Chapter 19 Economic Infrastructure: Transport

It is good to be out on the road, and going one knows not where.

John Masefield, *Tewkesbury Road*

Use and Pricing of Transport Infrastructure ♦ Investment in Transport Infrastructure ♦ Role and Regulation of the Private Sector in Transport

There are many forms of economic infrastructure. In this chapter, we discuss transport as an example of the economic issues arising in the use, supply and management of infrastructure. With any infrastructure, there are two key issues: how to use the existing infrastructure and how to augment it.

Using transport infrastructure efficiently is essentially a pricing issue. When traffic volumes are low, use of infrastructure is non-rival and efficient prices may be low. However, prices should allow for externalities such as congestion, accidents, noise and air pollution. On the other hand, when capacity is limited, high prices are efficient. But introducing equity, low public transport prices may reduce social inequalities and exclusion (Stanley *et al*. 2011).

Regarding investment, standard cost-benefit principles apply. But special issues arise as transport infrastructure may provide economic benefits that are not reflected in user benefits.

Turning to management, most railways, ports, airports, pipelines and some major roads are natural monopolies. These market failures provide a strong case for government involvement. Nevertheless, there may be a role for private investment or management.

In this chapter, we discuss each of these issues in turn: use of existing infrastructure and the role of pricing; investment in transport infrastructure; and the role and regulation of the private sector in supplying and operating transport infrastructure.

Use and Pricing of Transport Infrastructure

Following our core principle for efficient use of capital, use of transport infrastructure is efficient when the net social benefit (NSB) from use is maximised. NSB is simply:

$$NSB = \sum_{i=1}^{m} B_i - \sum_{j=1}^{n} C_j \tag{19.1}$$

where B is benefits, there are i = 1…m benefits, C is costs and there are j = 1…n costs. Most benefits accrue to users of the infrastructure though there may also be third-party benefits. NSB is maximised and infrastructure is used efficiently when the marginal benefit of use equals the marginal opportunity cost inclusive of external costs.

Prices signal how transport infrastructure can be used most efficiently. General pricing principles for public services were described in Chapter 17. We summarise below the main points with special reference to transport infrastructure before describing some applications in the transport sector.

General pricing principles

The default principle for pricing public services is short-run marginal cost (SRMC) pricing, i.e. setting user charges equal to the marginal cost of the service (see Box 19.1). This maximises the NSB from use of infrastructure. If use is non-rival and there is no marginal cost of use, there would be no user charge. If there is a cost of use, including externalities such as congestion costs, the user charge should reflect the marginal social cost of use.

However, this assumes either no significant funding shortfall to cover fixed costs or that a funding deficit can be funded at little deadweight cost. If these assumptions do not hold, a transport agency may be required to meet a revenue target, such as full cost recovery. In this case, it would raise the extra revenue with minimum distortion of transport activity by adopting price mark-ups or multi-part tariffs as described in Chapter 17.

Price mark-ups may also be appropriate if competitive services are priced above SRMC. It would be inefficient to price public services at marginal cost and attract custom away from a private operator that can supply similar services at lower cost. Equation 19.7 in Box 19.1 provides the pricing formula that maximises NSB in these circumstances.[1] The price mark-up for the subject service depends on the size of the mark-ups and the cross-price elasticities of demand for the substitute services. Pragmatically, if private services are close substitutes (the cross-price elasticities are high), long-run marginal cost (LRMC) pricing is generally appropriate for the public transport service.

Road user charges

There are various kinds of road user charge. The purest form is a charge per vehicle kilometre for road use at a regulated time. Closely related charges are road tolls and cordon charges that permit access into a specified area at regulated times. Other charges include area licences that permit entry into an area over a longer period or fuel taxes.

The alternatives to road user charges are generally less efficient ways to ration road space because they do not focus on the spatial and timing problems of congested roads. Selective toll roads and cordon charges encourage drivers to choose less efficient routes or less preferred destinations. Fuel taxes are not related to use of congested roads. Other charges such as vehicle registration fees or sales taxes may contribute substantially to road construction and maintenance, but they are fixed charges and not charges for road use. Fixed charges may have a minor effect on vehicle ownership, but have little effect on road use.

Following our previous discussion, the efficient pricing principle is that the road user charge should reflect the marginal social cost of use. If a road is not congested and use is non-rival, efficient use requires no charge unless there is road damage or negative external effects (such as safety risk or air pollution). Most passenger vehicles do minimal pavement damage. The damage increases at the third or fourth power of axle weight. Given that the axle weight on a car is typically about 450 kilograms and the axle weight on a truck is typically some 8000 kilograms, a truck axle causes about 6000 (18^3) times as much road damage as a car axle. Trucks are also involved in a disproportionate number of fatal accidents. Thus, there is a strong case for road user charges for trucks even when there is no congestion.

[1] This is equivalent to Equation 17.2.

Box 19.1 Pricing principles for transport services[2]

Suppose that net social benefit (NSB) from a transport service is expressed as:

$$NSB = \int_0^Q P(X)\,dX - Q.AC(Q) \qquad (19.2)$$

where $P(X)$ is the inverse demand curve, X is a variable of integration, Q is the volume of trips and $AC(Q)$ is average cost which varies with trips made. The first term on the right hand side is customer willingness to pay for services (the area under the demand curve) and the second term is the cost of the service, which is here borne by the service provider.

To maximise NSB, the first derivative of Equation 19.2 with respect to Q must equal zero. Rearranging the result:

$$P = AC + Q.(dAC/dQ) \qquad (19.3)$$

Equation 19.3 states that the price of a service must equal the sum of the average cost of servicing a user and the change in this cost from servicing an extra user.

Now suppose that, as with a road, the users provide and pay for their own services and bear these costs. *NSB* can now be expressed as:

$$NSB = \int_0^Q P(X)\,dX - Q.AUC(Q/L) - LCC(L) \qquad (19.4)$$

where AUC is the average cost to a user of operating a vehicle on the road, L is the number of units of capacity (such as lanes on a highway), CC is the average amortised cost to provide a unit of the facility and the other variables are as before. User costs rise with congestion and so depend on the ratio of (Q/L).

To maximise NSB from use of the infrastructure, the level of capacity is held constant and the first derivative of Equation 19.3 with respect to Q is set to zero. Rearranging the result, the following is obtained:

$$P = UC + Q.(dAUC/dQ) \qquad (19.5)$$

The efficient road user charge is equal to the second term on the right-hand side of the equation. This is the change in average user cost as a result of serving an additional user, which is the marginal cost at the efficient level of road use.

If these efficient prices do not cover total cost, Ramsey's formula to raise sufficient revenue at least cost (see Chapter 17) is:

$$(P_i - MC_i)/MC_i = k/\eta_{di} \qquad (19.6)$$

where P_i is the mark-up price for customer i, MC_i is the marginal cost of serving customer i, η_{di} is the compensated price elasticity of demand for customer i and k is a constant determined by the revenue required to meet the budget.

Price mark-ups may also be appropriate in second best cases where competitive transport suppliers charge in excess of marginal cost. The efficient price here is:

$$P_i = MC_i - \Sigma_j\,[\eta_{ij}/\eta_{di}).(Q_j/Q_i).(P_j - M_{Cj})] \qquad (19.7)$$

where i is the subject service, j is the competitor(s) service, η_{ij} is the cross-price elasticity of demand for good j with respect to the price of good i and the other variables are as above.

When roads are congested, efficient road use requires that the user charge should reflect the congestion cost caused by the marginal road user. This is demonstrated in Figure 19.1 overleaf. The generalised cost (GC) of a trip is shown on the vertical axis. GC is the sum of costs borne by the trip-maker, including the cost of travel time, trip time reliability, vehicle operating costs and any road tolls. The demand for trips rises as GC falls. The marginal private cost (MPC) per trip, measured in terms of generalised trip costs, increases when congestion occurs. When congestion occurs, the marginal social cost (MSC) of a trip exceeds MPC because each vehicle entering the traffic stream imposes a cost on other vehicles, which increases with traffic volume.[3]

In the absence of road pricing or controls on road use, there will be Q_M trips per period at an average trip cost of \$*D*. But at this traffic level, MSC exceeds MPC. The efficient quantity of trips is Q_E, the point at which the marginal travel benefit equals *MSC*, where the average trip cost is \$*G*. The deadweight loss of congestion equals area *BCF*. This is the difference between the marginal social cost of the excess trips and the benefits of these trips. This can be avoided by setting a road user charge equal to the marginal external cost *at the efficient level of traffic* (equal to *BH* in Figure 19.1). Ideally the charge would vary with changes in congestion over time.

2 The presentation in Box 19.1 is based on Gomez-Ibanez (1999). This presentation does not include externalities.
3 For ease of exposition, other externalities of trip making are ignored here.

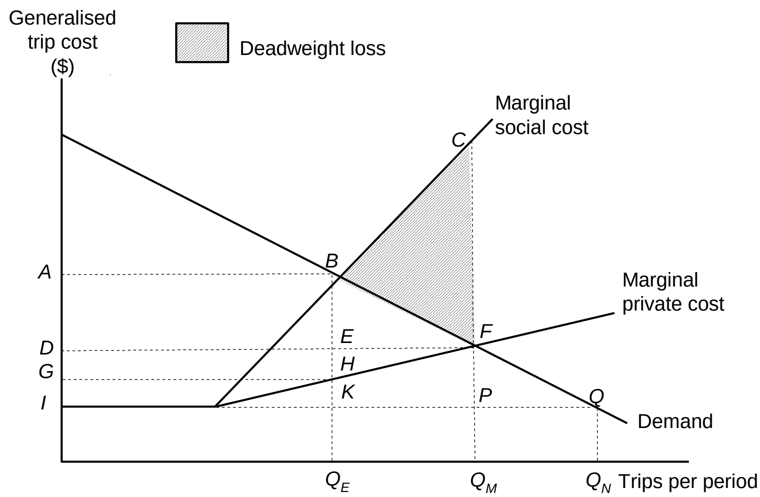


Figure 19.1 Traffic congestion costs

Efficient pricing has two main benefits. First, it eliminates excess congestion costs equal to area *BCF*. Second, it may defer the need for new infrastructure. When existing capacity is used inefficiently, there is greater need for increased capacity. In relation to Figure 19.1, suppose an investment in road capacity increases road use from an inefficient level Q_M to Q_N and that there is no congestion at Q_N. The benefits would equal areas *DFPI* + *FPQ*. On the other hand, if road use is efficient, road use increases from Q_E to Q_N. The benefits of the increased road capacity now equal areas *GHKI* + *BQK*. Although there is a greater increase in road use, the benefits from the increase in capacity are lower because the existing capacity is better utilised. Accordingly, new investment may be deferred. On the other hand, if inefficient road use is endemic, increasing capacity does produce the benefits shown above, albeit as a second-best solution.

However, road user charges present two challenges. One is to introduce a network-wide electronic tagging (GPS) system. Selective charging for roads encourages inefficient road use as some road users choose inferior free roads. The second challenge is political. If there is high congestion, the revenue from road user charges (area *ABHG* in Figure 19.1) may greatly exceed the benefits to continuing road users (area *DEHG*). This does not affect the NSB from a road user charge because, in cost–benefit accounting, the revenue is a transfer payment rather than a resource cost. However, the excess of revenue over benefit may create a political problem unless a popular strategy can be devised such as using the funds for public transport or reducing vehicle registration fees.

Prices for public bus and rail services

Pricing features prominently in debates about public bus and rail services. This reflects concerns about access, large deficits and efficiency. A central issue is again whether prices should be set equal to short-run marginal cost or perhaps even below SRMC or to recover all costs. Here we consider three arguments for SRMC or even sub-SRMC pricing: scale economy, second-best and equity arguments. We then note some arguments against subsidies.

Where there are fixed costs and some scale economies, efficient use of bus or rail would require SRMC pricing. However, the efficiency benefit of SRMC may be small for two main reasons. First, scale economies may not be large and marginal cost may not be much lower than average cost. Gomez-Ibanez (1999) estimated that the marginal cost of rail service was usually at least 70 per cent of average cost. For buses, including congestion effects of buses, marginal social cost may exceed average financial cost. Second, efficiency gains from SRMC pricing are usually small because demand for most bus and rail services is price inelastic. Typically, the direct price elasticity for urban public transport is about –0.3 (Goodwin, 1992).

Turning to second-best issues, the NSW Independent Pricing Tribunal (2015) proposes that rail and bus services should be priced at below average cost when private road users do not pay for the marginal social cost of road use especially in peak hours. Substitute services should be under-priced to attract road users and reduce the negative externalities of road use. However, this argument should be treated cautiously. In many places, private motorists pay high fuel excise changes, road tolls and parking charges which may partly, or wholly, substitute for road user charges. Importantly, the cross-price elasticity of demand with respect to private vehicle use is usually low. When a rail or bus operator lowers the price of their services, they attract patronage mostly from other public transport modes and some new passengers (which creates a deadweight loss); they may attract few road users. When roads are under-priced, raising road user charges is more efficient than a second-best solution (reducing bus and rail fares) and, if feasible, should be preferred.

Equity considerations may support transport subsidies, including pricing below SRMC. Many bus and rail users have low incomes. And some trips, to work or medical services, may be viewed as necessities. However, if government wishes to subsidise some public transport users, travel vouchers to targeted individuals would be more efficient than general trip subsidies. Vouchers allow trip makers to choose their preferred mode. Also, they do not distort price relativities for most users. Reducing transport fares below marginal cost for all users is not efficient and may not be equitable.

On the other hand, there are arguments for full-cost recovery pricing of public transport services. Loss of revenue may reduce quality of service unless government supplies adequate deficit funding. Management disciplines are hard to maintain. If subsidies are provided, they may be absorbed into higher cost structures rather than improved services. Moreover, revenue deficits must be financed from tax revenues that themselves have deadweight costs.

In summary, SRMC pricing for rail or bus services is generally efficient because it maximises use of services. Second-best and equity considerations may justify below-SRMC pricing in some places. However, low prices may encourage inefficient use. Also, SRMC pricing may create management issues and generally involves raising taxes to fund deficits. Where services are competitive and cross-price elasticities are high, competitive neutrality (efficiency) principles suggest public supply should be based on full cost recovery.

Ideally pricing policy would be based on estimates of the benefits and costs of alternative pricing structures. Glaister and Lewis (1978) estimated optimal second-best fares for London rail and bus services separately for peak and off-peak services. Allowing for operating costs, external pollution and congestion costs of all modes, own-price and cross-price elasticities of demand, they found that peak fares for London rail and buses were too low and that off-peak fares were too high. Similar exercises could provide useful information to decision makers elsewhere. However, this is a complex exercise.

Charging for airports

The principles for efficient use of airports are similar. However, their application involves different issues.

Airports provide various runway, terminal and related services. Airport managers recoup expenses by landing fees, rental charges for use of terminals and parking charges which reflect monopoly powers. Landing fees are usually based on aircraft weight. This may be reasonable in off-peak periods. Aircraft weight reflects potential runway damage, although this is small for all aircraft, and may be viewed as a proxy measure for the number of passengers and amount of freight using the terminal facilities.

However, critically, in peak hours runway capacity is the limiting factor on airport capacity. Weight-based costs are not then an efficient basis for airport charges. When an aircraft lands or takes off in congested conditions, it imposes delay costs on other traffic. Slow aircraft, usually small aircraft, impose more delays and larger costs because they reduce the landings and take-offs that can be achieved in peak periods. When one aircraft displaces another, with no change in congestion, the real cost of using a time slot is the value of the slot to the displaced aircraft. If the displacement increases delays for other aircraft, the efficient charge is the sum of the price the displaced aircraft would pay for the slot and the additional congestion cost. Accordingly, the first step towards efficient use of scarce runway capacity is to establish charges that reflect these costs. An auction for time slots would ensure that the airline that places the highest value on a landing or take-off slot will obtain this slot.

These peak-hour landing fees may exceed average runway and terminal costs. However, from an efficiency perspective these fees should rise with demand for runway use until the excess demand for runway capacity justifies the cost of a new runway (see below). Efficient pricing produces an efficient use of existing capacity and creates signals for efficient investment in additional capacity.

A further issue is the monopoly position of most airports. This enables airport authorities (public or private) to extract economic rents from airlines and air passengers. This may warrant some form of price regulation (see the discussion below).

General conclusions

Marginal social cost pricing is an important efficiency principle. When transport facilities are not fully used, this pricing principle encourages use of the spare capacity. The inclusion of externality costs (see Box 19.2 on p. 331) ensures that trip makers consider the full costs of their travel. When transport facilities are heavily used, and especially in congested conditions, marginal social cost includes congestion costs. These prices allocate scarce capacity efficiently and provide an important signal for investment.

On the other hand, full cost recovery pricing ensures efficient allocation of resources when substitute services are charged at full cost, is competitively fair, encourages efficient management and avoids taxation funding of deficits.

Investment in Transport Infrastructure

Investment in infrastructure is efficient when the expected present value of the benefits of increased capacity exceeds the present value of the costs, inclusive of externalities. Benefits typically include cost savings for existing users of the infrastructure and benefits to new users. More precisely, the most efficient level of capacity is achieved when the present value of an extra unit of capacity just equals the cost of providing that unit. However, marginal adjustments are difficult to achieve when investments are lumpy (indivisible) as they are with road lanes, rail tracks and airport runways. Efficient timing is achieved when the net benefit of increased capacity exceeds the opportunity cost of capital employed.

Efficient investment is linked to efficient pricing. Low prices may create excess use of infrastructure before additional capacity is justified. Figure 19.2 shows existing capacity (S_1) and new capacity (S_2) with a lumpy asset such as an airport runway. Charging at SRMC, there may appear to be sufficient demand for a new runway even with D_1 level of demand. Actually, a new runway will be financially viable only when demand rises to D_3. In this case the efficient price with existing capacity would be P_3. However, a new runway could be justified with less demand than D_3, but with more demand than D_2. At a market clearing price for the new runway slightly below P_2, some users could gain consumer surpluses. If these surpluses were large enough to offset the loss to the airport authority, the net social benefit would be positive. Of course, the evaluation should allow for shifts in demand over time.

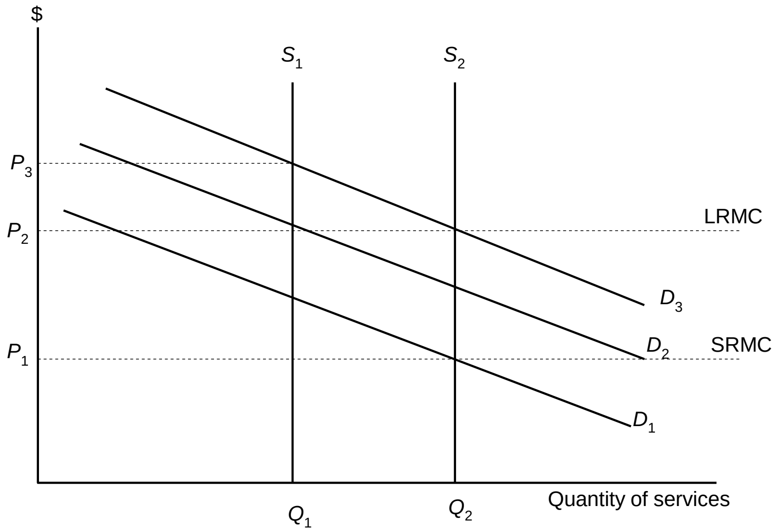


Figure 19.2 Pricing and investment

We describe below how to evaluate road, rail and airport infrastructure along with some examples. Slightly different issues occur in each mode.

Evaluation of road investment

The major costs of a road upgrade or new road are land, property construction and maintenance costs. There may also be environmental costs. On the other hand, nearly all the benefits are associated with lower road user costs, especially travel time savings.

As shown in Figure 19.3a overleaf, the demand for road trips between any two points in a period is a function of the generalised trip cost (GC). By far the largest component of GC is travel time. The value of travel time was discussed in Chapter 8 (Box 8.3). The UK Department of Transport (2010) and Hensher (2011) provide more detailed discussions of the value of travel time savings. GC also includes vehicle operating costs (VOC); these are the marginal VOC associated with the trip, which are largely fuel costs. User costs also include road user taxes and charges, if applicable.[4]

Figure 19.3a shows two sets of generalised cost (GC_1 and GC_2) representing user trip costs before and after a road upgrade. As trip cost falls from GC_1 to GC_2, trips increase from T_1 to T_2. Approximating demand by a linear schedule, the change in consumer surplus (ΔCS) is:

$$\Delta CS = T_1\,(GC_1 - GC_2) + (T_2 - T_1)\,(GC_1 - GC_2)/2 \qquad (19.8)$$

Equation 19.8 includes gains to existing users and estimated benefits of generated trips. The latter include trips diverted from other routes, modes and destinations, as well as new trips.

However, Equation 19.8 does not allow for changes in producer surplus. Generalised cost includes transfer payments, such as fuel excise and road tolls that do not consume resources. Panel (b) introduces ‘resource cost’ schedules (RC_1 and RC_2) before and after the road upgrade respectively. As the term implies, resource costs reflect the use of resources. This includes travel time and vehicle operating costs, but not transfer payments. The shaded area in

---

4 In principle, generalised cost should include an allowance for accident costs. However, these costs are often treated separately in road evaluations.

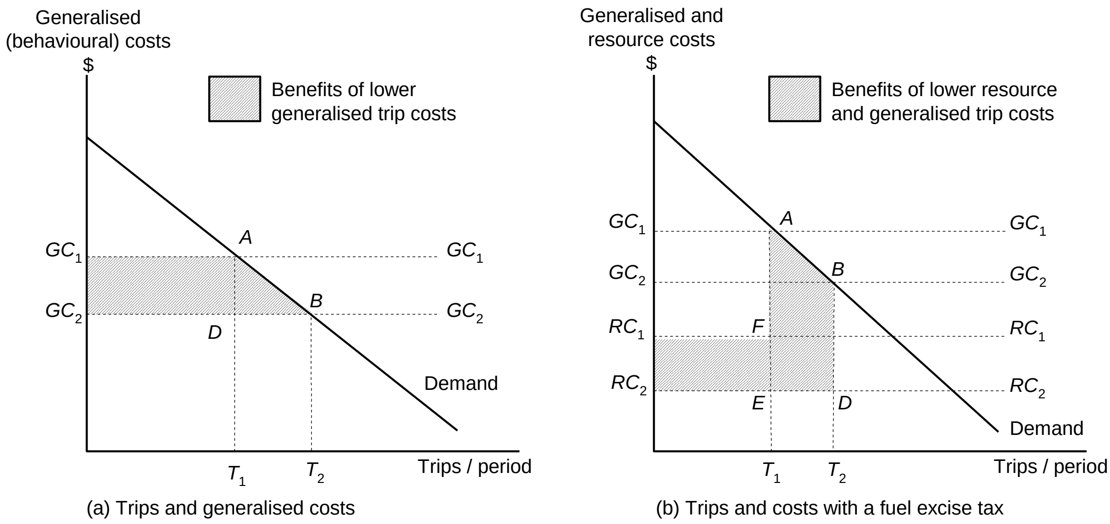


Figure 19.3 Valuation of road user benefits

panel (b) shows the net benefit of lower trip costs. The area equal to T_1 ($RC_1 - RC_2$) shows the savings in resource costs for existing trips.

Area *ABDE* is the benefit for generated trips. This equals the amount that trip makers would be willing to pay for the generated trips less the resource costs incurred. Including changes in producer surpluses, the benefit can be expressed as:

$$\Delta CS + \Delta PS = T_1\,(RC_1 - RC_2) + \Delta T\,(GC_2 - RC_2) + \Delta T\,(GC_1 - GC_2)/2 \qquad (19.9)$$

where ΔT equals ($T_2 - T_1$).

When a road is part of a network, these benefits are estimated over the network or at least the relevant parts of the network. Ignoring externalities, the estimated present value of surpluses (summed over all relevant network links) would be compared with the present value of the capital and operating costs of the improved network. Introducing a toll just for the new road would not change RC_1 or RC_2 but it would raise GC_2 and so reduce generated trips and the benefits from the new road. In practice, economic evaluations are often based on estimated cost savings for a given set of trips across a trip network. The benefits then include only the first term on the right-hand side of Equation 19.9 and exclude any benefits from generated trips. On the other hand, externalities such as traffic noise and air pollution costs may also be included in the evaluation (see Box 19.2 and case study below).

Cost-benefit example

A cost–benefit analysis (CBA) of a 3.4 km road tunnel (the Lane Cove Tunnel) connecting two motorways in north-west Sydney exemplifies some of these issues.[5] The cost and benefits and overall results are summarised in Table 19.1. The evaluation was based on forecast trips over the regional road network with and without the proposed tunnel over 29 years following a five-and-a-half-year construction period. No residual value was included.

---

[5] This summary is based on *Lane Cove Tunnel and Associated Road Improvements*, Working Paper 14, *Economics*, prepared by Sinclair Knight Merz (2001) for the NSW Roads and Traffic Authority.

Box 19.2 Environmental values in transport studies

There is a strong case for factoring the costs of environmental impacts into (1) the use of transport infrastructure via pricing and (2) investment decisions via cost-benefit analysis. These impacts include air pollution, climate change due to the emission of greenhouse gases, noise, impacts on nature, soil and water deterioration, severance and visual intrusion. However, valuations of damage vary with study and circumstances. Two estimates of the costs per passenger car vehicle km in cities from air pollution, climate change and noise are cited below.

Austroads (2003), a major Australian reference, draws on two international studies. The estimates by Bickel *et al.* (2006) are averages of results found for seven European cities. These indicate lower but still substantial costs. Costs per vehicle km will vary with fuel consumption which depends especially on the type of car and the amount of stopping and starting. The estimated costs per heavy goods truck are about 20 times the costs per passenger vehicle.

Estimated average costs of environmental damages in cities: cents per vehicle km by passenger cars

	Austroads (2003)	*Bickel et al. (2006)*
Air pollution	2.37	0.50
Global warming	1.69	0.80
Noise	0.79	0.35

Table 19.1 Economic evaluation of Lane Cove Tunnel: summary of results ($m in 2001-02 prices)[a]

	2002	2003	2004	2005	2006	2007	2008	---	2035
Costs									
Capital	10	31	232	271	194	77	0		0
Recurrent	0	0	0	0	0	7	7		8
Total cost	10	31	232	271	194	84	7		8
Benefits: savings in									
Vehicle operating costs						67	69		111
Travel time						211	215		294
Vehicle accidents						8	8		9
Environmental costs						27	28		37
Total benefit						313	320		451
Net benefit	-10	-31	-232	-271	-194	229	313		443
Discount rates	4%	7%	10%						
NPV ($m)	463	2711	1646						
BCR	6.6	4.7	3.5						

(a) The full results included annual estimates for each year between 2008 and 2035.

Source: Sinclair Knight Merz (2001).

As is common in such studies, but not very satisfactory, forecast trips were a function of fixed land uses and did not allow for any generated trips. In effect, the CBA estimated only the fall in total trip costs for a given overall trip matrix. To model choice of road, a $2 toll was assumed. All costs and benefits were estimated in 2001–02 prices. Costs and benefits were discounted at a real discount rate of 7 per cent, with 4 and 10 per cent rates used in sensitivity tests. Capital costs include property take, project development and construction.

The benefits include savings in vehicle operating costs, travel time, vehicle accidents and environmental costs across the regional road network due to lower congestion. Savings in vehicle operating costs exclude tax components and were estimated separately for trucks and cars. Savings in travel time were estimated separately for people travelling privately or for work purposes and for commercial vehicles and in peak and off-peak hours. Savings in environmental costs were based on reductions in emissions of NO_X, particulates and CO_2 and on reductions in noise impacts on households. Drawing on various studies, the report valued total savings in environmental costs at 4.2 cents per vehicle kilometre (vkm), which included air pollution costs at 1.4 cent per vkm, CO_2 emissions at 2.25 cents per vkm and noise costs at 0.55 cents per vkm.

Not surprisingly, given the strategic nature of this tunnel in the Sydney transport network, the study reported high estimated net present values and benefit–cost ratios. The positive results were not affected by alternative estimates for key variables such as construction costs, trip forecasts or values of travel time savings. The report did find that the proposed road toll, which would not apply to alternative routes, would reduce traffic using the tunnel by over 25 per cent, but it did not indicate how the toll affected the estimated net present value.

Evaluation of rail or bus services

The evaluation of rail or bus services is similar in many ways to evaluation of road services. Will user and third-party benefits justify the incremental capital and operating costs of the improved service? However, two special problems arise. One arises when the quality of the transport service improves in say comfort (reduced crowding) or frequency. There is often no market price measure of the value of this improvement. In this case, stated preference surveys (see Chapter 11) may be needed to determine what users would be willing to pay for this quality improvement.

Related evaluation issues arise with mode switching. Again, the appropriate valuation approach is to attempt to measure what mode switchers are willing to pay for the improved service. Evaluations often simply compare the costs of door-to-door trips by different modes on the assumption that the modes provide equivalent service levels. This tends to underestimate the benefits of private car travel, such as freedom to choose the time of travel,

Ideally user benefits of rail and bus services should be valued, like other services, by the amounts that people are willing to pay (WTP) for them. This avoids use of ad hoc premiums (or discounts) for service features such as frequency or comfort of service.[6] Suppose that WTP values for a rail and bus trip and marginal vehicle operating costs are as follows:

Mode	*WTP value ($)*	*Marginal cost ($)*
Rail	9.0	7.0
Bus	7.0	4.0

If passengers switch from rail to bus, there is a $3 saving in operating cost per passenger. If passengers value the services of the two modes equally, this represents the net benefit of the switch in mode. However, if passengers would be willing to pay $2 extra for rail, a switch to bus represents a net benefit of only $1. When a new mode is less attractive than an existing one, savings in resource costs inclusive of travel time overestimate the benefits of the new mode. Conversely, if the new mode is more attractive, savings in resource costs underestimate the benefits.

In general, where other modes set price (P) equal to marginal cost (MC), the net benefit of a new service is the difference between the amount that a user is willing to pay for it and the marginal resource cost (RC) of the service. If $P > MC$ in a substitute mode, the net benefit of a trip diverted to say rail from another mode is the difference between the benefit of the rail trip and the benefit of the trip on the substitute mode:

6 However, in the absence of data on willingness-to-pay, WTP values may have to be derived from estimates of ‘door-to-door’ travel times and other trip features. In any case, trip forecasts may be modelled on ‘door-to-door’ travel times and other trip features.

$$NB_{DR} = (WTP - RC)_R - (WTP - RC)_0 \qquad (19.10)$$

where NB_{DR} is the net benefit of the trip diverted to rail and subscripts R and O represent rail and other modes respectively.

Figure 19.4 shows the benefits of trips due to an improved rail service. In panel (a) demand for rail trips increases from D_1 to D_2 and rail trips rise from Q_1 to Q_2. Panel (a) also shows the consumer surpluses of existing and generated rail users and the operating producer surplus (increase in net revenues) when the fare exceeds marginal operating costs. Note that panel (a) does not include any fixed costs associated with the improvement in services.

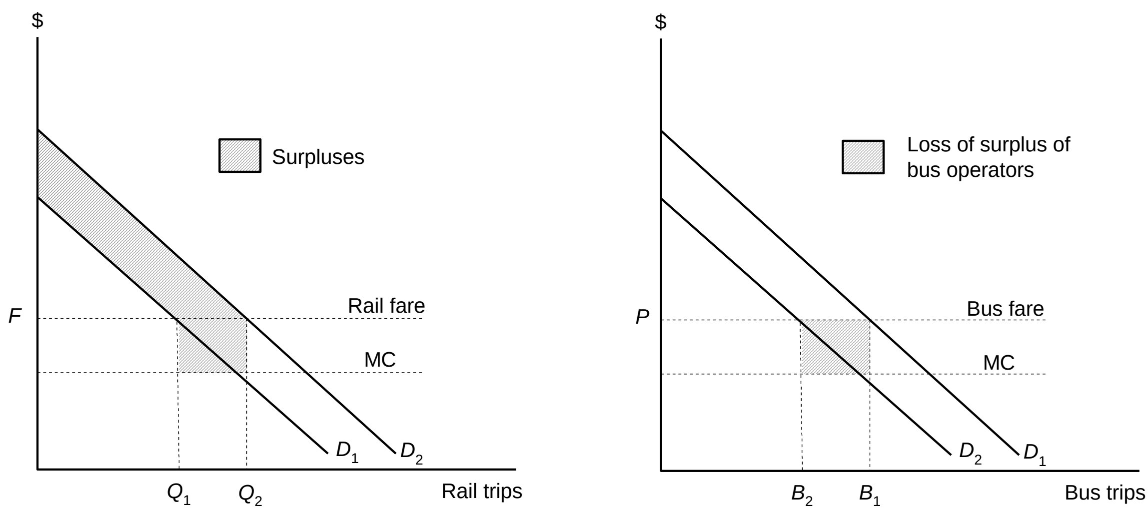
Panel (b) shows the shift in demand for bus trips as the new rail operates. The shaded area in panel (b) shows the surplus lost by bus operators as bus trips fall from B_1 to B_2. The net benefit is the surplus gained in the rail sector less the surplus lost in the bus sector.

Similar losses of producer surplus could occur with generated trips, when individuals transfer to rail from a non-transport activity. However, most non-transport sectors are assumed to have competitive markets with prices close to marginal cost. Thus, the benefits of generated trips are similar to those of diverted trips in panel (a). Deductions are rarely made for any loss of surplus in other sectors.

Box 19.3 shows possible costs and benefits associated with investment in rail services. Clearly the importance of the effects varies with the context. Freight traffic is critical to the Adelaide to Darwin rail service via Alice Springs, but of little importance between Sydney and Canberra. However, the benefit to freight, like that to passengers, is better estimated by what customers are willing to pay for the freight service itself rather than by ad hoc estimates of the values of travel time savings or other service attributes for freight traffic.

Evaluation of airport location

New airports are major economic investments, can transform the local region and create major third-party impacts. In 1971 the Commission of Inquiry into the Third London Airport produced a pioneering cost–benefit study of potential new airport sites for London.



(a) Surpluses in rail sector

(b) Loss of surplus in bus sector

Figure 19.4 Benefits of improved rail service

Box 19.3 Possible costs and benefits of investment in new rail services

Possible costs	*Possible benefits*
Land and property costs	Passenger willingness to pay for services
Construction costs (rail infrastructure)	Freight user willingness to pay for services
Construction costs (rail stations)	Reductions in transport accidents
Costs of rolling stock	Reduced congestion on roads
Fixed rail operating costs	Net reduction in noise impacts
Variable rail operating costs	Net improvement in air quality
Rail maintenance costs	Net reduction in greenhouse gas emissions
Lower surpluses of other transport services	Savings from deferring other transport infrastructure
	Development benefits (very occasionally)

Actually, the Third London Airport study (TLAS) was essentially a cost-minimisation exercise. The study aimed to find a new four-runway international airport close to London that would, with the other two major airports (Heathrow and Gatwick), minimise the costs of the proposed three-airport system for the forecast air traffic (passengers and freight). The costs included property take, site development, airport construction and operating costs, aviation flight costs, the costs of relocating defence and civil airports, costs of providing access to the airports, costs of access for air passengers, noise costs and costs of urbanising areas around the new airport. The study departed from cost-minimisation only in so far as one of the three-airport options provided superior access to air travellers and was forecast to generate more air trips, which provided a small offset against costs of that option.

Despite its technical quality, the TLAS was a practical failure. None of the four sites on its recommended short list has been developed as an airport. Instead, Heathrow and Gatwick airports along with Stansted (then a small airport) have expanded into a three-airport system.

In 1972, the NSW government commissioned a similar cost–benefit study for a second Sydney airport. The aim was to find the best site for an international two-runway airport with long wide-spaced parallel runways for international aircraft and a cross runway and Badgery’s Creek was recommended. Over the next 40 years, several more studies, but no new airport. Only in the last few years has the Australian government announced that it will finally develop a second Sydney Airport. Where? At Badgery’s Creek!

Political factors play a large part in decisions about airports which can affect thousands of households. These households cannot be easily compensated for an adverse airport location. However, there are also instructive economic issues. These relate to the nature of the problem to be solved, the pricing of airport services and the timing of new infrastructure.

Both London and all Sydney airport studies suffered from two related flaws. They failed to identify the infrastructure problem, namely a lack of capacity in peak hours. In Sydney, there was excess demand for runway space in peak hours because a quarter of the peak traffic were slow-moving, small aircraft providing local services, which create congestion and slow down faster and larger aircraft. Only a fifth of the flights were international. Sydney then required extra runway capacity to service these small aircraft. It did not need a new two-runway international airport. In the 1970s, London likewise needed some extra runway capacity, but not a new four-runway airport.

Related to this, the TLAS and most Sydney studies failed to consider pricing and timing issues. Peak-period pricing sorts out the aircraft that need to land or take off at a central airport in peak hours from those that can move to other airports or hours. Until there is considerable excess demand, the amounts that airlines are willing to pay for extra peak hour services are usually well below the costs associated with a new airport. Incremental investment is justified when there is sufficient excess demand *given efficient peak pricing.*

The optimal timing for infrastructure occurs when the service benefits exceed the benefits of deferment. A project that is viable in the long run should not necessarily be started today. Suppose that *K* is the capital cost of investment and *r* is the annual rate of return on capital. The benefit from deferring an investment for a year is *rK*. Investment should be deferred until $b>rK$, where *b* is the benefit of the infrastructure service in that year. Figure 19.5 illustrates the principle. In this case, the project should open in year 10.

Wider economic benefits

In recent years, there has been extensive literature on a suite of impacts not included in the standard traditional economic appraisal of transport projects, which have the generic title of wider economic benefits (WEBs). These fall into six main categories:

1. Agglomeration economies: the impacts of increased employment density on productivity
2. The impacts of transport infrastructure on labour supply
3. The value of business travel time savings in imperfectly competitive markets
4. The impact of transport infrastructure on competition
5. Benefits arising from dependent (or induced) development, and
6. Possible further economic impacts of investment in transport infrastructure.

The first four WEBs were identified in the influential paper on WEBs published by the UK Department for Transport (2005) and have been the focus of most analysis since then. WEBs 5 and 6 above were discussed as possible additions to the standard transport appraisal in the 1980's and 1990's and have re-surfaced. We discuss these WEBs briefly in turn.

Agglomeration economies occur if a transport project increases employment densities and this raises productivity (and income) per worker. This may happen, though transport projects can also disperse employment. Contentiously, it is sometimes argued that lower transport costs increase "effective density" without actual changes in employment density and that this also raises productivity. However, this form of agglomeration economy is implausible and not strongly supported by evidence.

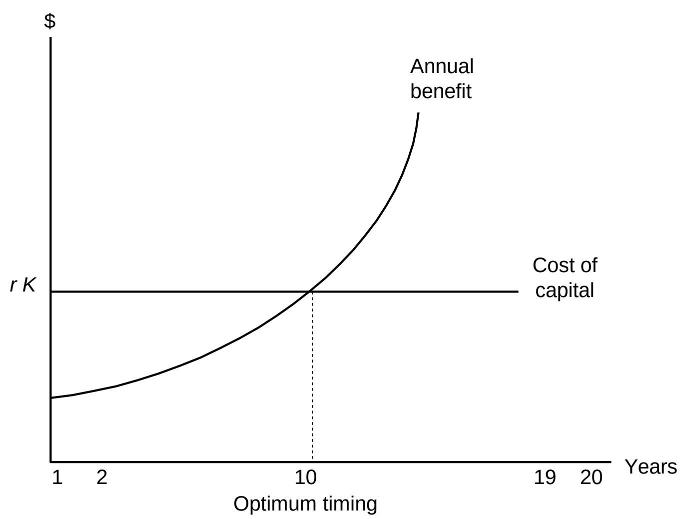


Figure 19.5 Optimal timing of infrastructure

Labour supply. Transport projects may encourage labour supply by reducing commuting costs. Technically, the labour benefit is equivalent to the consumer surplus of new trips shown in Figure 19.3 and Equation 19.9. However, there may be some additional tax revenue that is not picked up in Equation 19.8.

Value of output. In imperfectly competitive markets, goods are sold with a mark-up over factor cost. To allow for this, the value of business time saved may be marked up by 15 or 20 per cent over the wage rate.

Increased competition. UK DfT (2005) suggested that transport projects could increase the competitiveness of cities linked by the new infrastructure. However, this is regarded as implausible where transport networks are already well developed.

Induced developments. If new transport infrastructure induces non-marginal development, such as a resource project in a rural area or major new housing in an urban area, this may create producer surplus that is not captured in travel benefits. This surplus is the market revenue from the new project less the sum of all related public and private costs.

Impacts on the economy. All expenditure has an opportunity cost. It is generally concluded that investment in transport infrastructure does not have special additional effects on employment, productivity and output that are not captured by the standard method of evaluation.

However, these are strongly contended positions, in part because they have major impacts on whether transport projects pass standard cost-benefit criteria. Suggested extra references are Dobes and Leung (2015), Douglas and O'Keefe (2016) and UK DfT (2016a, 2016b and 2016c).

Role and Regulation of the Private Sector in Transport

Because transport infrastructure is often a natural monopoly, government has traditionally financed, owned and operated most such infrastructure. Nevertheless, private firms may be able to supply and operate some of the infrastructure more efficiently than the public sector.

There are many examples of private involvement in transport infrastructure and services and these examples take many forms. In its fullest form, government cedes to the private sector the right to finance, build, own and operate the infrastructure. For example, a private company (Eurotunnel) owns and can operate the Channel Tunnel without a time limit. The UK Government has also privatised most rail services in the UK while maintaining ownership of the rail infrastructure. In some cases, for example French autoroute operations, the operation of the transport infrastructure is a joint public–private venture.

In Australia, several major airports and most shipping ports are privately owned. In Sydney, several bus and ferry services are run privately under contract with the state government. Likewise, most intra-urban and coastal ferry services in Queensland are run by private companies under contract with the Queensland government. And there are many privately-run toll roads in Melbourne, Sydney and Brisbane.

We discuss below some of the generic issues in public-private partnerships and some regulation issues of private firms in the transport sector.

Roles of public and private sectors

As discussed in Chapter 18, the effectiveness of public–private partnerships (PPPs) depends on the circumstances, the structure of the arrangements and the nature of any regulations. In principle, PPPs should facilitate explicit objectives, clear and detailed contracts and competition in supply and use of infrastructure. However, poor management of a complex process can have poor outcomes, especially when the private agency is a *de facto* monopoly.

Theory and evidence suggest that productivity is likely to rise under PPPs. Private firms have more incentive than public agencies to complete construction faster (as revenues start earlier) and to a higher quality (because this affects patronage). In Sydney, the Harbour Tunnel and major motorways (M2 and M5) were constructed much faster under PPPs than was expected. There have also been savings in the order of 20 per cent in construction and operating costs of major roads in Australia, the UK and France (Bureau of Transport and Communications Economics, 1996; Quinet and Vickerman, 2004). While some savings may have been achieved by lower wages or inferior working conditions in the private sector, almost certainly there were savings from enhanced productivity.

Two related issues are the allocation of risk and the cost of finance. Here the arguments for, or against, PPPs are less clear, but they may influence the preferred structure of a PPP. Risks may be political, technical or commercial. Political risk is the risk that government will change regulations or introduce competition to the infrastructure element under concession. Technical risk relates to construction and operating costs. Commercial risk is market risk.

Government is the only party that can deal with political risk. It can also spread risks over millions of taxpaying households. On the other hand, the private sector is generally more skilled at dealing with technical and commercial risk than government agencies. Also, markets can spread risk to parties that are willing to take it on, sometimes because the new transaction provides diversification and even reduces portfolio risk. In a well-structured PPP, the party that can best deal with the risk should bear it.

One way that markets deal with risk is to charge a higher price of capital. Private financing costs are often several percentage points higher than government's borrowing rate. The actual differential depends on the extent to which government underwrites an infrastructure project with support of various kinds. However, the real cost of public borrowing depends on the extent to which repayments are funded from taxes. If funded by taxes the deadweight loss is often at least 20 per cent of revenue raised. This deadweight loss may offset partly or wholly any differential in the nominal cost of capital.

The main potential cost of a PPP is the concession of a monopoly service to private interests with poor contract design and regulations that fail to safeguard the public interest. Government may concede to a private firm the right to own and operate an infrastructure asset before it finalises the contract. In other cases, as with the East-West cross-city Sydney tunnel, to maximise the financial return to government, government agreed to close some existing roads and thereby protect the private firm from route competition. This reduced the economic benefits from the investment.

Critically, contracts are often renegotiated without the benefit of competitive pressures. In a study for the World Bank, Guasch (2000) found that, of over 1000 concessions awarded in Latin America and the Caribbean, over 60 per cent were renegotiated within three years, nearly always at the initiative of the concession holder. Eighty per cent were in transport or water sectors. In many cases contracts were renegotiated not because of changes in economic conditions but because the concession holder had submitted a low bid to gain the concession in the expectation of renegotiation. This means that the most efficient firm may not have won the contract.

Managing transport infrastructure monopolies

The main public policy issues arising in management of privately owned infrastructure are vertical integration, conditions of access, pricing and quality of service, such as frequency. The issue of vertical integration relates to whether the owner should also provide the services that use the infrastructure. This issue arises most acutely for rail services and pipelines. Issues of access and pricing are related as it is generally agreed that service providers should have access to essential infrastructure, but the sticking point is the price.

Vertical integration and access. There are three main options for the relationship between the owner and users of the infrastructure.

1. Full vertical integration of infrastructure and services provided with one owner-operator.
2. Partial integration with both infrastructure owner and other operators using the services.
3. Structural separation where all users of the infrastructure are independent of the owner.

Vertical integration (option 1) may be the most efficient industry structure when there are economies of scope as well as of scale. Economies of scope exist when a single agency can supply the infrastructure and the services using the infrastructure at lower unit cost than can separate agencies. Some rail services may be an example of this. However, it is difficult to prove this without market testing. To ensure that services are provided efficiently it is generally desirable to allow competition among potential users and to ensure as far as possible that they compete on equal terms. Thus option (2) is usually preferred to option (1).[7]

However, if the infrastructure owner competes against outside service suppliers, the competition may be unfair and hence inefficient. Accordingly, a regulator is required to monitor that the terms of access are efficient (and not contain monopoly profit to the owner) and that all competitors have access to the network on equal terms to each other and the owner. However, the regulator will have to rely on information on costs and contracts from the regulated infrastructure operator.

Complete structural separation (option 3) sidesteps an owner's incentive to deal unequally with agencies tendering for use of the infrastructure. These considerations have led several European countries and the state of Victoria in Australia to break up vertically integrated rail systems and introduce competitive bidding for the right to run rail services over the network.

Price regulation. Given the natural monopoly character of most of these private operations, the charges are generally regulated.[8]

Whichever industry structure is adopted, there are efficiency and equity arguments for regulation of charges for use of the infrastructure. In off-peak periods, when there is underutilisation of capacity, user charges should reflect the marginal costs of use. When infrastructure use is at or close to capacity, the owner should be generally be encouraged to charge market-clearing prices which ensure that those who place the highest value on use at that time gain access. Overall, charges may have to be modified to allow for a fair return on equity capital. Also, fairness considerations are likely to influence the regulated prices.

As we saw in Chapter 14, the case for regulating infrastructure user charges or other service provision is based on the presumption that the regulator (either government or an independent regulator) will determine a more efficient and fairer set of charges than would the monopoly owner. Detailed fare and service regulation is highly complex. Also, regulation may stifle service innovation and deter investment. Thus, there are potential costs as well as benefits from regulation of monopoly-owned infrastructure.

In principle, road tolls are straightforward. However, there are significant equity implications because of the wide range of road users. For rail, the major policy issue is the amount of access to allow private operators. Once access is agreed, the allocation of access may be determined by an auction process. Rail operators may in turn be subject to separate price and quality controls. However, few governments regulate prices for access to airport runways, the critical monopoly feature of airports.

---

[7] Regulation may involve service regulation. For example, Eurotunnel is required to provide a minimum level of service at all times of the day and year.

[8] In Queensland, the government reimburses some ferry operators for agreed operating costs and takes the revenue risk rather than the operator taking it.

Summary

- In this chapter, the transport sector exemplifies important general issues for economic infrastructure, much of which is supplied by natural monopolies.
- The key issues are (1) how to use existing transport infrastructure efficiently and equitably and (2) how to achieve an efficient level of transport infrastructure.
- Efficient prices create efficient use. Use of infrastructure is efficient when the marginal social benefit of use equals the marginal social cost. However, equity may also be important.
- When infrastructure is under-utilised, short-run marginal cost pricing encourages efficient use. However, cost recovery may be important for efficient management and competitive neutrality. When infrastructure is heavily used, market-clearing prices ensure that those who most highly value access can gain it.
- Investment in infrastructure capacity is efficient when the estimated present value of the benefits of increased capacity exceed the present value of the costs, inclusive of third-party benefits and costs.
- Optimal capacity is achieved where the marginal benefit of extra capacity just equals the marginal cost.
- To achieve optimal timing, new capacity should be available when the net benefit from use over a period such as a year exceeds the opportunity cost of capital employed.
- Some recent literature contends that transport infrastructure produces wider economic benefits that are not captured in standard appraisals, but this is currently an open question.
- Government traditionally financed, owned and operated most transport infrastructure. However, competitive private sector involvement can often provide services more cost-effectively. These benefits may not be achieved by a private monopoly supplier.
- Two main issues arise with privately owned infrastructure. One is whether the owner should also provide services that use the infrastructure. Where this is allowed, private operators should generally be allowed access to the infrastructure on a competitively neutral basis.
- Second, there are generally efficiency and equity reasons for regulating both the services provided and the prices charged by private monopolies. However, detailed regulation is complex and may stifle service innovation and deter investment. Public policy must determine where the balance lies.

Questions

1. What are the basic principles for efficient use of transport infrastructure?
2. What are the general principles for efficient investment in transport infrastructure?
3. Why is efficient pricing of services a key link between efficient use and efficient investment in transport infrastructure?
4. Would these efficient pricing and investment principles apply equally to the other major infrastructure networks (water, power and telecommunications)? If not, why not?
5. How might equity or social factors influence the use of transport infrastructure?
6. How might equity or social factors influence investment in transport infrastructure? Give examples.
7. How would you determine the role of the private sector in the ownership and operation of shipping, ports and airports?
8. The text distinguishes between political and market risks for PPPs and suggests that government should be responsible for the former and private firms for the latter. Is this distinction practical?
9. Suppose that the construction cost of a new road is $100 million and that the opportunity cost of capital is 7 per cent. Construction takes one year and the road has a life of 30 years. The estimated benefits of the road would be $5 million in year 2 and would rise by 4 per cent per annum.
   i. What would be the net present value of the road if it was constructed today?
   ii. What would be the optimal year in which to construct the road?
   iii. If constructed then, what would be the net present value of the road?
10. Is it possible to account efficiently for environmental factors in transport pricing and investment decisions?

Further Reading

Clark, H. and D, Prentice (2010) 'Reforming taxes and charges on Australia road transport' pp. 389–414 in C. Evans, R. Krever and P. Mellor (eds), *Australia's Future Tax System: The Prospects after Henry*, Thomson Reuters, Sydney.

Douglas, N. and B. O'Keeffe (2016), 'Wider Economic Benefits – When and if they should be used in evaluation of transport projects', *Australian Transport Research Forum Proceedings*.

Gomez-Ibanez, J.A. (1999) 'Pricing', pp. 99–135 in J. Gomez-Ibanez, W.B. Tye and C. Winston (eds), *Essays in Transportation Economics and Policy, A Handbook in Honour of John R. Meyer*, Brookings Institution Press, Washington DC.

Guasch, J.L. (2000) *Concessions: Bust or Boom? An Empirical Analysis of Ten Years of Experience in Concessions in Latin America and Caribbean*, World Bank Institute, Washington DC.

Hensher, D.A. (2012), 'Transport Economics: An Overview', pp. 1-48 in D. Hensher (ed.) *Major Works in Transport Economics*, Vol.1, Routledge, UK.

Hensher, D.A. (2011) 'Value of Travel Time Savings', pp. 135–158 in A. de Palma, R. Lindsey, E. Quinet and R. Vickerman (eds), *Handbook in Transport Economics*, Edward Elgar, London.

Small, K.A. (1999) 'Project Evaluation', pp. 136–177 in J. Gomez-Ibanez, W.B. Tye and C. Winston (eds), *Essays in Transportation Economics and Policy, A Handbook in Honour of John R. Meyer*, Brookings Institution Press, Washington DC.

Newbery, D.M. (1995) 'Pricing and congestion: economic principles relevant to pricing roads', pp. 345–372 in T.H. Oum et al. (eds), *Transport Economics: Selected Readings*, Korea Research Foundation for the 21st Century, Seoul.

UK Department for Transport (2016), *Wider Economic Benefits Appraisal*, TAG Unit A2.1.