

Shifting the Transport Infrastructure Debate



INTRODUCTION TO ROAD ECONOMICS

The economics of economically efficient pricing and investment for roads is complex and has some unusual features compared with other network industries. Economically efficient (or optimal) pricing and investment aims to promote allocation of resources in a way that achieves the greatest overall benefit to society from a given amount of resources. If people pay economically efficient prices, the decisions they make considering only their own private interests will be the best decisions from the point of view of society as a whole. Economically efficient investment is about providing the amounts and types of road infrastructure that maximise the benefits to society, based on people's willingness-to-pay, minus the costs to society.

When the standard approaches to economically optimal pricing and investment are applied to roads, a number of difficulties arise that are either not found at all, or are less pronounced than, in other network industries.

- There are two dimensions to consider:
 - capacity, which affects to the level of congestion users experience, and
 - pavement durability or strength, which affects to the amount of wear and tear users impose on pavements.

Fortunately, for most theoretical and practical discussions, the two dimensions of road economics can be addressed separately.

- The economically optimal amounts and types of investment and maintenance vary widely for individual lengths of road depending on demand and costs. A huge range of service qualities, from earth tracks to multi-lane freeways, coexist in a road system. Somehow, the economically optimal service level for each part of the road system has to be determined.
- As explained below in the paper, economically optimal pricing in its purest form leads to major under-recovery of capital and maintenance costs for most of the road system. If cost recovery is an objective, questions include:

- How to estimate the costs to be recovered, given that most were incurred in the past?
 - What proportion of costs should be recovered from users, with the remainder from taxes?
 - In what level of detail should cost recovery applied? At one extreme the cost of the whole network could be averaged over all users. At the other extreme, cost recovery could be pursued for each small homogeneous length of the road in the system.
 - How to charge users above economically optimal prices in a way that minimises the negative economic efficiency impacts?
- On routes where road and rail compete for freight, charges on heavy vehicles will affect modal split. Price–cost relationships between comparable modes should promote an economically efficient modal split.

The present paper presents the issues under four headings:

1. Economically efficient pricing for congestion and road wear
2. Cost recovery under efficient charging and investment
3. Ways to recover costs
4. Applicability of commercial investment criteria.

The focus throughout the paper is on the economic efficiency objective, leaving for elsewhere discussion of equity, safety, technology, planning, political acceptability and other considerations that enter into government decision making about road provision and pricing.

Economically efficient pricing

The starting point for economically efficient pricing is to set prices equal to ‘marginal social cost’ — the cost imposed to the entire community by a single additional user. On congested roads, an additional user experiences the average cost of congestion (more fuel consumed, longer and more variable trip times) while at the same time, increasing the costs to existing users. The marginal social cost therefore consists of the cost that is incurred by the additional user, plus the ‘external cost’ imposed on others. Costs of emissions, noise and road damage also form part of the marginal social cost. Economically optimal road pricing involves levying charges so that road users incur full marginal cost they impose on society.

In making decisions about if, where, how and when to travel, people compare the costs of the alternatives. Naturally, they consider only the costs they incur themselves. Consumer decision making will be distorted if the costs to individuals of the alternatives are different from the costs to society as a whole, resulting in individuals making decisions that are sub-optimal from a whole-of-society perspective.

For example, in the absence of efficient road congestion charges, some individuals will make socially sub-optimal decisions to drive on congested roads during peak times, while others may make socially sub-optimal choices between alternative routes or modes of transport, or to travel too

often. In the long-run, congestion charges can affect locational decisions, even the level of urban sprawl. Similarly, for damage to road pavements, if trucks are under-charged for the damage they do, they will travel too much on roads with weaker pavements and use axle configurations and carry payloads that are too damaging, such that the cost to society as whole exceeds the benefits to the individuals causing those costs.

Charging above marginal costs can be just as bad because it leads to under-use of the roads with high charges and over-use of alternatives.

Optimal congestion pricing

Figure 1 shows the average cost curve for a road — the relationship between generalised cost and the traffic volume per period time (usually an hour). ‘Generalised cost’ is a way of combining all the impacts on users into a single dollar amount. In the simplest case, generalised cost would consist of vehicle operating cost + the time taken × the value of time. Average cost rises as more traffic uses the road until capacity is reached, typically 2,000 passenger car equivalents per hour per lane, with a truck counting as three passenger car equivalents.

Figure 2 illustrates why marginal cost exceeds average cost by the amount of the external cost. The initial average cost is \$1 for four vehicles. The addition of a fifth vehicle increases average cost to \$1.05. The external cost imposed by the additional cost imposed by fifth vehicle on existing four vehicles is therefore $4 \times \$0.05 = \0.20 . The marginal social cost of the fifth vehicle is \$1.25 — its own cost of \$1.05 plus the external cost or \$0.20. This explains why the marginal cost curve lies above the average cost curve.

Figure 1 Average cost curve for a length of road

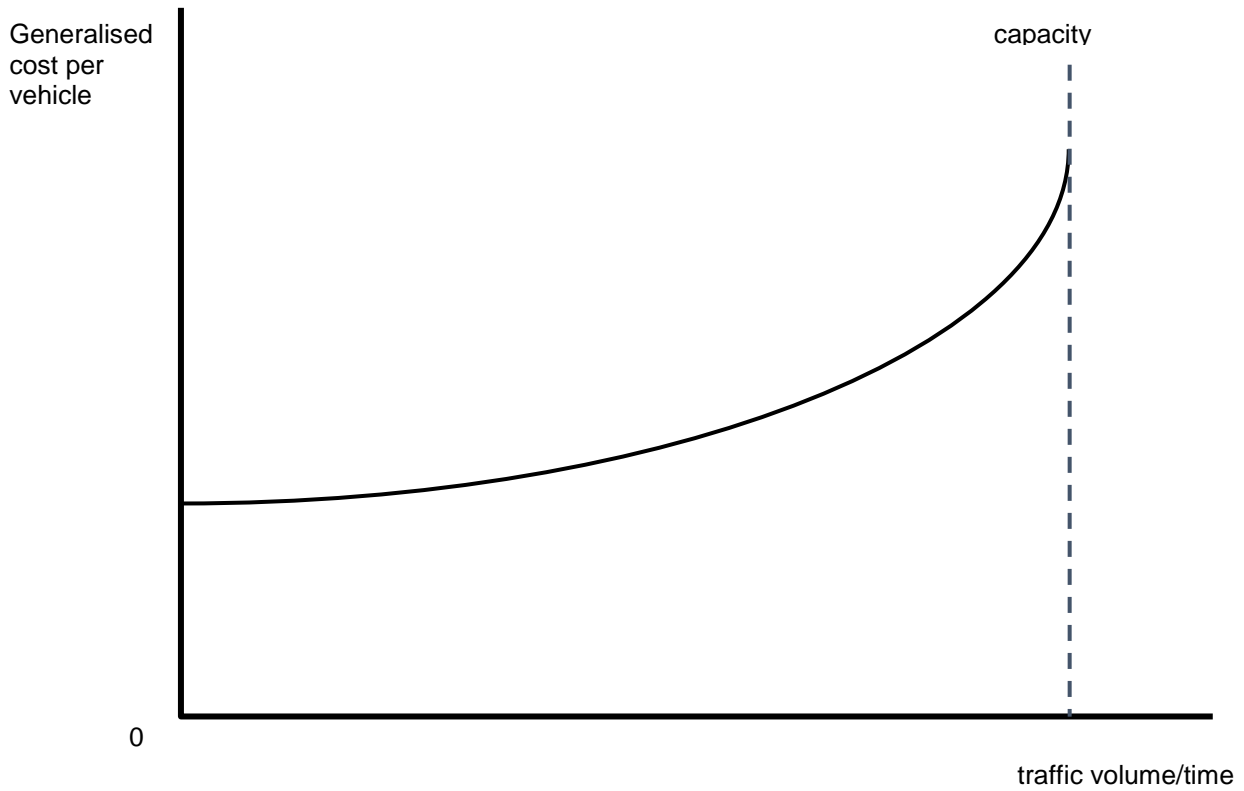


Figure 2 External cost of congestion

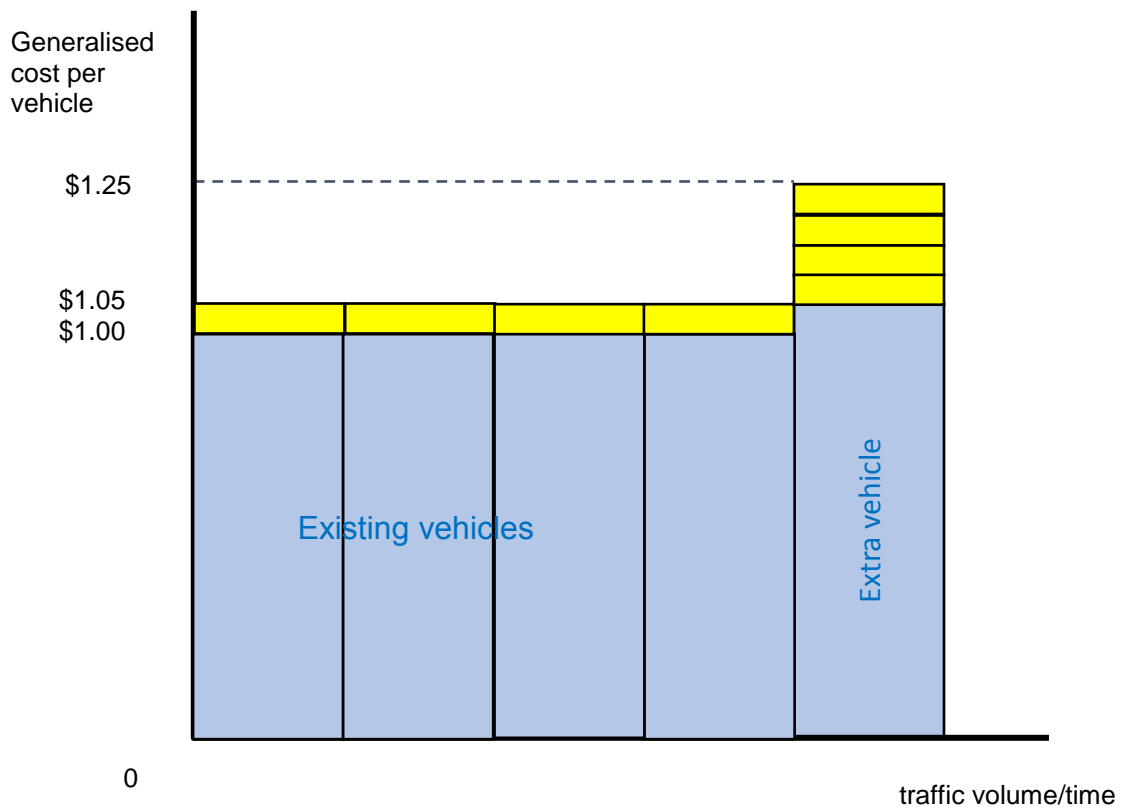


Figure 3 shows the average and marginal cost curves together with a demand curve. In the absence of congestion charging, road users incur the average cost and the traffic volume is found at the point where the demand curve crosses the average cost curve. However, the economically optimal traffic volume is lower, being found where the demand curve crosses the marginal cost curve. For the additional traffic between the optimal and actual levels, the valuation to users (the area under the demand curve) is below the cost to society (the area under the marginal cost curve) by the amount of the triangular coloured area. Economists call this area as a 'deadweight loss' to society. It is what BITRE aims to measure for its cost of congestion estimates (BTRE 2007).

Figure 4 shows the optimal congestion charge. It occurs at a level equal to the gap between the marginal and average cost curves at the point where the demand curve crosses the marginal cost curve, that is, at the economically optimal level of traffic. It reduces traffic to the point where the deadweight loss is fully eliminated. The government gains revenue equal to the area of the coloured rectangle — the traffic volume multiplied by the congestion charge.

The demand for road use fluctuates over the day, being highest in the morning and evening peak times, and varies by location. Ideal implementation of congestion pricing would have

- prices varying continuously by time of day
- prices varying at each point along the road where traffic level and/or road capacity changes;
- prices varying between vehicle types in proportion to passenger car units (1 car = 1 passenger car unit (PCU), 1 large truck = 3 PCUs)
- for some major roads, prices varying between lanes to give users a choice between high price–high speed and low price–low speed options, and
- no exemptions from charges at all — not even for emergency vehicles (BITRE 2008).

Inevitably, any practical congestion charging scheme will involve compromises depending on technical factors, implementation and operating costs, community acceptance and politics. Experience suggests that operating costs are far greater than implementation costs and amount to at least 20% of revenue annually (Oehry 2010).

Figure 3 Cost of congestion

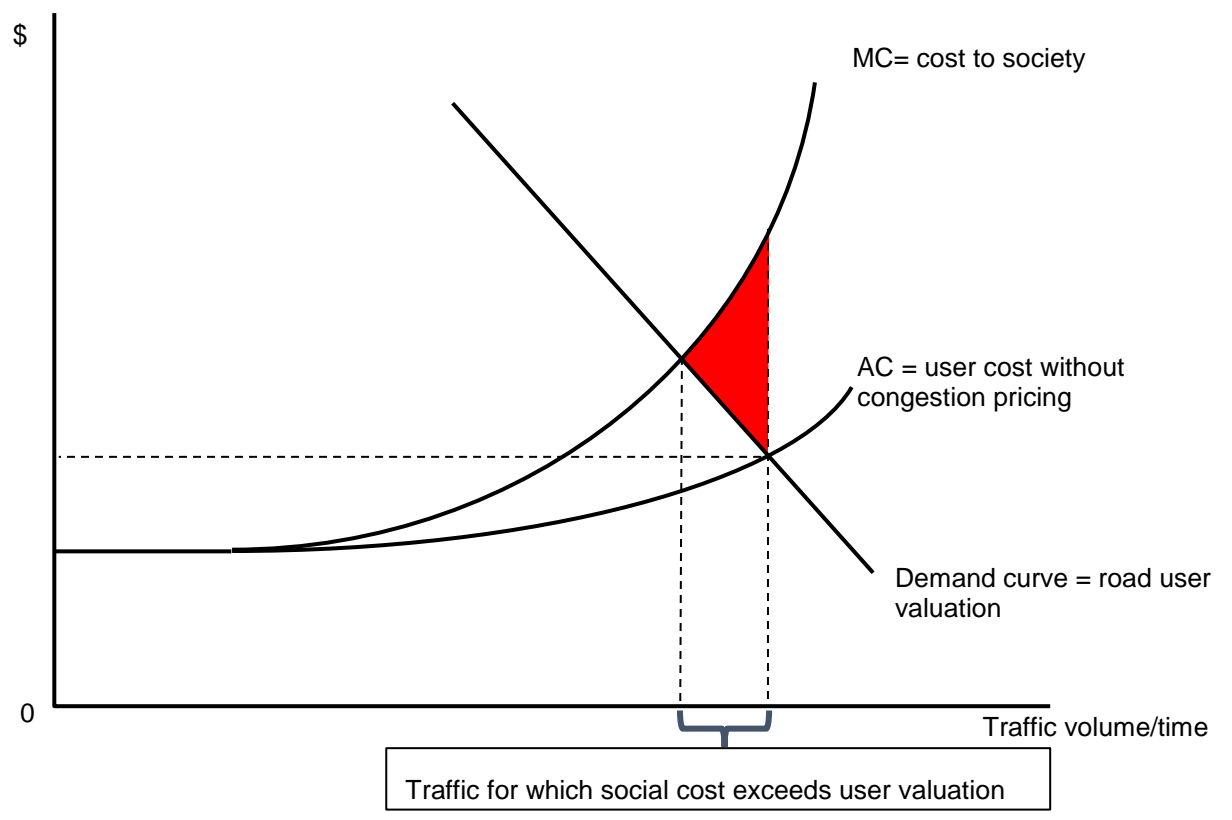
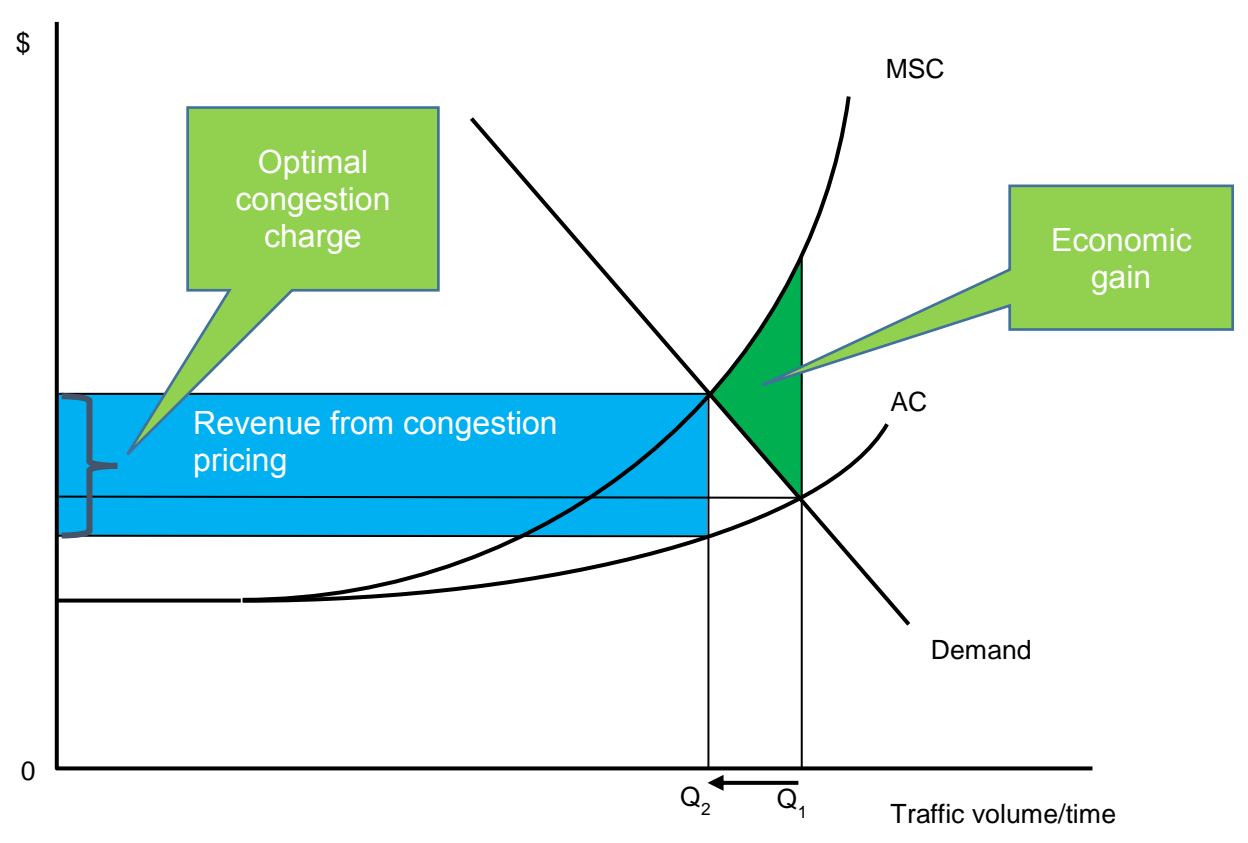


Figure 4 Optimal congestion charge



Optimal road wear charges

Road maintenance has 'routine' and 'periodic' components. Routine maintenance comprises tasks such as cutting grass, clearing drains, repairing or replacing signs, sealing cracks and patching potholes. These are independent of the traffic level and are fairly constant from year to year. Periodic maintenance tasks for bitumen pavements are resealing, rehabilitation and reconstruction. These are incurred at intervals of 7 to 15 years for resealing and 20 to 50 years for rehabilitation. The term 'rehabilitation' covers a range of treatments including digging out and replacing the upper layers and surface of the pavement, and putting additional layers and a new surface on top of an existing pavement.

Bitumen pavements deteriorate with both time and axle loads. The initial strength of the pavement, the climate, the quality of routine maintenance, and the frequency of reseals are also important factors. The wearing impact on a pavement by a vehicle is measured in 'equivalent standard axles' (ESAs), which is defined as a single axle with dual tyres loaded to 8.2 tonnes. The damaging power of a loaded axle rises exponentially with the weight, generally understood to be a third or fourth power relationship. For example, a six-axle semi-trailer is estimated to account for 1.14 ESAs in total with no load and 4.96 ESAs fully loaded. As larger trucks spread their loads over greater number of axles, the number of ESAs for a truck is not simply proportional to the number of axles.

Additional trucks travelling on a pavement will bring forward the time at which future rehabilitations become necessary. The discounted present value of future maintenance costs then becomes greater. The increase in present value divided by the additional ESAs giving rise to the increase is a measure of the marginal cost of an additional ESA caused by a heavy vehicle.

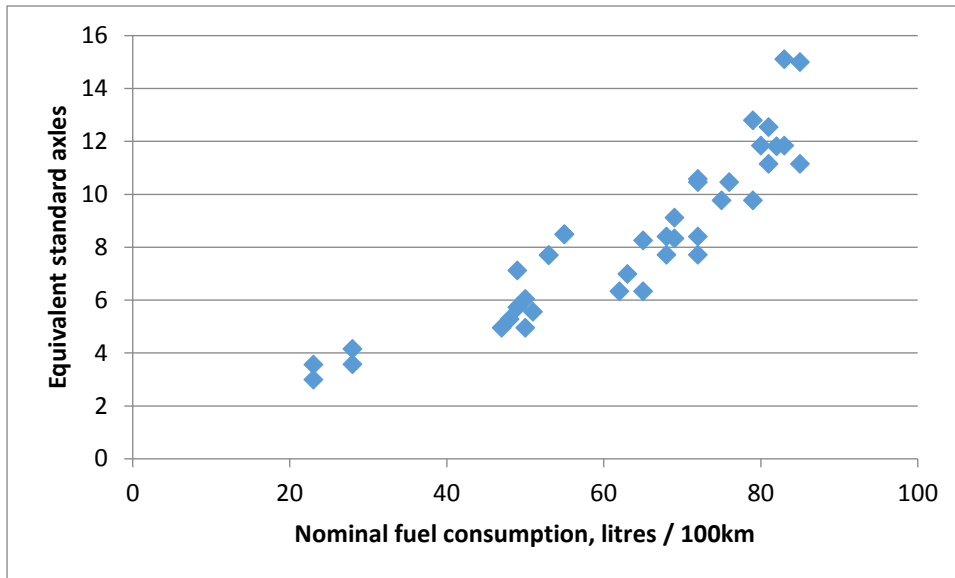
Cars do negligible damage to paved roads so their marginal cost of road damage is practically zero.

The ideal system of road damage charges for trucks would have charges varying by location, depending on the susceptibility to damage of the pavement, and the number of ESAs of the vehicle. ESAs will vary with the number and types of axles the vehicle has, vehicle weight, and the payload carried.

Fuel excise is not a good way to charge for road damage. As well as not varying with location, it is not greatly affected by load carried, and there are substantial economies of scale in fuel consumption as truck size increases while ESAs rise at an increasing rate. Figure 5 shows a plot of the ESAs for a range of truck types, starting with two axle rigid trucks and finishing with BAB-Quad road trains, against fuel consumption for each type. The ESAs are for fully laden trucks. For fuel excise to be proportional to road damage, the relationship would have to be linear. The convex shape of the relationship means that a fuel excise would over-charge smaller trucks and under-charge larger

trucks. The solution in Australia, at present and for a long time past, is to close the gap with an annual registration charge that is great for larger trucks.

Figure 5 **Equivalent standard axles (fully laden truck) plotted against fuel consumption for different truck types**



Source: Australian Trucking Association, Truck impact chart, June 2010

As the registration charge is time-based, not distance-based, vehicles that travel longer distances over year are advantaged relative to vehicles that travel shorter distances. The larger trucks are more advantaged as they tend to make longer round trips and therefore spend less time each year loading and unloading.

Cost recovery under efficient charging and investment

The economically optimal charges for congestion and road damage discussed above are ‘short-run’ marginal costs (SRMC). SRMC is the cost imposed on society, *taking the existing infrastructure as given*. In the ‘long-run’, when infrastructure can be changed, road providers have more options available to accommodate an additional user. An additional or marginal vehicle could be accommodated without adding to congestion by expanding road capacity. An additional truck could be accommodated without causing extra road damage by building a stronger pavement. The long-run economic optimisation problem is to determine the amounts of capacity and pavement durability to provide, balancing the capital costs against the associated savings in congestion and road damage costs.

Long-run marginal cost (LRMC) is the cost imposed by an additional user when the infrastructure can be changed in a way that accommodates the additional user at the lowest possible overall cost to

society. The LRMC on a congested road may comprise a combination of additional infrastructure costs and additional user costs. There is no conflict between prices being set equal to SRMC or LRMC. The economically optimal price, at any point in time, is the SRMC because this is the actual cost an additional user imposes on society. In the long-run, when infrastructure can be changed, it is axiomatic that with optimal investment in capacity and pavement strength, short-run and long-run marginal costs will coincide.

Optimal investment and cost recovery

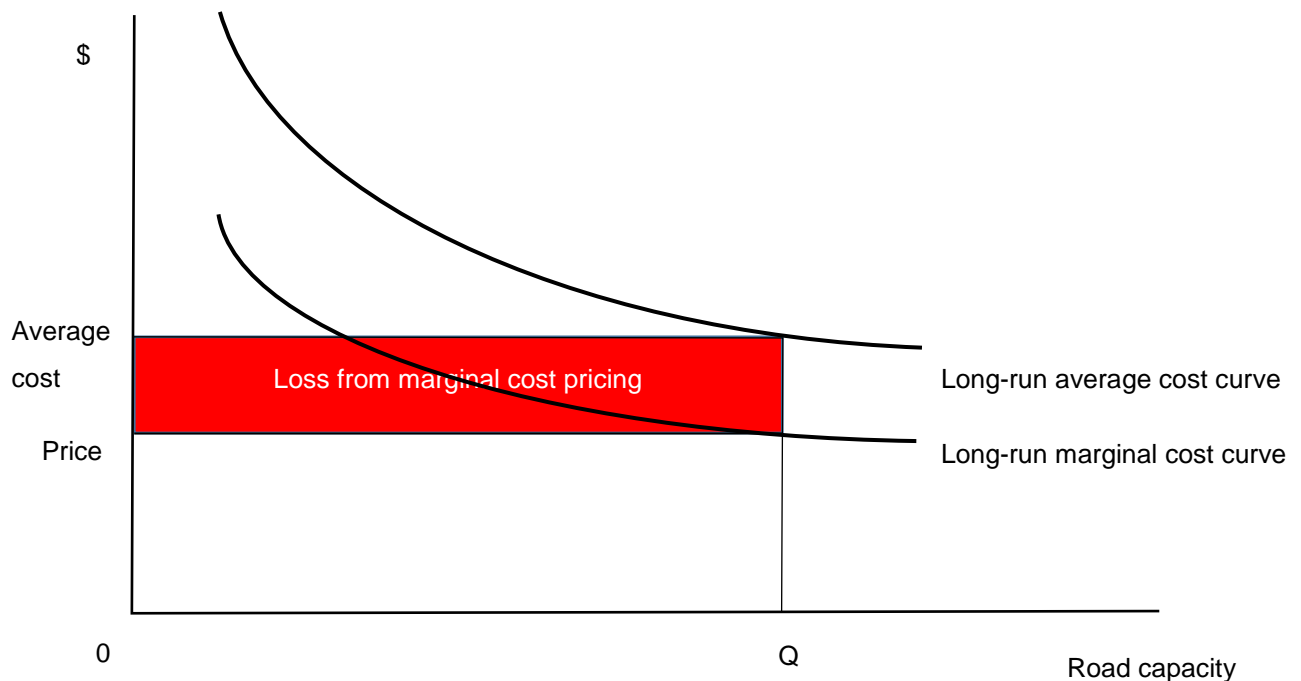
As mentioned above, determining the optimal level of investment in road infrastructure requires comparing the benefit to society in the form of savings to roads users and others with the cost of small (marginal) changes in road infrastructure. For example, the benefit from a small increase in road capacity would be a reduction in congestion. The benefit of improved pavement durability would be a saving in pavement damage. With optimal road congestion charges, a reduction in congestion due to a capacity increase would result in a small reduction in the congestion charge, and improved pavement durability would result in a small reduction in the charge per ESA.

As the service of standard provided by the road (volume–capacity ratio, speed limit, straightness, smoothness and safety) is increased and congestion and road damage charges fall with greater investment, users' valuations of each small increase in investment falls (in other words, the market moves down the demand curve). The optimal point is found where the benefit from a further small increase in investment in road standards equals the marginal cost of that increase. In other words, the benefit–cost ratio of a further expansion falls below one. This ignores the lumpiness of road investment, assuming instead that road investment is finely divisible. But the approach just described is valid for exploring the theoretical issues.

Under the assumption that road investment is finely divisible, whether economically optimal charges lead to over-, exact or under-recovery of capital costs depends on how the total costs of the system (users and the road provider) change as more is invested. If there are economies of scale, long-run average costs fall as road standard is increased. Falling average costs imply that marginal cost lies below average cost. The cost of an additional (marginal) unit of capacity or pavement strength has to lie below the average cost in order to pull the average cost down. A price equal to long-run marginal cost (also equal to SRMC) will lie below average cost and there is under-recovery of costs. (Figure 2 above explains the converse case that when average costs are rising, marginal cost lies above average cost.)

In figure 6, the downward sloping long-run average cost curve is associated with a downward sloping long-run marginal cost curve, which lies below it at all capacity levels. If price is set equal to long-run marginal cost, there is a loss equal to the area of the coloured rectangle.

Figure 6 Under-recovery of costs with optimal pricing and economies of scale



Individual roads are subject to economies of scale. Costs of infrastructure along the sides of roads (shoulders, signs, guide posts, drainage ditches) are the same regardless of the number of lanes. Also, because of the greater passing opportunities, a four-lane road has more than twice the capacity of a two-lane road (Small 1989). For rural roads considered as a network, there is an additional source of economies of scale — as a network becomes denser, having more interconnections, more direct routes become available between origin–destination pairs. There are enormous economies of scale in pavement strength. For the bitumen pavements used in Australia, the rule-of-thumb is that a 10 per cent increase in pavement thickness results in a doubling of the traffic loading required to produce a given amount of damage.

Congested urban roads

For major urban roads, the economies of scale are offset by diseconomies of scale. The number of intersections increases faster than the number of lane-kilometres of roads in a network in a given area. Intersections are land-intensive and often require traffic signals or grade separation. It is more expensive to build roads in densely populated areas because of obstructions from buildings, other roads, railway lines, pipes and cables. There is a consensus in the road economics literature that the

economies and diseconomies of scale for major roads in urban areas roughly cancel out giving rise to approximate constant returns to scale. (See Small and Verhoef 2007, ch 3 for a literature survey).

Constant returns to scale implies that optimal charges will cover capital costs. The Mohring–Harwitz theorem shows that for a single road with exact constant returns to scale, optimal congestion pricing combined with optimal investment in capacity leads to exact cost recovery. The theorem holds more generally where there is growing traffic, heterogeneous users, time-varying demand and networks of roads. (See Small and Verhoef (2007) and Verhoef and Mohring (2009) for literature surveys.)

The theorem applies only in the long-run when existing road infrastructure needs to be replaced or expanded. If congestion pricing were applied to an existing network that has already been paid for, it may be many years before the congestion pricing revenue is required to be spent on the network. So introduction of congestion pricing on an existing network should yield a surplus to spend on public transport and to compensate the public in other ways for many years to come.

The theorem assumes roads are finely divisible. For congested roads, lumpiness in capacity expansion means that optimal congestion charging will only lead to approximate cost recovery for individual roads over the very long-term or averaged over a large number of roads. Where capacity can only be expanded in discrete jumps, the combination of optimal pricing and optimal investment means that, with traffic growing, there will be periods of time when positive profits are made (when the road is more congested leading to expansion) and periods of time when losses are made (after an expansion when there is less congestion).

Uncongested roads

A very different model applies for uncongested roads. By length, most of the road system consists of uncongested or low-volume roads, that is, roads with volume–capacity ratios such that optimal congestion prices would be zero or low most or all of the time. Most non-urban roads be categorised as uncongested, as well as suburban streets and minor arterials in urban areas.

Charging a price to use an uncongested road will inhibit some people from taking advantage of the road, which is wasteful because their use of the road imposes no cost on society above what users incur themselves. The reason for the ‘public good’ nature of uncongested roads is a combination of economies of scale, which, as discussed above, lead to under-recovery of costs with optimal pricing, together with lumpiness in supply (not finely divisible). The basic two lane road is economically optimal over a wide range of traffic levels. Supply is also lumpy because:

- a road must be at least as wide as the narrowest car plus a margin for safety, and wider still to allow trucks to pass over it
- the number of lanes must be a whole number
- there is a minimum pavement depth, and

- a road must have an earth surface, a gravel surface or a sealed surface. Walters (1968)

If supply were perfectly divisible, capacity could be reduced sufficiently to induce congestion, forcing the price up to the point where the level of cost recovery was commensurate with the extent of the economies of scale. However, economies of scale mean there would still be under-recovery of costs.

A further consideration mentioned in the economics literature for low-volume rural and inter-urban roads is that service quality cannot be improved without also improving capacity and conversely. Investment to improve road quality by building a wider, smoother, straighter road with more passing opportunities is often found to be economically warranted in cost–benefit analyses (CBA) due to benefits from savings in travel time, vehicle operating costs, and crash costs. However, these improvements also add to capacity, potentially reducing any congestion price to practically zero. In other words, for low-volume roads, the reason for investing to improve the road is not that greater capacity reduces congestion as is the case for congested roads, but that a straighter, wider, smoother road gives users a faster and safer ride (Walters 1968).

Ways to recover costs

Why cost recover?

The previous section summarised the economics literature, showing that economically optimal pricing of congested roads in urban areas can be expected to lead to approximate cost recovery over the long-term and averaged over a large number of roads. For low-volume roads, the pure economic approach is to only charge for the costs of pavement damage to heavy vehicles. As most of the road system fits into the latter category, the notion that it is somehow economically efficient for road users to pay for exact cost of the road system as a whole has no basis in economic theory.

There are arguments in the economics literature that user pays is desirable, even in cases of significant economies of scale. One argument is that user pays provides a market test of whether users' willingness-to-pay is sufficient to justify continued supply (Coase 1946, 1970).

Another argument for recovering costs in full from users is that recovering costs from general taxation is costly, because of the deadweight losses to economic welfare caused by taxes that affect incentives to work, invest and consume. That being the case, one should compare the deadweight losses from higher levels of road cost recovery from road users with the reductions in deadweight losses from the associated tax cuts. The main economic argument for road users to pay the exact cost of the roads appears to be the 'user pays' concept of equity — that is, it is equitable for road users to pay for the full costs of roads.

Determining the cost to recover from users is problematic. In Australia, the pay-as-you-go or PAYGO method is used to calculate heavy vehicle charges. The annual cost of investment and maintenance for each year is taken as the cost for the year. If the distribution of pavement ages across the system is reasonably uniform, the annual PAYGO amount will be fairly stable. The annual PAYGO amount will be greater to the extent that the system is growing because, under PAYGO, capital expenditure accrues in the year in which it is incurred. If the rate of growth is expected to slow in the future, there might be a case for deferring some cost recovery to reduce the amount of cross-subsidy from present to future users. Having a set of accounts with annual depreciation charges is a way to spread capital costs over time. However, there is a great deal of arbitrariness in setting the values of existing assets, for example whether the historical or depreciated replacement cost method is used. There are arbitrary decisions to be made about how to treat assets such as earthworks that never have to be replaced and how far back in time to go.

In calculating the costs to recover from users, there is no economic justification for adding on a 'return on capital' to charge to users on top of depreciation. Indeed, it is debatable whether all road assets should be included in a cost recovery calculation. Unlike for privately produced and sold goods and services, the user has no control over what is supplied and the costs of such supply. If a private firm produces a product that consumers are unwilling to pay for, the private firm incurs a financial loss. High cost private producers can be undercut by competitors and not recover all their costs. The private sector may not be able to recover costs of mistakes and inefficiency. Hence, a system that passes all capital costs onto road users cannot be considered to emulate commercial supply.

A significant amount of investment and maintenance spending on roads, particularly in rural and remote areas, is likely to be 'community service obligations', where uneconomic roads are provided at higher levels of service than could be justified on economic grounds for social, equity and safety reasons. Ideally, such spending would be made transparent and paid for out of general revenue so governments can weigh up the social value of the spending against other priorities.

Road–rail competition

For most of the road network, there are no parallel railway lines and no modal competition. But for the major corridors in the national road network, the two modes are, to some extent, substitutes (BITRE 2009). As these roads are also the most heavily trafficked non-urban roads for trucks, they are built with the strongest pavements and hence have the lowest road-wear costs.

Like roads, rail infrastructure has strong economies of scale and lumpiness of investment; however, it is operated on a broadly commercial basis. It must charge above its short-run marginal costs to

achieve cost recovery. The Productivity Commission (2006) found that, in practice, rail fails to cover its full economic costs on many routes. Requiring trucks to meet only the short-run marginal costs of the infrastructure they use would raise concerns about competitive neutrality with rail. If trucks were charged only for the damage they do roads, there could be a significant shift in mode share from rail to road on routes where they compete. In the short term, the increase in road freight could lead to greater congestion, environmental costs and accident costs. In the longer term, there may be increased need for investment in road capacity and closure of railway lines.

To solve the problem, one approach would be to subsidise rail and have both modes charge at short-run marginal cost. However, subsidising rail could lead to other problems. In particular, it could reduce incentives for efficient management and operations. Another solution is to charge heavy road vehicles above the costs of road-wear to the extent necessary to ensure a reasonably efficient modal split for freight, subject to the constraint that rail covers its full costs (AFTS

In practice, it is extremely difficult to estimate the amounts to add on to heavy vehicle road-wear costs to achieve the most efficient modal split. An approximate rule is that the ratios of the prices paid for infrastructure usage to short-run marginal costs should be similar for the two modes. In other words, the percentage mark-ups on vehicle/train-related wear and tear on roads/tracks to contribute to capital and fixed infrastructure operating costs should be not be too different.

The existing heavy vehicle charging system allocates part of the fixed costs of the road system to heavy vehicles on a vehicle-kilometre basis, in addition to the attributable road-wear costs. While the resultant mark-up on top of road-wear costs is not in any sense 'economically optimal', it probably leads to better intermodal allocation of resources on routes where road and rail compete than would pure short-run marginal cost pricing approach applied to road alone. Development of any alternative approach to heavy vehicle charging should take into account the implications for resource allocation between road and rail.

Ways to cost recover

Taxes on land are efficient because land is immobile. Unlike labour or capital, it cannot be shifted out of supply to escape tax. A broad land value tax reduces the price of land but it does not affect how land is used or how much is used. Local roads, funded by rates paid to local governments are, in effect, already funded in part by land taxes. For local roads that provide access to properties, there is likely to be a high level of willingness-to-pay by property owners for high standard roads with curbs and drains that would not be picked up by the standard CBA methods that take into account only traffic-related benefits. Land taxes are therefore an equitable and efficient way to pay for local access roads.

If road users are to pay for part or all the cost of roads above economically optimal charges for congestion and road damage, the most economically efficient way to do it is through 'Ramsey pricing' or 'charging what the traffic (or the market) will bear' (Baumol and Bradford 1970). Users are divided into a number of classes and higher charges are levied on classes with that can afford to pay more. Ramsey pricing spreads the charging across users with the aim of minimising the negative effects on welfare from reduced consumption.

Where the demand curves for the user groups are independent, the formula for the optimum set of prices is:

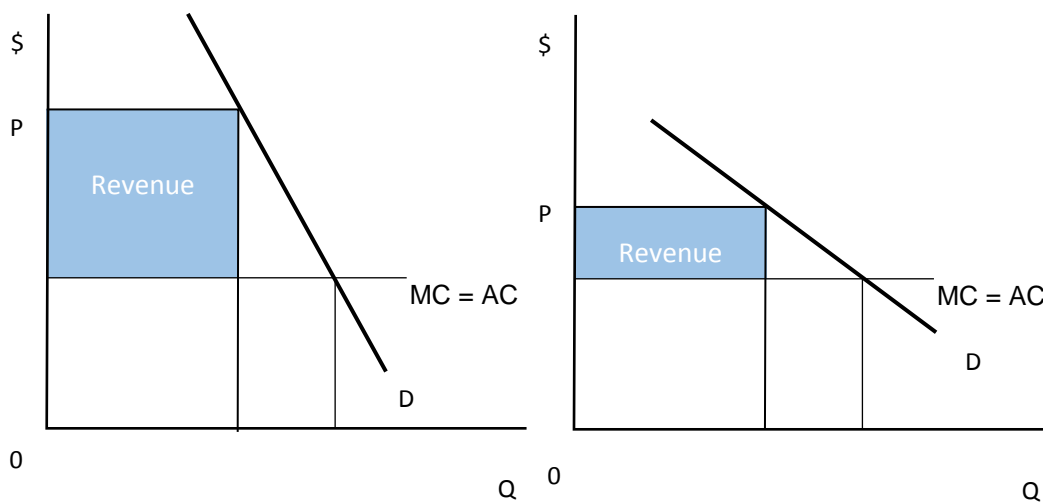
$$\frac{p_i - MC_i}{p_i} = -\frac{k}{\varepsilon_i}$$

for all i , where the subscript refer to the i th group of consumers, ε_i is the elasticity of demand (a negative number) for the i th group, and k is a constant set above zero to the degree necessary to raise the additional required revenue. The value of k cannot be set above one as this is maximum amount of net revenue that can be raised. Above k , the negative effect of higher prices on quantities demanded becomes dominant.

Figure 7 shows demand curves for two groups of users, the one on the right being more elastic or price sensitive. Costs are assumed to be constant, so marginal and average costs are the same for all quantity levels. The Ramsey price is higher for the group with the less elastic demand because they can absorb a higher charge with a smaller reduction in quantity demanded.

If the demand curves are not independent, that is, an increase in the price for one user group causes some leakage of demand to other user groups, Ramsey prices are set at the levels that cause each group to reduce its consumption by the same proportion.

Figure 7 Ramsey prices for two groups of consumers



The idea of Ramsey pricing is sometimes dismissed on the grounds that it requires knowledge of the demand elasticities for different user groups (Interstate Commission 1990). The counter argument is that rough Ramsey prices are preferable to setting prices on the basis of ‘fully distributed’ costs in terms of economic efficiency (Productivity Commission 2006). Fully distributed costs involves averaging costs not directly attributable to particular classes (total costs minus vehicle-related road damage) over all users by some measure such as vehicle–kilometres, gross vehicle mass tonne–kilometres, vehicle-years or gross vehicle mass years. Fully distributed cost allocations such as these are arbitrary accounting rules that have no justification in economic theory. They may be politically attractive because of their simplicity and seeming precision. A vehicle–kilometre cost allocation is used in Australia for setting heavy vehicle charges.

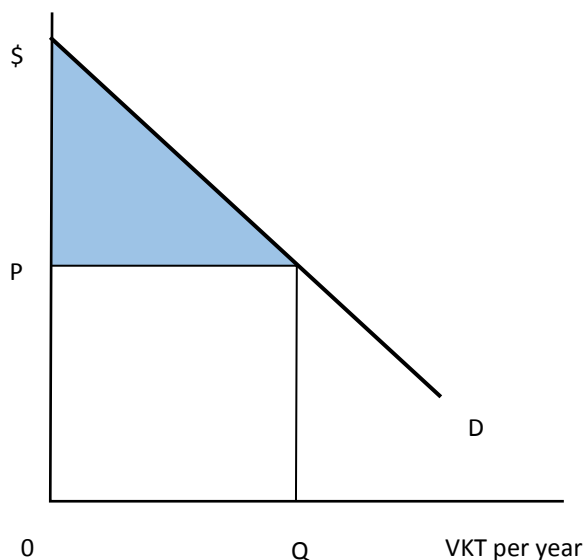
Demand elasticities would be lower, and hence Ramsey prices higher, for larger vehicles because they cost more to purchase and operate. A charge that is a constant percentage of costs would rise with vehicle size but at a decreasing rate due to economies of scale. Charging systems that try to approximate a uniform percentage of vehicle costs could be considered as a basic form of Ramsey pricing. An example is the Meteorological Service Charge levied by Air Services Australia on aircraft. For aircraft with a maximum take-off weight (MTOW) above 20 tonnes, the charge is proportional to distance flown and the square root of MTOW. This would approximate the cost curve for aircraft with respect to aircraft size.

Fuel taxes could also be considered as crude Ramsey prices. As well as being related to distance travelled, they bear some relationship to ability to pay because larger vehicles with higher operating

and capital costs pay more in absolute terms. However, fuel consumption is also affected by other factors unconnected with ability to pay, and the economies of scale for total vehicle costs would be greater than for fuel costs alone (for example, a vehicle requires one driver regardless of size) (Harvey 2015).

Raising a certain proportion of the costs of roads through annual registration charges, which are the same regardless of distance travelled, could improve the efficiency with which revenue is raised. Registration charges extract part of the ‘consumers’ surplus’ of individual consumers. Consumers’ surplus is the amount consumers are willing to pay for a given quantity of a good or service, over and above what they actually pay. Figure 8 shows the demand curve for a single road user, with vehicle–kilometres travelled (VKT) per year plotted on the horizontal axis. The price per kilometre, P , consists of vehicle operating costs, time costs and distance-related charges. For each VKT between 0 and Q , the surplus is the vertical distance between the demand curve (what they would be willing to pay) and the price (what they actually pay). The consumers’ surplus triangle is the sum of these vertical distances for all units from 0 to Q . As long as the registration charge for an individual does not exceed their consumers’ surplus area, there will be no impact on VKT, apart from a small across-all-commodities reduction in consumption caused by their reduced income after paying the registration charge.

Figure 8 Consumers’ surplus for an individual road user



Once the annual charge exceeds the consumers’ surplus for an individual, they will decide not to own a vehicle. For some low income earners and people who drive only a small number of VKT each year, even a small registration charge could affect their decisions. So there will be some economic welfare loss from registration charges. The welfare loss can be minimised by taking a Ramsey

approach and having the charges vary with ability to pay. Examples of how this might be implemented include higher rates for larger cars as measured by the number of cylinders, higher charges for trucks by size category, and discounted registration for pensioners, the unemployed and low income earners. The welfare losses from registration charges need to be balanced against the welfare losses from distance-related charges in determining the proportions of revenue to be raised from registration and distance-related charges (Brown and Sibley 1986).

Applicability of commercial investment criteria

Proponents of road pricing reform sometimes claim that revenue raised from individual roads will provide signals to guide investment, providing a check that CBA is working well in guiding decision-making or replacing CBA altogether.

The combination of constant returns to scale and the Mohring–Harwitz theorem for major urban roads means that the goal of using commercial investment criteria to guide investment decisions may be achievable, in theory, for those roads. To avoid charges being set at monopoly levels, a commercial supplier would have to be regulated. If a regulator required a commercial supplier of congested urban roads to charge economically optimal prices, the optimal level of investment could, in principle, be found at the point where costs are exactly recovered. As traffic volumes grew on a particular road, the regulator would allow congestion charges to move upward. If revenue exceeds interest and maintenance costs for the road, it would signal that the road should be expanded. If revenue falls short of the interest and maintenance costs, traffic should be allowed to increase until congestion costs and hence the price charged rises to cover the costs (Newbery 1994).

For uncongested roads, there is no link between profitability and optimal investment. CBA is the only way to determine optimal investment. The right sides of figure 9 show a relationship between spending on the road and road user cost per kilometre travelled. The more that is spent on a road, the smoother, wider and straighter it becomes, causing users' time and vehicle operating costs to fall (C per vehicle) until no further cost reductions to users are possible, at which point the curve becomes flat. A charge τ per vehicle is levied to raise sufficient revenue to cover annualised expenditure on the road so that annualised expenditure equals τQ . The left sides of figure 9 show how the number of vehicles on the road, Q , is determined by the cost incurred by users, $C + \tau$, via the demand curve.

Figure 9a shows the road with a high charge and a high amount of revenue collected, which is sufficient to provide a high standard road with low user costs. Figure 9b shows the road with a low charge and low revenue, which is only sufficient to provide a low standard road with high user costs. Both are commercially viable as users pay for the exact cost of the road.

Indeed, figure 9 could be drawn with a large number of possible combinations of user charge and road standard at which revenues exactly equal costs. Over the feasible range, any given charge will be associated with a road standard at which costs are exactly covered, or conversely, any given road standard will be associated with a level of charge at which costs are exactly covered. Without undertaking a CBA, there is no way to tell which road standard–charge combination is best from an economic efficiency viewpoint. It is therefore impossible to make economically optimal investment and maintenance decisions for the uncongested roads that comprise most of the length of the road system using financial or commercial criteria (Harvey 2015).

Figure 9a Uncongested road — high price and high standard

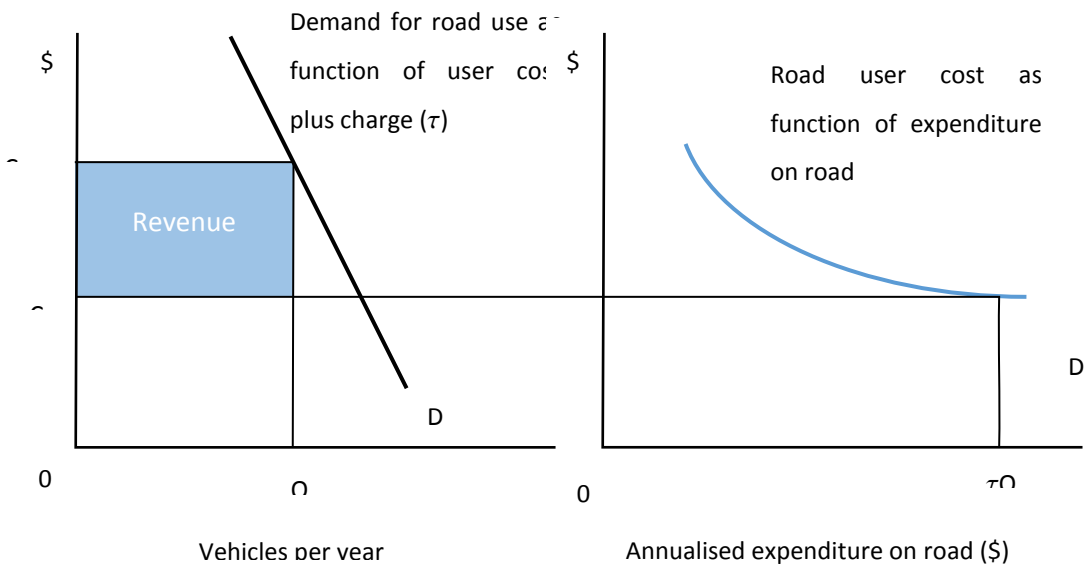
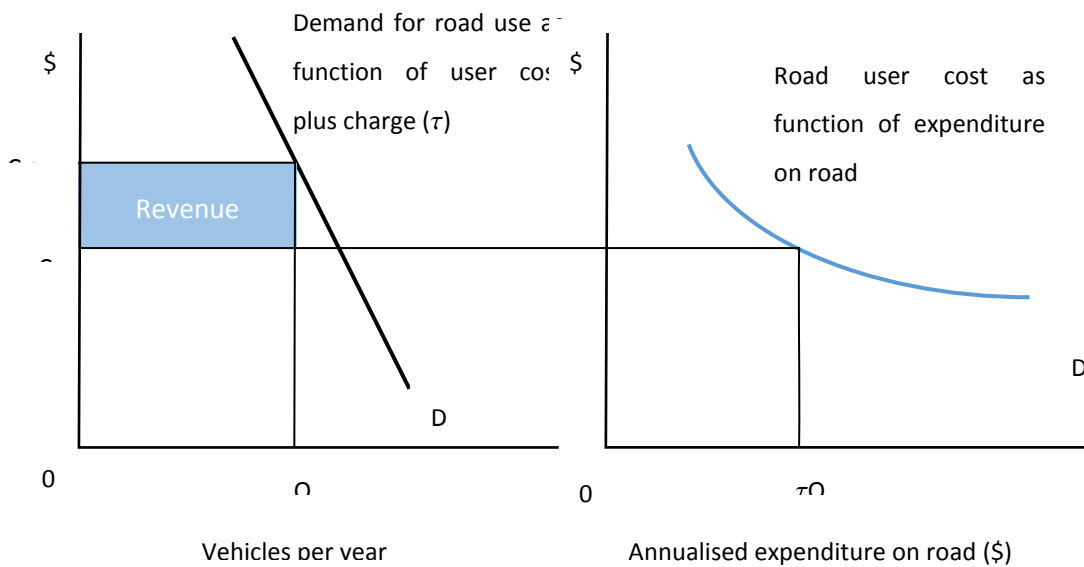


Figure 9b Uncongested road — low price and low standard



Commercial models of road supply

Once the optimal standard of road is known, whether through CBA or by comparing revenues with costs on congested urban roads, the question can be considered about commercial models of road supply. The term 'commercial' here means a supplier who aims to maximise profits.

To date, road commercialisation has been limited to newly constructed toll roads that are either major urban arterials or interurban highways — roads at the highest level in the hierarchical structure of road networks. There are no examples of widespread commercialisation of existing roads or of lower-level roads. The complexity of the legal and administrative arrangements that govern road supply could make widespread private ownership impractical, but a public utility with commercial objectives might be possible if good incentives for efficient management and investment could be put in place.

In addition to price, the economic problem for road commercialisation is to secure a good outcome in terms of quality. For other network industries (electricity, gas, water, telecommunications), regulation is primarily concerned with limiting monopoly pricing. Regulating service quality is of secondary importance because there is not a large range of quality levels that are economically efficient and commercially viable under different circumstances. For individual roads, a very wide range of service qualities can be warranted under different circumstances, from an earth track to a multilane freeway.

In the absence of regulation of charges, a commercial supplier would take advantage of the monopoly position of a road or a network of roads and charge above costs. It is not clear how the standard of road supplied by an unregulated profit maximising monopolist would compare with the socially optimal standard. The reason for the lack of certainty is that the standard of road supplied by an unregulated profit-maximising monopolist is affected by the distribution of users' values of time and the extent of economies of scale.

There is no lack of certainty about the road standard supplied when, in order to prevent monopoly charging, a regulator fixes the charge. Fixing the monopoly pricing problem by having a regulator set the user charge creates another problem, under-provision of service quality. A monopoly supplier of a single product will always supply less than the economically optimal quality when the firm is required to sell its product at a fixed price because a price ceiling prevents the firm from capturing any of the benefits to consumers that a higher service quality would engender. For a congested road with enforced optimal charging, the problem is worse because at lower capacities, the optimal congestion charge would be higher. The supplier would be rewarded for under-supply of capacity with a higher charge, as well as saving costs.

Hence, any road supplier with commercial objectives would have to be regulated for both price and service quality. Regulation of service quality could be undertaken by specifying infrastructure characteristics such as road widths, shoulder widths and lane numbers, curvatures, gradients and maintenance works. Modern approaches to service quality regulation specify performance levels

and leave it to the supplier to determine how best to meet its performance targets. Contracts include financial penalties for under-performance, and possibly also bonuses for meeting or exceeding targets. In the case of roads, performance targets might include minimum average actual speeds achieved by users (up the legal speed limit), maximum crash rates and maximum roughness levels (Harvey 2015),

Conclusion

The main message of this paper is to challenge ideas that economically efficient prices for roads in general are associated with cost recovery, and the roads could be managed efficiently on a commercial basis without a high degree of regulation.

Even calculating the amount to cost recover from users involves arbitrary assumptions about the treatment of costs incurred in the past and questionable inclusion of spending to promote equity objectives or the result of poor decision making. The only credible justification for attempting to recover 'exact' total costs from users is the 'user pays' concept of equity. Because a loss-making road system necessitates increased taxation, which has efficiency costs (land taxes aside), there is a strong case for raising road charges above marginal costs to cover at least part of the deficit. Ideally, the charges on road users would be levied in a way that minimises efficiency costs taking account of the administrative costs of greater complexity. If revenue from road users could be raised with a lower deadweight loss than from other sources of government revenue, there might even be a case for earning a surplus so that other taxes could be reduced.

Any attempt move toward commercial road supply, for example, to set up a road supply agency with some degree of independence from government and having a cost recovery objective, requires a great caution. The road supplier has strong incentives to take advantage of its natural monopoly position to over-charge on price and under-provide on service quality. A detailed, well-designed regulatory system would be needed.

References

Australia's Future Tax System (Henry Tax Review) 2010, *Final Report*, Section E3 Road transport taxes. <http://taxreview.treasury.gov.au>

Baumol W.J. and Bradford D.F. 1970, Optimal departures from marginal cost pricing, *American Economic Review*, 60(3):265–283.

Brown S.J. and Sibley D.S. 1986, *The theory of public utility pricing*, Cambridge University Press, Cambridge.

Bureau of Transport and Regional Economics [BTRE], 2007, *Estimating urban traffic and congestion cost trends for Australian cities*, Working paper 71, BTRE, Canberra ACT.

Bureau of Infrastructure, Transport and Regional Economics [BITRE] 2008, *Moving urban Australia: can congestion charging unclog our roads?*, Working paper 74, BITRE, Canberra.

Bureau of Infrastructure, Transport and Regional Economics [BITRE] 2009, *Road and rail freight: Competitors or complements?*, Information Paper no. 34, BITRE, Canberra.

Coase R.H. 1946, The marginal cost controversy, *Economica*, 13(51):169–182.

Coase R.H. 1970, The theory of public utility pricing and its application, *Bell Journal of Economics and Management*, 1(1):113–128.

Harvey M.O. 2015, Commercial road supply with incentive regulation, *International journal of sustainable transportation*, 9(4):241-53.

Inter-State Commission 1990, *Road use Charges and Vehicle Registration: A National Scheme: Volume 1*, March, AGPS, Canberra.

Newbery D.M. 1994, The case for a public road authority, *Journal of Transport Economics and Policy* 28(3):235-53.

Productivity Commission 2006, *Road and rail freight infrastructure pricing*, Report no. 41, Productivity Commission, Canberra

Oehry B. 2010, *Critical Success Factors for Implementing Road Charging Systems*, Prepared for the ITF/OECD Round Table of 4-5 February 2010 on Implementing Congestion Charging, Joint Transport Research Centre Discussion Paper No. 2010-3, International Transport Forum/ OECD, Paris, France.

Small K.A., Verhoef E.T. 2007, *The Economics of urban transportation*, Routledge London

Small K.A., Winston C. and Evans C. 1989, *Road Work: A New Highway Pricing and Investment Policy*, The Brookings Institution, Washington, DC.

Verhoef E.T., 2007, Second-best road pricing through highway franchising, *Journal of Urban Economics*, 62(2):337-361.

Verhoef E.T., Mohring H. 2009. Self-financing roads, *International Journal of Sustainable Transportation*, 3(5):293-311.

Walters A.A. 1968, *The Economics of Road User Charges*, International Bank for Reconstruction and Development, World Bank Staff Occasional Papers No. 5, Johns Hopkins Press, Baltimore.

