" (mimeographed, University of Adelaide, 1984).

Tait, A.A., and D.R. Morgan, "Gasoline Taxation in Selected OECD Countries, 1970-79," Staff Papers, International Monetary Fund (Washington), June 1980, pp. 349-79.

Tanzi, Vito, "Quantitative Characteristics of the Tax Systems of Developing Countries" (Washington: International Monetary Fund), DM/83/79 (1983), forthcoming in Newbery and Stern (1987).

Thomson, J.M., Great Cities and Their Traffic (London: Gollancz, 1977).

Transport Research Board (TRB), Applications of Traffic Flow Theory, in Modeling Network Operations (Canada: Waterloo University, 1976).

U.S. Congress Committee on Public Works and Transportation, The Status of the Nation's Highways: Conditions and Performance, Report of the Secretary of Transportation (Washington: USGPO, 1985).

U.S. Federal Highway Administration (USFHWA), Final Report on the Federal Highway Cost Allocation Study (Washington: USGPO, 1982).

Vickery, W., "Congestion Theory and Transport Investment," American Economic Review, Vol. 59 (May 1969), pp. 251-60.

_____, "Automobile Accidents, Tort Law, Externalities, and Insurance: An Economist's Critique," Journal of Law and Contemporary Problems (1969), pp. 464-84.

Vincent, R.A., A.I. Mitchell, and D.I. Robertson, "User Guide to TRANSYT Version 8," TRRL Report No. LR 888 (Crowthorne, England, 1980).

Walters, A.A., "The Theory and Measurement of Private and Social Cost of Highway Congestion," Econometrica (October 29, 1961), pp. 676-99.

_____, The Economics of Road User Charges (Baltimore: Johns Hopkins, 1968).

Winston, C., "Conceptual Developments in the Economics of Transportation: An Interpretive Survey," Journal of Economic Literature, Vol. 23 (March 1985), pp. 57-94.

World Bank, Urban Transport (Washington, 1986).

Zahavi, Y., "Travel Characteristics in Cities of Developing and Developed Countries," World Bank Staff Working Paper 230 (Washington: World Bank, 1976).