DISCUSSION PAPERS

ROAD USER CHARGES, COST RECOVERY AND ROAD-RAIL COMPETITION

P.J. Forsyth

Discussion Paper No. 125

July 1985

G.P.O. Box 4, Canberra 2601, Australia
The Centre for Economic Policy Research was established in 1980 as one of a number of University initiatives. It was provided with an annual grant for a period of five years and given a mandate to foster policy oriented studies of the Australian economy. The Centre intends to work closely with other economic research groups - both within the Australian National University and in other Australian universities.

The Discussion Papers of the Centre are intended to make available to a wider audience a series of economic research studies. These studies will have been either commissioned by the Centre (for instance in conjunction with a conference held under the Centre's auspices) or undertaken by its research workers.

Discussion Papers will also be published where economic research work with relevance to policy has been undertaken by individual academics within the Australian National University - or where the Centre is able to act as a focal point for such research.

The Centre will not have any views on policy; individuals will.

---

ROAD USER CHARGES, COST RECOVERY AND ROAD-RAIL COMPETITION

P.J. Forsyth
Department of Economics
Research School of Social Sciences
Australian National University

DISCUSSION PAPER NO. 125
July 1985

ISBN: 0 949293 33 4
ISBN: 0725 430X

Revised version of a paper presented at the 14th Conference of Economists, University of New South Wales, May 1985. The author is grateful to the following people for comments on an earlier draft: Colin Cronin, Howard Dick, George Doohra, Christopher Findlay, Fred Gruen, Ted Kolten, Phil Laird, Gordon Mills, Tony Shaw, David Starkie, Margaret Starrs, John Taplin and Graham Teale.
Abstract

The question of how road users should be charged, to achieve economy efficiency, is examined. Several subsidiary questions are considered. The first concerns the measurement of marginal cost. It is concluded that many discussions in Australia underestimate marginal costs because they ignore congestion costs. A consequence of this is that the cost recovery problem is exaggerated. A second question concerns the importance of road-rail competition in setting road charges. To solve this, it is necessary to model railway behaviour - alternative plausible models of railway behaviour give widely divergent results. Current presumptions about the relative size of cross and own price elasticities, which are based on incomplete reasoning, overestimate the importance of the road-rail issue. Another question is whether revenues are equal to total measured cost, whether they determine current total expenditure (including investment), or whether current expenditure determines revenue. The costs of imposing revenue constraints depend on this. The fourth and fifth questions involve he charging structure - on different users and on different aspects of vehicle use. While Ramsey/Boiteux inverse elasticity rules are appropriate, current information is insufficient to allow their application with any confidence. A policy which is likely to minimise expected distortion cost is also a simple one. Users should be levied charges in proportion to marginal cost, and a mix of road charges (on registration and fuel) should be levied.
1. Introduction

The problem addressed in this paper is that of setting taxes or charges for road use by freight vehicles. The answer presupposes that the charging problem for other road users is being solved jointly. Thus the structure of charges for all users needs to be considered, though attention here will be focussed upon freight vehicles. The road charging solution for freight may have implications for more than just trucks - for example, the allocation of traffic between road and rail may be affected. The primary focus is on long distance transport.

It will be taken that economic efficiency is the primary objective. It is likely, in reality, that this objective will be compromised. Specific constraints may be imposed, such as the one that revenues from road charges cover total costs, or are equal to expenditure, capital and current, on roads over a period. The objective will then be taken to be to achieve the most efficient solution consistent with these constraints. The problem is thus one of finding a second best solution.

Five key questions can be identified. They are as follows:

1. How can marginal cost of road use be measured?
2. How should the road-rail competition issue affect road charges?
3. What revenue constraints are imposed on the road system, and what is their cost?
4. How can the costs of revenue constraints be minimised?
5. How can an efficient structure of charges, on different aspects of vehicle ownership and use, be derived?

The first is one of determining the nature of efficient charges for road use. In general, charges equal to marginal cost will be efficient. It is then necessary to determine the level of marginal cost in this context. Much recent discussion ignores the role of congestion costs, and hence it misstates the problem. Once marginal cost has been defined and measured, it is necessary to examine whether charges set at marginal cost will cover total cost - this raises the issue of returns to scale in road provision.

The second question concerns the way in which road charges affect the road-rail traffic mix. If rail charges do not reflect marginal cost, and road is a substitute for rail, does it make sense to set road charges equal to marginal cost? The answer depends on several factors: the degree to which road is a substitute for rail; the degree to which demand for freight in general depends on freight prices; and the way in which rail freight charges vary in response to road charges. It will be argued that the long term elasticity of demand for freight in general has been underestimated and that this has meant that the road-rail problem has been exaggerated.

A third, though less important question, concerns what constraints are to be imposed on road charging. It has been
suggested by many, including, most recently, the Road Freight Industry Inquiry Report (The May Report, 1984) that charges be set so as to cover current capital and current expenditures. The merits of this policy, as compared to those of imposing no revenue constraint, or of covering the costs of provision for a given year, need to be examined.

Once a revenue constraint has been set, how can it be met at minimum cost to efficiency? The outlines of a solution have been given in recent discussion, but some important issues have been overlooked. For example, rules which are appropriate for final consumption goods have been applied to road freight, an intermediate good. Adjustments need to be made which take this into account.

Finally, there is the question of the structure of charges. For the most part, road use will not be charged for directly, nor will road freight be taxed as such. The inputs into the production process will be taxed as an indirect way of charging for road use. Different charging structures may have an impact on efficiency of production; it is necessary to ask whether these are significant. If they are not, the question of the structure of charges may be unimportant.

There is a considerable literature in Australia, dating back perhaps thirty years, on several of these issues. At a policy level, road charging has often been debated. It would be difficult to claim that the results are particularly satisfactory, since many confusions remain, important aspects are ignored, and conflicting propositions are advanced. Thus, when it investigated these questions, the May Report (1984) had little in the way of a synthesis to draw upon. This, however, does not justify the strong, though questionable, conclusions which it put forward.

The object of this paper is to apply the relevant theory to clarify the questions which have been outlined above. Given the current lack of empirical information in some critical areas, it is not possible to answer all of the questions conclusively. Nevertheless, it is possible to sketch how answers might be found, and what empirical questions need to be resolved. In the meantime, there remains the problem of what to do in the face of inadequate information; solutions to these will be suggested. A long run perspective is adopted. This requires that structures of charges be devised which will be efficient over a number of years, and which do not require substantial revision at frequent intervals.

In this paper, the issue of whether roads, road freight or rail freight are provided at minimum cost is not discussed, except in a few instances. It is not suggested that this issue is unimportant — arguably it may be more important than those discussed here. The focus of attention is on allocative efficiency, not production efficiency. There may be substantial gains from increasing the production efficiency with which road and rail services are provided.
2. Efficient Pricing and the Cost Recovery Problem

2.1 Determining Marginal Costs of Road Users

Suppose that it is desired that road charges, for freight and private vehicles, be set so as to achieve economic efficiency. To what extent will this lead to a problem of revenues from road user charges falling short of the costs of providing the road system? The usual starting point for achieving economic efficiency is one of setting price equal to marginal cost. It may be thought that the marginal cost of a vehicle using a road is small, perhaps even trivial, in comparison to the overall costs of providing the road. This, however, ignores one of the major elements of the cost of road use -- congestion cost. In fact, it is quite likely that charges that are equal to marginal cost (including congestion costs) for all users will cover, or nearly cover, the total costs of providing the road system. There be no contradiction between efficiency and cost recovery.

Roads involve construction, maintenance, and congestion costs. These last may be paid directly by the road user, but they are no less real than the others. Congestion will slow travel, and, for leisure travellers as well as trucks carrying goods, time has a value. Congestion may also add to vehicle operating costs. The role of congestion costs in road pricing has long been recognised (for good discussion, see Walters, 1968). Congestion is most noticeable in urban areas, though it is present on rural roads. There is a trade-off between road capacity, which has a cost, and congestion, which can be alleviated by additions to capacity. Efficiency requires that the additional cost of capacity be equal to the value of congestion reductions enabled by that additional capacity.

A little reflection should confirm that congestion is a problem on major rural highways. Congestion may be understood broadly to include all negative aspects of quality which vary with the relationship of capacity to use. It would be possible to reduce rural road construction costs substantially by constructing only minimum width roads. In general, this does not happen. Where roads are busy, they are upgraded to four lane standard or better, or overtaking lanes are provided. Not all road expenditure is directly related to congestion. Other aspects of quality, such as alignment and gradient, have a cost, as does safety. However, at the margin, the road builder always has a choice of the level of capacity to provide, and hence the congestion that will be present on the road. There do exist roads on which congestion is minimal, such as little used rural or suburban roads. However, most vehicles, including freight vehicles, use congested urban roads and congested intercity highways, which account for most of the spending on roads.

Ideally, congestion charges would form part of the overall charge for road use. When an additional vehicle uses a congested road, it faces costs, in the form of delay and higher operating
costs, itself. But it also makes congestion worse for other users. The marginal cost of using the road thus exceeds the cost borne by the user. Efficiency will be achieved if a congestion toll is levied equal to the difference between overall marginal cost, and the cost faced directly by the user. The presence of congestion in a road system, or any other system for that matter, does not affect the likelihood of efficient prices covering costs. Whether they do depends on the returns to scale in the system.

2.2 Returns to Scale and Cost Recovery

If there are constant returns to scale, the optimal congestion toll will yield sufficient revenue to cover capital costs (see Mohring, 1976). Economies of scale will result in prices set equal to marginal cost yielding less than sufficient to cover capital costs. The degree of cost recovery depends on the extent of scale economies; if costs per unit fall sharply as output increases, the difference between revenue and total cost will be large.

The situation is illustrated in Fig. 1. Long run average costs (including financial and congestion costs) (LAC) are falling with output, and long run marginal costs (LMC) lie below them. For simplicity, suppose that the only variable costs are user costs, and that these increase with congestion. For a given capacity level, SMC and SAC show Short Run Marginal and Average Costs respectively. Suppose that the correct level of capacity
is constructed - i.e. the demand curve D passes through the point of intersection of short and long run marginal costs. Average user costs (shown by SAC) are automatically paid by the users - these are shown as distance A. The efficient congestion toll is shown as B. The difference between this distance, and that showing average fixed costs, C, is the difference between LAC and LMC, i.e. E. If excess capacity is provided, as would be the case if the capacity shown in Fig. 1 were provided for a demand level of D, the efficient price would be short run marginal cost, and the cost recovery problem would be greater. If capacity were too low, it is possible that the price which efficiently rations the available capacity would more than cover costs - for example, if the demand curve cuts SMC to the right of F. Thus, whether cost recovery and efficiency are compatible depends on two issues. The first concerns the shape of the long run average and marginal cost curves, and the second concerns the position on the diagram to which the actual situation corresponds (whether there too much, too little, or just the right amount of, capacity).

While there is some literature on the problem, it is not possible to resolve the empirical issue of whether there are economies of scale in the provision of roads. At certain low levels of capacity and use there may be. The presence of both cars and trucks on most roads may have several explanations. Possibly there are indivisibilities such that separate provision of roads for cars and trucks would be costly. Another possibility is that there are no indivisibilities, but economies of scope exist - this would mean that it would be costly to provide separate roads no matter how heavy the traffic (on economies of scope, see Baumol, Panzar and Willig, 1982, Ch. 4). Alternatively, there may simply be economies of scale, which would mean that aggregation of different traffics would be desirable.

Several sources suggest that scale economies exist. By doubling the width of the road, capacity (defined in terms of the number of vehicles which can be handled at a given level of congestion) will be more than doubled. If costs are proportional to width, scale economies exist. Meyer, Kain and Wohl (1965, pp.199-213) find scale economies with US urban roads. Walters (1968, pp.179-186) examines evidence for congested urban roads and uncongested rural roads which is consistent with scale economies. Mohring (1976, Ch. 12) discusses theoretical reasons for scale economies. Bureau of Transport Economics evidence (1984, pp.76-7) suggests scale economies. A recent study of upgrading roads in Victoria by Starkie (Starkie, 1982, pp.61-75) contradicts this evidence. A study which examines evidence in the U.S. concludes that constant returns to scale prevail (see Keeler and Small, 1977).

Given the conflicting evidence, it is not possible to be certain whether the Australian road system exhibits constant or increasing returns to scale. If the latter is the case, full cost recovery is not consistent with efficiency in pricing. The
shortfall in revenue which would come about with marginal cost pricing depends on the degree to which average and marginal costs fall with increases in capacity. Few suggest that scale economies, if they do exist, are particularly significant. Thus, if revenues fall short of costs, they are unlikely to do so by very much. The view, put forward in several recent discussions (see Dept. of Transport, Tasmania, 1981 and the May Committee, 1984) that setting charges equal to marginal or avoidable costs would result in substantial revenue shortfalls cannot be supported, and arises because congestion costs have been ignored.

2.3 Practical Problems Applying the Model

The second issue is even more difficult to obtain evidence on. It is possible that about the right level of capacity for the demand has been provided in Australia. For sometime now, additions to capacity have been subjected to cost benefit analysis, though these analyses have not been the sole criterion for actual investment. Road authorities probably would like to eliminate congestion entirely - that is, they would over-invest - but governments have constrained them. For some congested roads, especially in cities, it might be the case that too little capacity has been provided. This was the conclusion of Keeler and Small (1977) when investigating congestion and road provision around San Francisco.

There are a number of other practical problems which need to be recognised. A model, such as that summarised in Figure 1, can capture the essence of a problem, yet real world complications may mean that results are not so clear cut.

One of these problems is the existence of peaks in demand and congestion. The marginal costs of using roads, both rural and urban, can vary widely throughout the day. Ideally, prices should be set for each period equal to marginal costs in that period. At this stage, there is little likelihood of road user charges reflecting peak and off-peak cost differences. It would be difficult to set different prices for different periods. From an efficiency point of view, setting a charge which is an average of marginal costs in different periods may be a poor second best. The inefficiencies which arise from peak variations in congestion may be significantly greater than those which arise from covering costs when scale economies are present.

Another concerns the rural/urban road mix. Vehicles which intensively use urban roads impose greater costs per kilometre than do those which use rural roads. Since fuel economy is less in urban areas, and some taxes are levied on fuels, these vehicles would pay more, though the additional payments would most likely be much less than in proportion to costs. There may be some scope for setting different charges for different user classes, according to their expected use of urban and rural roads. (This is discussed in Section 6 below).
Thirdly, it is necessary to recognise that what is being priced is not a single stretch of road, but a road network. Some links on this network are congested, others not. Again, it would be desirable to charge for the use of specific links separately, but this is unlikely to be feasible. An average pricing system could be much inferior.

In summary, it is not possible to answer all the key questions surrounding efficient pricing and cost recovery, but it is possible to identify them. Firstly, evidence on the level of marginal cost, including congestion costs, for different road users, needs to be gathered. Much of the confusion surrounding efficient pricing arises from ignoring the integral aspect of congestion. Secondly, the returns to scale question must be answered; it is this which determines the efficiency cost of imposing cost recovery. Thirdly, it is desirable to obtain evidence on the variation in marginal costs, from peak to off-peak, urban road to rural road. While the charging structure is likely to be too limited to discriminate, this evidence is of use in designing a second best structure of charges, and in measuring the cost of simplicity.

3. The Road-Rail Competition Problem

3.1 Road-Rail Issues

Road-rail competition will present a "problem" if it is the case that pricing on one mode is inefficient, and there is a desire to set efficient prices on the substitute mode. This will be an example of the traditional second best problem in economics, where, for some reason, it is not possible to eliminate all the distortions present in related markets. In the context here, since we are interested in road pricing, it may be supposed that rail pricing is inefficient. In practice, this need not be the case, though it is certainly a possibility which should be considered. This second best problem has been considered by several (e.g. Kolsen, 1968, 1980). The usual solution advocated has been of one equal price-marginal cost ratios on road and rail.

If rail prices are set inefficiently, the obvious first best solution is to estimate, and charge, efficient prices. This is, for example, what the May Committee suggests (1984, Ch. 10). Efficient prices on rail could involve losses, though this may not be so. It is possible that there may be constraints on rail pricing, such that costs be covered (or that the deficit be no greater than a particular amount). If so, this will need to be taken into account in the road pricing formula. A further possibility is that there may be a cost recovery constraint imposed over road and rail together (the case examined in Taplin
and Waters, 1984). The solution of this would involve subsidies from road to rail or vice versa.

It will be assumed here that rail freight rates, for goods which can be carried by rail or road, are set at less than marginal cost. Rail may or may not be subject to a cost recovery/deficit constraint. The analysis can be easily adapted to the price above marginal cost case. Granted this, there are two issues which are central to the road-rail second best problem. The first concerns the response of the railways to a change in demand brought about by a change in road freight charges. The second concerns the measurement of the change in demand for rail, and road, freight in response to road freight price changes. This is both a theoretical and empirical problem, and it will be argued later that the Australian discussion of it has been misplaced.

3.2 Models of Rail Behaviour

It is best to consider the rail response to demand shifts first. Demand for rail increases as a result of a rise in road freight prices (perhaps because road user charges have been raised to equal marginal cost). Will the distortion present on rail be increased or decreased? If more traffic is carried at less than marginal cost, there will be an increase in the distortion. However, there are several responses that the railways might make: some might worsen, others lessen, and other

leave unaffected the distortion. Five cases are considered. They are not an exhaustive list. Some may be more plausible than others, though the truth probably lies in between them all.

Suppose that constant returns to scale occur on rail. The first and simplest case is where rail accepts additional traffic at the going price. In determining the second best pricing rule for road, it is correct to take the rail price marginal cost ratio as given. This model of behaviour presumes that the rail deficit can be as large as desired. In Fig. 2, price remains at OA as demand rises from O to D, quantity rises from OQ to OQ and the deficit rises from ABDX to ABFE. This is probably quite unrealistic; rail may be subject to a limit on its deficit.

Another possible case would be one where the rail authorities kept the deficit constant by keeping both price and quantity constant, in spite of increasing demand, by rationing capacity by non-price means. In Fig. 2, the price remains at OA, and the quantity at OQ. This case may not be entirely unrealistic, especially if the railways give preference to regular customers (see Joy, 1980, p.129). In this case, even though the rail price is less than marginal cost, this ratio is irrelevant for road pricing, since the ratio of price to marginal cost on the road has no effect on the road-rail mix of traffic.

A third case is one which lies in between these two. Suppose that the railways are willing to accept more traffic, but they are constrained to incur no greater a deficit. They can do this
when demand expands by charging a price OA'. Output rises by GL, which is less than GH, the shift in demand. The new deficit is A'BKJ, which is the same as before. The rise in output is inversely related to the own price elasticity of demand for rail.

With the first and third cases, a rise in road price increases the distortion on rail; in the third case by less than the first. The fourth case is one in which this distortion is greatly increased. Suppose that cost padding takes place on the railways. By this is meant that, given deficit constraints and the prices which can be charged, railway authorities allow costs to absorb any profit that might be present. The rationale for this would be that managers and employees cannot share any profit, but they can direct expenditure in ways beneficial to themselves — e.g. through perks or easier manning arrangements (see Albon and Kirby, 1983). If they are required to operate at no larger deficit than before, and demand shifts, they may allow costs to rise. Thus, they may keep output at G, allow prices to rise to OA'', and allow costs to rise to OB'' (MC'' = AC''). In this case, the addition to distortion cost would be the addition total costs — shown by the areas AODA'' = BOMB''. Compared to the distortion cost changes before, this would be very large. While there may be an element of reality in this case, it is unlikely to be such a blatant case of waste as indicated in the diagram.

In contrast, with the final case, the distortion cost is reduced. Rail carries several traffics; some are competitive with
road, and other are not. Suppose that rail is subject to a deficit constraint, but that the non-competitive traffics are charged above marginal cost, while competitive traffics are charged below - as shown in Figure 2. When demand increases for the competitive freights, the railways may be satisfied to preserve their market share. The price rises at OA", the quantity stays the same, and the deficit on competitive traffic falls by AQWA'. The distortion cost is unaffected. The railways then keep the overall deficit constant by reducing the prices for non-competitive traffics. This brings them close to marginal cost, and thereby reduces the distortion cost. This is by no means an unrealistic representation of the railways' behaviour.

The correct second best rule for road pricing thus depends on the response that railways are expected to make to a change in road prices. It may be that even though rail charges are less than marginal cost, a price above marginal cost is appropriate for road. It would be correct to use a given price-marginal cost ratio and an unadjusted cross-elasticity only if the first case were the closest approximation to reality. In some respects it is the least likely of the five cases in so far as with it alone the rail deficit is unconstrained.

3.3 Estimating Demand Elasticities

The second aspect of the road-rail problem is to obtain evidence on the relevant cross- and own-price elasticities. Granted the importance of these elasticities in determining an efficient pricing policy for roads, it might be thought that much effort would have been spent in deriving empirical estimates of them. Such is not the case. There are few estimates, for Australia, or for overseas, and most of those which have been derived are from misspecified models.

Two studies of elasticities for the road-rail case in Australia are Fitzpatrick and Taplin (1972), and the BTE (1979). Fitzpatrick and Taplin estimate the cross elasticity of road with respect to rail price as 2.16 using a group of routes. This estimate is from an equation with a price ratio; this implies the own price elasticity of road is -2.16. The BTE estimated long term cross elasticities, again using a ratio formulation, as 0.7 for the Melbourne-Sydney route and 0.9 for a group of routes. Fitzpatrick and Taplin also estimate elasticities of road freight demand by including the road and rail prices separately. As will be argued below, this is the correct procedure, though the authors discount their results. They find a cross-elasticity of road with respect to rail price of 3.25, and an own-price elasticity of -2.54. These estimates are on the high side, especially the former, though there is nothing inconsistent about them.

There are problems with the estimates. Use of ratios implies that cross elasticities are equal to the negatives of the own price elasticities. This is unlikely, as argued below. Secondly, there should be no presumption that cross elasticities would be
the same across different routes. Suppose two routes; on one rail has a 60% share of traffic, and on another it has a 10% share (perhaps rail is more indirect). A given percentage increase in the rail price might induce a large relative increase in road freight in the first case (where road has 40% of the market), but a small relative increase in the second (road has 90% of the traffic already). Cross elasticities are related to market shares, and these differ widely across markets.

Thus, the available evidence cannot be relied upon. There is additional information of an informal kind. For example, several economists who have considered the road-rail problem have concluded that, in general, cross elasticities are low (see, for example, Joy, 1980). Many of the studies undertaken elsewhere are subject to the same problems as raised above. One which does not is that of Friedlaender and Spady (1980). This study, for the US, finds that the cross elasticity of road with respect to rail price is low, and negative (indicating that they are complements). These results cannot be generalised to Australia, though they are indicative.

On own price elasticities, there appears to be widespread belief in Australia that these for freight transport in general are very small, say, less than 0.1. This seems to stem from the discussion in Fitzpatrick and Taplin (1972). It implies that most of the response to a price change will be in the allocation of traffic between modes, and that overall traffic will be virtually unaffected. This justifies the use of price ratios in demand equations. Fitzpatrick and Taplin justify this assumption by examining the share of transport in total output. Since this is small (less than 10%), they argue that the demand elasticity is very small.

There are problems with the formula that Fitzpatrick and Taplin use—it ignores an important element of the response to higher freight charges. We are dealing with the derived demand for a factor—this demand depends on several variables. An expression for the own price elasticity of demand for a factor is given by Allen (1938):

$$\eta_F = - \left( \eta_m \cdot \sigma - \rho_F . \eta \right)$$

where

- $\eta_F$ = Own price elasticity of demand for freight
- $\rho_m$ = Share of other factors in Total Cost
- $\rho_F$ = Share of freight in Total Cost
- $\sigma$ = Elasticity of substitution between Freight and other factors.
- $\eta$ = Elasticity of demand for the final product.

The first problem arises in the measurement of the elasticity of demand for final product. If we take goods in general, the uncompensated elasticity will be unity. Since income is constant, a change in price is consistent only with an equal proportionate change in quantity. However, it is not the
uncompensated elasticity which is relevant. Taking the economy as a whole, when, say, road charges are increased, the government revenues rise - this additional revenue ultimately must be returned to taxpayers in the form of lower taxes, or spent, thereby increasing the real income of individuals. If a general equilibrium perspective is taken, it can be seen that the compensated elasticity is relevant. When there is a single, aggregate, commodity, this will be zero.

However matters are not so simple. In reality, goods differ as to their freight intensity. When freight rates rise, the prices of some goods rise relative to others. Even though there is no overall income effect, there are substitution effects. These lead to the demand for freight being sensitive to its price. It is not possible to make reliable a priori forecasts about how significant these might be. The way is to estimate the proportion of freight in total costs for disaggregated outputs, and to obtain information on compensated elasticities for these outputs. Together with the shares of each output in total expenditure, it will be possible to use this information to estimate the responsiveness of freight demand to its price. This is a good example of how excessive aggregation can be misleading. It would be much easier to estimate the elasticities directly.

The second problem is that the factor substitution term has been ignored. Freight is a substitute for other inputs. As freight becomes cheaper, more of it is used. Indeed, the history of Australia has been interpreted in terms of falling transport costs (see Blainey, 1966). For example, when freight prices fall, it becomes possible to locate production at one rather than many sites and to take advantage of scale economies, thus reducing the input of other factors (see Mohring, 1976, Ch. 11). Falling surface transport costs, recently enabled by more efficient road transport, have made a major impact on the industrial structure of Australian cities. In the short term, the substitution elasticity between freight and other inputs may not be high, but in the long term, evidence suggests that it is significant.

This is the finding of Friedlaender and Spady for the US. Own price elasticities for road were found to be around -1.0, and for rail, to be well above -1.0. This is in spite of their finding that cross elasticities were very low. These results cannot be generalised to Australia, though they should indicate how incorrect the prevailing Australian assumption might be. If they are accepted, they suggest that the road-rail competition aspect in the US is unimportant, and that pricing rules can be derived for the modes separately without much inaccuracy.

Whatever the model of road pricing being used, the own price, and rail cross price elasticity of demand are potentially relevant. It is thus worth exploring the relationship between them. Suppose that both rail and road transport are used to produce an intermediate good, freight. In Australia, this is not an unrealistic assumption, where shippers may use the services of a freight forwarder, who may use rail, road or some combination of the two to ship goods. Using the relationships in Allen
the demand elasticity of road, and cross-elasticity of demand for rail with respect to road price, may be expressed as follows:

\[ \eta_{tr} = (\rho_r \sigma_1 + \rho_r \eta_r) \]  
\[ \eta_{tr} = \rho_r (\eta_r - \sigma_1) \]

where
- \( \rho_r \) = Share of rail in total freight cost
- \( \rho_r \) = Share of road in total freight cost
- \( \sigma_1 \) = Elasticity of substitution between road and rail.

By substituting in the expression for \( \eta_{tr} \), we get:

\[ \eta_{tr} \]
\[ = -\rho_r \sigma_1 - \rho_r \sigma_2 + \eta_r \rho_r \eta \]
\[ = +\rho_r \sigma_1 + \rho_r \rho_r \rho_2 - \rho_r \rho_r \eta \]

In most of the discussion of elasticities for road and rail in Australia, attention has focused on the first terms of (5) and (6). The assumption is made that the third terms are very small, and the second terms are ignored. Thus the rail cross elasticity is simply the negative of the road own price elasticity weighted by shares. It should be no surprise that rules stating that prices be proportionate to marginal costs emerge when substitution between road and rail is the only effect of price changes. As indicated above, we know little about the size of the third term, depending as it does on elasticities of demand for different final goods. The second term is quite significant, over the long run, and breaks the simple relationship between own price and cross elasticities.

These expressions are based on a specific method of aggregation (of road and rail into something called "freight") which can be objected to. It may be that, in the production process, either rail or road may be used. If the road price rises, rail demand may not fall because of substitution of other inputs for freight. If other factors, rail, and road all entered independently as inputs for final good production, slightly different expressions would be derived. In particular, the second term in the cross elasticity expression would disappear. If, as is likely, \( \rho_r \rho_r \rho_2 \eta \) is small, then the difference in own and cross-elasticities is summed up in \( \rho_r \rho_r \rho_2 \), which will be large if \( \sigma_2 \) is large. In the longer term, this is likely.

The importance of the road-rail competition aspect of road pricing can be summarised as follows. Suppose that rail is underpriced; should road charges be less than costs to compensate? Take first the case where there is a fixed freight task, and the only variable is the share going to road or rail. In this case, lowering road charges towards the level which would equate price/marginal cost ratios will have an effect which depends on the railway's response. With some responses, it may be desirable, though it is possible that there will be no effect at all, and if the railways carry more than one traffic, some of
which road cannot carry, the effect may be to increase, not decrease, the distortion. Thus, even in this somewhat unrealistic case, equal price-marginal cost ratios may or may not be desirable.

The case where the total volume of freight is not invariant to the road (or rail) price is more realistic. Even if the effect of a road price reduction is to lower the distortion on rail, equal price-marginal cost ratios are not desirable. It is a matter of balancing two effects — prices equal to marginal cost are desirable to achieve an efficient level of freight overall, and road freight in particular, and prices in the same proportion to marginal cost are desirable to allocate freight between road and rail efficiently. A price somewhere in between these two levels will be efficient. When the rail response to a road price fall is negligible, or such as to increase distortion, there is no case for charging road prices less than marginal cost — prices above marginal cost may be called for.

The case for equating price-marginal cost ratios on road and rail is a weak one. It relies on improbable assumptions about the relevant elasticities, and plausible, though very restrictive assumptions about railway behaviour. To answer the question of what road charging policy is most appropriate, better evidence on the the relevant elasticities is essential, and an explicit model of railway behaviour is needed. As a working rule, marginal cost pricing on roads is more likely to produce efficient results than charging prices equal to the same proportion of marginal cost on both modes.

4. Recovering the Costs or Spending the Revenues?

4.1 Three Models of Revenue/Expenditure Constraints

The notion of cost recovery seems a simple one, though its practical application often proves complex. Three cases may be identified. Firstly, charges may be raised at such a level as to cover some measured total cost. This is simple cost recovery. The next two cases are where there is earmarking or hypothecation of funds. One includes choosing a level of expenditure (including investment expenditure) and raising revenues to finance it. The other involves choosing a level of revenue to be raised, and spending whatever is available.

All of these are possibilities for the handling of road funds. Simple cost recovery is a possibility, but there is often pressure to see that revenue collected from roads is spent on roads. The question then arises of what is determining what. Is it the available revenue which determines spending, or vice versa? All of these problems have implications for efficiency. They are summed up in Figure 3.

Along the horizontal axis, revenues and spending are measured. On the vertical axis, the marginal cost, per dollar, of raising revenue, and the marginal benefits of expenditure are measured. (Typically, these will exceed one dollar — see Findlay and Jones, 1981, Stuart, 1984.) The curve MC shows the marginal cost of raising funds, primarily through taxation, generally
throughout the economy. It is horizontal, though it need not be (the road budget is assumed to be only a small proportion of the total budget). The curve B shows the marginal benefits from expenditure on roads. The curves MCR and MCR show, for the two separate cases, the marginal cost of raising revenue from road charges. The vertical line at C shows the costs of providing and operating the road system (determined primarily by decisions taken in earlier periods).

If prices are set efficiently, and MCR is the marginal cost of revenue from roads curve, revenues will be R*. Below this it is cheaper to raise revenues from roads than elsewhere, since in this area, raising road charges reduces distortion. Beyond R*, road charges are being pushed to inefficiently high levels. R* may be above or below c. In the constant returns to scale case, if the marginal cost of raising one dollar is one dollar, R* and c coincide. Alternatively, MCR may be the appropriate curve, and R* the resultant revenue. The efficient level of spending on roads is given by B*, where the marginal cost of spending equals the marginal cost of funds. If MCR is operative, it would be efficient to raise R*, which falls short of costs C, spend B*, and finance the difference out of general revenue. If MCR is operative, it would be efficient to spend less than the amount raised from roads.

Suppose that B* is chosen as the level of spending, and that road spending must come from road charges. If MCR is operative,
charges will exceed the efficient level. The marginal cost of funds will be raised to the level of B F. The welfare loss is shown by the area BFG. Simple cost recovery would entail a welfare loss, but only of the level BDE. If the MCR curves are flat, raising revenue so as to finance expenditure will not create a large welfare loss. The MCR curves would be flat if the demand for road use is inelastic.

Imagine instead that it is revenues which determine spending. If R\textsuperscript{k} is raised, only R\textsuperscript{k} is spent. Worthwhile investments in roads will be foregone. This will entail a welfare loss of HFS. Alternatively, if MCR is the operative curve, efficient pricing raises more revenue than can be efficiently spent on roads. If all of it is spent, there will again be a welfare loss, this time of the size shown by FK1. If the spending curve, B, is flat, the welfare losses will be low (and if steep, the losses will be high).

The relative merits of the different policies depend on the shapes of the curves. Simple cost recovery has been shown in this diagram to be relatively efficient. It need not be, and would not be so if \(\varepsilon\) lay to the right of B\textsuperscript{k}. Whichever constraint, (which is, in the final analysis, arbitrary), is imposed there will be a cost. The cost depends on whether spending determines revenue or vice versa. If demand elasticities with respect to road charges are low (see the following section), then the distortions imposed by raising too much or too little revenue are low. The costs of investing too much or little could be high.

This means that a policy of letting road revenues determine road expenditure could be a relatively costly one. Getting the level of road expenditure correct is a more important problem, in this situation, than raising the right revenue.

4.2 Some Related Issues

Some comments on related issues are worthwhile. It is possible to make some suggestions about the relative size of \(\varepsilon\), \(\varepsilon\) and B\textsuperscript{k}. If constant returns to scale are present, and if the marginal cost of raising \$1 is \$1, then R\textsuperscript{k} and \(\varepsilon\) coincide. If, as is likely, the marginal cost of funds exceeds \$1 per \$1, then R\textsuperscript{k} exceeds \(\varepsilon\) – in other words, it is efficient to levy a tax on roads as on other goods and services with scale economies in the provision of roads, R\textsuperscript{k} tends to fall short of \(\varepsilon\).

The relationship between spending B\textsuperscript{k} and revenues R\textsuperscript{k} is analogous to the relationship between deprecation provisions and investment. In a steady state, static economy, efficient deprecation provisions will just cover the gross investment. For a steadily growing economy, they will fall short. This suggests that, for a growing road system, the efficient level of spending will exceed revenues raised under efficient pricing. In Australia, if anything, the road system is still growing.

Another consideration is the time pattern of costs and spending. If cost recovery is the objective, recovery can be
spread over many years, meaning that charges in any one period (say year) can be comparable to those in other years. With hypotheccation of funds, this evening out year by year is more difficult. When intended road spending rises, it can only be financed by higher charges. For a few years, road use charges may be relatively high. Alternatively, if revenues determine expenditure, the latter may fall short sometimes, and exceed at other times, an efficient level of spending. There may also be problems of intertemporal equity, when some generations are levied high charges to provide for the roads of the future.

Hypotheccation of funds, just as simple cost recovery, creates inefficiency losses. It is not possible to say, a priori, which policy would result in the lesser loss. There are, however, more things that can go wrong with relating current spending to revenues. This is especially true if desired spending levels are subject to change. Over the recent decades in Australia this has been the case, and careful matching of revenues and expenditures would have resulted in significant welfare costs.

5. Achieving a Revenue Target

5.1 Revenue Constraints and Inverse Elasticity Formulae

There may be a requirement that a level of revenue be raised which is not consistent with efficient pricing of roads. This problem need not be as great as is often supposed — it has been argued in Section 2 that efficient prices will probably yield either enough, or nearly enough revenue to cover costs. However, as recognised in Section 4, exact cost recovery may not be the objective. Whatever the revenue requirement, it is desirable that it is met by charges which create the least distortion cost.

The solution to this problem is well known, and it been discussed in relation to road charges (for example, by Taplin (1980, 1982) and the May Committee (1984)). It stems from the treatment of optimal taxation by Ramsey (1927) and that of covering deficits of public enterprises by Boiteux (1956, 1971). In its simplest form, it involves setting the ratio of price to marginal cost in inverse proportion to elasticity. This minimises the distortion between the efficient level of sales and the actual level.

In this section, the problem of achieving a given revenue from charges on road users is considered. As Taplin (1982) and Taplin and Waters (1984) point out, the cost of raising such revenue is less if the task is spread over a wider range of outputs. It would be more efficient to cover combined rail and
road deficits with an integrated system of charges than it would be to cover the deficits separately, even if none of the services of one mode was a substitute for any service of the other. An integrated system would undoubtedly involve subsidies being paid by one mode to cover the deficit of the other. It will be taken here that such an integrated approach is unlikely. It will be argued later that the benefits from integration, while positive, are likely to be very small.

Notwithstanding this, the interaction between road and rail can be of critical importance. If both road and rail seek to cover costs by imposing Ramsey/Boiteux type rules in isolation of each other it is doing, a distinctly inefficient result could result. This is because the high elasticities, from the point of view of the individual mode, are likely to occur on competitive traffics. Simplistic application of inverse elasticity formulae would result in competitive traffics bearing very little of the burden of cost recovery, even though their elasticity of demand for freight in general might be low. The correct way to handle this problem is to analyse the road and rail cost recovery problems jointly, though with separate revenue requirements. An alternative, and more practical way, is to allow for rail’s behaviour in setting road charges (and vice versa). This means recognising the second best problem as discussed in Section 3.

Another general point about the application of Ramsey/Boiteux rules concerns the role of commodity taxation in an economy. Government revenue may be raised through taxation of goods, and the rules for efficient taxation are set out in, for example, Ramsey (1927) and Baumol and Bradford (1970) – these are the same type of inverse elasticity rules as occur in the revenue constraint problem. One requirement, however, is that intermediate goods should not be taxed (unless it is not possible to tax the final output). Production efficiency is desirable (see Diamond and Mirrlees (1971) and, for the transport context, Forsyth (1977)), and intermediate goods would distort the choice of inputs.

Some of the outputs of road users (e.g. leisure car journeys) are final, and other (road freight) are intermediate outputs. If taxes are to be levied on commodities generally, they should be levied on car trips, or on the road use which makes such trips possible. However, it would be inefficient to levy taxes on road freight. Thus, it would be expected that car users (strictly, leisure car users) would pay marginal cost, plus an amount calculated according to Ramsey/Boiteux formulae to achieve cost recovery, plus a further amount for tax, whereas road freight would be charged marginal cost plus an amount for cost recovery.

An alternative perspective on how taxes affect the cost recovery problem is as follows. Final goods may or may not be subject to taxes, and these taxes need not be optimal. In the presence of final good taxation, the derived demand for a factor will not measure, in a competitive environment, the value of its
marginal product; it will be something less than this. In application of Ramsey/Boiteux rules, to minimise the welfare cost of diverging from marginal cost pricing, the value of marginal product functions, not the demand functions, are relevant. A slight increase in price above marginal cost for the intermediate good will not just produce an infinitesimal welfare loss — the loss could be large relative to the revenue gain. This is because an existing distortion (final good taxation) is being added to. The welfare cost of a change in price to an intermediate good exceeds that to an untaxed final good, ceteris paribus. Thus, to this extent, we would expect efficient cost recovery to result in higher price marginal cost ratios for cars than freight vehicles.

5.2 Applying Pricing Rules

To impose inverse elasticity rules on pricing, it is obviously necessary to have information about the relevant elasticity. As indicated in earlier sections, our knowledge of these is limited. It is necessary to identify, firstly, which elasticities are required. This links in with the question of road charging structure, to be examined in the next section. One point that may be foreshadowed here is that charges will be levied on inputs to affect output and this complicates the problem. The second problem is to estimate the likely magnitude of the elasticities.

It is quite possible that the demand elasticities with respect to road charges (though not necessarily with respect to overall road use costs) are low. Subject to the qualifications to be made in the following section, it can be noted that the proportion of total costs faced by road users accounted for by road user charges is small (less than 10%). It is likely that the impact of a rise in charges on road use will also be small. This is true both for freight and passenger vehicles. If this is true, the distortion costs imposed by levying charges which result in prices greater than marginal cost may be quite low.

The facts that (a) little is known with any degree of reliability about the elasticities for different classes of road used and (b) to the extent that there is a presumption, based on rough evidence that relevant elasticities are quite low, have strong implications for differential pricing of different road users. The case, on efficiency grounds, for differential prices for cars and heavy goods vehicles as suggested by the May Report (1984), is very weak. For example, the Report allocated "joint" costs such that they form 93% of "attributable" costs for articulated trucks and 54% of "attributable" costs for cars. Elasticity estimates are not sufficiently reliable to draw such distinctions. Furthermore, within reasonable limits, the allocation of "joint" costs is unlikely to have any perceptible efficiency impact, granted that elasticities are low, and consequently the distortions from pricing at other than marginal cost are low. A situation whereby both cars and articulated trucks were allocated "joint" costs in proportion to
"attributable" costs would probably be no less efficient. What is more, any statement about degrees of cost recovery from individual classes of user must be regarded with suspicion. The evidence is simply not good enough to state that articulated trucks are not paying their way.

Thus, with current empirical information, it is not possible to solve revenue constraint problems to determine uniquely efficient charging structures. The Ramsey/Boiteux rules are appropriate, but the information available is too limited to apply them with any confidence. Rules of thumb, which are more likely to be accepted by road users, will result in equally efficient outcomes. As shown in Section 2, the magnitude of the revenue constraint problem, and the importance of second best rules, may have been much exaggerated. Later empirical evidence could contradict this tentative assumption. A more significant problem may be one of determining which inputs of road users to levy charges upon.

6. Road Charging Structures

6.1 Pricing Inputs Rather Than Outputs

It is very likely that road charges will be related to inputs, rather than outputs. While some charges have been based on tonne kilometres performed, the likelihood is that charges will be on fuel, operation of vehicles (registration) and on purchase of new vehicles (sales tax). The elasticities which are relevant are not the elasticities with respect to the price of car operation or road freight, but the elasticities of demand for the inputs with respect to the charges.

A substantial rise in road charges might make little difference, in the long term, to road use by freight or passenger vehicles. The distortion costs could be high, however. Road charges form only a small proportion of the costs of operating vehicles - perhaps 5-10%. As shown above, it is dangerous to assume that elasticities are small if the proportion of cost that an input represents is small. Inputs are substitutes. For example, a fuel tax tends to encourage the use of more fuel efficient vehicles and the elasticity of demand in the long run for fuel need not be low. Road charges distort the use of inputs, and result in diminished production efficiency. This impact on efficiency may be much more important than the distorting effect on consumption patterns.
The road charging problem can be illustrated by means of a simple model. The output $X$, or road use (upon which road costs depend), is a function of inputs, such as fuel $w$, vehicle ownership $y$ and other inputs $z$. Thus:

$$X = f(w, y, z)$$  \hspace{1cm} (6)

Charges may be levied on fuel and vehicle ownership, but not other inputs. In the main, the inputs are substitutes for one another - for example, fuel consumption might be reduced by use of more expensive materials in vehicle construction. The charging problem is one of setting charges on $w$ and $y$ so that the best combination of (a) overall level of $X$ - the allocative efficiency problem and (b) the combination of inputs $w$, $y$ and $z$ - the production efficiency problem, is achieved. It is quite possible that one aspect of the problem may be much more important than the other. For example, if the demand for road use were not very elastic, yet the substitution between inputs were high, the input choice problem would dominate. The empirical evidence that is needed is on the response of input choice to charges.

This simple model illustrates how the indentification of particular charges with particular elements of cost is quite arbitrary. One such identification might be to attempt to recover marginal or "attributable" costs by fuel taxes and "joint" or "fixed" costs by sales taxes or registration fees. All types of charges will affect road use and input mixes. Of course, it may turn out that one type of charge has a large effect on road use whereas another does not. If so, when there is simply a desire to set prices equal to marginal cost, this charge will be relied on heavily. When there is a revenue constraint which is effective, both charges will be used. It is also possible that different charges may have much the same effect. If, for example, vehicle utilisation were fixed, a registration charge or fuel tax would have the same impact on total road use (through affecting the number of vehicles used).

6.2 Possible Charging Structures

It is worthwhile briefly examining the more popular forms of charges. These are fuel taxes, registration fees and sales taxes. The first of these will raise vehicle operating costs directly. Over time it will lead to a substitution for less fuel intensive vehicles. In a country such as Australia, which imports vehicles, or builds them to overseas designs, substitution is limited to the range of options available internationally. Fuel taxes, to be effective, will normally be linear - levied at a given rate per litre.

Registration fees will be a charge on ownership or use of a vehicle for a specified period. They will affect the number of vehicles operated, and the intensity of use of vehicles. If registration fees were low, older vehicles would not be scrapped so soon. Registration fees need not be linear. They can, and do, vary as to the class of vehicle, and other characteristics, such
as size.

Sales taxes, which occur once on the purchase of vehicles, will affect lifespans of vehicles, and indirectly, their utilisation. Higher sales taxes increase lifespan, and also encourage greater utilisation during the life. They also need not be linear. Different rates of sales tax can be levied on different classes or values of vehicles.

6.3 Setting Charges When Information Is Limited

The problem of choosing the most efficient mix of charges on different inputs to road use is analytically comparable to that of choosing the charges on different users. Inverse elasticity rules are appropriate. In this area, however, our knowledge of the elasticities is no better than for different road users. Given this lack of information, and that distortion costs rise more than in proportion to rates of charges, it is probable that an efficient structure will be one where all types of charges are used to an extent. The practical solution is to preserve fuel taxes, registration charges and sales taxes. If reliable evidence becomes available on elasticities the balance may shift. However, the econometric problems are likely to mean that this situation is a long way off.

There may be some scope for using charging structures to promote efficiency. Particular user classes might be estimated to impose costs in excess of others. For example, small trucks may use highly congested city roads, while large trucks may use less congested highways. It may be necessary, on practical grounds to charge the same fuel tax, but the higher per kilometre costs of the smaller trucks might be covered by a higher registration fee. It is through estimation of the marginal costs of different user classes, and through setting charges equal to marginal costs, that the charging structure may become important.
7. **Conclusions**

The approach taken here has been to clarify the theory as it applies to road charging issues. It is reasonable to rule out some possibilities and suggest how certain problems can be solved. But theory takes us only so far — in the end it is necessary to resolve questions by empirical analysis. The priorities for analysis have been identified here. It should be noted that some of the econometric problems are difficult. It is unlikely that reliable estimates of many of the relevant elasticities are going to be available in the short term.

In some cases, our empirical knowledge is distinctly limited. The approach to road charging must take this into account. There is little merit in developing sophisticated structures for road charges which require large amounts of data, or reliable estimates which are not available. In some cases, some presumptions about relevant parameters do exist, but it has been shown here that they rely on tenuous reasoning. Simple structures of road charges may be both easier to implement, and impossible to refute on grounds of economic efficiency, given current information. In future, more elaborate and efficient structures may become feasible.

Several implications are worth noting. There need be little conflict between cost recovery and efficiency charging structures. If there is any conflict, it is likely to be small. It is incorrect to ignore congestion costs in estimates of marginal costs. The priority here is to make estimates of the marginal costs of road users — this is a relatively simple task, though it has been neglected.

Secondly, the road-rail competition issue is relevant, but it is probably of secondary importance. The derivation of appropriate second best rules presupposes that a model of railway behaviour has been developed. This is hardly the case. The conventional wisdom about cross and own price elasticities greatly exaggerates the importance for road charging of the road-rail competition issue. Own price elasticities are relatively more important than supposed, especially in the longer term, and this indicates that the priority should be to devise efficient charging structures for road users, and only secondary attention should be given to the road-rail issue.

In a practical context, charges may not be set equal to marginal costs. Revenues may need to be raised to cover costs, or expenditure, current and capital, on the road system. Here the main problem is one of identifying what is determining what. Of the various options, a situation where whatever revenue is raised is then spent on roads, i.e. where revenues available determine expenditure, is likely to prove least efficient.

When revenue constraints are to be met, inverse elasticity formulae are relevant. However, knowledge of the elasticity of demand by different road users with respect to charges, or of
demand with respect to charges on different inputs (e.g., fuel or registration) is very restricted. This suggests that, until better information becomes available, expected distortion costs will be minimised if different users are treated similarly. Cost allocations based on application of inverse elasticity formulae cannot be trusted. Furthermore, there is no case for identifying some road charges (e.g., fuel charges) with marginal cost, and other (e.g., registration) with fixed costs. All charges will, in the long run, affect road use, though it is difficult to determine how much. Again, lack of information implies that the structure of charges which creates least expected distortion will be one which relies on a mixture of charges, and thus does not affect the use of any one input by too much. Complexity in road charges is unlikely to increase efficiency.

References


Blaikie, G. (1966), The Tyranny of Distance: How Distance Shaped Australia's History. (Melbourne: Sun Books).


Kolsen, H.M. (1968), The Economics and Control of Road-Rail Competition (Sydney: Sydney University Press).


Tasmania, Department of Transport (1982), Pricing Tasmania's Roads, Hobart.