CHARGING FOR ROADS

David M. Newbery

The industrial market economies collect about 6 percent of total government—revenues from taxes and charges on road users. World Bank data suggest that road taxes\(^1\) are an even bigger share of government revenue in developing countries. The (unweighted) average figure for a sample of twenty-four countries was 11 percent in 1982–84 (Newbery 1987a, pp. 1 and 2). In both groups of countries, taxes on motor fuel accounted for about half of the total.

These averages conceal wide variations among countries. Some countries collect more than twice what they spend on roads; some appear to balance revenue and expenditure; others collect less than they spend. Which is correct? Should road users contribute to general revenue over and above road expenditures, or is the highway system best considered a necessary infrastructure benefiting all, to be financed out of general tax revenue? What about the interest on the capital stock, which is usually ignored in calculating the level of road expenditure?

The structure of road taxes varies even more widely than the level of such taxes from country to country. At one extreme, Denmark and Japan charge only three to four times as much per year for a sixteen-ton truck as for a medium-size automobile; others, including the Federal Republic of Germany, New Zealand, Switzerland, and the United Kingdom, charge more than ten times as much. The reasons for such differences include wide variations in the purchase tax on automobiles and trucks, in the taxes on diesel fuel and gasoline, and in license fees. In some countries diesel is taxed as heavily as gasoline; in others diesel is subsidized while gasoline is heavily taxed. The relative importance of taxes on use (such as fuel taxes) and taxes on access (such as license fees and purchase taxes) has often varied as much over time within a country as it has across countries (see Tait and Morgan 1980).

The appropriate structure of road taxes is at issue in several coun-
tries. The U.S. Federal Highway Authority published its Highway Cost Allocation Study in 1982. New Zealand recently introduced a sophisticated system of distance-weight charges (Starkie 1984), and several states in the United States have introduced ton-mileage taxes. Singapore has had an area licensing system since 1979 (World Bank 1986), and Hong Kong has recently completed successful experimental trials of electronic road pricing (Dawson and Catling 1986).

The problem of designing road taxes can be broken down into various subproblems. First, what is the marginal social cost (that is, the extra cost to society) of allowing a particular vehicle to make a particular trip? Part will be the direct cost of using the vehicle (fuel, wear and tear, driver's time, and so forth) and will be paid for by the owner. This is the private cost of road use. Other costs are social: some will be borne by other road users (delays, for example); some by the highway authority (extra road maintenance); and some by the society at large (pollution and risk of accidents). These are called the road use cost—the social costs (excluding the private costs) arising from vehicles using roads. It seems logical to attempt to charge vehicles for these road use costs, so as to discourage them from making journeys where the benefits are less than the total social costs (private costs plus road use costs). The first task, therefore, is to measure these road use costs.

The second question is whether road users should pay additional taxes above these road use costs. One argument is that road users should pay the whole cost of the highway system, not just the extra cost of road use, either to be “fair” in an absolute sense or to achieve parity or equity with, say, rail users who must cover the total costs of the railway system. Another argument is that the government needs to raise revenues and some part of this revenue should be collected from road users, since to exempt them would be to give them an unreasonable advantage over the rest of the population. Both arguments appeal either to the desire for equity or fairness, or to the need for efficiency in the allocation of resources (road versus rail), or both.

The first serious study of the problem of designing a set of road user charges was done for the World Bank by Walters (1968). Since then, the Bank, in collaboration with the Brazilian government and the United Nations Conference on Trade and Development (UNCTAD) has completed a study of roads in developing countries. The theory of public finance has also been transformed, so we are now well placed to reexamine both the determination of road use costs and the appropriate level of additional road taxes.

Relevant Principles of 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. Broadly, the conditions for this result, set out formally in Diamond and Mirrlees (1971), are (a) that production efficiency is feasible and (b) that any resulting private 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.

The theory has immediate implications for road taxes. Road users can be divided into two groups: those who transport freight, which is an intermediate service used in production, and those who drive their own cars or transport passengers, who enjoy final consumption.

Within that framework freight transport should pay the road use costs to correct externalities and to pay for the marginal costs of maintenance, and additional taxes on passenger transport can be set, using the same principles that guide the design of other indirect taxes. We shall show below that one would expect a close relationship between road use costs and total road expenditures, though in most developing countries road use costs are likely to fall short of highway expenditures. There is no logical reason to attribute the taxation of passenger transport to the highway budget, since it is a component of general tax revenue. But if all road taxes and charges are taken together, there are good reasons to expect that they will exceed total highway expenditure. In short, in a well-run country no conflict need arise between the goals of designing an equitable and efficient system of road use charges and taxes and the desire to cover the highway system's costs.

These distinctions suggest the following terminology. The efficient road user charge is the amount road users should pay to ensure efficiency; it will be equal to the road use cost. The pure tax element is the amount by which road taxes exceed the road use cost. Under the conditions described earlier, freight transport should pay the efficient road user charge; passenger transport may be subject to additional pure taxation.

The theory provides a useful framework for the study of road user charges. The first step is to identify the road use costs. The second is to see what methods are available for levying charges and how finely they can be adjusted to match these costs. The third step is to examine how far these methods have repercussions outside the transport sector and, where these occur, how to take them into account. At the same time, one should ask whether the economy satisfied the conditions of the Diamond-Mirrlees theorem—that is, whether production efficiency is feasible. If not, then it may be desirable to change transport taxes to improve the overall efficiency of the economy. If, for example, rail charges are too high and cannot be lowered, then it may be right to raise road taxes in order to improve the allocation of freight (or passengers) between the two systems.
These three steps will suffice for freight transport. For passenger transport, one other step is needed: to determine the appropriate level (and method of levying) the pure tax element.

The Derivation of Road Use Costs

Vehicles impose four main costs on the rest of society—accident externalities, environmental pollution, road damage, and congestion.

Accident externalities arise because extra vehicles on the road increase the probability that other road users will be involved in an accident. Even when insurance companies pay the full cost of such accidents, in cases where blame is joint each participant pays only for his own accident and, hence, fails to allow for the increased risk to others. Although accident externalities are potentially large (see Newbery 1987b for estimates for the United Kingdom), in the present state of knowledge it is almost impossible to estimate them. According to 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).

The main difficulty is that, as traffic increases and driving becomes more hazardous, drivers take more care and the authorities introduce more safety features, with the paradoxical result that accident rates per kilometer appear to be falling over time. Accident costs are, therefore, ignored in this article.

Environmental pollution by noise and exhaust emissions may be important in some cases, but estimates of its magnitude are not readily available. Where they have been quantified (for example, in the United States), they appear to contribute less than 10 percent to road use costs. They are likely to be closely correlated with the quantitatively more important congestion costs and can be subsumed into these costs, so they will not be discussed further here.

Road Damage Costs

Road damage falls into two types: pavement costs, which cover repairing the road and are borne by the highway authority, and the road damage externality, which is the increased operating cost for subsequent vehicles of driving on the rougher road.

The damaging power of a vehicle is measured by the number of equivalent standard axles (ESAs), where one ESA is the damaging power of an 18,000-pound single axle. Damage increases at approximately
the fourth power of the axle load—which means, in practice, that almost all road damage is caused by heavy vehicles. A truck may do 10,000 times the amount of damage of an automobile.

Newbery (1988) has shown that, if the road network has a uniform age distribution and roads are overlaid or restored when their roughness reaches a predetermined level ("minimum tolerable standard," in the terminology of U.S. highway engineers), then road damage externalities, when averaged for roads of different ages, are identically zero in an important special case and negligible in all reasonable cases. The special case has no 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 the volume of traffic. In the special case, what might be called the fundamental theorem of road damage states that the efficient charge is the road damage cost exactly equal to the average cost of maintenance per ESA kilometer.

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 sample case, $R$ is a function of cumulative ESAs since the 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 state, $R_0$. The "age" of the road can be measured by cumulative ESAs. Imagine a road between two points but of uniform age distribution. If its average lifetime is $N$ ESAs before overlay, then a fraction $m/N$ will have an age of $m$ or less (see figure 1). Initially, suppose that the youngest surface is at the start of the road, and the oldest, just requiring overlay, is at the end. Each year, if annual traffic is $n$ ESAs, a fraction of $n/N$ will be overlaid at a cost of $C$ per kilometer, so that the total annual cost is $Cn/N$ per kilometer, or $C/N$ per ESA kilometer. As time passes, the "age" of the surface at each distance will change, but the age distribution of the surface will remain unchanged. Variations in the annual flow of traffic will alter the rate at which the age of a particular piece of road changes but not the age distribution of the whole road. The cost of traveling on 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. This result does not require an optimally set maintenance policy, only a consistent policy in which \((R_o, \bar{R})\) are predetermined and consistently applied.

Another way to understand this result is to examine the time path of vehicle operating costs (and roughness) shown in figure 2. The effect of an extra ESA is to raise subsequent operating costs by the vertically shaded amount, to advance the date at which roughness reaches the critical level \(\bar{R}\) and the road is resurfaced, and to lower subsequent vehicle operating costs as a result by the amount of the horizontally shaded area. Averaged over roads of all ages, these two areas are identical when expressed in present discounted value.

In practice, roads deteriorate not just because of traffic; weather and other environmental factors are also important. Since we assume that roads are resurfaced once they reach a predetermined standard of roughness, the effect of weather is therefore to reduce the proportion of surface damage that is attributable to traffic. It is no longer possible to appeal to the general arguments of the fundamental theorem to determine the road damage cost. What is needed is an estimate of the relationships between road damage (measured by roughness) and traffic. Fortunately, this relationship has been estimated as part of the World Bank's Highway Design and Maintenance Research Project, and the results described in Paterson (1985). His formulas make it possible to calculate the proportion of total road damage attributable to traffic, which depends on the severity of the climate (through the durability of the surface and its sensitivity to moisture and temperature variations) and on the stringency of the maintenance standards (in terms of the roughness range, \(R_o\) to \(\bar{R}\)); in the general case it depends also on the level of traffic loading in relation to the road's strength. In arid subtropical climates the proportion of damage attributable to traffic ranges typically between 60 percent and 80 percent; the same goes for humid subtropical or nonfreezing temperate climates; and in freezing temperate climates, between 20 percent and 60 percent (stringent maintenance standards or low traffic levels giving the lower values).

The formula for the fraction of maintenance costs attributable to traffic, \(\mu\), is

\[
\mu = (1 + \phi)^{-1}, \quad \phi = mT/(1 - R_o e^{mT/\bar{R}})
\]

where \(T\) is the interval in years between overlays (typically ten to
twenty-five years), and $m$ is the rate of increase of roughness in the absence of traffic (about 0.01 for arid tropical areas, 0.025 for humid subtropical areas, and as high as 0.05 for temperate freezing climates). As a rough guide, $T$ could be taken as fifteen to twenty years and $R/R_o$ as three years on heavily used, well-maintained roads; and twenty to twenty-five years and four years, respectively, on little-used or less well-maintained roads.

For the empirical application of the research project, Tunisia was chosen; it has rapid traffic growth and sizable damage from the weather. Charging for road damages on lightly used roads lasting twenty years would recover an estimated 55 percent of normal maintenance costs; on more heavily used roads lasting fifteen years, some 65–75 percent of costs would be recovered. For the country as a whole, allowing for the greater incidence of heavy traffic, an average figure of two-thirds of maintenance costs recovered from road use charges seems reasonable (Newbery 1986b).

The same arguments apply to paved and unpaved roads. But, whereas it is normal for paved roads to be maintained or resurfaced according to their degree of roughness, 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 it is no longer true that road damage cost is some stable fraction of maintenance cost; road damage externalities may be appreciable. If, however, the period between maintenance is optimally chosen, then the earlier fundamental result holds—this time as a consequence of an envelope theorem (Newbery 1988).

A special case 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 costs of road damage may be even greater in such cases as extra vehicles increase the costs of repair when the road is eventually repaired (Newbery 1986c).

The insights offered by the fundamental theorem are of considerable practical value. They are robust as to the exact relationship between road damage, roughness, and ESAS 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 (1986a) was able to suggest that estimates of road damage costs on U.S. rural interstate highways were too high. The reason appears to be that calculations were done for a representative road, which was assumed to be seven years away from its next resurfacing. Extra traffic was assumed to raise vehicle operating costs over the next several years.
but no credit was given for the fall in these costs after resurfacing. Thus, in figure 2, the extra operating cost (shaded vertically) has been counted, but not the benefits (shaded horizontally).

The main limitation of the theorem is that it does not apply if the maintenance policy is not geared to the condition of roads. If roads are repaired only when funds are available or according to a fixed schedule, then road damage externalities may be appreciable. Offsetting this factor, traffic damage will have no effect on the timing of maintenance expenditure, though it may have a considerable effect on the amount spent, especially if reconstruction is required. In such cases road damage costs may be appreciably higher than suggested by the fundamental theorem and will need to be explicitly calculated as in Newbery (1986c).

**Congestion Costs**

In the case of road damage costs, researchers initially estimated the relationship between traffic and road damage and then identified a special theoretical case in which costs would be estimated without knowing the form of the relationship. For congestion costs, the same sequence happened. The early work, dating from Walters (1961), attempted to quantify congestion externalities and estimate the optimal congestion tolls, using relationships estimated by traffic engineers. The costs appeared impressively large for urban streets, which stimulated a great deal of subsequent research, surveyed by Winston (1985).

Walters had been chiefly 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. 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 cover the costs of providing the optimal amount of roads. This theorem 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 came from a model in which roads were infinitely durable (and continuously adjustable to capacity). But roads deteriorate under traffic pressure, and although congestion is an increasing function of traffic (or the ratio of passenger car units (PCU) to capacity), the wear on the roads depends on cumulative ESAs. Thus, maintenance costs depend on cumulative ESAs 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 ESA miles per vehicle),
including the externalities, which involves calculating congestion costs. This raises some questions. If the road network is to be optimally designed, how should the costs of strengthening surfaces to take heavier vehicles 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 the volume of traffic did not grow, then charging the road damage cost would exactly recover the maintenance costs. By analogy, it is tempting to conclude that optimizing the highway capacity will lead to additional congestion charges, which will recover the capital costs, leaving the road budget exactly balanced. But is it correct to allocate all the capital costs to congestion charges on the basis of PCUs when much of the capital cost is required to strengthen the road to take heavy trucks? It is common to allocate the minimal expenditure needed for a road suitable for automobiles on a PCU basis and the balance, needed for heavier vehicles, on an ESA basis. But is this correct, especially given the increasing returns from strengthening the surface?

These questions are resolved in Newbery (1987b, c). If there are constant returns to highway expansion, if all lanes are built to the same strength and repaired at the same time, if all heavy trucks are confined to the slow lanes, and if the road capacity is optimally adjusted to the traffic flow—if all these conditions are met—then the optimal congestion charge will cover the total costs of the highway. These total costs include interest on capital plus all traffic-related costs, including maintenance. Heavy trucks should pay an additional charge equal to their road damage cost, so the highway budget will run at a surplus. If, however, trucks spread out and use all lanes equally as the road is widened, then the optimal congestion charge will recover the overhead costs—that is, the interest on capital and all running costs except the allocable fraction of road damage costs. This latter should be collected by charges per ESA kilometer. The difference between these two results arises because, if trucks use the whole road width equally, there will be fewer ESA kilometers per year on each lane, so maintenance costs will be lowered. If trucks remain in the same lane, the cost of widening the road includes having to incur the same maintenance on a wider road. This will be justified only if the benefits of reducing congestion are large enough to cover the maintenance costs as well, so they should be included in the congestion charge.

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) which are easier to obtain than observations on traffic flows on different roads. Though the theoretical approach remains a useful guide for rough orders of magnitude, how-
ever, it does have three problems. The first is that road capacity 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 large economies of scale from expanding rural roads up to four lanes (but not much thereafter); whereas expansion costs are often very high in urban areas, where there may be decreasing returns to spending more. Economies of scale are important, because they reduce the fraction of revenue to costs collected by optimal tolls. For example, if doubling the investment (the number of lanes) quadruples the capacity (vehicles per hour), then the optimal toll would collect only half the average cost. The third, and most serious, objection to the theoretical approach is that there is no guarantee that roads have been optimally adjusted, short of measuring the congestion costs directly and comparing them with expansion costs. Since the theorems suggest that congestion costs are likely to be larger than road damage costs, their measurement is critical for devising efficient road user charges.

The measurement of congestion costs is fraught with conceptual and empirical difficulties. The rationale for charging vehicles for congestion is to cause their owners to weigh the benefits of driving against the total (social) costs of a trip. If each journey could be separately charged (like metered phone calls), then the right price would be the marginal congestion cost caused by the journey to all other traffic, after the traffic patterns had adjusted to the set of congestion tolls. If more drivers use a particular road at a certain time, some of the other drivers will take different routes or choose a different time of the day or decide not to make the journey. These responses would mitigate the increase in congestion, thus lowering the optimal toll. If it is not possible to meter each journey, the congestion charge will work much more bluntly. Higher congestion charges will reduce traffic and congestion on average, but traffic will be undercharged on congested streets and overcharged on uncongested roads.

In both cases information is needed on how the time taken for journeys changes when an extra vehicle makes a trip. This will be the marginal time cost ($MTC$) of the trip, usually measured in vehicle hours per vehicle kilometer. When $MTC$ is multiplied by the value of time per vehicle, the result is the marginal congestion cost ($MCC$) in say, cents per vehicle kilometer. The $MTC$ is a technical relationship that may be similar on similarly congested roads in different places and at different dates. The $MCC$ is a less useful measure, since the value of time varies across countries and over time.

Most engineering studies of traffic congestion investigate the effect of extra vehicles on the traffic flow on a given road. Though useful, such a measure does not allow for all the responses listed above. Furthermore, the relationships measured appear to be highly sensitive
to road conditions (number and type of intersections, frequency of turns against the traffic flow, and so forth), and the MTC derived are sensitive to the functional forms fitted to the data (Newbery 1986d, 1987a).

Fortunately, some highway studies do attempt to measure the relationship between average traffic speeds and volumes, and so come closer to the measure of MTC needed for estimating congestion charges. A good recent example is provided by Harrison and others (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 ±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" (p. 141). The average speed and flow for an area were found by averaging over the links in the area.

Another method is to simulate traffic flows through a network using a model of the delays at traffic signals. Such models are intended mainly to optimize the timing of traffic signals in a network, but they can be used to simulate the effect of extra traffic to links in that network. Dewees (1979) reports the results of applying such a model to Toronto. Harrison and others (1986) tested their predictions against the results from a TRANSYT program—the detailed analysis of junction delays described in Vincent and others (1980). The Harrison study concluded that the TRANSYT results were too sensitive to flow changes, because they allowed no rerouting of the traffic—a weakness of every single link-specific traffic flow model.

The British Transport and Road Research Laboratory has attempted to measure average traffic speeds by "floating car" methods (in which a car remains in the traffic stream for a period of time), estimating the average speed of the stream over some distance. The most useful of these (U.K. Department of Transport 1978) studied thirteen towns and cities at both peak and offpeak hours; the results can be compared with those of similar earlier studies. Subsequent analysis by Duncan and others (1980) found a reasonably stable and precisely identified linear relationship between average vehicle speed, \( v \) kilometers per hour, and traffic flow, \( q \), measured in PCU per lane hours:

\[
    v = a - \beta q.
\]

This relationship gives an MTC of \( \beta q/v^2 \), so observations of \( v \) and \( q \) and a knowledge of \( \beta \) allow one to estimate the MTC. The estimated
figures for $\beta$ was 0.035, similar to the figures for Hong Kong in Harrison and others (1980).

All these studies estimate the short-run response of speed to flow. The long-run response, after drivers have adjusted to the new traffic and charging system, can be deduced from the short-run MTC if the elasticity of trip demand (that is, the response of traffic to the private costs of travel) can be estimated. Although it is hard to measure these demand elasticities at all accurately, the errors involved may not be too serious. If a congestion charge can be finely adjusted to traffic flows (as discussed below), then the traffic response can be monitored and the charges adjusted. If only crude methods of charging are available, then the resulting average congestion charges will be relatively insensitive to the elasticities.

Charging for Road Damage and Congestion

Ideally, vehicles should be charged for the road use cost of each trip. Road damage costs can be calculated once the load is known and the choice of route decided: in New Zealand, trucks pay for licenses to take a given load over a given distance. This seems unnecessarily precise because the damage done by a particular vehicle does not vary much by type of road. Assuming that legal weight limits are enforced to a predictable degree, the average ESA of a truck is readily estimated, and then simple distance charges would do. The Tunisian experience, discussed below, suggests that ESA kilometers and ton kilometers correlate reasonably well, and some states in the United States impose charges on the basis of ton miles. A combination of vehicle-specific and distance-related charges (on fuel, tires, parts, and vehicle purchase) can also be fairly closely related to road damage costs.

The real difficulty lies in charging for congestion. There are huge differences between congestion costs in urban and rural areas and between peak and offpeak times. Electronic metering is technically feasible in certain circumstances, such as a highly congested city state like Hong Kong (see Dawson and Catling 1986), but for most countries it is unlikely to be practical. Traffic flow can be speeded up considerably by using parking charges, restricting access, and developing public transport (World Bank 1986). But even with the best traffic management schemes, congestion will remain. The challenge is to design a system of charges that matches the congestion.

The crudest system is to charge for distance driven and for access. Distance charges (like fuel taxes) are blunt instruments, because they barely discriminate between congested and uncongested roads (and indeed discriminate against driving on uncongested roads, being a larger fraction of total vehicle operating costs on such roads). Access charges have more potential, especially if they are area-specific, as in
Singapore. They bear more heavily on urban vehicle owners, because urban journeys are typically shorter and alternative forms of transport more readily available. There are good reasons to make license charges depend on the address of the owner (as revealed for insurance purposes), but even if all vehicles of a given type pay the same access charge, the ability to charge for access and distance is some help in covering congestion costs.

Road damage costs can vary across different vehicles by a factor of between 100 to 1 and 10,000 to 1 (depending mainly on the proportion of vehicle kilometers traveled on unpaved roads, for which the variation is quite low). Congestion costs vary relatively little across different types of vehicles, and typically are more than half of all road use costs. Together they make up road use costs that vary across vehicle types by a factor of between 5 to 1 and 20 to 1. This range is comparable to the variation in fuel consumption and other input costs—which suggests that input taxes may be a reasonable way of levying road user charges (at least if electronic pricing is ruled out).

Table 1 illustrates the position for commercial vehicles in Tunisia. The variation in road damage costs is about 200 to 1, but in total road use costs it is 5 to 1. Column 5 shows the tax rates needed on diesel fuel to charge the road use costs. They vary across vehicles by a factor of 2.8 to 1, less than the variation of road use costs. Fuel taxes are thus a natural choice for road user charges. Column 6 shows the effect of taxing diesel at six U.S. cents per liter (at 1983 prices), charging a purchase tax of 10 percent on tires and parts, and collecting the rest of the road use costs by a special purchase tax for vehicles. Taxing tires is in principle a good way of charging for distance and weight, but too high a tax will encourage excessive and possibly dangerous use. Purchase taxes (nonrebateable) on vehicles and spares are a good method of charging trucks, because 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. A special purchase tax for vehicles would fall more heavily on vehicles that are not used very much (since the interest costs on the tax would be higher for longer-lived vehicles) and might restrict entry somewhat (though truck leasing would reduce the force of this objection).

**Fuel Taxes**

Taxes on diesel fuel are a potentially attractive way of charging trucks for using roads. Gasoline taxes are a good way of charging private cars for congestion, though they discriminate inefficiently in favor of small cars. It is doubtful that this discrimination is warranted on distributional grounds in industrial countries; not so in developing
Table 1. Illustrative Road User Charges for Commercial Transport in Tunisia, 1983

<table>
<thead>
<tr>
<th>Vehicle class</th>
<th>Road use costs (U.S. cents per kilometer)</th>
<th>Diesel tax rate (U.S. cents per liter)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Tax scheme A purchase tax rate (percent)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Tax scheme B&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Purchase tax rate (percent)</th>
<th>Balance (U.S. dollars per year)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Road damage (1)</td>
<td>Nonurban congestion (2)</td>
<td>Urban congestion (3)</td>
<td>Total (4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Utility (pickup)</td>
<td>0.03</td>
<td>0.15</td>
<td>1.46</td>
<td>1.64</td>
<td>18.22</td>
<td>26</td>
</tr>
<tr>
<td>Truck class</td>
<td>Light</td>
<td>0.12</td>
<td>0.21</td>
<td>1.94</td>
<td>2.27</td>
<td>15.13</td>
</tr>
<tr>
<td>Truck class</td>
<td>Medium</td>
<td>0.51</td>
<td>0.33</td>
<td>0.44</td>
<td>1.28</td>
<td>6.40</td>
</tr>
<tr>
<td>Truck class</td>
<td>Heavy single</td>
<td>2.18</td>
<td>0.39</td>
<td>0.52</td>
<td>3.09</td>
<td>7.92</td>
</tr>
<tr>
<td>Truck class</td>
<td>Heavy tandem</td>
<td>4.46</td>
<td>0.39</td>
<td>0.52</td>
<td>5.37</td>
<td>12.07</td>
</tr>
<tr>
<td>Truck class</td>
<td>Articulated</td>
<td>5.64</td>
<td>0.41</td>
<td>0.44</td>
<td>6.49</td>
<td>13.33</td>
</tr>
</tbody>
</table>

a. The rate needed to fully cover road use costs through a tax on diesel.
b. Tax scheme A stipulates six cents per liter diesel tax, 10 percent tire tax, and 10 percent parts tax; purchase tax rate is rate needed to equate charge to cost.
c. Tax scheme B stipulates three cents per liter diesel tax, 20 percent tire tax, 20 percent parts tax, and purchase tax as set; balance is amount needed to equate charge to cost (possible license fee).

Source: Newbery and others 1986.
countries, which have fewer effective ways of taxing income. Hence, the rate of a gasoline tax can be set without regard for its impact elsewhere, except insofar as high gasoline taxes encourage the use of diesel-fueled substitutes (and this tendency can be offset by higher license fees on diesel-powered alternatives).

Diesel and its close substitutes—kerosene, gas oil, and other middle distillates—are used extensively outside transport, both as intermediate inputs and as final consumption goods. This need cause no concern, if diesel fuel for road use can be differentially taxed (as in Germany and the United Kingdom, for example) and the illegal use of untaxed substitutes in vehicles effectively prevented. But in most developing countries, this is not feasible. Before adopting a road user charge on diesel, governments should explore its impact on the rest of the economy. The impact depends on four factors: the extent to which middle distillates are used in production rather than transport; the degree of substitutability in production and consumption; the amount of kerosene used by the poor; and the structure of the tax system.

Hughes (1986a) has studied the impact of raising the price of diesel (and its close substitutes) in Tunisia, where 60 percent of such fuel is used outside the transport sector. (This required an input-output table with flexible coefficients, a proper model of tax shifting, and a survey of consumer budgets to examine the resulting impact on welfare). The tax structure, together with the demand responses, determine the impact on government revenue and allows one to calculate approximate measures of deadweight loss. (The deadweight loss of a tax is the amount by which the loss caused to the taxpayers exceeds the revenue gain to the government; it is a measure of the tax's inefficiency.) Values for the substitution elasticities in production and consumption were taken from empirical studies elsewhere, as there are few studies available for developing countries. They give only a feel for the importance of allowing for substitution responses, but are nevertheless of considerable interest.

Hughes found that the long-run derived demand elasticity for diesel was quite high, so that diesel taxes would lead to quite large deadweight losses. For instance, imposing a United Kingdom level of tax on diesel leads to a deadweight loss of more than 50 percent of the revenue collected, so would be highly undesirable compared with other more broadly based taxes. The distributional impact of the tax was also adverse, because kerosene is used heavily by the (mainly rural) poor in Tunisia, as in many developing countries. (But subsidizing kerosene for distributional reasons and taxing diesel usually leads to massive adulteration of diesel, as the Indian evidence shows. Motor vehicles can run unharmed on a fuel mixture of 30 percent kerosene and can tolerate even higher levels.)
The other result was that the rest of the tax system modifies the impact of raising the tax on diesel, as one would expect. Revenue from gasoline taxes increases, though there are many less obvious revenue effects that depend sensitively on economic and tax structures.

The conclusion is that heavy diesel taxes are likely to be undesirable in developing countries, so license fees and purchase taxes on vehicles are important for charging commercial vehicles for using the roads. 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.

To quantify this conclusion, the last two columns of table 1 show the effect of halving the fuel tax rate to three U.S. cents per liter, doubling the taxes on tires and parts to 20 percent, and adjusting purchase tax rates to leave room for an annual license fee. (Such a fee offsets the tendency of purchase taxes to encourage owners to keep their vehicles for too long.) The structure of charges is matched fairly well to the road use costs, except for medium-size trucks, which would be overtaxed by any reasonable system of license fees and purchase taxes. Pickup trucks are also overtaxed, but this is desirable for reasons discussed in the next section.

**Personal Transport Taxes**

Because personal transport is a final consumption good, it is a legitimate object for indirect taxation, over and above the collection of road use costs. Budget studies reveal that spending on gasoline, car purchase, and maintenance is among the most income-elastic expenditures in developing countries. It is thus an attractive proposition for countries where income tax is limited in coverage and effectiveness (Deaton 1987). Vehicle and gasoline taxes are easy to administer, and high rates of tax can readily be justified on distributional grounds. However, heavy taxes on gasoline, together with light taxes on diesel, will encourage people to buy diesel-engined alternatives. This can be mitigated by heavy license fees on diesel-powered private cars (at a level such that purchasers would choose the same vehicle as they would in the absence of all fuel taxes and extra license fees). And purchase taxes and license fees should be made a function of the value of the vehicle, rather than of its power or cubic capacity. The main snag will be that high automobile taxes encourage people to buy pickups, which are also used for commercial purposes (and for which they should be subject to lower taxation). This is probably a serious problem in several developing countries such as the Philippines, Thailand, and Tunisia. It might be possible to reduce this substitution by taxing pickups with more
than, say, two seats at automobile rates. If this proves difficult, there
will be less scope for taxing private cars heavily without also causing
potentially costly distortions. Nevertheless, European levels of gaso-
line taxes are quite easy to justify.

The Impact of Road User Charges on the Rest of the Economy

Once a set of charges and taxes has been designed, a government
will need to answer two more questions: 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 (1986b) can be employed to
answer these questions, though the questions themselves have to be
carefully formulated. On the reform of road user charges, 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,
deciding how to boost it is a separate issue, not one that should be
prejudged by raising road taxes.

The results are reassuring and appear to be robust, as they come
from three developing countries (Indonesia, Thailand, and Tunisia).
On balance, raising transport taxes and lowering sales taxes or import
duties lowers the price level, because transport taxes are partly shifted
back to factor incomes, whereas sales taxes are shifted forward onto
final consumers. Raising the cost of the freight of an exported good
(such as rice in Thailand) will not raise the world price, but it will
lower the farm gate or ex-factory price and hence reduce farm wages
or land rents (or both).

The effect of higher road user charges on income distribution de-
pends on which tax is increased. Gasoline taxes are quite progressive,
diesel plus kerosene taxes (at the same rate) are somewhat regressive,
and taxes on transport are virtually neutral. Overall, the effects are
small (even for quite large tax changes), but very "noisy." Although
the average effect may be small, some households may suffer a lot
while others may benefit; these impacts, however, are poorly correlat-
ed with income. Politically, this poor correlation is a drawback: the
losers will presumably be more vociferous than the gainers. The cost
of the reform may therefore appear higher than it really is. Nonethe-
less, fluctuations in the world price of fuels dwarf the tax changes that
are likely to be desirable (particularly given the arguments for rather
low taxes on diesel fuel). The 1986 falls in oil prices have made it
possible to raise tax rates without raising domestic prices. The evi-
dence from many countries is that quite large increases in fuel prices
are possible. It is also important to realize that road use costs are a
modest fraction of vehicle operating costs.
Conclusions

Recent theoretical advances have clarified the nature and measurement of road damage costs and externalities. It now appears that road damage externalities are negligible and that charging for road damage costs will recover between one-half and three-quarters of the costs of road maintenance, almost entirely from heavy vehicles. On the measurement of congestion costs, the theoretical and empirical state of the art is less advanced, but if roads have constant expansion costs per unit of capacity and are optimally adjusted, then congestion charges will recover capital costs, other current overheads, and a large part 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 full coverage rather unlikely.

In the absence of electronic pricing, road user charges will have to cover both road damage costs and congestion costs. They cannot do this perfectly, but it is not difficult to devise a satisfactory system of input taxes, purchase taxes, and license fees. If fuels for transport use cannot be differentially taxed, then diesel is not a 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, a reasonable compromise for trucks is purchase taxes on vehicles and parts, together with taxes on tires, and the balance recovered from license fees. 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 while lowering other indirect taxes is mildly counterinflationary and, with the exception of diesel taxes, has little effect on income distribution. Taxes on private cars are quite progressive.

Abstract

The article discusses two theoretical methods of measuring road use costs and designing a system of road user charges. The first states that road damage externalities are zero and road damage costs are equal to the traffic-related fraction of maintenance expenditure. The second states that, with constant returns and optimal road capacity, congestion charges should recover the remaining total overhead costs. Vehicles should be charged these costs, and additional pure taxes on passenger vehicles should be guided by principles of indirect taxation. Although road user charges alone may fail to cover the total highway budget, the additional pure taxation is likely to more than cover the shortfall. The article argues that an appropriate system of taxes and charges can be devised to meet these requirements without adversely effecting the rest of the economy.

Notes

This article is based on a World Bank research project, “Pricing and Taxing Transport Fuels in Developing Countries” (RPO 672-38), reported more fully in Newbery and others (1986), and on subsequent research supported by the International Monetary Fund while I was a visiting scholar in the Fiscal Affairs Department, reported in Newbery (1987a). I am indebted to my research collaborators, Esra Bennathan, Gordon Hughes,
and William Paterson, and to Vito Tanzi of the International Monetary Fund for his support.

1. Road taxes are to be interpreted broadly to include charges such as license fees. The distinction between road user charges and the pure taxation of road users is discussed below.

2. For a recent study of the extent to which the theory is relevant to developing countries, see Newbery and Stern 1987.

3. Commuting to work is traditionally treated as consumption, not an input into production, on a par (and substitutable) with housing expenditures.

4. Across paved roads, the variation may be 10 to 1; by its nature a relatively small fraction of ESA kilometers occur on the weak, low-volume roads for which the damage costs are highest.

References


Harrison, W. J., and others. 1986. “Some Advances in Model Design Developed for the Practical Assessment of Road Pricing in Hong Kong.” Transportation Research 20A (March): 135–44.


——. 1986c. “Road Damage Costs in Conditions of Road Crisis.” International Monetary Fund, Fiscal Affairs Department, Washington, D.C.


