Australian Transport Council

2006

National Guidelines for Transport System Management in Australia

5 Background material

ISBN: 0-9802880-4-5

This publication is available free of charge by downloading it from: http://www.atcouncil.gov.au/documents/NGTSM.aspx

or from the Bureau of Transport and Regional Economics (BTRE)

GPO Box 501, Canberra, ACT 2601,

by phone 02 6274 6806, fax 02 6274 6816.

Feedback on the Guidelines can be forwarded to the Executive Director, BTRE on the contact details above.

Printed by the Department of Transport and Regional Services

Contents

TABLES

FIGURES

Foreword

This document presents background material to the *National Guidelines for Transport System Management in Australia* (2nd edition) endorsed by the Australian Transport Council (ATC) in November 2006. It is part of a series of five documents that comprise the Guidelines. The other documents cover an introduction, a detailed framework for undertaking strategic transport planning and development, detailed information on the appraisal of initiatives and an analytical approach for urban transport proposals.

I gratefully acknowledge the contributions made by committee members towards this very significant piece of work. All of the members have given generously of their time and competencies, over an extended period of time, to make the Guidelines a comprehensive and user-friendly manual that will assist all jurisdictions in the complex business of transport system planning and management. In particular, I acknowledge the significant contribution of the Chair of the Committee, Dr Anthony Ockwell who directed and managed the project throughout its entire process. A list of members is presented elsewhere in this publication.

The Guidelines support transport decision-making and serve as a national standard for planning and developing transport systems. They are a key component of processes to develop and/or appraise transport proposals that are submitted for government funding. Potential users of the Guidelines include governments, private firms or individuals, industry bodies and consultants.

The Guidelines have been endorsed by all Australian jurisdictions. They were developed collaboratively over several years by representatives from all levels of government in Australia through the Standing Committee on Transport (SCOT), in consultation with SCOT modal groups (Austroads, Australian Passenger Transport Group, SCOT Rail Group). The Guidelines have been endorsed by ATC and the Council of Australian Governments (COAG).

This is the second edition of the Guidelines. It is an expanded and revised edition that reflects directions from SCOT, ATC and COAG as well as feedback from users. The revision has focused on making the material more cohesive, accessible and user-friendly, while maintaining rigour. These improvements will help to facilitate the widespread adoption of the Guidelines that has been specified by COAG.

The terms assessment, appraisal and evaluation are often used interchangeably in practice to mean the determination of the overall merits and impacts of an initiative. In the Guidelines they are used as follows:

- *Assessment*: A generic term referring to quantitative and qualitative analysis of data to produce information to aid decision-making.
- *Appraisal*: The process of determining the impacts and overall merit of a proposed initiative, including the presentation of relevant information for consideration by the decision-maker.
- *Evaluation*: The specific process of reviewing the outcomes and performance of an initiative after it has been implemented.

The current focus of the Guidelines is land transport—road, rail and inter-modal. There is scope to further broaden the Guidelines to cover other modes and transport issues in the future.

It is envisaged that the experiences of users who apply the Guidelines will continue to provide useful insights into areas requiring further improvement. The Guidelines should therefore be seen as an evolving set of procedures and practices. The agencies involved in the development of the Guidelines welcome feedback that will contribute to the process of revision and improvement.

Michael J Taylor
Chair
Standing Committee on Transport
December 2006

Committee members for 2nd edition

AUSTRALIAN GOVERNMENT	
Anthony Ockwell (chair)	Department of Transport and Regional Services
Mark Harvey	Bureau of Transport and Regional Economics
Kym Starr	Department of Transport and Regional Services
Caroline Evans	Bureau of Transport and Regional Economics (until December 2005)

STATE AND TERRITORY GOVERNMENTS		
Peter Tisato	Department for Transport, Energy and Infrastructure	South Australia
Russell Fisher	Department of Main Roads	Queensland
Jon Krause	Department of Main Roads	
Renny Phipps	Queensland Transport	
Rhonda Daniels	Department of Planning	New South Wales
Julieta Legaspi	Roads and Traffic Authority	
Fotios Spiridonos	Department of Infrastructure	Victoria
Andrew Trembath	Department of Infrastructure	
Ed McGeehan	VicRoads	
Steven Phillips	Department for Planning and Infrastructure	Western Australia
Philip Petersen	Department of Infrastructure, Energy and Resources	Tasmania
Michael Hogan	Department of Infrastructure, Energy and Resources	
Greg Scott	Department of Planning and Infrastructure	Northern Territory

NEW ZEALAND GOVERNMENT	
Simon Whiteley	Land Transport New Zealand
Ian Melsom	Land Transport New Zealand.

AUSTRALIAN CAPITAL TERRITORY & AUSTRALIAN LOCAL GOVERNMENT

The Australian Capital Territory, Department of Urban Services and the Australian Local Government Association were consulted throughout the development of the Guidelines.

Introduction

Volume 5 of the *National Guidelines for Transport System Management in Australia* contains detailed background material that underpins the methodology outlined in Volume 2 and Volume 3 of the Guidelines. Volume 5 is relevant for those who wish to understand, at a detailed level, the logic behind the Framework. In some cases, the text in Volumes 2 and 3 is a summary of the corresponding material in Volume 5. In other cases, Volume 5 extends the coverage.

Volume 5 is divided into four parts:

- Part 1 contains supporting material relating to Volume 2—'Strategic Transport Planning and Development' and Volume 3, Part 1—'Appraisal Process and Methodology'. Part 1 of Volume 5 opens with a general discussion of strategic planning for transport. Questions addressed include the nature of strategic planning in a transport context, the benefits of strategic planning, cautions to note and the current state of the art. It also covers aspects of strategic planning and analysis at the network, corridor and area levels—equity, data needs, economic methodologies and maintenance.
- Part 2 contains supporting material relating to Volume 3, Part 2—'Analytical Techniques'. Part 2 of Volume 5 is designed to be used in conjunction with Volume 3, Part 2. The section numbers and topics in Volume 5, Part 2 correspond to the same section numbers and topics in Volume 3, Part 2 to make it easy for users to move between the two volumes.
- Parts 3 and 4 are papers commissioned by the Working Group from consultants at the University of South Australia and Arduus Pty Ltd. In developing a set of multi-modal guidelines, the Working Group strived to give equal treatment to the road and rail modes. While significant amounts of material exist on the evaluation of road projects, for example in Austroads publications and the guidelines of road agencies, there is a shortage of equivalent material on rail. The two papers in Parts 3 and 4 are intended to be a first step towards redressing the imbalance. Part 3 is a paper on demand forecasting and modelling. Part 4 is a paper on the estimation of impacts of rail infrastructure improvements.

1 Transport system management: Background material for Volume 2 and Volume 3, Part 1

1.1 Strategic planning for transport infrastructure

1.1.1 Nature of strategic planning in a transport infrastructure context

Strategic planning provides the direction necessary to develop an investment program that will deliver the desired outcomes in a range of likely futures. Generically, strategic planning can be described as:

> … a disciplined effort to produce fundamental decisions and actions that determine strategy. It involves understanding the present environment and anticipating the future environment. It involves asking questions such as, 'Are we doing the right thing?', 'What are the most important issues to respond to?' and 'How should we respond?'. It entails attention to the 'big picture' and willingness to adapt to changing circumstances. Although described as 'disciplined', strategic planning does not flow smoothly from one step to the next. It is a creative process that involves iteration. It can be complex and challenging.[1]

In the transport context, strategic planning provides a framework for ensuring that transport decisions are responsive to the expectations of users and the community generally. These expectations, in total, usually translate to transport investment wants that exceed the government's capacity to fund them, given the competing priorities of government. The government therefore has to make choices and decide on transport investments that are both affordable and responsive to future needs. The strategic planning process provides a rigorous framework to narrow the choices about the options that are given priority. The process balances many competing considerations, often involving value judgments, subjective assessments and political considerations that cannot be reduced to quantitative measures. For example, sealing a road to, or within, a remote Aboriginal community, based on traffic volumes, could not normally be justified strictly on economic grounds. However, in the context of government objectives for indigenous communities and equity considerations, this investment may rank highly from a broader strategic fit perspective.

The major elements of a strategic plan that might be developed for a transport network include:

- a context scan
- objectives
- strategies and actions to achieve the objectives, and
- a monitoring and review framework.

1 Adapted from http://www.nonprofits.org/npofaq/03/22.html.

Not all strategic plans will contain every one of these major elements; some strategic plans will contain additional elements.

The context scan element involves gaining an understanding of government and community expectations, issues, the likely future, policy instruments at the government's disposal and constraints.

A statement of objectives is an operational way of expressing the outcomes, or vision, the strategic plan aims to achieve. The objectives should be aligned with the values of the community as a whole. The objectives of a government are numerous, varied and likely to be competing, more so than for a private organisation. For a private firm, maximising shareholder value is likely to have primacy; while, for a transport network, the objectives are along the lines of economics, the environment, safety, equity and security. The community's expectations are diverse. Economic considerations such as travel time and vehicle operating costs are obviously important to transport users. However, some stakeholders will give emphasis to environmental factors (e.g. vehicle emissions, noise) or equity considerations (e.g. access by remote communities to essential goods and services). Not only does strategic planning have to identify the objectives, it also has to determine the relative importance of the objectives.

Objectives often need to be treated hierarchically because economic and environmental factors are multi-dimensional and leave a lot to interpretation at the operational level. Hence, within each broad category, there are more narrowly defined objectives further down the hierarchy. For example, growth and trade come under the heading of economics, and air quality, greenhouse gas emissions and noise come under the heading of the environment.

Strategies express how an organisation plans to achieve its objectives in broad terms (e.g. invest in infrastructure to improve transport logistics). In parallel with objectives, strategies are set out hierarchically. Strategies can be implemented in many ways so the list of actions under each strategy sets out, in detail, what is to be done (e.g. the specific types and locations of investments). The strategic plan may not necessarily detail actions. Instead, it may provide assessment criteria by which alternative actions such as investment proposals can be compared in order to select the actions that best achieve particular sub-objectives. A simple version of this is to specify intervention standards that will trigger actions.

To monitor and evaluate the success of a strategic plan, transport objectives need to be measurable through performance indicators.

Alignment between the outcomes produced by strategic planning and subsequent actions, and the outcomes originally envisioned by the government and stakeholders is described throughout the Guidelines as 'strategic merit' or 'strategic fit'.

1.1.2 Benefits of strategic planning

The strategic planning approach has many advantages over piecemeal decision-making.

Democratic values

When the objectives of strategic plans are derived from community values and when strategic plans are developed with extensive stakeholder consultation, the process is a genuine attempt by governments to be responsive to community needs and expectations.

Coordination, integration and options

In any interconnected system, actions taken in one part of the system have impacts elsewhere in the system. For example, upgrading the transport infrastructure on one site may alter flows in other parts of the network, including other modes. Failure to recognise how a decision may affect options for the future could lead to inferior outcomes. An integrated strategic plan provides a discipline to ensure that interactions are taken into account.

For transport planning, better outcomes are likely to ensue when governments make use of all available policy instruments, and where the policy settings are developed in a coordinated manner with a clear understanding of the objectives, and of how each instrument impacts on the objectives.[2] The strategic planning process provides a framework within which this can occur.

Responsibility for transport may be shared across levels of government and the private sector. A government making best use of ways to achieve its objectives will negotiate with, and persuade, other parties who have control or influence over the transport network.

Consideration of the future

Strategic planning encourages disciplined thinking about the future, in terms of both the future implications of present decisions and the implications of future events.

However, this does not mean developing a single view of the future and preparing a response for this future. Rather, strategic planning facilitates identification of a range of likely futures, or scenarios, and the preparation of candidate responses. These candidate responses are then assessed through a process of risk management. This includes assigning probabilities to particular futures, identifying areas of commonality that need to be addressed and assessing the consequences of not addressing the needs of some futures.

Skill is required to maintain constancy of the important outcomes and aspirations, but, at the same time, there must be flexibility to respond to changing futures.

Application of intelligence and creativity to problem solving

In developing a strategic plan, people identify problems and consider solutions in a rational manner. At the same time, there is opportunity to apply lateral or out-of-the-box thinking to propose innovative solutions. Data, analysis, judgment and intuition can be brought together.

Communication

A single individual can grasp only a limited part of the range of information available about any real-world system, the factors that impinge on that system and the complex interactions involved. Strategic planning requires a collegiate approach across people in the relevant system and in related systems. It facilitates information transfer and development of a shared vision within the government.

Engagement of stakeholders extends information and vision sharing beyond the boundaries of the government. Stakeholder engagement becomes a tool for influencing, and even managing, the expectations of stakeholders. This is important when it is impossible to satisfy stakeholder expectations because of conflicting interests (one stakeholder can only be appeased at the expense of another) or budget constraints.

Finally, having a well-articulated strategic plan is an invaluable aid for a government to inform the community about its activities.

Learning from the past

If the planning process includes a system to monitor the performance of past decisions, this also provides an opportunity to learn lessons that can improve future decisions.

2 The findings of economic theorists about the relationships between policy instruments and objectives support this. Tinbergen's rule stipulates that, in order to advance a number of declared (conflicting) objectives concurrently, the number of policy instruments available to the government must be at least equal to the number of objectives. Mundell's rule takes the argument further by providing the optimal set of policy settings for the various instruments. Each instrument should be employed to promote the objective for which its effectiveness, relative to the other instruments, is greatest. Source: http://psteger.free.fr/regbe-tinbergen-gb.htm

1.1.3 *Cautions to note in strategic planning*

The benefits from strategic planning will only be realised if the planning is done well and is actually implemented. Strategic planning performance reviews provide valuable learning. Shortfalls in strategic planning occur when fundamental activities such as those listed below are not completed.

Gaining commitment to the plan

Strategic plans, once developed, must be supported with full commitment. Ways to increase the likelihood of implementation are:

- having a high level of stakeholder involvement in development of the plan to avoid potential future disagreements and to give stakeholders a sense of ownership
- ensuring that the necessary framework is in place to implement the plan, and
- implementing systems to monitor actions and associated effectiveness in achieving the desired outcomes.

Ensuring balanced stakeholder influence

Stakeholders who are better organised, or can access better resources, have a greater ability to influence plans, often at the expense of other stakeholders. Situations can arise where the benefits of a government decision accrue to a small number of individuals, while the costs are widespread, with the total costs exceeding the benefits. The opposite situation is also common, where the costs are concentrated, and the benefits dispersed, but the total benefits outweigh the costs. Governments can be pressured to side with the minority interest at the expense of the community as whole. In these cases, if decision-making that favours the most influential stakeholders becomes the rule rather than the exception, the effects are cumulative, and society as a whole is poorer.

In transport, externality costs are often borne by a local community through which a road or railway passes, while the benefits are widespread—this typifies the 'not in my backyard' or NIMBY problem. Yet these same communities are users of roads and consumers of products that are transported on the roads that impose externality costs on other communities.

Planners need to understand where these situations are likely to arise and to look for ways to mitigate, or avoid, negative effects on local communities. Planners may seek to ensure that local communities do not have an exaggerated view of the costs. Often, there is no choice for governments but to accept the political reality. But when the benefits to society as a whole are particularly large relative to the costs, or the position of influential stakeholders is unreasonable, it is worth arguing the case.

Making informed decisions

The decisions made in developing strategic plans need to be based on good information. In the absence of good information, decision-making will be guided by perceptions that are often unreliable and biased. Inadequate, or misleading, information may be the result of deficiencies in data or interpretation of data.

Data issues include:

- quantitative data may not encompass important non-economic and qualitative factors
- data may be too aggregated to be useful
- there may be a lack of data, and
- unreliable data.

Data requirements include characteristics of the physical infrastructure, current utilisation of the infrastructure, demand, performance, environmental and heritage considerations, safety, plans and priorities of stakeholders, and current and proposed initiatives. A detailed list of data requirements is provided in Section 1.3. Techniques for analysing quantitative data to predict likely investment and maintenance needs are discussed in Sections 1.4 and 1.5.

To improve the data available for transport planning, particularly at the strategic level, the National Transport Data Framework aims to promote comparability in data, sharing of data and addressing of gaps in data. It will develop arrangements to collect, assemble, harmonise, house and manage data for the joint purposes of all jurisdictions across Australia.

A literature review of previous work related to the network under consideration can be a valuable source of data. It might cover:

- previous strategic plans
- previous studies that relate to the network or parts of it, including corridor, area and regional studies
- previous assessments of initiatives, and
- published and unpublished documents prepared by stakeholders.

Stakeholder consultation is another source of information and insight, in particular, for qualitative data.

Interpretation of data involves using both intuition and analysis to draw out salient points, to test whether options proposed during the planning process are viable and to compare alternatives. For example, if the strategic plan includes objectives in the form of quantitative targets, analysis is necessary to check whether the targets are:

- feasible in a technical sense (e.g. road vehicles and trains can only travel so fast)
- feasible in a behavioural sense (e.g. many shippers would not change transport modes willingly without a substantial change in relative prices or service qualities)
- feasible within constraints (e.g. budgetary, environmental), and
- achievable at an acceptable opportunity cost in terms of other objectives (e.g. if achieving an environmental target requires a level of investment well beyond the economically efficient level).

The same applies to actions specified in a plan such as proposed investment projects. As part of the iterative approach to developing a strategic plan, targets and actions can be modified in light of new information.

Forecasting plays a major role in strategic planning as strategic planning inevitably involves uncertainty. Unjustified optimism needs to be avoided to ensure that forecasts represent expected, or central, values. Even then, the likelihood of deviation about the expected, or central, value must be understood. This requires a sound understanding of factors contributing to future growth in demand.

Extrapolation of current trends does not recognise the possibility of major discontinuities or one-off events. Scenario building is a tool that has attracted attention in recent years and should be applied where there is a reasonable probability of futures that differ markedly from the present. Probabilities, whether explicit or implicit, need to be attached to each scenario so attention can be focused on the more likely scenarios. Hedging involves making plans that yield reasonable pay-offs under more than one scenario. However, hedging comes at a cost, because the chosen plan may not be the best for the scenario that is actually realised. A balance should be struck between planning based on a single view of the future and hedging for a variety of futures. Erring too much on one side or the other can be costly. The probabilities of different scenarios need to be balanced against the costs of hedging.

Developing plans that are responsive to changing futures

As one author states, '…plans are commitments, or should be, and thus they limit choice. They tend to reduce initiative in a range of alternatives beyond the plans' (Mintzberg 2000, p. 183 quoting Steiner 1979, p. 46). If the future proves to be very different from the forecasts on which a strategic plan is based, or new information comes to light that shows the plan will not achieve its desired outcomes, urgent revision may be required. On the other hand, having total flexibility is akin to not having a plan at all. Again, a balance is required.

Strategic plans can be constructed in a way that has a certain amount of in-built responsiveness by:

- keeping them at a high enough level to allow decision-makers to consider a range of options within the broad parameters of the plan, and
- making some actions in the plan conditional on future events taking place or certain information coming to light.

Means to plan and deliver

The capacity of governments and industry to professionally plan and deliver initiatives depends on having the necessary knowledge and expertise. Accordingly, investment in government and industry capability should be provided for:

- data
- analytical and modelling tools
- professional expertise and development
- research and development, and
- management systems.

1.1.4 Austroads principles

Austroads (1998) lists ten principles of good strategic planning:

1. **Focus on outcomes**, rather than on the outputs normally delivered by an organisation.
2. **Tailor the process** to the problem, taking into account its nature and complexity, political and community imperatives, and the availability of resources and time.
3. **Generate possible futures**, with a view to identifying the kind of future we would like to move towards.
4. **Consider the full range of means** available to achieve intended outcomes.
5. **Consider all stakeholders**, including organisations whose activities impinge on achieving the outcomes, and all who have an interest in the fulfilment of the outcomes.
6. **Reveal the choices** to be made in light of the anticipated consequences of the options considered.
7. **Use 'iterations'** by reviewing and, if necessary, modifying the results of earlier stages of the planning process in light of feedback from subsequent stages.
8. **Decide when to commit** to important strategic choices and avoid prematurely closing off options.
9. **Support transparency and accountability** so that it is clear how, why and by whom decisions are made, and how and to whom responsibilities and accountabilities for implementation are allocated.
10. **Monitor the strategies and actions** by measuring their effectiveness in achieving the desired outcomes.

1.1.5 Current state of strategic planning for transport

Transport planning at the strategic level has evolved to reflect current approaches to pubic policy development and changes in community values, including increased concern about the environment and technological changes. Table 1.1 explores the differences between earlier and current approaches to strategic planning for transport.

Table 1.1: Evolution of strategic planning for transport systems

FROM	→	TO
Focusing on delivering transport outputs.	→	Focusing on achieving multiple outcomes.
Making decisions that best meet the needs of this generation.	→	Making decisions that also consider the needs of future generations.
Devising solutions that can perform well in a predicted future.	→	Devising solutions that can perform well in a range of possible futures.
Understanding the transport system.	→	Understanding the transport system and the broader systems that transport fits within.
Recognising how land use affects the way the transport system works (and vice versa).	→	Planning land use to protect and help the transport system to work better (and vice versa).
Planning transport and land use separately.	→	Planning transport and land use concurrently and iteratively.
Responding to demands on the transport system.	→	Influencing and managing the demands on the transport system.
Accepting or mitigating the negative impacts of transport on the natural environment.	→	Seeking ways to conserve and enhance the natural environment.
Focusing on mobility (the movement of people and goods).	→	Focusing on access (to people, places, goods and services) and consequent mobility.
Considering transport planning measures in isolation from each other.	→	Selecting the best package of measures.
Planning a number of separate modal systems.	→	Planning one interconnected transport system that capitalises on the strengths of each mode.
Supplying new transport infrastructure and services.	→	Making best use of existing infrastructure and services first.
Consulting governments, industry and the community.	→	Engaging and developing partnerships across governments, industry and the community.
Planning separately based on who owns and operates infrastructure and services.	→	Planning collaboratively based on achieving good system-wide outcomes.

Source: Queensland Government (2003).

This evolution encourages planners to explore an extended range of options. It is imperative for planners to identify options beyond traditional supply solutions that involve expanding the network infrastructure. For example, a traditional solution to congestion during peak hour on a major bridge structure is to construct a duplicate bridge. Contemporary solutions that should be explored include:

- managing demand through improved logistics or use of other modes of transport
- congestion pricing that varies with time of day so that it spreads peak demand, and
- managing residual peak demand through tidal-flow arrangements, ramp metering, and electronic tolling.

This is not to say that traditional solutions are necessarily inappropriate. Rather, community values have evolved and modern technology, together with contemporary legislation, provides an expanded range of available tools. Application of these new tools can lead to more cost effective and sustainable solutions.

1.2 Addressing equity objectives

In most economic and social contexts, equity issues involve variations across the community in terms of income, wealth and access to services. For transport planning, equity is largely concerned with stakeholder expectations about accessibility and social cohesion, together with the impact of the allocation of funds between regions with different population densities.

The economic benefits from transport initiatives are strongly dependent on traffic levels. If funds are allocated purely on the basis of economic efficiency (i.e. proposed initiatives are prioritised solely according to benefit–cost ratios (BCRs)[3]), the proportion of funds going to relatively less-populous areas, and the resulting service standards, would be too low to gain stakeholder support.

For the national land transport network, there are three types of equity issues involving population density:

- issues between the more populous eastern states, where forecast population growth is strongest, and the less populous states and territories
- issues between sections of the network located in metropolitan areas and the rest of the network, and
- issues between the network itself and off-network infrastructure.

Decisions about equitable levels of infrastructure provision are, inevitably, subjective. They should take account of stakeholder expectations, economic efficiency costs and funding constraints.

Departing from the most economically efficient allocation of funds to accommodate equity objectives imposes an economic efficiency cost on society as a whole. Therefore, a balance is required between economic efficiency and equity.

Leaving the efficiency–equity trade-off to be determined on an ad-hoc, initiative-by-initiative basis is likely to lead to outcomes that are inconsistent across decisions and across regions. In the Transport System Management Framework (the Framework) detailed in the Guidelines, the impacts of each appraised initiative are presented in a comprehensive Business Case. Final decisions are made by the appropriate minister, taking into account all relevant factors. Efficiency and equity are considered in every decision.

Equity considerations can be incorporated in policy decisions by specifying required infrastructure standards. More systematic options involve funding shares, weightings and a combined approach.

Infrastructure standards

The trade-off between efficiency and equity becomes apparent when infrastructure standards within the network are considered. Standards can be varied across many aspects of infrastructure such as road roughness, road width and sealing of shoulders.

Figure 1.1 provides an indicative[4] illustration of some of the options for relationships between infrastructure standards and demand for infrastructure usage. The options include:

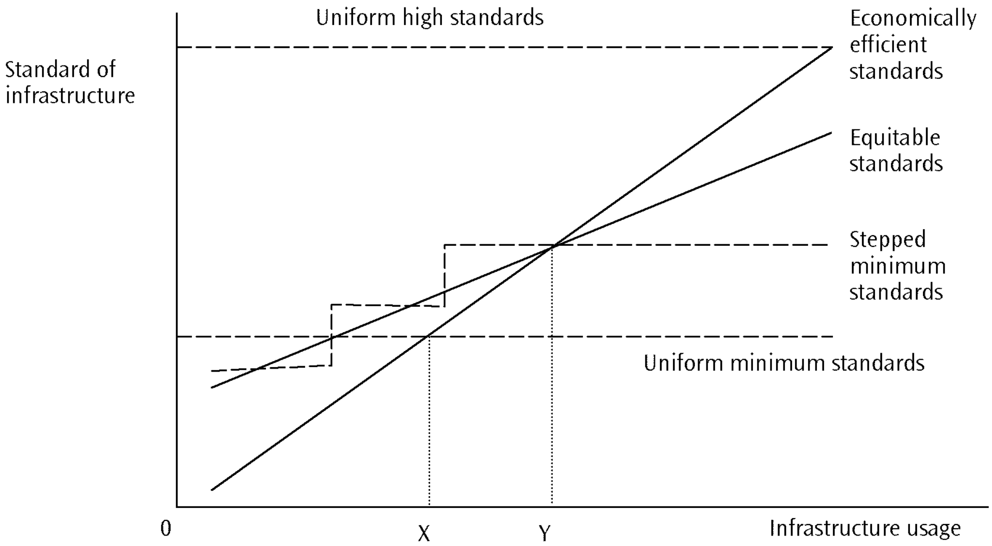
- *Uniform high standards across the whole network*. In reality, this approach is prohibitively expensive and wasteful from an economic efficiency viewpoint.

3 The BCR is an indicator of economic merit of an initiative and is determined in benefit–cost analysis (BCA). See Section 2.10.4 in Volumes 3 and 5 for a detailed discussion.

4 A key qualification is that the lumpiness of infrastructure means the real relationships are not as linear as suggested by the diagram.

- *Economically efficient standards*. Because these standards rise with demand for usage (as indicated by the economically efficient standards line)[5], they are generally considered to be unacceptably low for low-demand infrastructure.
- *Equitable standards*. This approach is a compromise between equity and economic efficiency. However, funds may be insufficient to achieve efficient standards for highly utilised infrastructure, with the shortfall exacerbated by the diversion of funds to provide equitable standards for poorly utilised infrastructure.
- *Uniform minimum standards*. While uniform minimum standards can be a valuable planning tool for advancing network objectives, they fail to recognise that stakeholder perceptions about equitable standards do take account of infrastructure utilisation. The minimum standard may be considered excessive for the least-trafficked parts of the network (left of point X) but do nothing to guarantee equitable outcomes for the remainder of the network.
- *Stepped minimum standards*. This approach can ameliorate the problems of uniform minimum standards. With individual links grouped into categories based on usage or purpose, different minimum standards can be applied to different categories. At the lower end of the utilisation range, minimum standards rise with utilisation in a series of steps, and, once point Y is reached, there is no need for minimum standards.

Figure 1.1: Efficient, equitable and minimum standards of infrastructure



Funding shares

Funding shares may be specified ex-ante for different categories of initiatives. Then, initiatives in different categories will not compete with each other for funds. Examples of these categories include:

- initiatives in different states and territories
- on-network versus off-network initiatives (as in the case of the specified national land transport network)
- different types of infrastructure (e.g. bridges, pavements), and
- maintenance versus capital.

5 This is a simplification. The level of usage is not the only factor that determines economically efficient infrastructure standards. Standards are also affected by the costs of providing the standards, the composition of demand in terms of vehicle mix, and the pattern of usage on a daily and weekly basis.

Within each funding category, initiatives can be prioritised using BCRs or another prioritisation method (see discussion of Phase 6 in Volume 2).

Ex-ante specification of funding shares automatically produces an allocation that is likely to be politically acceptable in view of expectations created by funding outcomes in the recent past. However, it can result in the implementation of low-priority initiatives in some categories, while high-priority initiatives in other categories are not undertaken. Ways to manage this issue include:

- revising the funding shares in each period in light of information about the merits of initiatives undertaken in previous periods
- setting a cut-off BCR for initiatives undertaken in each funding category, with the ratio possibly being lower in less-populous regions, and
- capping the level of funds for initiatives with BCRs less than 1.0, with the result that a specified percentage of funds go to such initiatives.

While funding on the basis of strategic national importance is relevant, equity is the main reason for having separate funds for different categories of initiatives. A second reason is that off-network and on-network initiatives can be assessed against different weightings of objectives. The designated network features the highly trafficked trunk routes that link the major centres of population and carry freight to cities and ports. Off-network initiatives, which are generally small-scale, are aimed more at providing access to the network for people and freight in regional and outer metropolitan areas.

Distributional weighting

Equity issues can also be incorporated in policy decisions by weighting the impacts on different groups in society. For example, impacts on people in less-populous areas can be given a higher weight than impacts on people in more-populous areas. This would introduce a systematic bias in favour of initiatives in non-metropolitan areas.

It should be noted that some of the affected people may not reside in the area where an initiative is implemented. For example, infrastructure upgrading in an area may benefit shippers living in other areas through reduced transport costs arising from travel time savings.

Adjusted BCA, which is discussed in Section 2.12 in Volumes 3 and 5, uses a weighting approach. The advantage of this method is that it balances efficiency and equity objectives in a consistent and transparent manner.

A major difficulty with the weighting approach is that it can be extremely difficult to identify the distribution of the benefits and costs generated by an initiative. Depending on levels of competition, benefits to industries may be passed down the production chain and may ultimately be spread widely, but thinly, throughout the whole economy. For example, an initiative that facilitates increased exports from one region may cause the exchange rate to be higher than otherwise, making imports cheaper for all Australians, but disadvantaging an import-competing industry in another region. Benefits from improved infrastructure accrue to local traffic and to through-traffic, which has disparate origins and destinations, resulting in some of the benefits being widely spread.

A disadvantage of the weighting approach is that the resulting allocation of funds might not accord with expectations based on recent history. Ways to overcome this disadvantage include:

- testing proposed weights on actual initiatives from previous periods
- revising the weights in each period in light of experience with allocations of funds in previous periods, and
- setting upper and lower limits on the amounts of funds for each category.

A simple, practical scheme for applying distributional weights is proposed in Volume 3.

Combined approach

Equity issues can also be handled by using a combination of the minimum standards, pre-specified funding shares and weighting approaches.

For example, a separate fund could be set aside for off-network initiatives. Weightings and minimum standards could be used to address on-network issues, involving equity between states and territories and between metropolitan and regional locations. Furthermore, equity-based weights could be employed when prioritising initiatives within a funding category.

Accessibility indexes

Relative levels of access to other people and to services can be measured objectively using accessibility indexes. The accessibility of one place to all other places can be defined as a function of a measure of the attractiveness of the other places, and of the costs of getting there. The simplest accessibility index, for place *j*, is based on the gravity model formula:

$$A_j = \sum_i \frac{P_i}{f(d_{ij})}$$

where

- P_i is a measure of the importance of origin *i*, generally measured by population for passengers and economic activity for freight
- d_{ij} is the resistance for transport between origin *i* and destination *j*, usually measured by distance, transport cost or time, and
- $f(d_{ij})$ is the 'impedance function', which often takes the form of d^{β} or $e^{\beta d}$ where β is a constant.

Any change in the network that affects the costs that residents at *j* incur to reach other points in the network, or any changes in land use affecting the distribution of population or economic activity, will alter the index.

Unlike a net present value or BCR, the value for a single accessibility index, viewed in isolation, is meaningless. An index for a particular place only has meaning if it is compared with an index calculated in the same way for a different location, time or scenario. A key issue is the definition of a minimum standard that balances economic efficiency and equity.

A practical drawback of accessibility indices is that they are quite data intensive.

For an extended discussion of accessibility indices in the Australian context, including different formulations and practical uses, see Austroads (1999).

1.3 Data for system planning

The following is a detailed list of data that would support strategic planning for a transport system, including undertaking the deficiency and economic assessments described in Sections 3.5.4 and 3.5.5 in Volume 2.

- Population in regions, current and projected.
- Economic activity in regions, current and projected, broken down by industry. This includes new developments, underway or planned, which could affect transport demand on parts of a network or corridor.
- Transport infrastructure:
  - For homogeneous segments of road: location details; name of road or identification code; state or territory legal classification; Australian Government legal classification; location of towns and major centres on the route; segment number (in sequential order); length of segment; grid coordinates; divided or undivided; carriageway identification; carriageway direction; number of traffic lanes; lane widths; pavement width; surface type; surface width; number of shoulders; sealed shoulder average width; unsealed shoulder average width; roughness; year of measurement of roughness; cracking; year of measurement of cracking; rutting; year of measurement of rutting; terrain characteristics (e.g. flat, undulating or mountainous); curvature; gradient; design speed and legal speed limit.
  - For homogeneous segments of rail track: distances, passing loops, speed restrictions, curvature, gradient, rail weight, sleeper type, track condition and terminals.
- Vehicle numbers and traffic composition:
  - For road segments: average annual daily traffic (AADT); year of AADT count; percentage of commercial vehicles in the AADT (preferably by type—rigid, articulated, road trains, B-doubles); growth rate (preferably growth rates for each vehicle type) and hourly volume distribution.
  - For rail links: train numbers by type (freight or passenger, priority, speed, length) and time of day.
- Performance indicators (times taken, reliability).
- Flows of passengers by mode and origin–destination. BTRE has developed a Non-Urban Passenger Travel Model. Inter-regional passenger numbers are forecast using a population-based gravity model derived from Bureau of Tourism Research data. Intra-regional passenger numbers are forecast to move in line with regional population projections.
- Flows of freight by mode, origin–destination and type. BTRE, with funding from Austroads, has developed a series of models for forecasting freight flows.
- Forecast flows of passengers and freight need to be converted into numbers of cars, trucks and trains on each segment of infrastructure.
- Access to and from the corridor. Origin–destination data shows quantities of traffic and trains entering and leaving the corridor at each access point. Information may be collected about the off-corridor infrastructure at these points to indicate access problems.
- Terminals (inter-modal and modal). The terminals from which freight enters and leaves the corridor may be identified, and data obtained on quantities handled and physical characteristics (e.g. storage capacity, rail track lengths, crane capacities).
- Bridges: location, length, width, superstructure material, type of superstructure, simply supported or continuous, span, length of maximum span, age, design load and condition.
- Environmental and heritage considerations. Information may be collected on issues and sensitive sites along the corridor.
- Safety: crash rates by type and location.
- Role of the corridor in national and state economies and the broader transport network. Major industries using the corridor, including the parts they use and the importance of those

industries in regional, state and national economies (output, employment). The extent to which the corridor is an alternative for, or feeder into, other corridors may be identified as well as the relative importance of corridors (passenger-kilometres, tonne-kilometres of freight, passenger and freight densities).

- Social factors. Populations dependent on the corridor and the purposes for which they use it. Populations affected by externalities created by the corridor, amenity of life and access to essential services.
- Australian Government, state or territory, local and private sector plans and priorities that relate to the corridor, land use and other plans, as well as plans for expansion or contraction of industries.
- Existing and proposed initiatives and policies on, or relating to, the corridor, including infrastructure initiatives under construction or at an earlier stage of development.

1.4 Economic assessment of networks

In BTCE (1994b; 1995)[6], a deficiency analysis of mainline rail infrastructure was undertaken to identify potential upgrading initiatives in an assessment of the adequacy of the infrastructure. Simple BCAs were undertaken on the potential initiatives identified by making estimates of investment costs and savings in train operating costs based on likely savings in train operating times. The BCAs did not incorporate improvements in reliability and benefits from traffic shifting from road to rail.

In an interstate rail network audit, the Australian Rail Track Corporation took a similar approach, except the BCAs were more detailed (ARTC 2001). In addition to savings in train operating costs, they included benefits to customers from increased reliability and lower transit times as well as benefits arising from modal shift.

The BTRE's Road Infrastructure Assessment Model (RIAM), which was employed in a rudimentary form for BTCE (1994a; 1995) and used in BTCE (1997), undertakes strategic assessments of non-urban road infrastructure. It is designed to rapidly process large databases of road segments containing road characteristics and traffic forecasts to estimate future needs for capacity upgrading, town bypass construction and maintenance for non-urban roads. RIAM incorporates a road-user cost model and undertakes BCAs of hypothetical investment and maintenance initiatives.

To assess capacity upgrading needs, RIAM rounds off each segment of road in the database to the nearest of a series of 'road standards' specified by the user. Examples of road standards are 'narrow two-lane with unsealed shoulders', 'wide two-lane with unsealed shoulders', 'wide two-lane with sealed shoulders', 'wide two-lane with overtaking lanes every 5/10/20 kilometres' and 'four/six/eight-lane divided'. The user specifies investment costs for upgrading from one standard to another. Different upgrade costs may be given for different terrains and regions.

RIAM hypothetically upgrades road segments to higher standards and employs the first-year rate of return criterion to determine whether the optimal time for the investment occurs in the past. It continues to upgrade the segment to find the highest warranted standard. Where there are alternative paths through the standards, RIAM tests all possibilities to find the optimum path. Given traffic forecasts for, say, the year 2020, RIAM can determine all the upgrades that have their optimal time before 2020 and provide an estimate of likely future spending needs.

Non-urban roads are relatively easy to assess at a strategic level because small sections can usually be considered in isolation. For railways where there is congestion, an entire corridor has to be analysed as a single entity to estimate the impacts on train timetables. Computer models can

6 Alternative sources of information about this study are Harvey (1995), Harvey and Miller (1995) and Collins et al. (1995).

be used for this work if levels of detail and accuracy, greater than those obtainable by expert judgment, are required.

A rail corridor can be represented as a single line. For urban roads, the situation is more complex because the network is two-dimensional, usually represented as a series of zones with an origin–destination matrix. As traffic between each origin–destination pair seeks the path with the lowest trip time, any change to the network will have complex ripple effects as traffic redistributes itself.

The methodology used by RIAM for strategic-level assessment of large amounts of transport infrastructure can be applied more generally, including to urban and rail infrastructure. The methodology is to:

- obtain current and forecast future levels of infrastructure usage
- identify a significant number of potential upgrading initiatives
- estimate the investment costs and benefits of the initiatives at a 'rapid BCA' level of accuracy
- for specified points of time in the future (for example, the present and then at five-year intervals), use the first-year rate of return criterion to find all the initiatives that have their optimal times in the past[7], and
- develop a schedule of warranted infrastructure spending for each time interval based on the first-year rate of return criterion.

Further analysis could involve:

- assuming the existence of budget constraints for each time interval and allowing the less attractive initiatives in each time interval to slip into subsequent time periods[8], and
- adjusting the selection of initiatives in each time period for non-economic considerations such as non-monetised benefits and costs and equity considerations.

1.5 Maintenance assessment at network level

Deficiency assessment

Maintenance standards for deficiency assessments of roads may include roughness, cracking, seal age, pavement strength, rutting, ravelling, potholing and skid-resistance. Rail maintenance deficiencies are often measured by indexes derived from visual inspection of the condition of rails, sleepers and ballast. An objective measure for rail is the track quality index (the sum of standard deviations of each rail surface, rail line, cross level and twist) used by the ARTC.

The standards applied in practice are usually higher for more highly trafficked roads and railways because economically optimal maintenance standards are closely related to traffic levels. In their strategic plans, state and territory road agencies often split networks into sub-networks, reflecting traffic volumes and relative economic importance, and apply higher maintenance standards to the more important sub-networks.

Assessment of road maintenance needs at a network level for roads is a highly developed science. Computer models such as HDM4 and dTIMS are used by road agencies to assist development of asset management strategies based on achieving predetermined maintenance standards in the presence of budget constraints. The deterioration algorithms in these computer models recognise that delaying relatively inexpensive treatments such as reseals can add greatly to long-term maintenance costs by dramatically shortening pavement lives.

7 If the present time is included, the study may reveal a backlog of initiatives that are economically warranted right away.

8 By estimating the benefits forgone as a result of delaying initiatives beyond their optimal implementation times, it is possible to derive estimates of the benefits achievable by increasing the size of the budget.

Setting minimum maintenance standards is a simple way of achieving equity and connectivity objectives where low traffic levels lead to economically warranted low maintenance standards.

Economic assessment

Economic assessment of maintenance initiatives for individual road segments is treated in detail in Section 2.17 in Volumes 3 and 5. In brief, a large proportion of maintenance expenditure (e.g. cutting grass, clearing drains, replacing guideposts and signs) is of a routine nature and is treated as a fixed annual cost. Economic analysis of periodic maintenance is usually undertaken as a cost-minimisation problem. The objective is to find the set of maintenance treatments and timings that minimise the discounted present value of combined road-user and road-agency costs, subject to a budget constraint.[9] Computer models such as HDM4 and dTIMS, and PLCC and PLATO developed by ARRB Transport Research, are able to apply the cost-minimisation methodology to whole networks.

To assess maintenance needs, BTRE's RIAM estimates routine maintenance and resealing costs on a simple per square metre per annum basis. For rehabilitations, it finds a cost-minimising set of timings. RIAM is not set up to handle budget constraints.

An alternative approach to minimising all costs to society, often employed in practice, is to set the maintenance standards exogenously (to maximum allowable roughness levels) and to minimise the present value of road-agency costs only. Higher maintenance standards may be set for more highly trafficked roads, so the outcome can approximate the most economically efficient outcome, with adjustment for equity considerations or community expectations for low trafficked roads.

If a predetermined amount of funds is allocated to maintenance, taking routine maintenance as fixed, economic analysis of periodic maintenance needs at a network level can provide information that is useful for determining the optimal split of funds between capital and maintenance activities. When scarce funds are allocated between construction initiatives on an economic basis, initiatives are ranked in descending order of BCR until the funds are exhausted. The cut-off, or marginal BCR, shows the benefit to be gained (forgone) by increasing (decreasing) the funds available for capital works.

For maintenance, an equivalent measure can be found from a network computer model set up to minimise the discounted present value of combined road-agency and road-user costs subject to a budget constraint. The benefit to society from increasing the maintenance budget by a given amount (say $1 million) is the reduction in the present value of combined road-agency and road-user costs for the whole network. Dividing this benefit by the cost ($1 million in additional funds given to maintenance) yields a marginal BCR for maintenance spending.

With an economically optimal split of funds between capital and maintenance, the marginal BCRs for the two categories of spending will be identical. If they are different, economic efficiency can be increased by shifting funds from the category with the lower marginal BCR to the category with the higher marginal BCR.

In applying computer models that rely on pavement deterioration algorithms, it is important to be aware of the limitations of the algorithms as they are not noted for accuracy. See Section 2.17.4 for an explanation.

9 If there are significant costs imposed on society by maintenance works, such as delays to traffic, increased accidents, and noise and other externalities, these should be added to the road-agency costs.

References

ARTC (2001b), *Interstate Rail Network Audit—Final Report.* Report prepared by Booz Allen Hamilton, Melbourne. Available at: http://www.artc.com.au/library/final_reportv14.pdf Australian Rail Track Corporation, Adelaide.

Austroads 1998, *Principles for Strategic Planning*, AP-55/98, Austroads, Sydney.

Austroads 1999, *Assessment Techniques and Tools for Rural Accessibility*, AP-137/99, Austroads, Sydney.

BTCE 1994a, *Adequacy of Transport Infrastructure: Intercity roads*, Working Paper 14.1, BTCE, Canberra.

BTCE 1994b, *Adequacy of Transport Infrastructure: Rail*, Working Paper 14.2, BTCE, Canberra.

BTCE 1995, 'Adequacy of transport infrastructure', in National Transport Planning Taskforce 1995, *Building for the Job: Commissioned Work, Volume 1*, BTCE Report, AGPS, Canberra.

BTCE 1997, *Roads 2020*, Working Paper 35, BTCE, Canberra.

Collins, P., Harvey, M., and Heyhoe, E. 1995, 'Assessing the adequacy of intercity road infrastructure', *Road and Transport Research*, *4(2)*, pp. 20-32.

Harvey, M. 1995, 'Assessing the adequacy of national transport infrastructure: A methodology', *Road and Transport Research, 4(1)*, pp. 58-67.

Harvey, M. and Miller, J. 1995, 'The adequacy of Australian transport infrastructure', *Transport and Communications Bulletin for Asia and the Pacific*, Economic and Social Commission for Asia and the Pacific, no. 65, pp. 1–14.

Queensland Government 2003, *Integrated Transport Planning Framework*, Brisbane, September.

Mintzberg, H. 2000, *The Rise and Fall of Strategic Planning*, Pearson Education, London.

Steiner G. A. 1979, *Strategic Planning: What Every Manager Must Know*, Free Press, New York.

2 Analytical techniques: Background material for Volume 3, Part 2

Volume 5, Part 2 of the *National Guidelines for Transport System Management in Australia* contains supporting material for Volume 3, Part 2 of the Guidelines. The section numbers and topics in this part are parallel to those in Volume 3, Part 2 to assist users to move between Volumes 3 and 5.

2.1 Specify the initiative and analyse options

2.1.1 Description

Describe the initiative, including its location, physical characteristics, function, estimated cost, timing and main benefits. At the detailed appraisal stage, describe the initiative in much greater detail than at the rapid appraisal stage.

Most proposed initiatives involve expending resources to improve the speed, reliability or safety of some part of the transport system. This generates direct benefits to users and indirect benefits to the broader community and the economy. In some cases, initiatives may be aimed at reducing the unwanted side-effects of transport; that is, externalities such as noise and pollution. They may also be aimed at replacing worn infrastructure or extending the life of existing infrastructure. In most cases, the initiative involves making a capital expenditure in the short-term, in order to obtain some ongoing future benefits.

The detail in which the initiative should be specified is considerably greater for detailed benefit–cost analysis (BCA) than rapid BCA.

2.1.2 Objectives

Show how the initiative contributes to achieving government transport system objectives using as much available detail as possible.

Successful initiatives are developed within the context of a clear overarching goal and a limited set of defined objectives. The objectives must be derived from an understanding of the service provided—what, to whom, how, when and where—and what the proponent wants to achieve and, if applicable, charge.

More specific objectives should be used where these are available. If a corridor or area strategy has been developed, the objectives are likely to be more detailed. The proposal should show how the initiative fits in with the corridor or area strategy. For example, if targets have been set for vehicle, or train, performance in the corridor or for the standard of infrastructure, the proposal should show how the initiative helps to achieve these targets. Alternatively, if the strategy identifies specific challenges for the infrastructure or its performance, the proposal should show how the initiative contributes to providing a solution.

2.1.3 Scope

If the initiative consists of discrete or separate components, each one must be justified as if it were an independent initiative. Where the impacts of a series of initiatives are closely interdependent, consider grouping the initiatives together and then treating them as a single initiative.

Establishing the scope of an initiative is an important initial stage. There is often pressure to broaden the scope. If broadening the scope adds to an initiative's complexity, there is a greater chance that options for variations to the initiative will be missed because there are more possibilities to consider.

In some cases, one initiative produces little or no benefit until another initiative or initiatives are completed. For example, longer trains can only be used extensively along an entire corridor when all passing loops on a rail corridor have been lengthened. If a series of passing loop initiatives are being considered for a single corridor, it does not make sense to assess each passing loop initiative in isolation. The entire program of lengthening passing loops on the corridor should be treated as single initiative—it is a case of the whole being greater than the sum of its parts.

Initiatives should only be grouped together in this exceptional circumstance—if the net present value (NPV) of the group of related initiatives, assessed together as though they were a single initiative, is significantly greater than sum of the NPVs of the constituent initiatives assessed individually.

The grouped initiatives should not include 'free-riders' that do not pull their weight in contributing to the overall group result. An initiative should be added to the group only if the resulting increase in group NPV exceeds the initiative's stand-alone NPV.

The corridor and area planning approaches developed in Volume 2 of the Guidelines can help to identify where synergies exist between initiatives.

See Section 2.1.11 for a general discussion of synergies between initiatives.

2.1.4 Staging

No additions to the corresponding section in Volume 3 are made here.

2.1.5 Constraints

No additions to the corresponding section in Volume 3 are made here.

2.1.6 Base Case

The proposal should specify the Base Case, including any significant assumptions about actions that need to be undertaken in the Base Case, and one-off, future events that affect benefits or costs.

A BCA is a comparison between two alternative states—a Base Case and a Project Case. Benefits are always measured against the Base Case. For example, the benefits to existing users of a transport initiative are measured by taking the difference between the total costs they incur in the Base Case and the total costs they incur in the Project Case.

Usually, the Base Case is whatever would be done in the absence of the additional investment in the initiative, or business-as-usual. This is rarely the same as the do-nothing scenario. For example, where an upgrading initiative includes scrapping existing infrastructure, specifying the Base Case involves making assumptions about when, and at what cost, the existing infrastructure is replaced.

Specifying the Base Case requires forecasting, as does the Project Case; there is uncertainty and therefore assumptions are required.

One-off future events can affect the benefits and costs of an initiative, altering both the Base and Project Cases. Changes in land use, such as new suburbs or developing towns, can generate increased car traffic on a road. Changes in economic activity, such new factories, can lead to greater freight traffic. Major assumptions about future one-off events should be specified as part of the Base Case.

Implementing other transport initiatives can affect the benefits and costs of the initiative being analysed. For example, a major rail improvement could reduce the freight travelling by road. A major road improvement in one part of the corridor could have impacts on road traffic elsewhere in the corridor. There may also be effects on the cost side. For example, a new rail line could require a bridge or underpass in a new road initiative, or a road and a railway could share an alignment. The Base Case may, therefore, include assumptions about the nature and timing of other transport initiatives being implemented.

2.1.7 Options analysis

Demonstrate that different options for the initiative were considered. Provide the reasons for eliminating options.

Governments are keen to see innovative options considered as part of the proposal, including options that make better use of existing infrastructure, reducing the need for major capital expenditures.

A proposal may contain more than one option, deferring the choice between options until more detailed assessments have been undertaken.

From a purely economic efficiency point of view, the best option is the one with the highest NPV. As discussed in Section 2.10, the highest NPV is the decision rule to use for choosing between mutually exclusive options. An alternative is the incremental benefit–cost ratio (IBCR). But it should only be applied where options involving different sizes of initiative or standards of infrastructure are compared.

2.1.8 Pricing assumptions

Specify any assumptions about charging for infrastructure usage.

Pricing of infrastructure usage affects demand, which in turn affects benefits and costs. For the BCA and financial analysis, assumptions need to be made about the nature and level of charges. If there is question about tolling a road, BCAs should be undertaken with and without tolling. The implications of tolling on initiative benefits are discussed in Section 2.18.

Objectives for pricing, for example profit maximisation, economic efficiency maximisation, achievement of a revenue target or cost recovery, need to be specified. For a cost recovery objective, precisely specify the costs to be recovered and how they are to be measured.

2.1.9 Funding options

No additions to the corresponding section in Volume 3 are made here.

2.1.10 Financial analysis

No additions to the corresponding section in Volume 3 are made here.

2.1.11 Relationships between initiatives[1]

Relationships between initiatives can take the following forms:

- Independence: Implementation of one initiative has no effect on the benefits or costs of another initiative.
- Complementarity (or synergy): Implementation of one initiative increases the benefits or reduces the costs of another initiative. In other words, the existence of one initiative *increases* the need for the other initiative. This occurs where one initiative is upstream or downstream from the other. For example:
  - A highway upgrade that generates new traffic could increase traffic along other sections of the same highway, also increasing the benefits of initiatives located on these sections.
  - Upgrading a railway line could result in greater truck numbers along roads leading to rail terminals, increasing the benefits from upgrades to feeder roads.
  - The most extreme example of a complementary relationship is where the benefits from one initiative are zero unless a complementary initiative or initiatives are implemented. This could be the case for raising height clearances or lengthening passing loops on a section of rail track.
  - Complementary relationships can also occur for capital costs. For example, a new alignment could be constructed that is to be shared by a road and a railway line, saving on capital costs compared to the alternative of separate alignments for each initiative.
- Substitutability (or negative synergy): Implementation of one initiative reduces the benefits or increases the costs of another initiative. In other words, the existence of one initiative *reduces* the need for the other initiative. This can occur where one initiative is on an alternative route or relates to an alternative mode. For example:
  - A railway upgrade that causes freight to shift from road to rail reduces the benefits of, and hence delays the need for, upgrading initiatives along the road.
  - A new road may necessitate an additional bridge for a new railway line or vice versa, causing capital cost increases.

Complementarity between initiatives leads to the NPV for initiatives assessed collectively being greater than the sum of the NPVs of the initiatives assessed in isolation. In other words, the whole is greater than the sum of its parts. Substitutability has the opposite effect.

As for optimal timing, implementation of an initiative brings forward (or delays) the optimal implementation time for a related complementary (or substitute) initiative.

Where options exist for different sizes of an initiative, increasing the size of one initiative increases (or reduces) the optimal size of a related complementary (substitute) initiative.

Implications for appraisal include:

- Appraise initiatives as a single combined initiative where the level of complementarity between initiatives is high (as recommended in Section 2.1.3).
- Appraise initiatives as mutually exclusive options where the level of substitutability between initiatives is high.
- Compare options involving different combinations of related initiatives, different implementation times and, if applicable, different initiative sizes.
- Specify assumptions about whether related initiatives are implemented, when they will be implemented and, if applicable, their size in the Base Case (as recommended in Section 2.1.6).

1 See Austroads (2005a, pp. 11-12) for an alternative discussion of relationships between initiatives.

2.2 Benefits and costs

2.2.1 Secondary economic impacts

The corresponding section in Volume 3 uses the terms 'secondary impacts' or 'flow-on effects' to refer to benefits and costs passed on, or redistributed, within the economy. The most accurate measurement of benefits and costs is usually achieved by measuring them as close to their source as possible.

For transport improvements, a large part of the benefits usually appear initially as savings in user costs. Savings for freight may be passed on to suppliers and purchasers of the goods being transported. If the purchasers are producers, benefits may be passed on to other producers or retailers, and finally to consumers, in the form of lower prices. When spread throughout the economy in this way, measurement of benefits is impossible. Therefore, benefits from transport initiatives are estimated at their source, using models of road vehicle or train operating costs. If *both* savings in transport-user costs and reductions in prices paid by producers or consumers are included in a BCA, benefits are being double-counted.

An exception to the rule of measuring benefits as close to their source as possible is when a direct estimate of a benefit or cost is not available. A secondary impact can then be used to value the benefit or cost. For example, estimates of the costs of noise are sometimes obtained from consequent changes in land values (called hedonic pricing).

In all other circumstances, impacts on land values are excluded from BCAs. Improved access to properties as a result of a road upgrade may increase the value of the land. This is not a new benefit, simply the savings in time and vehicle operating costs reappearing in a changed form.

Where a transport initiative results in increased profits for certain businesses and losses for others, the losses in profits to the latter are not treated as costs of the initiative. From the point of view of society as a whole, these losses are exactly offset by gains to customers and to other businesses.

Often, the creation of employment is claimed as a benefit. But jobs associated with a particular initiative are seldom a social benefit because the employment is usually transferred from another part of the economy. From the point of view of society as a whole, the jobs created by an initiative generate benefits in addition to the value of the output produced, only where the initiative creates employment for people who would otherwise be unemployed. An example is where an initiative creates jobs in an area with high structural unemployment, with barriers to mobility that prevent potential workers from relocating to other areas where unemployment is lower. If an initiative creates employment for workers who would otherwise be idle, their opportunity cost may be less than the wage rate. Section 2.3.5 below shows how to estimate a 'shadow wage rate' in these cases.

There is no guarantee that employment benefits will stay in the region in which the initiative is undertaken. For example, fly-in-fly-out mining operations may provide jobs located in a particular region but, because of the nature of the operation, the employment generated is for workers from a capital city or larger centre. The income generated by that employment may be spent outside the region where the jobs are located.

Tourism benefits are particularly difficult to measure, because of the need to assess the extent to which an initiative is attracting new tourists (or increasing the length of stay or expenditure of existing tourists) or just changing the travel patterns of existing tourists so that other towns or regions lose tourism.

For a detailed discussion of issues relating to benefits and costs that can legitimately be included in BCAs, see BTE (1999b).

2.2.2 Depreciation

No additions to the corresponding section in Volume 3 are made here.

2.2.3 Point of view of the BCA

BCAs are usually undertaken from the point of view of society as whole. 'Society' is regarded as falling within national boundaries. For example, benefits accruing to foreigners as a result of paying lower prices for Australian exports are not counted in a BCA. Impacts on foreign tourists should not be counted, except where their activities have secondary impacts on the economic welfare of Australians. However, for land transport investments, the impacts on foreign nationals are usually a small proportion of the total, so can reasonably be ignored, especially given the difficulty in obtaining the necessary information.

2.2.4 Length of life of the initiative

It is usual to assume a 30-year life for road infrastructure initiatives and a 50-year life for rail infrastructure initiatives. Intelligent transport system initiatives have shorter lives. The life of an initiative is deemed to commence on completion of construction. Benefits and costs should be forecast for each year of the initiative's life.

When comparing initiative options with different lives, adjustments should be made to ensure a valid comparison. There are two ways to do this:

1. Find a common multiple of the lives (for example, 150 years for a 30-year road initiative and a 50-year rail initiative), and project the benefits and costs forward, assuming that the initiative is replaced at the end of each initiative's life. Discount the present value of benefits and costs, beyond the end of the first life (year 30 for the road and year 50 for the railway), to that time and treat the result as a residual value. The residual value concept is discussed in Section 2.3.3 in Volume 3.
2. Annuitise the present value of net benefits over the life of the initiative. This gives the same result as Method 1, but precludes the possibility of allowing replacement benefits and costs beyond the first life to grow with demand. An annuity is the constant amount received at the end of each year over a given number of years that discounts back to a given present value. The annuity formula is

   $$PV \times \frac{r}{1-(1+r)^{-n}}$$

   where PV is the present value, r is the discount rate and n is number of years over which the annuity is taken.

For further information, including a simple worked example, see Boardman et al. (1996, pp. 138–40).

Adjusting for different lives is necessary only for estimating NPVs to compare mutually exclusive options. It is not required for ranking initiatives by BCR to satisfy a budget constraint because the replacement initiatives are not competing for funds out of the current budget.

2.2.5 Real and nominal prices

In principle, BCAs and financial analyses may be undertaken in either real prices (constant dollars) or nominal prices (current dollars). Real prices represent the prices of the day, that is, price estimates without inflation. Nominal prices refer to prices including inflation.

Provided the analysis is undertaken consistently, the end results should be identical. It is vital to use the correct discount rate. The relationship between a real and a nominal discount rate is

$$(1+i)=(1+r)(1+f)$$

where i is the nominal discount rate and f is the inflation rate. Discounting nominal values at a nominal discount rate produces the same discounted result as discounting in real terms. The following equation demonstrates this

$$NPV = \sum_{t=0}^{n} \frac{R_t(1+f)^t}{(1+i)^t} = \sum_{t=0}^{n} \frac{R_t}{(1+r)^t}$$

where R_t is net cash flow in year t in real terms in year zero dollars, and $R_t(1+f)^t$ is the cash flow in year t in year t dollars. When working in financial terms, if the inflation rate changes during the life of the initiative, the nominal discount rate must be changed accordingly, so that it always aligns with the inflation rate assumed in the numerator.

It is usual to undertake BCAs in real terms and financial analyses in nominal terms. Proposals that include both a BCA and a financial analysis should make clear (by links within a spreadsheet) how the two analyses relate to each other, including the inflation adjustments made to convert between the BCA in real terms and the financial analysis in nominal terms.

When undertaking a financial analysis in real terms, fixed monetary amounts, such as the value of tax deductions for depreciation, need to be deflated.

2.3 Investment costs

2.3.1 Timeline

No additions to the corresponding section in Volume 3 are made here.

2.3.2 Cost estimation

No additions to the corresponding section in Volume 3 are made here.

2.3.3 Residual value

No additions to the corresponding section in Volume 3 are made here.

2.3.4 Land costs

Land as an investment cost

Land costs should only be included in BCAs where the land has an opportunity cost; that is, if it has an alternative use for which people are willing to pay.

Examples of land used for transport initiatives that have no opportunity cost include:

- land that is needed to access properties, regardless of the quality of the road on that land, and
- a corridor, acquired in the past for an initiative, that is not well suited for any other purpose and for which there is a significant willingness-to-pay (WTP) (e.g. housing, industrial use). The only alternative use of the land is open space. This could occur because of the shape of the land (e.g. too narrow; for example, the land acquired for widening a road), the cost of converting the land to an alternative use (e.g. relocating fences), changes in the surrounding infrastructure that have cut off access to the land, the high environmental value of the land (e.g. rare species of flora or fauna)[2] or political constraints such as community pressure to retain the land as public space.

2 Ideally, the value of the flora or fauna would be included as a cost to the initiative. To rule out alternative uses of land on environmental grounds is equivalent to assuming that the environmental value is high enough to outweigh the net benefits of all other possible alternative uses.

Examples of land having an opportunity cost include:

- land that has to be purchased in order for the initiative to proceed, and
- land acquired for an initiative in the past, which could be sold for another purpose.[3]

If the land required for an initiative has an opportunity cost, the land should be valued at its market price

The opportunity cost of land is the rent the land would earn over the life of the initiative and should be discounted to the present. The market price of land should equal the present value of expected future rents into the infinite future. Counting the market price of land as an investment cost is equivalent to treating forgone rent in each year of the initiative's life as a negative benefit.

Land as a residual benefit

The market price takes account of rents earned by land beyond the life of the initiative. So there may be a case for including the market price of land at the end of the initiative's life as a residual benefit. It hinges on whether the land is likely to have a zero opportunity cost at the end of the initiative.

If land is omitted from investment costs at the start of the initiative because it has a zero opportunity cost, then the same is likely to apply to the value of the land at the end of the initiative's life.

If land is included as an investment cost, and it *could* be sold for an alternative use at the end of the initiative's life, the land can legitimately be accorded a residual value at the end of the life of the initiative. Even if non-replacement of the infrastructure is improbable, the land may have a potential alternative use. Say, the infrastructure will need replacing in 30 years time. The opportunity cost of the land (forgone rents) for years 30 to 60 (the life of the replacement infrastructure) should be taken into account when making the replacement decision.

Land that cannot be converted to an alternative use at the end of the initiative's life has no residual value. The initiative has effectively rendered the land useless for any purpose other than a replacement project after the end of the initiative's life. When evaluating the replacement project, the land would have a zero opportunity cost.

Estimating the residual value of land

Estimating the residual value of land involves forecasting its value far into the future. As a rule, the value of the land can never be assumed to grow faster than the discount rate; otherwise, the present value of the residual would exceed the current value, implying that the initiative created a benefit by depriving society of alternative uses of the land over the life of the initiative.

That is not to say that land values can never rise faster than the discount rate in practice. If there was certainty that the price of a given piece of land would rise at a rapid rate, the current market price would rise to eliminate the prospect of profit-taking from holding the land. Then, the rapid rise over the life of the initiative would not occur. If, ex post, the land value does rise rapidly over the life of the initiative, from an ex ante perspective this is just one scenario out of a probability distribution of scenarios. The rapid rise scenario is at the high end of the probability distribution. At the low end of the distribution, there would be small increases in value and possibly even declining values. The current market value of a given piece of land is based on the mean of the probability distribution of possible future rent profiles over time. And, in BCA, it is the means of probability distributions that are required.

3 In this case, the proceeds from the sale are a benefit in the Base Case that is forgone in the Project Case, and so translate into a net cost to the initiative.

If the supply of land was perfectly inelastic, average rents could be expected to increase in line with real income. This is likely to be the case for inner metropolitan areas. The present value of future rentals and the value of the land will grow at the same rate as annual rents.[4] Where the supply of land is perfectly elastic, the rents, and therefore values, should stay constant over time. For rural and outer metropolitan areas, the situation is likely to be somewhere in between these two extremes. So as a general rule, for estimating residual values of land:

- for inner metropolitan areas, use a long-term forecast growth rate for GDP as the maximum, and
- elsewhere, assume that the value remains unchanged.

It may seem paradoxical that the higher the forecast increase in land price, the smaller the present value of the cost to society of the initiative occupying land over its life. However, the value of the land at the current time is quite sensitive to the expected growth rate (see formula in footnote). So, while a forecast high growth rate relative to the discount rate reduces the cost of occupying land in a BCA, the high current cost of the land should offset this.

2.3.5 Measurement of costs

This section is background material for Boxes 2.2, 2.3 and 2.4 in Volume 3.

Basic concepts

Willingness-to-pay (WTP)

BCA is based on a value judgement of consumers' sovereignty. All benefits and costs are valued in accordance with what consumers are willing to pay for them (or to avoid them if they are 'bads' rather than 'goods'). Willingness-to-pay (WTP) is the total value that consumers place on a given quantity of a good or service. It represents the maximum amount they are prepared to pay to have that quantity rather than do without it. It can be measured as the area under the demand curve between zero and the specified quantity—the total shaded area shown in Figure 2.1.

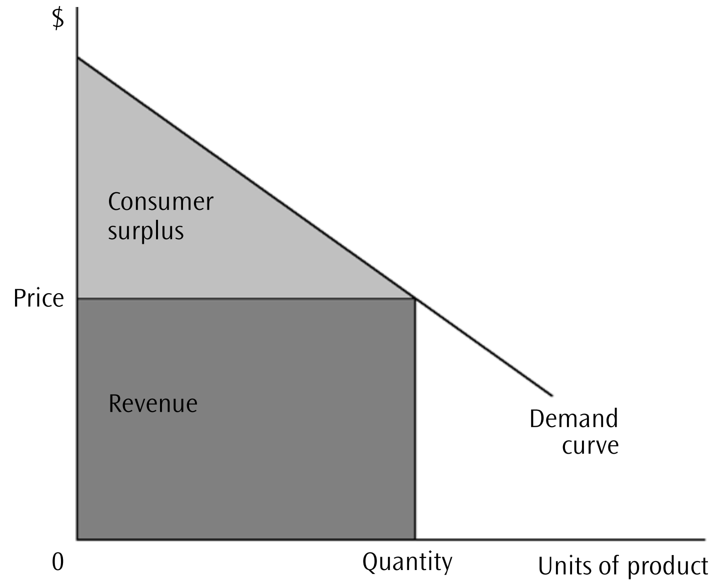
Consumers' surplus

WTP can be divided into two components, also shown in Figure 2.1:

- revenue—what consumers actually pay, and
- consumers' surplus—the maximum amount they are willing to pay, less what they actually pay.

4 The annual rate of growth in land value over time is equal to the annual rate of growth in rents (g). Provided g < r, (where r is the discount rate) the land value at year *t* is $V_t = a(1+g)^t \Big/ \left[\frac{1+r}{1+g} - 1\right]$ where *a* is annual rent in year zero.

Figure 2.1: Willingness-to-pay divided into consumers' surplus and revenue components



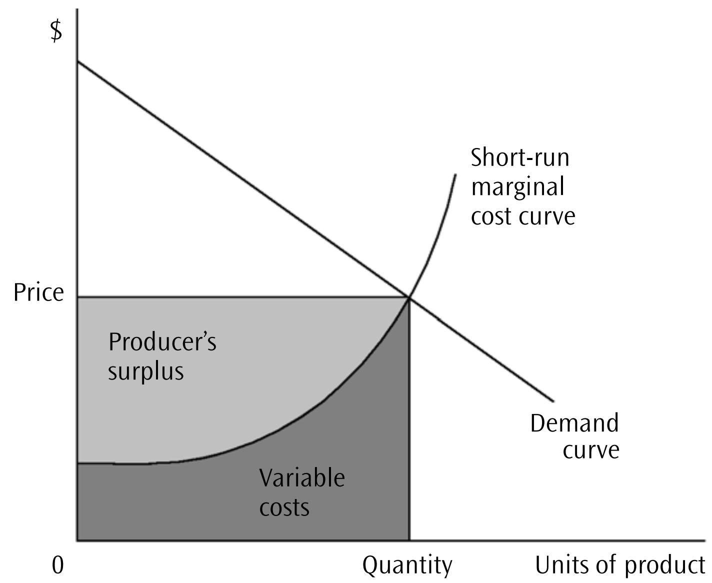
Producers' surplus

In the absence of taxes, revenue can be divided into two components:

- variable costs—costs of materials and direct labour, and
- producers' surplus—the residual that is split between profits and fixed costs, including overheads and interest (see Figure 2.2).

The producers' surplus concept is not relevant for valuing benefits and costs to road users because *individual* vehicles have constant usage costs. It is relevant for railways and other producers affected by transport initiatives.

Figure 2.2: Revenue divided into producers' surplus and variable cost components



Taxes on inputs

Taxes can be levied on outputs, inputs or as fixed amounts paid by producers. Taxes on outputs are represented in diagrams by either shifting up the cost curves or shifting down the demand curves. Taxes on direct inputs (payroll taxes, fuel excise) shift cost curves up. Taxes that are fixed amounts such as corporate income tax and registration fees for cars have no effect on the variable cost curves shown in the diagrams used in BCA.

Box 2.3 in Volume 3 recommends that to estimate resource costs, taxes, subsidies and tariffs should be removed from physical inputs such as fuel, tyres, vehicles and trains, but should be left in for labour. The underlying reasoning is based on the assumption that physical inputs are available in infinitely elastic supply and that the supply curve for labour is perfectly inelastic. These are special cases of a general rule.

Where the source of an input is additional supply (e.g. additional fuel, imported vehicles or manufactured concrete), the resource cost is the supply cost and excludes taxes, subsidies or tariffs. For an imported good, the supply cost is the cost of the foreign exchange required to purchase it and excludes tariffs. The resource cost is the area under the supply curve for the increased production.

Where the source of an input is reduced consumption by firms or households, because the initiative competes for resources by forcing up the price of the input (or wage for labour), the resource cost is the value placed on the input by those who can no longer consume it. The value is their WTP, measured by the price they pay. Taxes and subsidies should *not* be taken out. The resource cost is the area under the demand curve for the reduced consumption.

Increased demand for an input by an initiative (say, by Q units) results in a right shift of the demand curve for the input. Where the input is supplied under competitive conditions, with price determined by the intersection of an upward-sloping supply curve with a downward-sloping demand curve, the right shift of the demand curve leads to a price increase. The additional demand for the input by the initiative is met by a combination of decreased consumption (aQ units) due to the price increase, and increased supply ($(1–a)Q$ units) induced by the price increase. Using the notation of Sugden and Williams (1978, p. 105), the social cost of each unit of the input used by the initiative is

$$a(\text{gross-of-tax price}) + (1–a)(\text{net-of-tax price}).$$

Thus the shadow price is a weighted average of the gross-of-tax and net-of-tax prices, the weights dependent on the proportions of the initiative's inputs sourced from reduced consumption and increased supply. The value of a is determined by the relative slopes of the demand and supply curves. It can be shown that

$$\frac{a}{(1-a)} = \frac{-e_d}{e_s}$$

where e_d and e_s are the elasticities of demand and supply respectively (Sugden and Williams 1978, p. 106).

When evaluating a transport initiative, it is adequate to assume the physical inputs are available in infinitely elastic supply ($a = 0$) and labour in fixed supply ($a = 1$). For physical inputs that are internationally traded, the assumption of infinitely elastic supply is justified because Australia is a price-taker.

For labour, the demand curve by a competitive industry is the value-of-marginal product curve. Firms in a competitive industry will employ labour up to the point where the pre-tax wage rate equals the value-of-marginal product. Hence, the height of the demand curve, at any point, represents the value of a worker's contribution to the economy. The social opportunity cost of labour drawn away from other uses is, therefore, given by the pre-tax wage rate.

For a detailed discussion, see Perkins (1994).

Shadow wage rate

This section expands on the discussion of the shadow wage rate in Box 2.2 in Volume 3.

The model presented here applies to a very specific form of unemployment, where the number of workers who desire jobs at the wage paid in a particular labour market exceeds the number of workers employers are willing to hire at that wage (Boardman et al. 1996, p. 71). The market may be defined in terms of occupation, skills or location. The wage rate is assumed to be fixed by governments or unions at a level above the rate that is determined by a competitive labour market. The opportunity cost of employing an otherwise unemployed worker is the leisure forgone by the worker plus any presentation costs (transport, relocation, special clothing).

In Figure 2.3, the lowest upward sloping curve represents the value of leisure (*VL*). Presentation costs are added on to obtain the *VL + PC* curve. This represents the supply curve for labour in the absence of unemployment benefits and taxes. In such a situation, a person accepting a job forgoes their leisure and incurs the presentation costs. When unemployment benefits *(UB*) are available, the person accepting a job forgoes unemployment benefits in addition to leisure and presentation costs. No economically rational person would work for less than *VL + PC + UB*. So this becomes the supply curve for labour when unemployment benefits are available, but with the absence of taxes. Imposition of income tax *(IT*) and payroll tax *(PT*) creates a wedge between wage costs incurred by the employer and the wage received by workers, raising the supply curve further, to *VL + PC + UB + IT + PT*.

The Base Case demand curve for labour is D_1. With the wage rate paid by employers (including payroll and income taxes) fixed at *W*, there is an excess supply of labour of $L_3 – L_1$. The initiative being appraised demands $L_2 – L_1$ of labour, shifting the demand curve right to D_2. The demand is met by otherwise unemployed workers entering the workforce. The resource cost for those workers is given by the shaded area under the *VL + PC* curve: the value of their leisure forgone plus presentation costs.

Boardman et al. (1996, pp. 71–4) argue that there is no basis for assuming that the particular unemployed persons hired will value their leisure as measured by the height of the VL curve between L_1 and L_2. The particular individuals finding employment could have values of leisure anywhere between *e* and *f*. Boardman et al. suggest assuming that the workers gaining employment from the initiative have values of leisure uniformly distributed between *e* and *f*, making the value of forgone leisure *(f + e)/2*. Since the value of *e* is unknown, Boardman et al. suggest assuming that the value of leisure at this point is zero. The value of forgone leisure for the workers gaining employment from the initiative then becomes *f/2*. The opportunity cost of unemployed labour or the shadow wage rate is *f/2 + PC*.

Since $f = W – IT – PT – UB – PC$, the shadow wage rate is

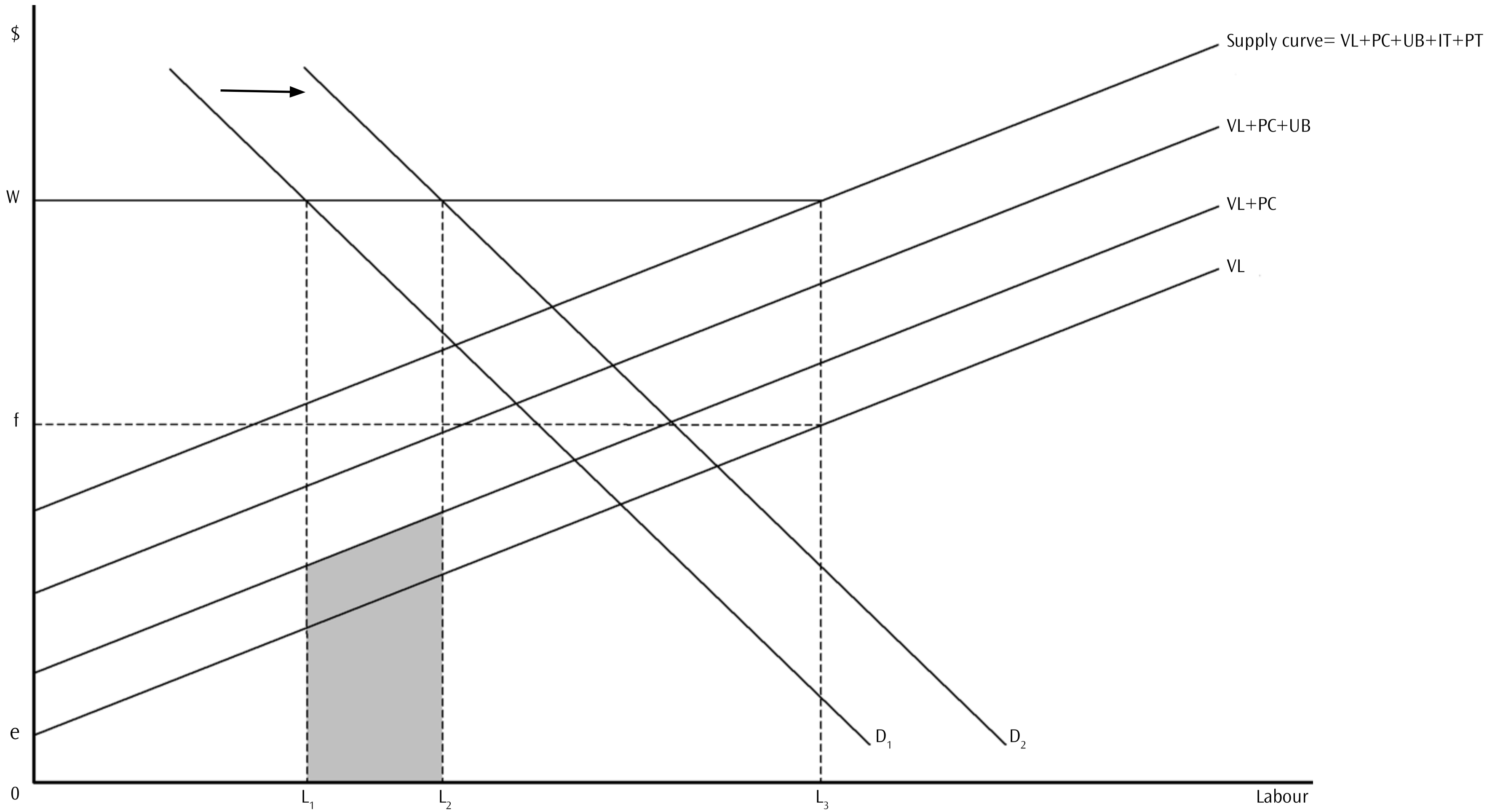
$(W – IT – PT – UB – PC)/2 + PC = (W – IT – PT – UB + PC)/2$,

that is

(pre-tax wage – income tax – payroll tax – unemployment benefit + presentation costs)/2.

Note that a shadow wage rate for unemployed workers should be used in BCAs only in exceptional circumstances: where there is high unemployment of workers of a specific type in a specific region. The shadow wage rate should be applied only to the proportion of workers assumed to be otherwise unemployed, not to all workers employed by the initiative.

Figure 2.3: Shadow wage rate for labour



Generalised cost

Transport users are not only concerned with the money costs of the service, but also with quality aspects such as speed, reliability, departure times and service frequency. For passengers, discomfort and lack of security can be concerns. For freight, probability of damage or theft and packing costs are considerations. Other costs include the money costs, time taken and other characteristics of the door-to-door journey as well as those for the primary mode of transport used (the line-haul): for passengers—walking between home and a bus stop, parking a car, transport to and from an airport or railway station; for freight—pick-up and delivery, packaging to minimise damage.

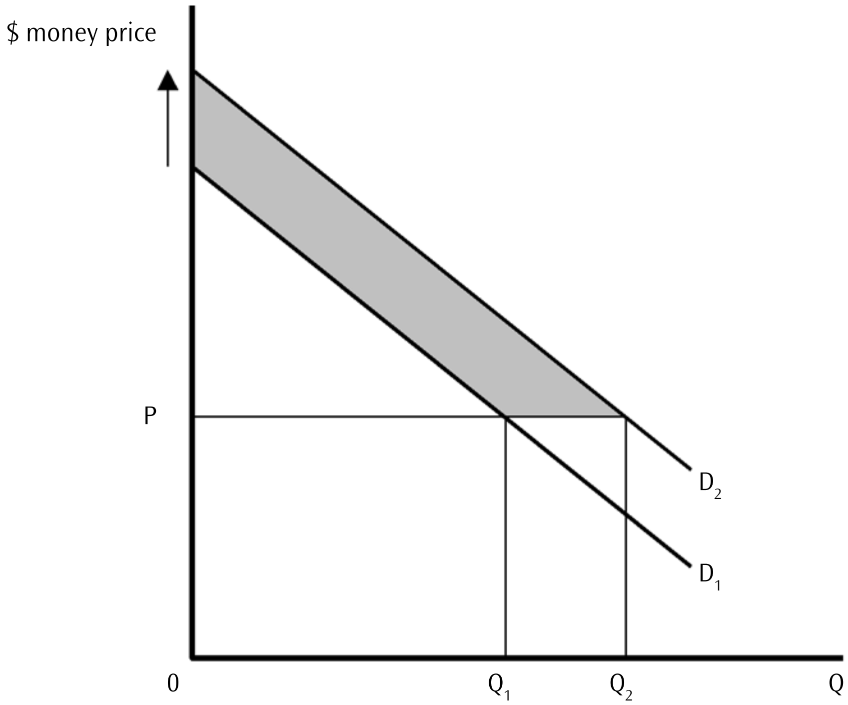
For analysing and forecasting transport demand and estimating benefits and costs, it is often convenient to combine these costs into a single index—the *generalised cost* (Button 1993, pp. 85–9). Some additional costs such as the cost of a taxi to the airport or the cost of pick-up and delivery for freight are already expressed in monetary terms and can be readily added onto the line-haul cost. Other costs, such as the time taken, level of discomfort for passengers, and unreliability of the service, are quality aspects that have no obvious market prices.

To incorporate quality aspects of a transport service into generalised costs, it is necessary to first define them in a negative sense as costs; for example, discomfort and unreliability are 'bads' that impose costs, rather than comfort and reliability that are 'goods' that create benefits. Secondly, it is necessary to have some measure of the quality attribute. Time is straightforward. Reliability can be measured in terms of standard deviations of trip times, or percentage of arrivals later than a set period of time after the scheduled time. Discomfort is problematic, and so is not usually included in generalised costs. Finally, it is necessary to attach a money value to each unit of the quality aspect, for example, a value of time. In practice, there is a distribution of values of time and other negative quality attributes across the population of transport users. The values used to estimate generalised costs must therefore be averages.

Demand curves for transport may be expressed as functions of either money or generalised costs. In the case of the former, demand is a function of price (money cost), time and quality attributes. A change in the quality of the transport service causes a shift of the demand curve. If demand is expressed as a function of generalised costs, then a change in any of the quality attributes included in generalised cost causes private generalised cost to change and a movement along the demand curve.

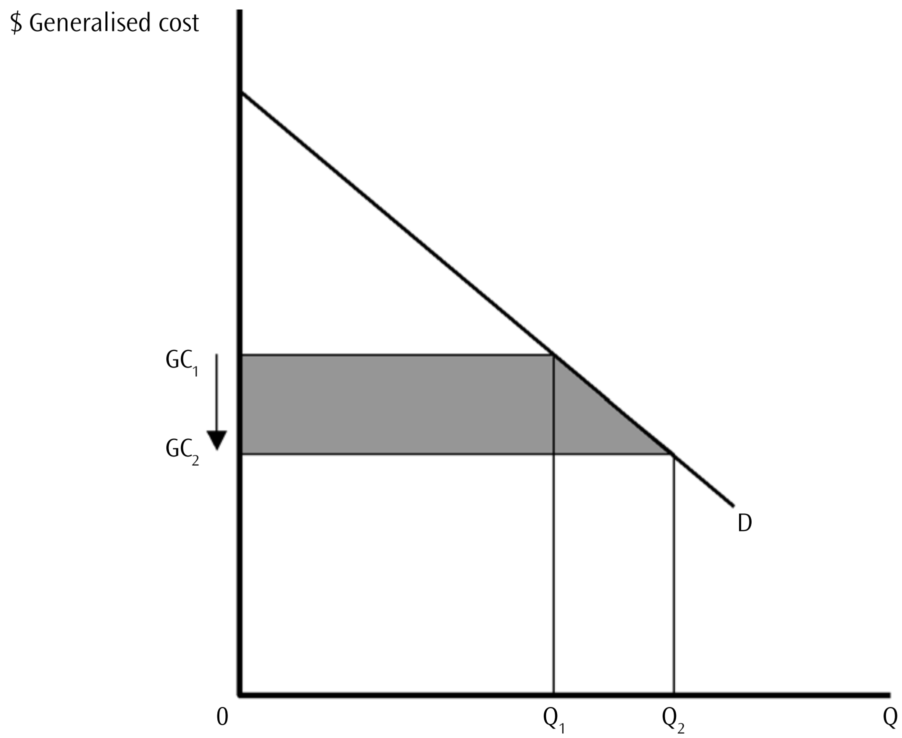
In Figure 2.4, the vertical axis measures money prices, not generalised costs. An improvement in service quality, such as saving trip time, shifts the demand curve up, increasing WTP and generating a gain to consumers equal to the shaded area.

Figure 2.4: Money cost measurement of benefit from a quality improvement



In Figure 2.5, the vertical axis measures generalised costs. An improvement in a service quality attribute included in generalised cost has no effect on the demand curve. The price falls instead, leading to an increase in consumers' surplus equal to the shaded area. The sizes of the welfare gain areas in Figures 2.4 and 2.5 are identical.

Figure 2.5: Generalised cost measurement of benefit from a quality improvement



If demand is defined in terms of generalised costs, it is necessary to use 'social *generalised* costs' instead of social costs in economic welfare calculations.

The conventional inverse demand curve can be defined as $Q = Q(P, a_1,a_2, \ldots, a_n)$, where P is money price and the *a*'s represent quality attributes. The corresponding WTP function is $WTP = WTP(P, a_1,a_2,\ldots, a_n)$.

Social welfare can be defined as

$$SW=WTP(P, a_1,a_2,\ldots, a_n) - TSC = WTP(P, a_1,a_2,\ldots, a_n) - [ACr + ACe]Q(P, a_1,a_2,\ldots, a_n)$$

where, SW = social welfare, TSC is the total social cost, ACr = the average resource cost of providing the transport services (e.g. vehicle operating costs, costs of infrastructure provision and operation, all valued in resource terms) and ACe is the average cost of externalities.

Generalised cost can be defined as

$$GC = P+\sum_i v_i a_i = P+ACu$$

where the *v*'s are the average values placed on quality attributes by users and ACu is the average user cost.

With demand specified in terms of generalised cost, the inverse demand curve is $Q = Q_{GC}(GC)$.

For this demand curve, there is a corresponding WTP function, $WTP_{GC} = WTP_{GC}(GC)$.

The conventional and generalised cost WTP functions are related as follows

$$WTP_{GC} = WTP(P,a_1,a_2,\ldots,a_n)+ ACuQ(P,a_1,a_2,\ldots a_n)$$

because the generalised cost demand curve is the conventional demand curve shifted up by *ACu* at each quantity level.

In the measure of social welfare, if *ACuQ* is added on to conventional WTP when demand is expressed in generalised costs, the same amount, *ACuQ*, must also be added on to social costs; that is

$$SW = WTP(P, a_1, a_2, \ldots, a_n) + ACuQ - [ACu + ACr + ACe]Q = WTP_{GC} - TSGC$$

where *TSGC* is total social generalised cost.

Perceived price

The terms 'perceived costs' and 'behavioural costs' are sometimes found in transport economics literature referring to the costs that road users perceive and therefore base their decisions on. They equate to generalised costs less costs that transport users do not perceive. For example, car users may not perceive the variable cost components of vehicle depreciation and maintenance. They may not be aware of the connection between speed and fuel consumption (Button 1993, p. 87). Where vehicles are provided by employers for private use, drivers may face no variable costs at all, so the only perceived cost of driving is time.

As it is usual to think of demand as being a function of price rather than cost, the Guidelines use the term 'perceived price' rather than 'perceived cost'.

Illustration

Table 2.1 shows how the cost concepts are related.

Table 2.1: Average variable car operating costs

(cents per vehicle-km, 1996 prices)

ITEM	MONEY	PERCEIVED	SOCIAL
Travel time[a]	–	40.0	40.0
Vehicle operating costs:			
Fuel	8.9	8.9	3.8
Tyres	1.0	–	1.0
Maintenance	7.3	–	7.3
Depreciation	4.8	–	4.8
Sub-total	22.0	8.9	15.5
Crashes	–	–	5.0
Environmental impacts	–	–	4.5
Road supply	–	–	3.1
Total	22.0	48.9	68.1

a. Based on \$12 per hour and 30 km/h.

Source: Adapted from Bray and Tisato (1997).

2.4 Demand forecasts

2.4.1 Unit of demand

No additions to the corresponding section in Volume 3 are made here.

2.4.2 Market segmentation

No additions to the corresponding section in Volume 3 are made here.

2.4.3 Base for projection

No additions to the corresponding section in Volume 3 are made here.

2.4.4 Base Case forecasts

Simple extrapolation

Extrapolation of past trends involves fitting a trend to data points from the past, usually with regression analysis. At least two data points are needed. The most common trends to assume are linear, $Y = a+bt$, and constant growth rate, $Y = a(1+g)^t$, where Y is the forecast variable, a and b are constants, t is time and g is the growth rate.[5]

Econometric models

More sophisticated forms of forecasting use an econometric model in which the variable being forecast is determined by explanatory variables such as population, real incomes, economic activity (gross domestic, state or regional product) and transport costs. The coefficients are obtained from regression analysis of past data. Forecasts are made using projections of the explanatory variables from sources such as the Australian Bureau of Statistics for population projections, or Commonwealth or State treasuries for forecasts of economic activity.

As example of the form of an econometric model is

Freight demand = $\alpha_1 \times GDP^{\alpha_2} \times \text{freight rate}^{\alpha_3} \times \alpha_4^{\text{time}}$

where the αs are constants, α_2 and α_3 are elasticities and GDP is gross domestic product. Time is included in this example as a proxy to allow for a trend caused by changes in tastes or technology. A function in this form can be linearised to facilitate regression by taking logs of both sides.

Gravity models

Gravity models are often used for predicting flows of freight or passengers on an origin–destination basis. The simplest form of gravity model is

$$T_{ij} = k \frac{P_i^{\lambda} P_j^{\alpha}}{f(d_{ij})}$$

where

- T_{ij} is the number of trips between origin i and destination j
- P_i and P_j are measures of the importance of origin i and destination j respectively, generally population and incomes for passengers, and economic activity for freight
- d_{ij} is the resistance to transport between origin i and destination j, usually measured by distance, transport cost or time, and $f(d_{ij})$, the 'impedance function', which often takes the form of d^{β} or $e^{\beta d}$, and
- k, α and λ are constants.

5 $Y = a(1+g)^t$ can be rewritten as $Y = ae^{bt}$ where $b = In(1+g)$. It can then be made linear by taking logs of both sides: $In\ Y = In\ a+bt$.

Constants can be estimated from current data, and projections of transport flows can be made using projections of population and income or economic activity.

Judgment

Extrapolation cannot be the sole forecasting technique employed in the likelihood of changes that bear no relationship to the past, such as when demand is strongly affected by one-off events such as land use changes, industrial developments and other transport initiatives. Projections may need to be adjusted to allow for one-off events and this requires judgment.

In situations where benefits of the initiative are strongly affected by the decisions of one, or a small number, of organisations, these organisations should be consulted about their plans and expectations for the future. Government strategy documents are also useful sources of information about future developments.

Appraisals should be specific about where judgment is applied in making forecasts, including specifying any assumptions made about the size of future developments (e.g. new factories, mines, agricultural activities and housing developments), and how the implications for demand levels are derived.

BTRE passenger and freight models

For non-urban infrastructure, the BTRE has developed models for projecting passenger and freight demand data using origin–destination data and projections of population and economic activity:

- the BTRE/TRA non-urban passenger travel model, OZPASS, for projections of local and inter-regional passenger travel (BTRE 2007), and
- the FDF/BTRE FreightSim model, for future growth in the inter-regional freight task (Austroads 2003d).

Growth in passenger travel is assumed to be related to growth in population and household income and inversely related to the cost of travel. Growth in the freight task is related to the production and consumption of 16 specific bulk commodities and non-bulk freight, and the freight flows necessary to balance the production and consumption of each commodity in each region.

2.4.5 Diverted traffic forecasts

Diverted traffic is traffic that switches from another route or mode to take advantage of the new initiative.

The first step in estimating diverted traffic levels is to obtain an estimate of the maximum volume of traffic that could potentially divert. For example, if the initiative is a town bypass, all traffic passing through the town with origins and destinations on either side of the bypass is potentially divertible. There may also be some traffic travelling from one side of the town to the other with origins or destinations inside of the bypass that would take advantage of the bypass. The second step is to estimate the proportion of potentially divertible traffic that is likely to divert. It may improve accuracy if cars and trucks are treated separately because the reasons for choosing to use the bypass or the town road are likely to be different.

For an initiative on an inter-urban rail corridor, only long-distance road freight that is not highly time-sensitive is likely to divert. Road freight in regions surrounding the rail corridor therefore needs to be broken down by origin–destination and type of freight to discover how much is potentially divertible to rail. The proportion forecast to divert would be greater for market segments for which freight travels longer distances and is less time-sensitive.

The simplest way to estimate the proportion of traffic likely to divert is to apply judgment to nominate a percentage. This should preferably be based on past experience with similar initiatives.

If a small number of shippers are responsible for much of the divertible freight, they should be consulted as to their probable reactions to the proposed change.

A simple quantitative technique is to use a cross elasticity. A cross elasticity could be defined as the percentage change in use of one alternative in response to a one per cent change in the price of another, all other factors being held constant. For example, the cross elasticity would be –2.0 if, for a particular commodity–origin–destination category, a one per cent fall in the price for rail, say, as a result of an infrastructure initiative, results in two per cent of the road freight in the category diverting to rail. Where there are quality improvements such as time savings, the price could be expressed as a generalised cost (see Section 2.6.4 for an explanation).

Logit models

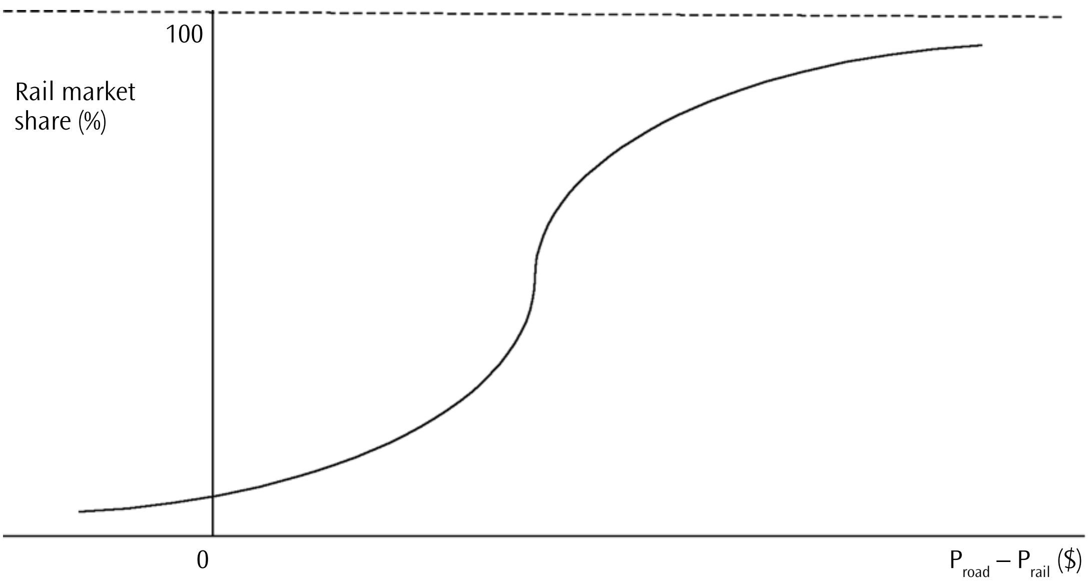
Logit models are a more sophisticated technique for predicting impacts of initiatives on mode shares. Where there are *n* modes of transport, the share (*S*) of total traffic using mode *j* is given by

$$Sj = \frac{e^{f(x_j)}}{\sum_{i=1}^{n} e^{f(x_i)}}$$

where x_i is a vector of attributes of mode *i* such as price, time taken and a measure of reliability. Values of time and reliability are incorporated into $f(x_j)$, usually in linear form.

Figure 2.6 shows the form of relationship implied by the logit model for prices. The horizontal axis shows the price charged by road minus the price charged by rail. The vertical axis represents the market share of rail. If both modes charge the same price $(p_{road} - p_{rail} = 0)$, the rail's market share will be quite small, because rail generally offers a slower, less reliable service than road. Rail's market share is higher the lower its price is relative to that of road.

Figure 2.6: Logit relationships for predicting rail market share as a function of prices



The mixed logit version of the model recognises the fact that transport users do not have a single value of time and reliability, but a distribution of values. 'Hierarchical logit' models allow for levels of choice; for example, a choice between road and rail, and, for road users, a choice between alternative routes. Further discussion of modelling impacts of initiatives on mode shares is provided in Volume 5, Part 3.

If there is insufficient data or resources to estimate coefficient values for the situation at hand, values are often assumed based on studies by others, including overseas studies.

2.4.6 Generated traffic forecasts

Generated traffic is new demand that is induced by the initiative.[6] For example, the initiative could lead to new industrial developments or land use changes that give rise to additional traffic using the infrastructure concerned. Information about the likely characteristics of these developments should underpin estimates of generated traffic levels in detailed BCAs.

Where the sources of generated traffic are more diverse, a price elasticity of demand could serve as the basis of estimation. Price elasticity is the percentage change in traffic caused by a one per cent change in price, all other factors staying constant. For example, a price elasticity of –0.5 implies that a one per cent drop in price results in a 0.5 per cent increase in traffic. See Part 3 for a more detailed discussion of elasticities. Generalised cost (explained in Section 2.6.4) might be used instead of price to account for changes in service quality.

2.4.7 Pricing assumptions

No additions to the corresponding section in Volume 3 are made here.

2.5 Infrastructure operating costs

2.5.1 Classification

No additions to the corresponding section in Volume 3 are made here.

2.5.2 Projections

No additions to the corresponding section in Volume 3 are made here.

2.6 User benefits

2.6.1 Estimating impacts on user costs

Figure 2.7 shows the process for estimating user benefits for rail. No equivalent diagram is drawn for road improvements because the only changes are the removal of the price reductions and financial analysis boxes.

Rail infrastructure improvements lead to a combination of reduced costs (savings in train operating costs, infrastructure operating costs or maintenance costs) and improved service quality; for example, reduced transit times, more convenient departure times for customers, increased reliability or reduced damage to freight. Part, or all, of any cost reduction may be passed on to customers as a price reduction. A quality improvement and price reduction should divert freight or passengers from road to rail and may generate some new demand; that is, freight or passengers that would otherwise not travel at all.

6 From an economist's point of view, all additional traffic using the project infrastructure in the Project Case compared with the Base Case is 'generated traffic' regardless of the source. Generated traffic may be divided into 'diverted traffic' and 'induced traffic' depending on its source. However, transport planners use the terms 'induced traffic' and 'generated traffic' interchangeably referring to traffic that is new altogether. Because the Guidelines are aimed at a broader audience than economists, the transport planners' terminology has been adopted.

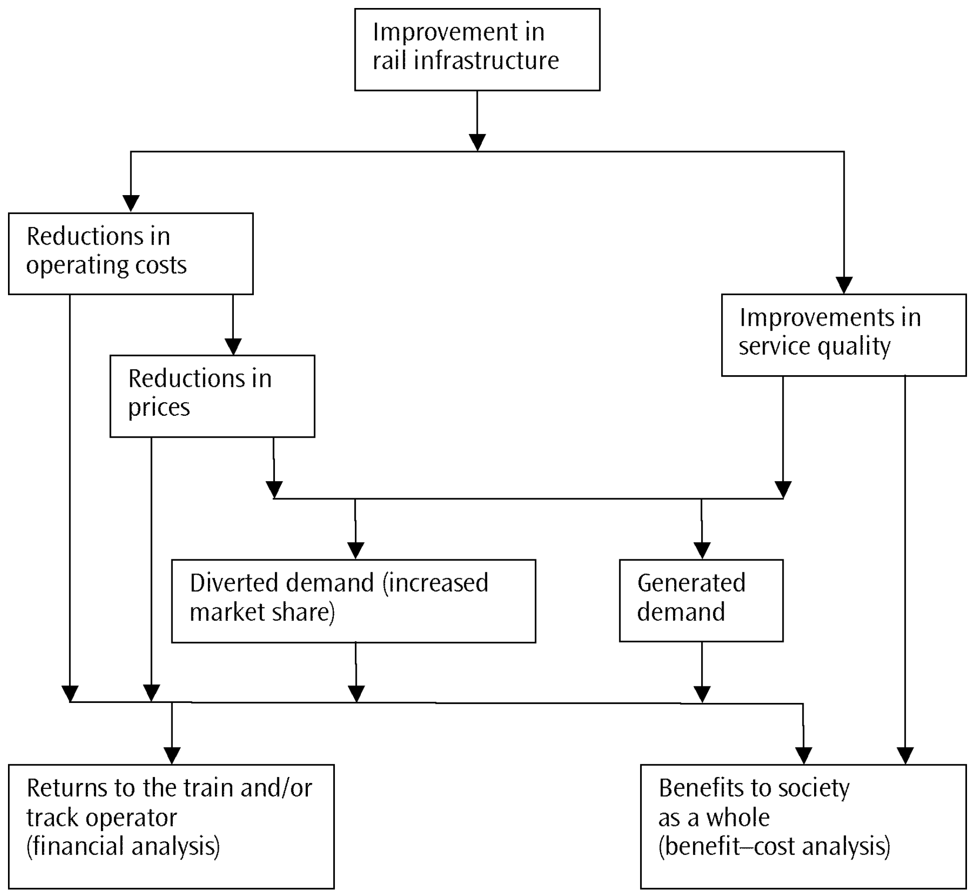
For the purposes of a benefit–cost analysis, the benefits from all this consist of:

- for existing passengers and freight
  - rail resource cost savings (which will be shared between the train operator, track manager and existing customers, depending on how much is passed on in the form of price reductions), and
  - the value of improvements in service quality (a benefit that accrues directly to customers)[7]
- for demand diverted from road to rail, the increase in WTP less the social generalised cost of meeting the diverted demand in the rail market, plus or minus any additional benefits or costs in the road market if prices differ from marginal social costs (see Section 2.7), and
- for generated demand, customers' WTP minus the social generalised cost of transport by rail (including externalities).

For the purposes of financial analysis, the train operator and track manager gain:

- for existing passengers and freight, cost savings minus revenue lost due to price reductions, and
- for generated and diverted demand, revenue minus costs of transporting the additional passengers or freight.

Figure 2.7: Benefits of improvement in rail infrastructure



Estimates of vehicle or train operating costs and times taken are required for the Base and Project Cases, for each year of the initiative's life for which benefits are estimated. For detailed BCAs, computer models should be used to estimate these costs. For rapid BCAs, simpler methods of arriving at estimates are acceptable.

7 Another scenario is for the rail operator to capture the benefit of a quality improvement through raising the price. For a BCA, this is just a transfer from the customer to the railway operator. However, it will affect the quantity of passengers or freight carried. The value of the quality improvement still has to be assessed and counted as a benefit.

Value of travel time savings (VTTS)

Work time for drivers and passengers is valued at the average hourly wage rate (including income and payroll taxes), on the assumption that the pre-tax wage rate measures the increased value of output from a saving in labour time. Estimates of non-work time are derived from statistical studies of data from situations where people trade-off time against money—the choice between a cheap, slow mode of transport and an expensive, fast mode. Typically, non-work time is valued at around 25 to 30 per cent of the wage rate.

The value of travel time savings for cars used in BCAs is a weighted average value for work and non-work time (proportions from the Australian Bureau of Statistics Survey of Motor Vehicle Use (SMVU)) multiplied by the average car occupancy rate (also from the SMVU). Austroads (2003b) provides values of time for freight derived from stated preference surveys, where consignors of freight were asked questions aimed at revealing their willingness to trade-off money against time.

Value of improvements in reliability

Research has shown that transport users not only value travel time savings but also improvements in the reliability of journey times. Where there is significant variability in journey times, transport users may have to allow more time for the journey in order to reduce the possibility of arriving late at their destination. If the travel time variability for a route is reduced, then transport users benefit by being able to reduce this extra time allowance (Transfund 2004).

Theoretical models of how transport users value reliability have tended to focus on the standard deviation measure (readily derivable from the coefficient of variation measure). Such models typically feature[8]:

- Travel time (T) from origin to destination as a random variable having a probability distribution with standard deviation σ. A number of different distributions are used, usually skewed towards a shorter journey time.
- A departure time (t_h) and a preferred arrival time (PAT). Travel time could be allowed to vary with departure time reflecting variations in congestion with time of day.
- A (dis)utility or generalised cost function containing the variables travel time, time, schedule delay early (SDE) (if one arrives before the PAT), schedule delay late (SDL) (if one arrives after the PAT) and the probability of arriving late (p_L). The function includes values for travel time (α), early time (β) and late time (γ). There may also be a constant amount (θ) added on when the transport user arrives late.

The optimal departure time is found that minimises the expected (dis)utility function

$$E[U(t_h)] = \alpha E[T(t_h)]+\beta E[SDE(t_h)]+\gamma E[SDL(t_h)]+\theta p_L(t_h)$$

Most authors make simplifying assumptions that travel time is independent of departure time (not a function of t_h) and that $\theta=0$; that is, there is no disutility from being late. In this case, if the departure time is optimised, the probability of arriving late is given by $p_L{}^*= \beta/(\beta+ \gamma)$[9], which implies that people choose a departure time to balance the consequences of early and late arrival.

Empirical work by a number of authors has shown that, with departure time optimised, for a wide range of probability distributions for T, the expression $\beta E[SDE(t_h)]+\gamma E[SDL(t_h)]$ is well approximated by a function, $H(\beta,\gamma)\sigma$, and that H can be considered constant for any given combination of β and γ. Hence, there is justification, in appraisals, for measuring reliability by standard deviation of arrival time and assuming a constant value of reliability per hour of standard deviation.

Generalised cost can be expressed as $\beta_t T+ \beta_\sigma \sigma$ where β_t is the value of time and β_σ is the value of reliability. The 'reliability ratio' is defined as $R = \beta_\sigma / \beta_t$. Making the aforementioned simplifying assumptions and assuming an exponential distribution for travel time, the reliability ratio is given by

$$R = \frac{\beta}{\alpha} \ln\left(1+ \frac{\gamma}{\beta}\right)$$

8 This section is based on Bates et al. (2001).

9 In this section, the superscript * means that the variable has been set to the optimal level.

On the basis of their literature survey and empirical work, Bates et al. (2001, p. 228) conclude that, 'Values (for the reliability ratio of) around 1.3 appear plausible for car travel: somewhat higher values may be appropriate for scheduled public transport, but values above 2 are unlikely'. In practice, this means that, if the value of time per vehicle assumed for a BCA is $10 per hour, then reliability improvements could be valued at $13 per vehicle per hour of reduction in the standard deviation of travel time.

Estimating the standard deviation of travel time for road initiatives

The Land Transport New Zealand Project Evaluation Manual, Transfund (2004), contains a detailed methodology for valuing reliability improvements for road initiatives.[10]

Trip reliability is defined as the unpredictable variations in journey times, which are experienced for a journey undertaken at broadly the same time every day. The impact is related to the day-to-day variations in traffic congestion, typically as a result of day-to-day variations in flow. This is not the same as the variations in individual journey times, which occur within a particular period. The procedure does not account for the delays that may result from major incidents on the road network.

Travel time reliability is, in principle, calculated for a complete journey, and the total network variability is the sum of the travel time variability for all journeys on the network. In practice, models may not represent the full length of journeys and this is accounted for in the procedure.

The level of variability in travel time is assumed to rise with congestion. The standard deviation of travel time, $SD(TT)_t$ (=σ), in minutes can be estimated from the function

$$SD(TT)=s_0+\frac{(s-s_0)}{1+e^{b\left(\frac{v}{c}-a\right)}}$$

where

s_0 = base uncongested level of SD(TT)

s = maximum level of SD(TT)

v/c = volume–capacity ratio estimated using passenger car equivalent units

a = v/c at the midpoint of the curve, and

b = a constant that describes the steepness of the curve.

The curve is a logistic-shaped curve that approaches s_0 at low volume–capacity ratios, equals $(s - s_0)/2$ at the midpoint where v/c = a, and approaches s as the volume–capacity ratio becomes large. Transfund's manual provides the calibration factors listed in Table 2.2.

Table 2.2: Coefficients to estimate standard deviation of travel time

CONTEXT	S	S_0	b	a
Motorway/expressway	0.90	0.083	–52	1.0
Urban arterial	0.89	0.117	–28	1.0
Urban retail	0.87	0.150	–16	1.0
Urban other (50 km/h)	1.17	0.050	–19	1.0
Rural highway (70–100 km/h) (2 lanes in direction of travel)	1.03	0.033	–22	1.0
Signalised intersection	1.25	0.120	–32	1.0
Unsignalised intersection	1.20	0.017	–22	1.0

Note: Evaluations of small retail areas on 50 km/h sections of rural highway should use the 'Urban other (50 km/h)' context.

Source: Transfund (2004).

10 This subsection reproduces material from Transfund (2004).

Benefits are estimated as the reliability ratio × value of travel time ($/h) × reduction in network variability (in minutes) / 60 × vehicle flow rate for the analysis period (vehicles/h) × correction factor (see Table 2.3).

Transfund recommends lower values of the reliability ratio than Bates et al. (2001). For a typical urban traffic mix, Transfund recommends 0.9. For initiatives with a significantly different vehicle mix, 0.8 is recommended for cars and 1.2 for commercial vehicles.

In many cases, an appraisal considers a defined area that does not represent the full length of most journeys. As a result, the changes in journey time reliability are overestimated. In these cases, the variability estimates need to be adjusted. A judgement must be made about the typical proportion of journey time variance likely to be incurred outside the study area, using the appropriate correction factor in Table 2.3. While the table gives some illustrative contexts of where different factors might apply, the decision about which correction factor to use should be based on an estimation of the variance of journey times, which occurs outside of the evaluation area (Transfund 2004).

Table 2.3: Correction factors for estimating reliability benefits

PERCENTAGE OF VARIANCE OUTSIDE STUDY AREA	CORRECTION FACTOR (PER CENT)	INDICATIVE TRANSPORT NETWORK MODEL COVERAGE
<20	100	Regional model
20	90	Sub-regional model
50	70	Area model
75	50	Corridor model
90	30	Intersection model, individual passing lane

The process for evaluating reliability benefits using the Transfund (2004) approach is:

1. Calculate standard deviation of travel time on each link between intersections and for each intersection movement or approach.
2. Square the standard deviations to produce variances.
3. Sum variances along each origin–destination path to obtain the total variance for journeys between each origin and destination.
4. Take the square root of the aggregated variance for a journey to give the standard deviation of the journey time.
5. Multiply the total trips for each origin–destination pair by the standard deviation of travel time and sum over the matrix to give the network-wide estimate of the variability cost.
6. Calculate the difference in variability cost between the initiative and do-minimum networks, and assess the benefits that can be claimed using the formula for claimable benefits above.

Indexation of the value of travel time savings and other unit costs

The VTTS is related to average earnings. Over time, real wages increase due to improvements in labour productivity. The VTTS for work time should therefore increase correspondingly.

Non-work time is valued as a proportion of the average wage rate. The proportion is estimated from statistical studies of how people trade-off travel time against money. Hence, it is a WTP value. The value of non-work time is not expected to increase as fast as real wages, because as incomes rise, people are likely to substitute additional leisure for some of the additional income. In other words, the income effect is, on average, stronger than the substitution effect (Mackie et al. 2001, p. 104). With people taking more leisure time, the marginal utility, and therefore value, is lower than otherwise.

Mackie et al. (2001) also suggest that a reasonable assumption is that the value of non-work time rises proportionately to per capita GDP growth, which is slower than real-wage growth. A lower bound is half that rate. Recent evidence suggests an elasticity of value of time with respect to income of 0.5 (Hensher and Goodwin 2003), but work performed for the UK Department for Transport suggests a figure of 0.8. The UK Guidelines recommend assuming that the value of non-work time increases with *per capita GDP* with an elasticity of 0.8, and the value of work time with an elasticity of 1.0 (UK Department for Transport 2006, see TAG Unit 3.5.6). Hensher and Goodwin (2003) conclude that the VTTS should grow at a rate less than income, but do not recommend a particular proportion.

If a growth rate for the VTTS is assumed over the life of an initiative, index all labour costs throughout the BCA; for example, the labour components of infrastructure operating costs and vehicle or train operating and maintenance costs. The value of statistical life used for estimating safety benefits and costs of externalities that impact on people can also be indexed to grow in line with forecast real income.[11]

In Australian BCAs, it is more usual *not* to increase the VTTS, the costs of labour, crashes and externalities in line with forecast growth in real income. Yet increasing these attributes is the correct approach. The difficulty is that if proponents choose their own growth rates for parameters, there could be a loss of comparability between appraisal results, and proponents may use over-optimistic growth forecasts to achieve more favourable BCA results.

The Guidelines recommend that parameters should not be increased over time unless the jurisdiction to which they are applying for funds specifies the growth rates to use for particular parameters. If Austroads specifies growth rates for road initiatives, jurisdictions should develop a consistent set of growth rates for rail initiatives.

Jurisdictions wishing to specify growth rates should base them on conservative long-term forecasts of growth in real wages or GDP per capita developed by reputable private- or public-sector bodies with specialist skills in macro-econometric modelling.

Growth in real wages should be used to index parameters that relate to labour—the value of work time, the labour component of infrastructure operating costs, crash costs estimated using the human capital approach and labour components of the health costs of externalities, including losses of earnings to the individuals affected and costs of medical personnel.

Growth in per capital GDP, multiplied by a factor below one, should be used for parameters that relate to peoples' WTP—the value of non-work time, crash costs estimated using the WTP approach, and externality values estimated using revealed or stated preference methods or hedonic pricing. The reason for preferring the per capita GDP growth rate over that for real wages is that a sizeable proportion of the relevant population whose WTP is being estimated are not workers e.g. retirees and students. The reason for multiplying the growth rate by a factor less than one is the argument discussed above in relation to non-work time—as income grows, people are assumed to consume more leisure, leading to decreased marginal utility of leisure and hence a lower value. This applies equally to safety and freedom from negative externalities. It is difficult be more precise about the factor by which to multiply growth in real wages. Forced to chose, a conservative factor is 0.5.

Although many expect the world's oil supply to run out causing the cost of fuel to rise, it is recommended that fuel costs are *not* indexed because there is so much uncertainty about the timing and size of future increases in prices for transport energy, and because any rises are likely to be accompanied by fluctuations.

11 In deriving the value of statistical life currently used in Australian BCAs via the human capital approach, BTE (2000, p. 27) assumed a two per cent annual growth rate in labour costs. However, this is not equivalent to indexing the value in a BCA. Indexation would require crash costs to be increased in each year of the project's life because the discounted present value of future earnings forgone extends into the future from the date of the crash, not from the start of the project's life. If a WTP approach is used for valuing statistical life, indexation is still required because with higher incomes, people are willing to pay more to reduce the risk of death or injury.

2.6.2 Valuation of user benefits

No additions to the corresponding section in Volume 3 are made here.

2.6.3 Valuation of user benefits

The valuation of benefits in a market can be approached in two ways:

1. the social welfare approach—net benefit = increase in WTP (defined in generalised cost terms) less increase in social generalised costs, and
2. the gainers and losers approach—net benefit = net gains to consumers + net gains to producers + net gains to governments + net gains to third parties.

The social welfare approach stems from the concept of social welfare that underlies BCA as total WTP minus social costs. Benefits are valued at the maximum amount that the beneficiaries are willing to pay for them. Social costs are, in effect, the WTP by members of society for the resources consumed to create the benefits.

The gainers and losers approach is based on the idea that BCA adds up all the gains and losses regardless of to whom they accrue. It implements the compensation test—a change is desirable if the gainers from the change could potentially compensate the losers out of their gains and still be better off.

Just as the two sides of an accounting balance sheet must have the same bottom line, both approaches should produce the same result. If they do not, the analyst has made an error. In practice, in BCAs, only one approach is ever employed, usually the simpler one under the circumstances. Both approaches are presented in the ensuing discussions to minimise confusion and to confirm the correctness of the conclusions reached.

In Volume 3, a simpler variant of the social welfare approach is presented that treats existing traffic and new (that is, generated and diverted) traffic separately. For existing traffic, the benefit is the saving in social generalised costs. For new traffic, the benefit is the gain in WTP minus the associated social generalised cost.

When applying the gainers and losers approach, it is important to ensure that the same gain or loss is counted only once. For example, if the direct consumer of a transport service is a producer, competition may force the producer to pass part or all of the gains on to customers. If the gain is counted as a gain to the producer, it should not be counted again as a gain to the firm's customers.

Since benefits and costs are often passed on to other parties, the division of benefits and costs for an initiative between consumers, producers, governments and third parties does not necessarily constitute a complete analysis of the distributional consequences of the initiative. For a brief discussion of the difficulties in estimating distributional consequences of transport initiatives, see Section 1.2.

Examples of net benefit estimation

A series of examples of estimation of net benefits in different situations is presented using the two approaches. The examples chosen represent situations most likely to be encountered when appraising transport initiatives. All involve reductions in the generalised cost of transport or increases in transport capacity. The notation employed on the diagrams, tables and equations is as follows:

D = demand curve

P = perceived price

Q = quantity per period of time

AC = average generalised cost = TC/Q

MC = marginal generalised cost = dTC/dQ

TC = total generalised cost = AC × Q = $\int_0^Q MCdQ$ + fixed costs

SC = social generalised cost

T = total tax revenue

t = tax per unit of output

X = total cost of negative externalities

x = externality cost per unit of output

WTP = willingness-to-pay

CS = consumers' surplus

PS = producers' surplus

subscript 1 = Base Case

subscript 2 = Project Case.

In all cases, the analysis is in terms of perceived prices and generalised costs.

Example A

Impact of initiative: cost reduction

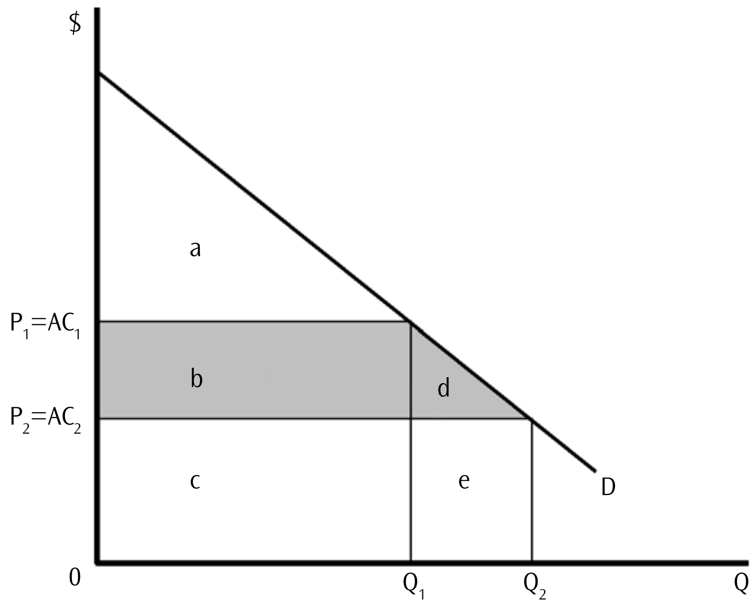
Cost curve: constant

Pricing: perceived price = average social generalised cost = marginal social generalised cost

Distortions: none

With constant costs, there are no effects on producers' surplus, government revenues or third parties to take into account.

Figure 2.8: Benefit from a cost reduction



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e	ΔWTP d+e	CS_2 a+b+d	ΔCS b+d
WTP_1 a+b+c		CS_1 a	
SC_2 c+e	ΔSC e–b		
SC_1 b+c			
Total benefit	ΔWTP – ΔSC b+d	Total benefit	ΔCS b+d

The welfare gain, areas *b+d*, is shaded. Area *b* is the saving in costs to existing users. There is no change in WTP for existing users. Area *d*, the consumers' surplus gain for new traffic, consists of the WTP for the new traffic (areas *d+e*) less the associated resource, area *e*. The new traffic may be diverted from other modes or routes (diverted traffic) or be traffic that did not exist before (generated traffic). To assume there is no new traffic is to assume that demand is perfectly inelastic.

For Example A, the gainers and losers approach is simpler, but, as the other examples show, this is an exception.

Demand curves are shown as linear in the diagrams throughout this part of the Guidelines for ease of drawing. In practice, the shape of the demand curve is unknown and will almost certainly not be linear. For initiatives that reduce user costs, it is not necessary to know the shape of the demand curve over the range $0Q_1$. Over the range of the new traffic, Q_1Q_2, the error introduced by assuming linearity is not likely to be large. So it is common to use the rule-of-a-half for valuing consumers' surplus benefits for generated and diverted demand, areas *b* + *d*.

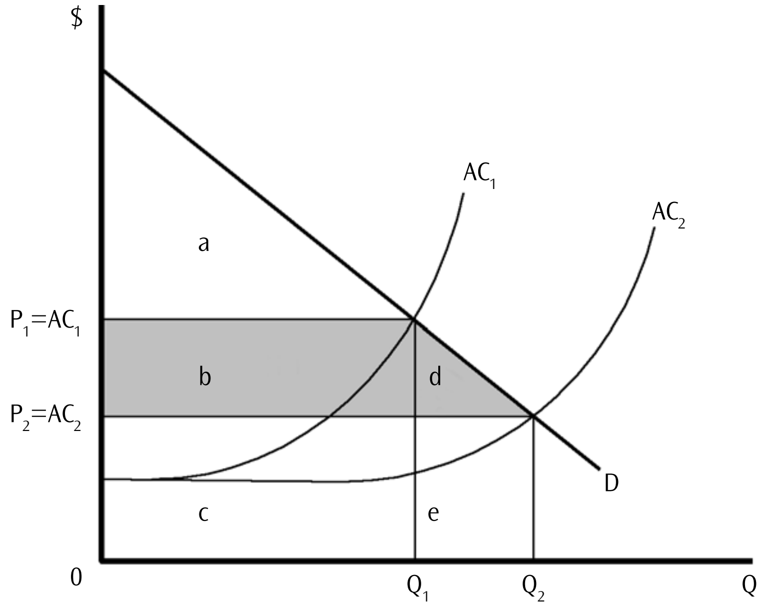
$$\text{Consumers' surplus benefit} = \frac{1}{2}(Q_2+Q_1)(P_1-P_2)$$

The rule-of-a-half can also be applied for estimating the change in WTP, areas *d+e* in Figure 2.8

$$\text{Change in WTP} = \frac{1}{2}(Q_2-Q_1)(P_1+P_2)$$

For traffic on congested roads, lack of congestion pricing means that users are paying prices equal to average private costs rather than marginal social costs. Short-run marginal and average costs rise as congestion slows traffic down. In this case, Figure 2.8 is still applicable because it is possible to measure total costs from an average cost curve by taking the rectangular area bounded by the price and quantity. In Figure 2.9, a capacity expansion of a road moves the average cost curve right. Assuming there are no externalities or taxes, the total social generalised cost is still *b+c* in the Base Case and *c+e* in the Project Case. Therefore, benefits can be measured the same way as for the constant cost case, areas *b+d*.

Figure 2.9: Benefit from a cost reduction: rising average cost curve



Example B

Impact of initiative: cost reduction

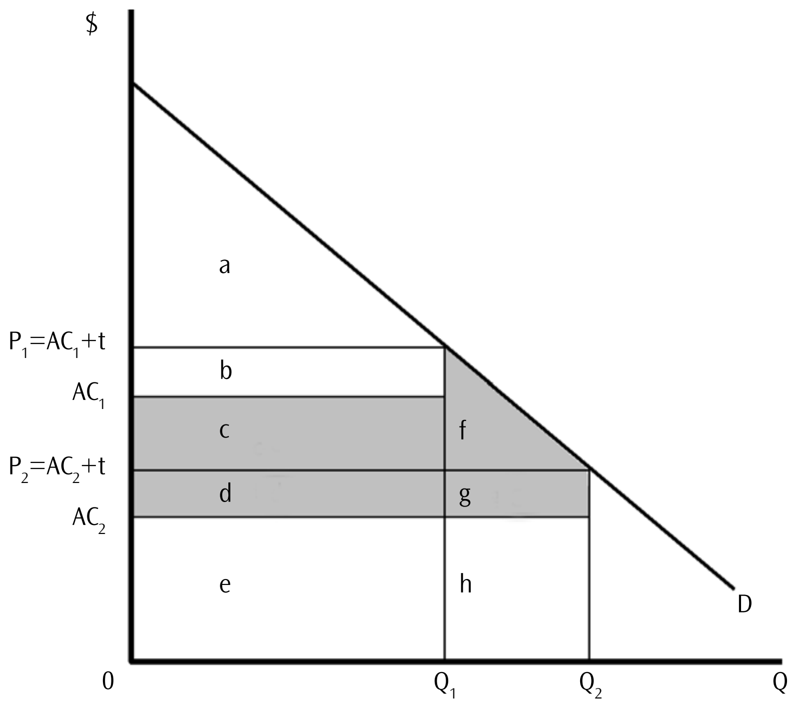
Cost curve: constant

Pricing: perceived price = average social generalised cost + tax = marginal social generalised cost + tax

Distortions: tax

In Example B, the fuel excise causes private costs to exceed resource costs.

Figure 2.10: Benefit from a cost reduction where there is a tax



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e+f+g+h	ΔWTP f+g+h	CS_2 a+b+c+f	ΔCS b+c+f
WTP_1 a+b+c+d+e		CS1 a	
SC_2 e+h	ΔSC h–c–d	T_2 d+g	ΔT g (since b=d)
SC_1 c+d+e		T_1 b	
Total benefit	ΔWTP – ΔSC c+d+f+g	Total benefit	ΔCS+ ΔT c+b+f+g =c+d+f+g

For existing traffic, there is no change in WTP, but the resource cost of carrying out the task falls by areas *c+d*. For the new traffic, there is a gain in WTP of areas *f+g+h*, but this is partly offset by the additional resource cost equal to area *h*.

The example can also be applied to a situation where a railway, charging a price above social costs, passes a cost reduction, in full, on to its customers. However, the excess of price over cost is profit, not tax. It could also apply to an improvement to a toll road where there is no congestion so that the price paid by users exceeds their resource costs.

The formula from Volume 3 for the benefit area is

$$Q_1(AC_1-AC_2)+\frac{1}{2}(Q_2-Q_1)(P_1+P_2)-(Q_2-Q_1)AC_2$$

which simplifies to

$$\frac{1}{2}(Q_2-Q_1)(P_1+P_2)-(Q_2AC_2-Q_1AC_1)$$

comprised of

increase in WTP	$\frac{1}{2}(Q_2-Q_1)(P_1+P_2)$	areas f+g+h
less increase in total social generalised costs	$-(Q_2AC_2-Q_1AC_1)$	– areas h–c–d
Total		areas c+d+f+g

Bray (2005), drawing on McIntosh and Quarmby (1972), writes the formula in the form

$$\frac{1}{2}(Q_1+Q_2)(P_1-P_2)+(Q_2P_2-Q_1P_1)-(Q_2AC_2-Q_1AC_1)$$

Bray calls the three components

A: user surplus (change in consumers' surplus)	$\frac{1}{2}(Q_1+Q_2)(P_1-P_2)$	areas b+c+f
B: increase in perceived user costs	$(Q_2P_2-Q_1P_1)$	areas g+h–b–c
C: less increase in resource costs	$-(Q_2AC_2-Q_1AC_1)$	– areas h–c–d
Total		areas c+d+f+g

As additions of the areas represented show, the formulas are equivalent. Note that these formulas apply when costs are rising as well as constant; that is, when there is congestion.

The Bray–McIntosh–Quarmby formula is intended for urban network models. The amount of *B* and *C* combined is called the 'resource correction', because it 'corrects' the raw estimate of benefit based on perceived prices for differences between perceived and resource costs. In Figure 2.10, the resource correction is comprised of areas *g*–*b*+*d*, which equals area *g*, provided area *b* equals area *d*.

Urban network models use perceived prices to predict traffic flows on a network. The benefit estimates they produce are therefore based on perceived prices. The resource correction is applied afterwards to obtain the true level of benefit, allowing for distortions such as taxes and failure of drivers to perceive the full costs of road use.

A network version of the formula is given in Example L.

Example C

Impact of initiative: cost reduction and increase in capacity

Cost curve: rising

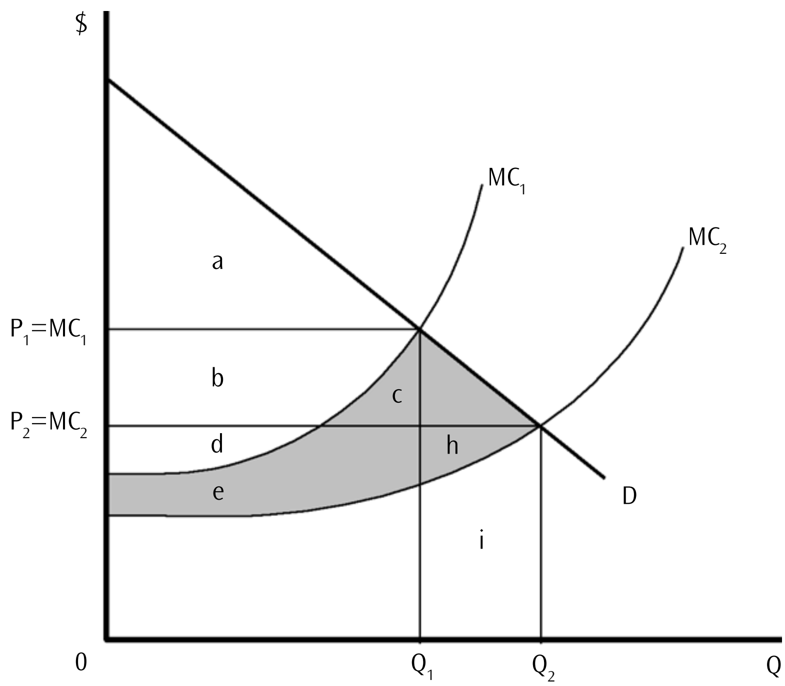
Pricing: perceived price = marginal social generalised cost

Distortions: none

Example C shows the textbook case of a firm participating in a competitive market, charging a price equal to marginal cost. Note that the actual prices charged by the firm and the prices paid by the users are different from the perceived prices and generalised costs shown in Figure 2.11.

This example shows that where price equals marginal cost, the benefit from a cost reduction or capacity increase can be measured as the area bounded by the Base Case and Project Case marginal cost curves and the demand curve. This follows from the fact that the area under the marginal cost curve equals total variable costs.

Figure 2.11: Benefit from a cost reduction: marginal cost pricing



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e+f+g+h+i	ΔWTP G+h+i	CS_2 a+b+c+g	ΔCS b+c+g
WTP_1 a+b+c+d+e+f		CS_1 a	
SC_2 f+i	ΔSC i–c–e	PS_2 d+e+h	ΔPS e+h–b
SC_1 c+e+f		PS_1 b+d	
Total benefit	ΔWTP – ΔSC C+e+g+h	Total benefit	ΔCS + ΔPS c+e+g+h

Example C could apply to the situation where a railway operator with congested infrastructure sets prices equal to marginal generalised cost to ration capacity. It is shown below that it also applies to a road with optimal congestion pricing in place.

Example D

Impact of initiative: cost reduction

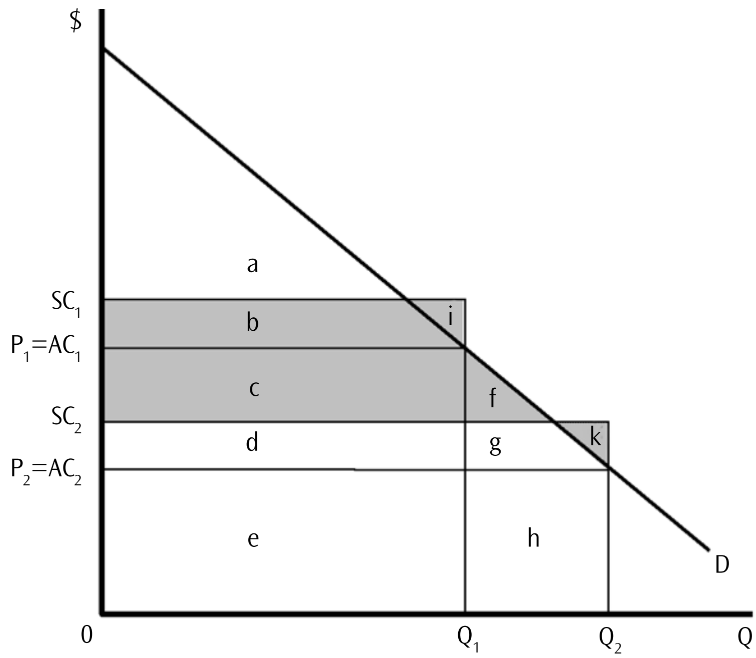
Cost curve: constant

Pricing: perceived price = average social generalised cost – externality = marginal social generalised cost – externality

Distortions: negative externality

The average resource costs are AC_1 and AC_2 for the Base and Project Cases respectively. The existence of an externality causes social costs in each case to lie above resource costs and prices paid.

Figure 2.12: Benefit from a cost reduction where there is a negative externality



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e+f+g+h	ΔWTP f+g+h	CS_2 a+b+c+d+f+g	ΔCS c+d+f+g
WTP_1 a+b+c+d+e		CS_1 a+b	
SC_2 d+e+g+h+k	ΔSC g+h+k–b–c–i	X_2[a] –d–g–k	ΔX –d–g–k+b+i
SC_1 b+c+d+e+i		X_1[a] –b–i	
Total benefit	ΔWTP – ΔSC b+c+f+i–k	Total benefit	ΔCS–ΔX b+c+f+i–k

a. Externalities are negative because they represent a cost imposed on the people bearing them.

Using the simplified social welfare approach in Volume 3, with existing and new traffic considered separately, for existing traffic, the social cost of carrying out the task falls by areas *b*+*c*+*i*. For the new traffic, there is a gain of WTP of areas *f*+*g*+*h*, but this is offset by the additional social cost equal to areas *g*+*h*+*k*. The net increase in economic welfare for the new traffic is then areas *f*–*k*.

In practice, one is more likely to estimate the change in consumers' surplus, (*c*+*d*+*f*+*g*), separately from the increase in externality costs, (*g*+*k*, assuming that *b*+*i*=*d*), using either default values as provided elsewhere in the Guidelines or by a case-specific method (see Section 2.9.2).

Example D can be applied to the situation where there is a subsidy. The subsidy is a cost to the government, just as an externality is a cost to third parties.

The example can also be applied to the case where the perceived price is below the actual price equal to social generalised cost because transport users are not aware of some of the costs they are incurring. Areas *b*+*i* and *d*+*g*+*k* then represent costs incurred by transport users.

Example E

Impact of initiative: cost reduction

Cost curve: rising

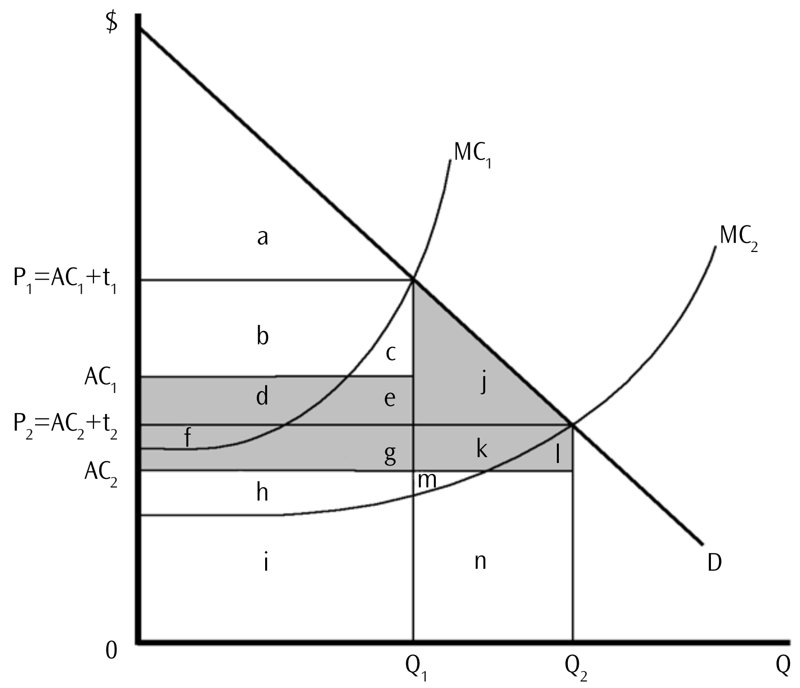
Pricing: price = marginal social generalised cost = average private / resource generalised cost + congestion tax

Distortions: none

Example E shows the benefit where economically optimal congestion pricing is in place. In the absence of the congestion tax, road users incur the average private generalised cost, which is assumed here to equal average resource generalised cost—there are no taxes on fuel and other inputs. On a congested road, the marginal road user imposes an externality on the other road users by slowing them down. Therefore, the marginal social generalised cost exceeds the average private generalised cost. A congestion tax brings the price paid by road users (average private generalised cost plus the tax) into equality with the marginal social generalised cost, improving the efficiency of resource allocation.

In Figure 2.13, the average cost curves, which lie below their respective marginal cost curves, are not shown, to avoid cluttering the diagram. However, the levels of average cost at Q_1 and Q_2 are shown.

Figure 2.13: Benefit from a cost reduction: optimal congestion pricing



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e+f+g+h+i	ΔWTP j+k+l+m+n	CS_2 a+b+c+d+e+j	ΔCS b+c+d+e+j
WTP_1 a+b+c+d+e+f+g+h+i +j+k+l+m+n		CS_1 a	
SC_2 i+n+l =h+i+m+n	ΔSC n+l–c–e–g–h =m+n–d–e–f–g	T_2 f+g+k+l	ΔT f+g+k+l–b–c
SC_1 c+e+g+h+i =d+e+f+g+h+i		T_1 b+c	
Total benefit	ΔWTP – ΔSC c+e+g+h+j+k+m =d+e+f+g+j+k+l	Total benefit	ΔCS + ΔT d+e+f+g+j+k+l

The social welfare approach shows that the welfare gain here can be measured in two ways. The first, benefit areas, *d+e+f+g+j+k+l*, is the same shape as that derived in Example B above, and is measured using the demand and average social generalised cost curves. The second, *c+e+g+h+j+k+m*, is the same shape as that derived in Example C, and is measured using the demand and marginal social generalised cost curves.

Example F

Initiative: taxation of a negative externality

Cost curve: constant

Base Case pricing: perceived price = average private generalised cost = marginal private generalised cost

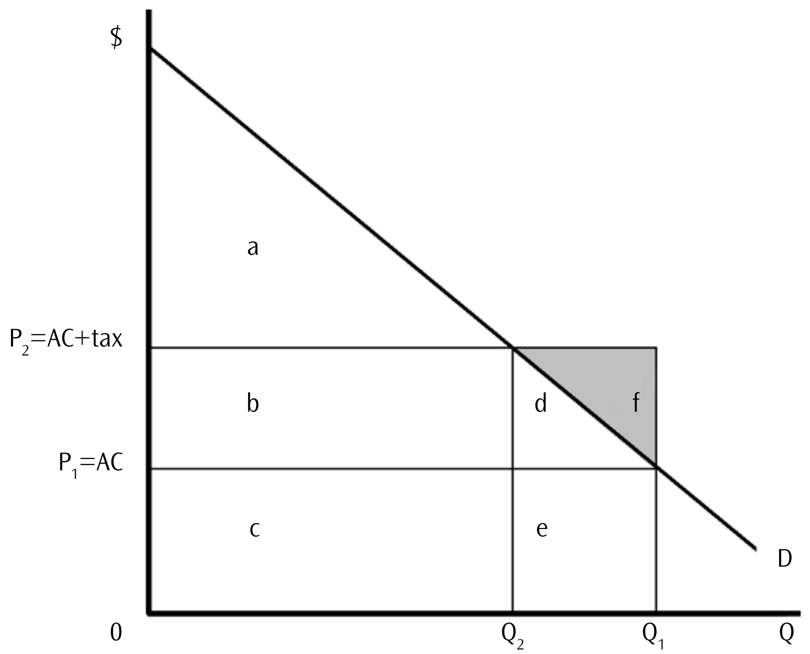
Project Case pricing: price = average (= marginal) private generalised cost + tax = average (= marginal) social generalised cost

Base Case distortions: unpriced externality

Project Case distortions: none

Examples F and G show how to estimate benefits for initiatives that involve charging for externalities. A tax or charge on a negative externality can improve resource allocation by bringing the price into equality with the marginal social cost. Example E shows the simple case of constant costs and Example F shows the case of rising costs applying to the introduction of congestion pricing.

Figure 2.14: Benefit from optimal taxation of a negative externality



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c	ΔWTP –d–e	CS_2 a	ΔCS –b–d
WTP_1 a+b+c+d+e		CS_1 a+b+d	
SC_2 b+c	ΔSC –d–e–f	T_2 b	ΔT b
SC_1 b+c+d+e+f		T_1 0	
		X_2 –b	ΔX d+f
		X_1 –b–d–f	
Total benefit	ΔWTP – ΔSC f	Total benefit	ΔCS + ΔX + ΔT f

From a social welfare point of view, for the units of output between Q_1 and Q_2, the social generalised cost exceeds consumers' WTP by an amount equal to area *f*. From a gainers and losers point of view, consumers lose areas *b* and *d* in consumers' surplus. Of this, area b is transferred to the government. Area *d* is offset by a gain to those suffering from the externality. The gain to sufferers of the externality of area f is the net gain to society.

Example G

Initiative: introduction of optimal congestion charging

Cost curve: rising

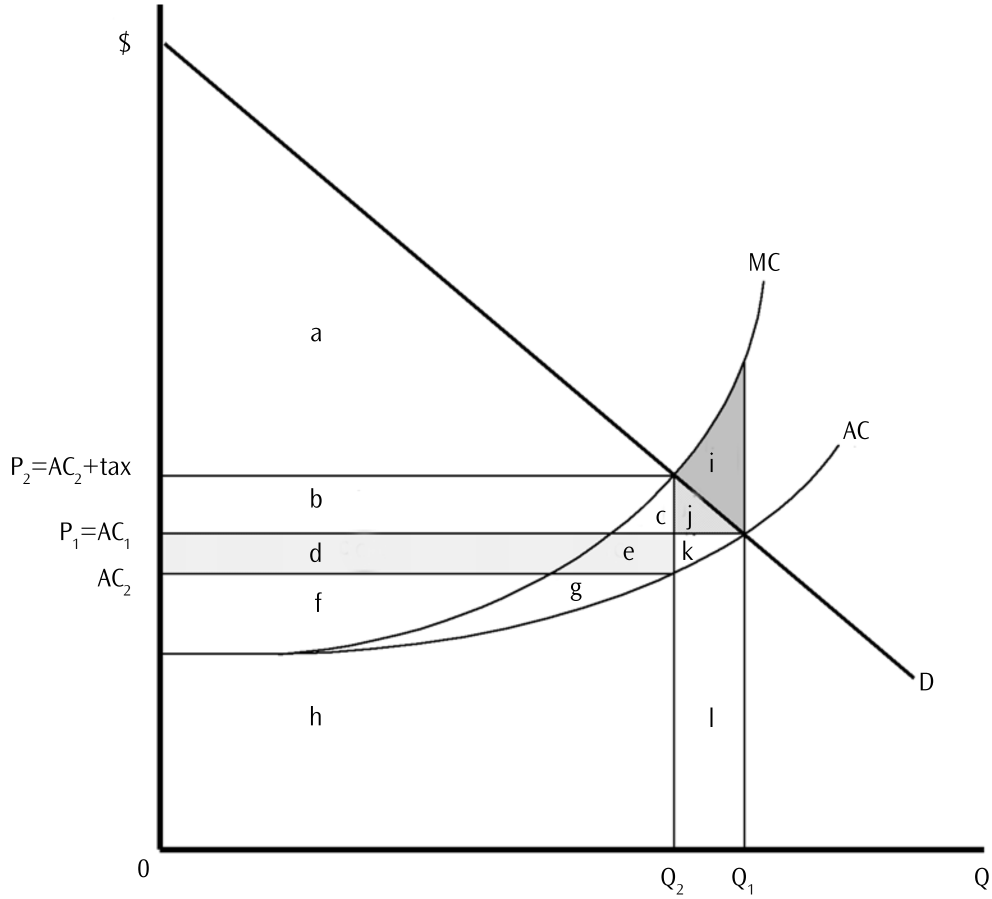
Base Case pricing: perceived price = average private generalised cost = marginal private generalised cost

Project Case pricing: price = average private generalised cost + congestion tax = marginal social generalised cost

Base Case distortions: unpriced externality

Project Case distortions: none

Figure 2.15: Benefit from optimal congestion pricing



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
WTP_2 a+b+c+d+e+f+g+h	ΔWTP –j–k–l	CS_2 a	ΔCS –b–c–j
WTP_1 a+b+c+d+e+f+g+h+j+k+l		CS_1 a+b+c+j	
SC_2 f+g+h =h+g+e+c	ΔSC d+e+k+l =i+j+k+l	T_2 b+c+d+e	ΔT b+c+d+e
SC_1 f+g+h+d+e+k+l =h+g+e+c+i+j+k+l		T_1 0	
Total benefit	ΔWTP – ΔSC d+e–j =i	Total benefit	ΔCS + ΔT d+e–j

Under the social welfare approach, consumers lose areas *j*+*k*+*l* in WTP. As shown in previous examples, social costs can be measured from either the average cost curve or marginal cost curve. Using the average cost curve, the reduction in congestion and traffic causes the total social cost rectangle (*AC*×*Q*) to shrink by areas *d*+*e*+*k*+*l*, which is a benefit to society. As areas *k*+*l* are offset by the loss in WTP, the net gain is the saving in resource costs for traffic that remains on the road, areas *d*+*e*, less the triangle of consumers' surplus for the lost traffic, area *j*. This same result is obtained under the gainers and losers approach. Consumers lose areas *b*+*c*+*j* in consumers' surplus. The government collects in tax the rectangle comprised of areas *b*+*c*+*d*+*e*. Areas *b*+*c* are a transfer, leaving areas *d*+*e* less the triangle of lost consumers' surplus, area *j*, as the net gain to society.

It is simpler to measure total social costs as the area under the marginal cost curve. The saving in total social costs is the area under the marginal cost curve between Q_2 and Q_1. The net gain to society is area *i*, the cost saving, areas *i+j+k+l*, less the loss in WTP, areas *j+k+l*.

Measurement of benefits in any real-world situation is more complicated. Firstly, there are taxes on inputs such as the fuel excise, so road users are already paying above the private resource costs of vehicle operation. Secondly, exact equality between price and marginal social cost is rarely achieved. This is particularly so for congestion pricing because marginal costs vary with time of day and location, though electronic charging systems make it technically possible to achieve a very close approximation to optimal charging.

For the purpose of estimating benefits of pricing initiatives in practice, the formula

$\frac{1}{2}(Q_2-Q_1)(P_1+P_2)-(Q_2AC_2-Q_1AC_1)$, given in example B, can be used.

Note that in this situation, the increases in WTP and total social costs will be negative.

2.6.4 Threshold cases

The term 'threshold case' is used here to refer to situations where the future of a major industrial plant, mine or other commercial venture depends on an initiative proceeding. It could be a new venture being established or an existing venture saved from closing down.

It can be difficult in threshold cases to determine whether the operator of the venture is bluffing. They have every incentive to encourage government spending on initiatives that benefit them. An expert assessment of an operator's market position and financial calculations can help, but is not foolproof.

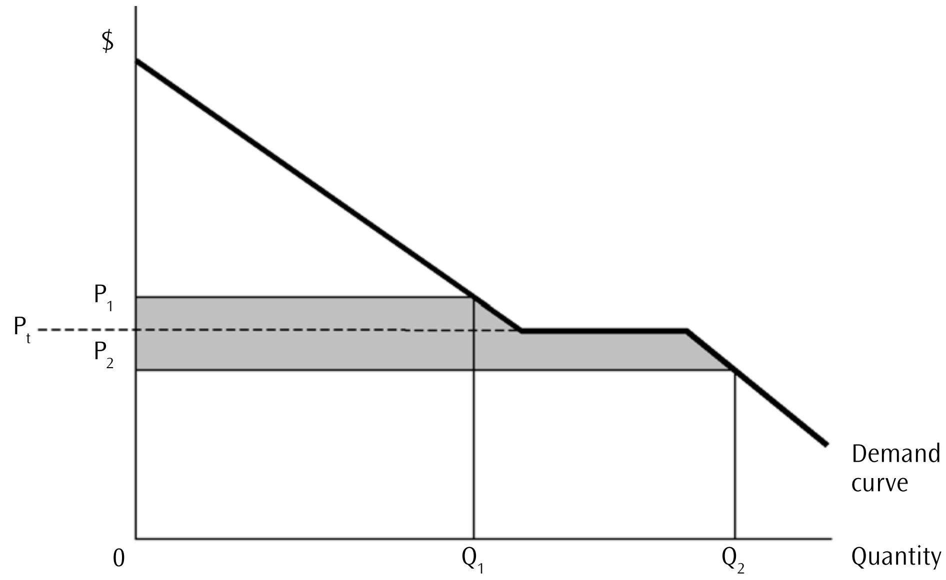
If the operator is the sole user of the infrastructure, there is unlikely to be any economic efficiency justification for public provision of the initiative.[12] The entire consumers' surplus benefit from the initiative accrues to the operator. If the present value of consumers' surplus exceeds the investment cost of the initiative, the operator should have the financial ability, and the incentive, to provide the initiative without any government contribution.

Where the initiative benefits other transport users in addition to the venture operator, the benefit can be measured in the usual way. However, if a threshold situation exists, the demand curve will have a flat section at some threshold level of perceived price (P_t) as shown in Figure 2.16. At levels of perceived price above the threshold, the operation does not proceed or shuts down. The length of the flat section equals the transport usage attributable to the operation.

The consumers' surplus benefit will be the shaded area just as for a more conventionally shaped demand curve. For estimating the consumers' surplus benefit, the level of the threshold private generalised cost for the operator is critical. It must lie between P_1 and P_2 for there to be a threshold issue in the first place. Paradoxically, the operator will want to make the threshold appear as low as possible to bolster the argument that a threshold situation exists, but this reduces the consumers' surplus benefit attributable to the initiative. An expert assessment of the operator's claims, as recommended above, should shed light on the value of P_t.

12 An economic efficiency justification could arise if the social cost for the operation is significantly greater than the private cost, for example, due to a positive externality generated or a shadow wage rate below the actual wage rate.

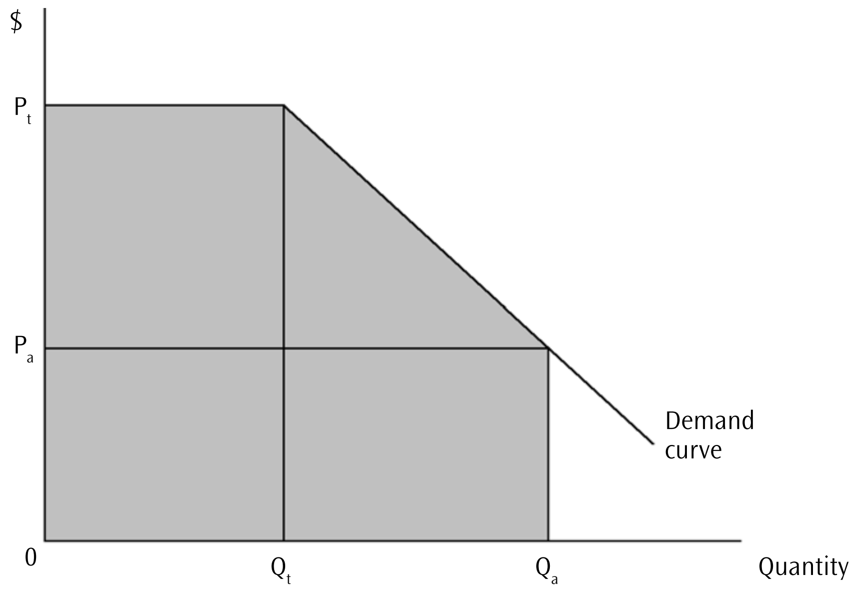
Figure 2.16: Benefit from a cost saving where there is a threshold case for a large user



Where the infrastructure constitutes a new service, and the operator shares the infrastructure with other users, the benefit is the total WTP area for the operator and for the other users together less social costs. Estimating this is discussed in Section 2.15 for transport users in general. Figure 2.17 shows the demand curve for transport for a single, large commercial user with a threshold minimum level of output (Q_t). One approach to estimating total WTP is to examine the operator's market position and costs and consult with the operator to ascertain the highest level of private generalised cost for transport at which the operation is viable, that is, the threshold level (P_t in Figure 2.17), together with the operator's level of transport usage at that cost (Q_t). It is also necessary to estimate the operator's level of transport use (Q_a) at the private generalised cost engendered by the initiative (P_a). These represent two points on the demand curve. The WTP area (shaded in Figure 2.17) could then be estimated making an assumption about the shape of the demand curve (linear in Figure 2.5).

If P_t and P_a are known and one of Q_t or Q_a are known, then the other quantity can be estimated using the elasticity of demand for freight transport derived as $f\eta_d$, where f is the proportion that transport costs bear to the selling price of final output, and η_d is the elasticity of demand for the product being shipped.[13]

Figure 2.17: Estimating total WTP benefit for a large single transport user in a threshold case



13 Note that this rule is based on assumptions that transport is not substitutable for other inputs and that the marginal cost of production is constant. See Equation 5.5 in Section 3.5 in Volume 5, Part 3 from Bennathan and Walters (1969) for the more general formula where costs are rising.

2.7 Cross-modal and network effects

2.7.1 Importance of cross-modal and network effects

Diverted demand for infrastructure improved, or built, as a result of some initiative may come from an alternative route or a competing mode of transport. For non-urban road initiatives, the greater part of traffic is local and inter-regional, so freight diverted from rail is unlikely to be an issue worthy of special consideration. For rail initiatives, freight diverted from road transport can be significant, but only for major initiatives or groups of initiatives that significantly improve rail's competitiveness with road transport. For urban roads, where there are alternative routes through a network, diverted traffic is usually critical. For example, some urban network models assign traffic to networks using Wardrop's principle. In broad terms, the principle implies that whenever there are alternative routes between an origin–destination pair, the individual drivers will always choose the quickest route (or the cheapest route in terms of generalised costs) so that, in equilibrium, travel time (or generalised cost) is equalised for all alternative routes.

Upgrade of part of a highway in a corridor could lead to increased demand for use of other parts of the same highway. Upgrade of a railway line could lead to increased traffic on roads leading to rail terminals. Flow-on effects in transport infrastructure can give rise to complementarities in demands for infrastructure usage.

Because the Guidelines have a multi-modal focus, considerably more attention is devoted to diverted traffic and flow-on effects than is usual for guidelines of this type.

2.7.2 Parallel infrastructure

The existence of parallel infrastructure creates a case of related markets—one service is a substitute for the other. The cross elasticity of the demand (see Section 2.4.5 and Part 3 in this volume for definitions) is a measure of the degree of substitutability, and can be used to estimate the amount of diverted traffic. A cost reduction, or service quality improvement, resulting from an infrastructure initiative causes a left shift of the demand curve for the substitute service.

Measurement of part of the benefits from diverted traffic has already been covered in Section 2.6.2 on valuing benefits in the market that the initiative directly affects. The new traffic induced by the fall in generalised costs comprises diverted as well as generated traffic. If the diverted transport users paid the full social cost of the service from which they have been diverted, there are no further benefits or costs to consider. This is demonstrated in Example H below. Unpriced, or under-priced, externalities, subsidies and the lack of congestion pricing on urban roads can lead to perceived prices for substitute services being below social generalised costs. Tolls on roads and taxes on inputs can lead to substitute services being priced above social costs.

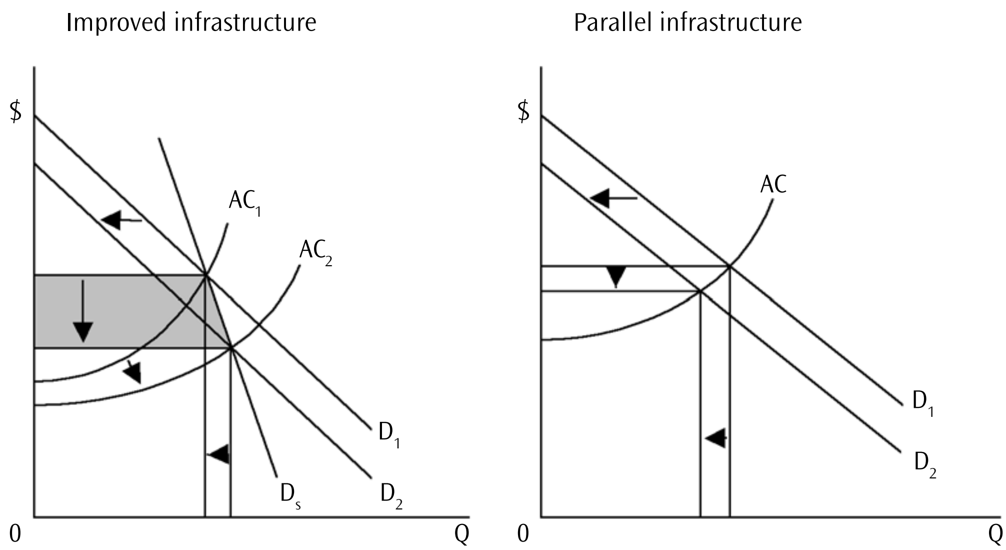
In Figure 2.18, the new initiative has caused a down and right shift of the cost curve for the infrastructure improved by the initiative (left diagram). This leads to a left shift of the demand curve for parallel infrastructure (right diagram) as traffic diverts. If the parallel mode or route is congested (that is, the average cost curve is upward sloping in the region where it crosses the demand curve), generalised costs could fall, lessening some of the traffic diversion. The less-congested parallel infrastructure becomes more desirable causing a left shift of the demand curve for the improved infrastructure. If there is congestion on the improved infrastructure, its generalised costs could fall causing a further left shift in the demand curve for the parallel infrastructure. The cycle would continue until the system converges to equilibrium.

For upstream or downstream infrastructure, there could be feedback effects from increased congestion elsewhere in the network. The effect is a reduction in the number of trips on the infrastructure on which the initiative occurs, because the increased congestion elsewhere would raise overall trip costs.

In a computer model of an urban transport network, there is a highly complex set of demand interrelationships between different links in the network. The model could be run through a series of iterations to converge to equilibrium.

As far as estimation of benefits is concerned, as long as the quantity/cost changes are not large enough to preclude an assumption of linearity of demand curves over the relevant range, only the initial and final positions should be taken into account. In the left diagram of Figure 2.18, a pseudo-demand curve (Ds) is drawn between the initial and final equilibrium points.[14] Benefits can be estimated from this pseudo-demand curve in the normal way (the consumers' surplus change shown by the shaded area) (Button 1993, pp. 182–4).

Figure 2.18: Network effects from a new initiative



The following set of examples show how to estimate benefits and costs from network effects. In all cases, the diagrams refer to the parallel infrastructure, not the infrastructure improved by the initiative. The demand curves shift left as traffic diverts to the improved infrastructure.

In all cases, prices are perceived prices and costs are generalised costs.

14 It is called a pseudo-demand curve because a true demand curve is drawn under a *ceteris paribus* assumption—prices of substitute goods or services, among other things, are assumed to be held constant. The pseudo-demand curve in Figure 2.18 could be called a 'general equilibrium adjustment schedule'.

Example H

Cost curve: rising

Pricing: perceived price = marginal social generalised cost

Distortions: none

Example H demonstrates the general rule that, if price equals marginal social cost in a related market, there are no additional benefits or costs to consider in that market. The gains and losses cancel out.

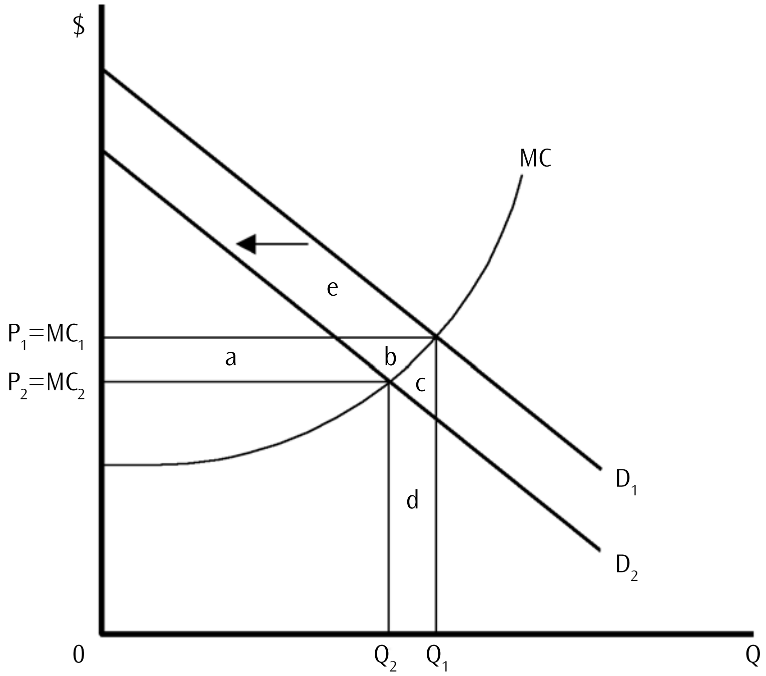
In Figure 2.19, for a very small left shift in the demand curve that changes quantity by $dQ < 0$, there is a loss of WTP given by the height of the demand curve multiplied by the change in quantity. Hence $dWTP = PdQ < 0$. As the demand curve continues to shift left, the equilibrium price follows the marginal cost curve down. Adding together the series of small changes in WTP (that is, integrating over quantity), the loss in WTP is the areas $c+d$.

Note that the area between the two demand curves, *e*, is not a loss of WTP. There is a drop in users' valuations because of an improvement in the alternative service due to the initiative, but users do not suffer any disbenefit. The marginal user, diverting to the improved infrastructure, does not experience a net loss in WTP either. Their WTP transfers to the market for the improved infrastructure.

As a result of a small change in infrastructure usage of *dQ*, there is a change in costs to society of *MCdQ*, which is a cost saving since *dQ* is negative. The sum of cost savings for the series of *dQ*'s from Q_1 to Q_2 (the definite integral) is the areas $c+d$. Hence, loss of WTP is exactly equal to the saving in social costs.

Using the gainers and losers approach, for a small change in price, $dP < 0$, there is an change in consumers' surplus of $-QdP > 0$. Adding up the gains from the series of small price reductions from P_1 to P_2, (that is, integrating over price), the total increase in consumers' surplus is the areas $a+b$. For each small change in price, there is a change in producers' surplus of *QdP*, which is negative since $dP < 0$. The losses of producers' surplus add up to the areas $a+b$ for the price reductions from P_1 to P_2. Hence, the consumers' and producers' surplus changes are exactly offsetting.

Figure 2.19: Network effect benefit: price = marginal social cost



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
dWTP PdQ	ΔWTP –c–d	dCS –QdP	ΔCS a+b
dSC MCdQ	ΔSC –c–d	dPS QdP	ΔPS –a–b
dWTP–dSC 0 (since P=MC)	ΔWTP–ΔSC 0	dCS+dPS 0	ΔCS+ΔPS 0

It can be shown that, with optimal congestion pricing, network effects do not create any net benefits or costs. From the point of view of the social welfare approach, with optimal congestion pricing, the congestion tax is always adjusted to keep $P = AC + t = MC$. With price always equal to marginal cost, changes in WTP and the social cost will be exactly offset. From the point of view of the gainers and losers approach, with optimal congestion pricing, there are two parties to consider: road users and the government as tax collector. For road users, $dCS = -QdP$. For the government

$$T = PQ - ACxQ$$

$$dt = QdP + PdQ - QdAC - ACdQ$$

$$dT = QdP + \left(P - Q\frac{dAC}{dQ} - AC\right)dQ = QdP + (P - MC)dQ \quad \text{since} \quad \frac{dAC}{dQ} = \frac{MC - AC}{Q}.$$

With optimal congestion charging, $P = MC$, so $dT = QdP$. The gain to road users from each small reduction in price, as the demand curve shifts left, is exactly offset by a loss in tax revenue to the government. The gains and losses cancel out.

Example I

Cost curve: constant

Pricing: perceived price = average social generalised cost – externality = marginal social generalised cost – externality

Distortions: negative externality

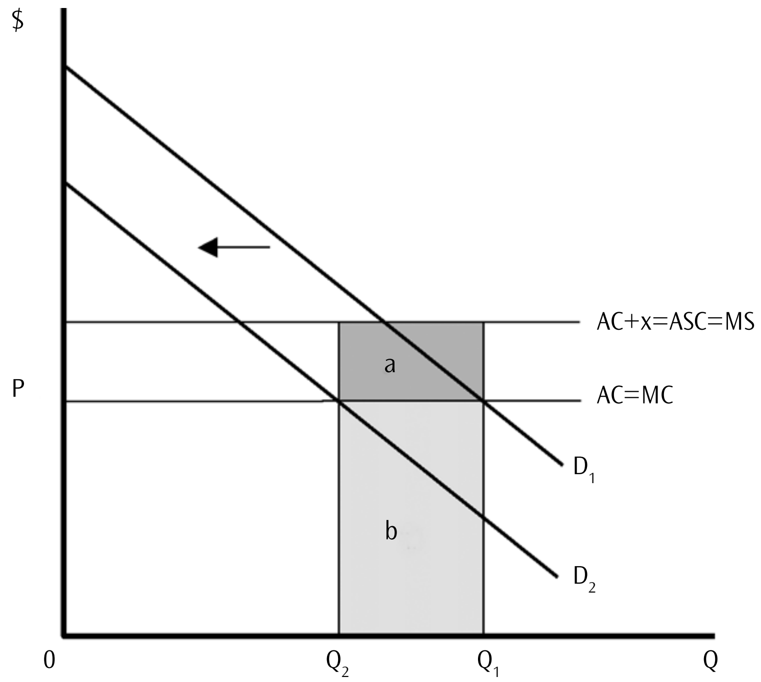
In Figure 2.20, the perceived price paid by transport users equals private costs, but there is a negative externality that costs society $*x* for each unit of transport use. From the point of view of the social welfare approach, a small change in quantity, $dQ < 0$, results in a change in WTP, $PdQ < 0$. But, there is also a change in costs to society of $MSCdQ < 0$. The net change in social welfare is $(P-MSC)dQ$, which is positive because $P-MSC < 0$. The net benefit of the quantity reduction from Q_1 to Q_2 is area *a*.

Following the gainers and losers approach, since there are no changes in prices, there are no changes to consumers' or producers' surpluses to consider. There is simply a gain to those suffering from the externality of area *a*.

Example I can be applied where there is a subsidy, causing the price to lie below marginal social cost. Under the gainers and losers approach, the gainer from a reduction in quantity is the government as payer of the subsidy.

A further application is where the perceived price is below the generalised cost due to a failure of users to perceive the full costs they incur. In this case, the gain of area *a* accrues to transport users diverting to the improved infrastructure.

Figure 2.20: Network effect benefit with an unpriced externality



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
dWTP PdQ	ΔWTP –b	dX xdQ	ΔX a
dSC MSCdQ	ΔSC –a–b		
dWTP–dSC (P–MSC)dQ	ΔWTP–ΔSC a		

This example is relevant for BCAs of improvements to non-urban rail infrastructure that lead to a shift of freight from road to rail. A potential error is to count the entire reduction in road externalities—crashes, wear and tear on roads, noise in urban areas, greenhouse gas emissions—as benefits. This is correct only if the assumption of Figure 2.20 holds—that the perceived price for transport of road freight equals the social generalised cost minus costs of externalities. It may be that part, or all, of the costs of these externalities are paid when all the charges imposed on the trucking industry are added in, namely fuel excise, registration and insurance. The benefit for each unit of freight that transfers from road to rail is the amount by which the full perceived price falls short of the social generalised cost, not the externality cost by itself.

Example J

Cost curve: constant

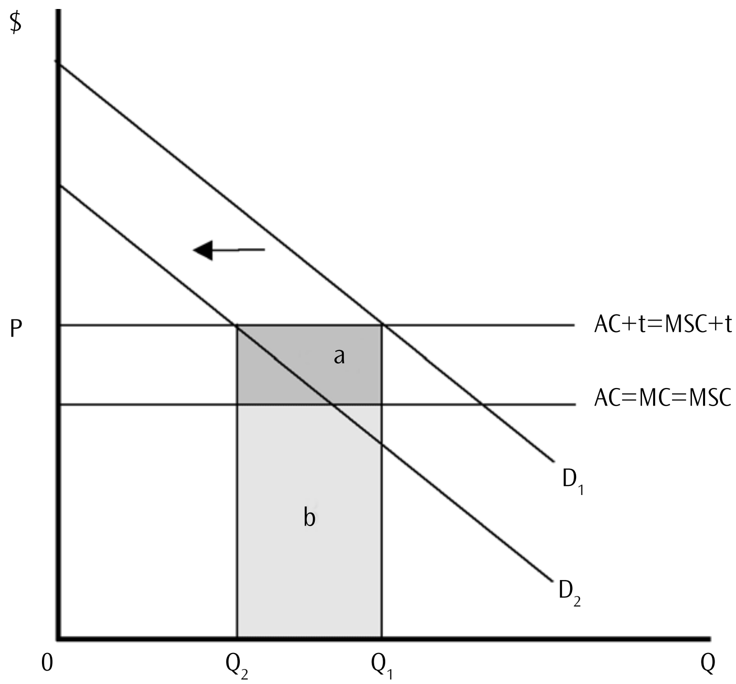
Pricing: perceived price = average private generalised cost + tax = marginal social generalised cost + tax

Distortions: tax

In Figure 2.21, the private generalised cost incurred by transport users exceeds the resource generalised cost by a tax. As in the previous examples, from the point of view of the social welfare approach, a small change in quantity, $dQ < 0$, results in a change in WTP, PdQ, together with a change in social cost, $MSCdQ$. In this case, the WTP loss for each reduction in quantity exceeds the saving in social cost, so there is a net loss to society, $(P–MSC)dQ < 0$. The loss for the reduction from Q_1 to Q_2 is the area *a*.

Following the gainers and losers approach, the change in government tax revenue is $tdQ < 0$ for each reduction in quantity with an overall loss of area *a*. No other parties gain or lose.

Figure 2.21: Network effect benefit with a tax



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
dWTP PdQ	ΔWTP –a–b	dT tdQ	ΔT –a
dSC MSCdQ	ΔSC –b		
dWTP–dSC (P–MSC)dQ	ΔWTP–ΔSC –a		

Example K

Cost curve: rising

Pricing: perceived price = average private generalised cost = average social generalised cost

Distortions: perceived price < marginal social generalised cost

Example K is important for evaluations of urban infrastructure initiatives because it applies to congested roads where there is no congestion charge. In Example H, it is shown that with optimal congestion pricing, network effects do not create additional net benefits or costs. This is not the case where congestion charging is lacking. Diverting traffic away from a congested road that is underpriced creates a benefit, and vice versa.

In Figure 2.22, four areas are identified. Using the social welfare approach, with each small change in quantity, the change in WTP is given by *PdQ*, which, in this case, equals *ACdQ*. However, resource costs incurred by society change by *MCdQ*. The net change in welfare is *(AC–MC)dQ*, the gap between the marginal and average cost curves. With $AC < MC$, there is a welfare gain for each quantity reduction. Adding the welfare gains for the changes in quantities from Q_2 to Q_1, the net benefit is areas *b+c*. This is the approach proposed by Harberger (1972, pp. 262–3). Provided the change in quantity is not too large, the area can be approximated by treating the cost curves as linear over the range between Q_1 and Q_2

$$\text{Benefit } \frac{1}{2}(Q_1-Q_2)[(MC_1-AC_1)+(MC_2-AC_2)] = \frac{1}{2}(Q_1-Q_2)[(MC_1-P_1)+(MC_2-P_2)]$$

Using the gainers and losers approach, for each small reduction in price, road users gain *QdP*, which is areas *a+c* between P_1 and P_2. This is the approach proposed by Neuberger (1971). It can be approximated by the formula

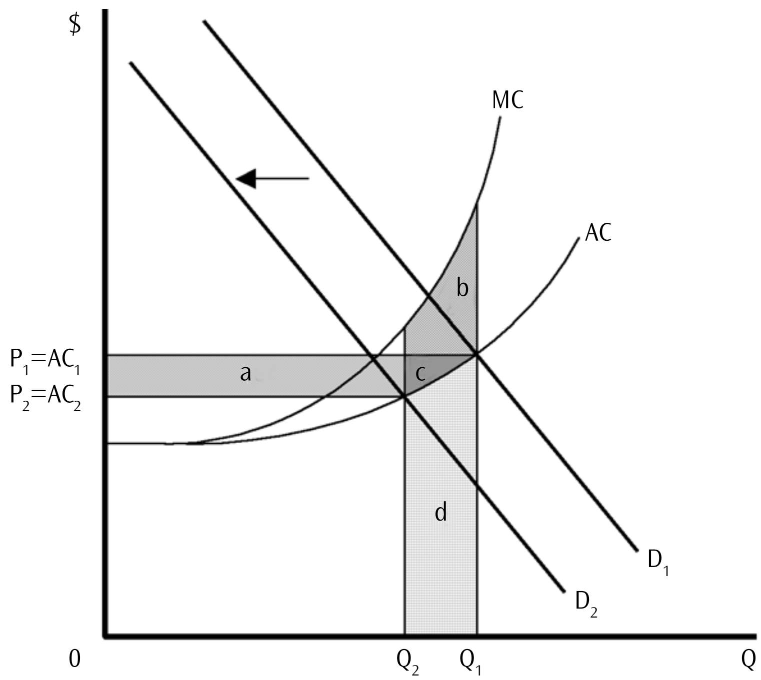
$$\text{Benefit } \frac{1}{2}(Q_2+Q_1)(AC_1-AC_2)=\frac{1}{2}(Q_2+Q_1)(P_1-P_2).$$

The Harberger and Neuberger areas are equal in size. Because the area under the marginal cost curve is total variable costs, areas *b+c+d* in Figure 2.22 must equal $TC_1–TC_2$. Hence the Harberger area is

$$b+c=TC_1-TC_2-d=Q_1AC_1-Q_2AC_2-d$$

which is the Neuberger area, comprised of areas *a+c*, as Figure 2.22 shows. With linear marginal and average cost curves, the two benefit formulas can also be shown to produce identical results.

Figure 2.22: Network effect without congestion pricing



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
dWTP PdQ	ΔWTP –d	dCS –QdP =–QdAC	ΔCS a+c
dSC MCdQ	ΔSC –b–c–d		
dWTP–dSC (P–MC)dQ =(AC–MC)dQ =–QdAC*	ΔWTP–ΔSC b+c =a+c		

* $\frac{dAC}{dQ} = \frac{MC - AC}{Q}$

The formulas also apply where the network effect is complementary, and the demand curve in the related market shifts right. With $Q_2 > Q_1$ and $AC_2 > AC_1$, the benefit formulas produce negative results consistent with the areas being costs instead of benefits.

When perceived prices equal average social generalised costs, the Neuberger formula for network effects is identical to the formula for estimating the direct benefit (the consumers' surplus increase) of the improved infrastructure (see Example A). Taking advantage of this, the following formula can be applied for estimating total benefits for an entire urban network

Benefit $\frac{1}{2}\Sigma_{ijkh}(Q^1_{ijkh}+Q^2_{ijkh})(AC^1_{ijkh}-AC^2_{ijkh})$

where the AC's are average social generalised costs and the subscripts refer to transport from origin i to destination *j*, using mode *k* and route *h* (See Neuberger 1971, p. 56). This formula correctly accounts for complementary network effects because $AC^1_{ijkh} < AC^2_{ijkh}$ making the welfare change a cost instead of a benefit.

Example L

Cost curve: rising

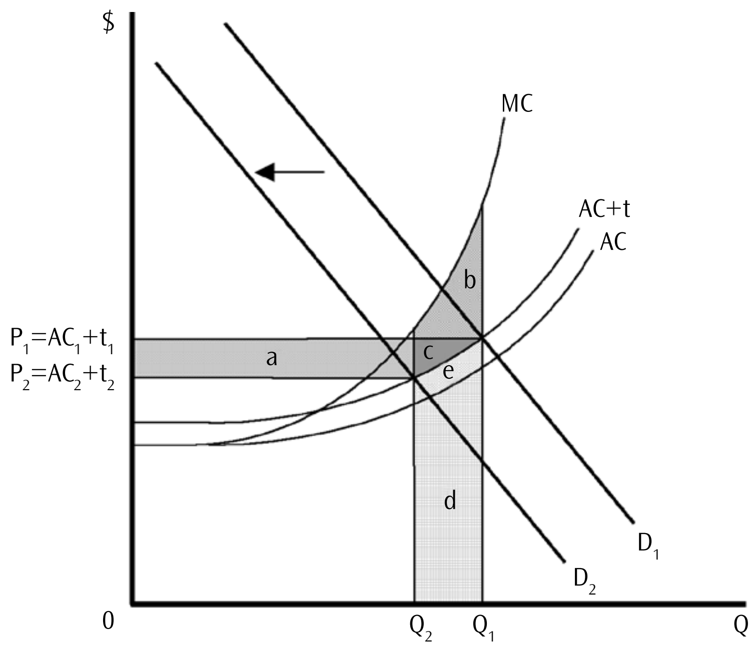
Pricing: perceived price = average private generalised cost = average social generalised cost + tax

Distortions: perceived price < marginal social generalised cost

Example L extends Example K by showing how to allow for the effect of a tax such as fuel excise, which is not set to replicate optimal congestion pricing. Using the social welfare approach, the benefit is the area between the marginal cost and average cost plus tax curves, areas *b+c*.

Using the relationship derived in Example H, $dT = QdP + (P-MC)dQ$, the gainers and losers approach yields an identical result.

Figure 2.23: Network effect without congestion pricing and with a tax



SOCIAL WELFARE APPROACH		GAINERS AND LOSERS APPROACH	
dWTP PdQ	ΔWTP –e–d	dCS –QdP	ΔCS a+c
dSC MCdQ	ΔSC –b–c–e–d	dT tdQ =QdP+(P–MC)dQ*	ΔT –e = –(a+c)+(b+c)= –a+b
dWTP–dSC (P–MC)dQ =(AC+T–MC)dQ	ΔWTP–ΔSC b+c	dCS+dT –QdP+tdQ =(P–MC)dQ	ΔCS+ΔT a+c–e =b+c

* See derivation in Example H above.

The formula for approximating the Harberger benefit area is

$$\text{Benefit} = \frac{1}{2}(Q_1-Q_2)[(MC_1-AC_1-t_1)+(MC_2-AC_2-t_2)] = \frac{1}{2}(Q_1-Q_2)[(MC_1-P_1)+(MC_2-P_2)]$$

In Neuberger form, the benefit area is given by the Bray–McIntosh–Quarmby formula provided in the previous section

$$\text{Benefit} = \frac{1}{2}(Q_1+Q_2)(P_1-P_2)+(Q_2P_2-Q_1P_1)-(Q_2AC_2-Q_1AC_1)$$

comprised of

A: user surplus (change in consumers' surplus)	$\frac{1}{2}(Q_1+Q_2)(P_1-P_2)$	areas a+c
B: increase in perceived user costs	$(Q_2P_2-Q_1P_1)$	areas –a–c–e–d
C: less increase in resource costs	$-(Q_2AC_2-Q_1AC_1)$	– areas –a–c–d
Total		areas a+c–e

For Part C of the formula to equal negative areas $a+c+d$, it is necessary to assume that the fall in perceived price is the same absolute size as the fall in average social generalised costs; that is, $P_1 - P_2 = AC_1 - AC_2$. The 'resource correction', Parts B and C of the formula, is negative, the area e.

For an entire network (see Example K for explanation)

$$\text{Benefit} = \frac{1}{2}\sum_{ijkh}[(Q^1_{ijkh}+Q^2_{ijkh})(P^1_{ijkh}-P^2_{ijkh})]+$$
$$\sum_{ijkh}(Q^2_{ijkh}P^2_{ijkh}-Q^1_{ijkh}P^1_{ijkh})-\sum_{ijkh}(Q^2_{ijkh}AC^2_{ijkh}-Q^1_{ijkh}AC^1_{ijkh})$$

Integrability condition

For a BCA that includes benefits or costs due to diverted demand estimated in this way, ensure that the assumed demand system satisfies the 'integrability' condition; that is

$$\frac{\partial Q_1}{\partial P_2} = \frac{\partial Q_2}{\partial P_1}$$

for all pairs of related markets *1* and *2*. If this condition is not met, and the initiative draws traffic away from a number of related markets, for example, different roads in a road network, the value of the total benefit could be affected by the order in which the welfare changes are evaluated. This is the result of the line-integral nature of multiple-good consumers' surplus. The integrability condition is the condition for path independence of line integrals.

The integrability condition is satisfied if the system of demand curves is derived from a utility function *and* either the income elasticities are unitary, or income effects are small enough to be considered negligible. The term 'integrability' refers to the fact that systems of demand curves meeting this condition can be sourced to a utility function, a process involving integration.

Demand systems based on logit models meet the condition. Wardrop's principle, and the four-step models commonly used for urban road networks, do not. In practice, the integrability condition is ignored when estimating benefits using urban network models because of the difficulty in building a model that meets the condition and the belief that the errors arising from not meeting the condition are unlikely to be large.

See Pressman (1970), Harvey (1996) and Johansson (1987) for further information on integrability. For further discussion of estimation of benefits and costs in related markets see Boardman et al. (1996, pp. 82–97).

2.7.3 Upstream and downstream infrastructure

The initiative being assessed can cause a right shift in the demand curve for usage on infrastructure that is downstream or upstream from the location of the new initiative. Measurement of benefits or costs is done in the same manner as for diverted traffic, except areas of benefit become costs and vice versa.

In a congested urban network, the Harberger and Neuberger formulas apply without change. The fact that Q_2 is greater than Q_1, instead of less, achieves the required reversal of sign for the estimated benefit—a right shift of the demand curve creates a cost when marginal social cost exceeds perceived price.

For complementary, the integrability condition is not a concern. For a given origin–destination (OD) pair, the total trip cost is the sum of costs for all links between the origin and destination. Say, a reduction in private generalised cost of $0.02 generated an additional trip between the OD pair. If the cost reduction occurred on Link 1, traffic on every other link between the OD pair, including Link 2, rises by one unit; hence, $\partial Q_2/\partial P_1=1/0.02=50$. A $0.02 cost reduction on Link 2 similarly generates an additional trip between the OD pair in question, which would increase traffic on Link 1 by one unit; hence, $\partial Q_1/\partial P_2=1/0.02=50$. Fulfilment of the integrability condition is not dependent on the specification of the demand curves representing behavioural assumptions.

Rather, it is fulfilled because of the simple physical relationships between OD flows and traffic on individual links.

2.8 Safety benefits

2.8.1 Crash rates

No additions to the corresponding section in Volume 3 are made here.

2.8.2 Unit costs

It is generally considered there are two alternative approaches to valuing fatality and injury costs from crashes: the human capital approach and the WTP approach.

The human capital approach involves estimating the discounted present value of all costs arising from a crash that can be directly measured, including loss of future earnings. For some individuals, in particular the very old and the very young, the present value of future earnings is small or zero. However, these effects are averaged out because the generic unit crash costs employed in a BCA are for an average individual.

The WTP approach involves estimating monetary amounts that people are willing to forgo to reduce the risk of death or injury (or willing to accept to tolerate an increased risk). Statistical analysis is employed to make estimates based on choices people make in situations where they trade-off risk of death or injury against money. The choices may be actual (revealed preference) or hypothetical (stated preference). The WTP approach tends to produce higher costs of crashes than the human capital approach.

The WTP method is more theoretically correct because it is consistent with the value judgment of consumers' sovereignty that underlies welfare economics. However, it produces a very wide range of answers, depending on the particular choice situation studied and the statistical techniques applied.

To ensure consistency in appraisal within a jurisdiction, it is important that transport and roads agencies adopt a single set of unit costs for crashes. The unit costs published by Austroads (2005c and 2006) set a nationwide standard for road crashes.

In Australia, the human capital approach is used. The most recent set of unit crash costs published by Austroads is based on BTE (2000). The BTE (2000) costs are higher than previous estimates based on a pure human capital approach because they include amounts for loss of quality of life based on monetary compensation awarded by courts. The current Australian approach can be described as a human capital approach with an element of WTP grafted on. The average cost of a fatality, according to BTE (2000), is $1.5 million in 1999 dollars. (Note that Austroads unit crash costs have been adjusted for inflation.)

By way of comparison, WTP values used in European countries range from $1.8 million to $4.2 million (1998 Australian dollar equivalents), while in the United States, a value of US$2.9 million (in 1994 prices) is used for all transport accident fatalities. New Zealand, as at June 1998, was using the value NZ$2.25 million (BTE 2000, pp. 21–3).

If Australia switched to the WTP approach, higher unit crash costs would alter the pattern of infrastructure expenditure to give higher priority to safety.

For further discussion, see BTCE (1996) and BTE (2000).

2.8.3 Estimate benefits

No additions to the corresponding section in Volume 3 are made here.

2.9 Externality benefits and costs

2.9.1 Default values

For rapid BCAs and detailed BCAs where externality costs are not critical, 'default externality parameters' or 'default values' can be used.

Default values are standard unit costs that can be applied across the board to obtain an estimate of externality costs. Rather than valuing environmental impacts at zero in a BCA, these values can be consistently applied where little information is available.

If, after undertaking an initial assessment using the default values, it is found that some externalities will have a significant impact within an appraisal, a detailed BCA should include an assessment that produces externality costs specific to the initiative being appraised, for example, using methods such as those in Section 2.9.2.

Where valuation of environmental externalities is not possible through default values or research specific to the initiative, a qualitative assessment should be conducted.

Values are provided in the Guidelines for:

- air pollution
- greenhouse gas emissions
- noise pollution
- water run-off from roads
- nature and landscape, and
- urban separation.

The default values are expressed in common units of 'dollar values per vehicle-kilometre travelled' (cents/vkm) for passenger vehicles and buses, and cents per net tonne kilometre (cents/ntk)[15] for road and rail freight transport. Values are disaggregated by passenger vehicle[16], freight type (light trucks, medium trucks, heavy trucks)[17] and buses, and some values are further divided by location (urban and rural) or chemical component (e.g. for air pollution).

A factor that makes externalities difficult to value is that externality impacts can vary with location and time. For example, the same quantity of a given externality will have a greater cost in a more densely populated area where there are more people affected. The affects of pollution can be dependent on weather conditions, which vary with time. Also, the total cost of an externality may not be proportional to the intensity of the externality (non-linear dose-response relationship). For example, if the amount of noise or pollution is doubled, the total cost is likely to be more than double. Therefore, some externalities are not valued in the Guidelines due to insufficient data. In other cases, it is inappropriate to assign overall default values.

Although, the default values provided in the Guidelines represent broad average values that can be applied to initiatives in all Australian jurisdictions, environmental valuation involves significant

15 Net tonne kilometre values are calculated using ABS (2000).

16 **Passenger cars** relate to all passenger vehicles that carry less than 10 passengers (including the driver). These consist of cars, station wagons, taxis, mini-buses, four-wheel drive passenger vehicles and forward control passenger vehicles. Passenger car values in the Guidelines are primarily petrol fuelled. Source: Cosgrove (2003).

17 **Light commercial vehicles** are trucks designed to carry goods and do not exceed 3.5 tonnes gross vehicle mass. These include utilities, panel vans, cab chassis and forward control load carrying vehicles. Values in the Guidelines represent LCV fleet (mixed petrol, diesel). **Medium duty trucks** are goods vehicles (including rigid trucks, articulated trucks and special purpose vehicles) with gross vehicle mass exceeding 3.5 tonnes but not exceeding 12.0 tonnes. Values in the Guidelines represent rigid truck fleet (mainly diesel). **Heavy duty trucks** are goods vehicles (rigid trucks, articulated trucks and special purpose vehicles) exceeding 12.0 tonnes gross vehicle mass (AGO 2004). Values in the Guidelines are representative of the articulated truck fleet (mainly diesel). Source: Cosgrove (2003).

uncertainty, and there are limitations in transferring these externalities from other countries to the Australian context. The values in Volume 3, Part 2, Appendix C should be treated with caution (particularly the values for buses, where further analysis is required).

Values in the Guidelines were obtained from Australian and overseas studies, and were selected after undertaking detailed research into factors such as the source and date of the data, the methodology used for measurement and the method of conversion from international literature.

Application of default values

As each appraisal differs in content and complexity, it is recommended that the default values are applied with data specific to the initiative being appraised.

Application of the default values requires the user to estimate a number of factors (such as kilometres and type of vehicles) prior to valuation. The valuation of the benefits may be broken into the following stages:

- Estimate the physical quantity of the externalities for the Base and Project Cases (e.g. consider urban and rural locations), using estimation techniques ranging from applying default parameters to more sophisticated transport models.
- Estimate the monetary impacts by applying valuation parameters to the quantity changes.
- Multiply quantity by dollars to estimate the stream of benefits and the key appraisal indicators (BCR, NPV and internal rate of return).

Proponents submitting proposals for consideration under the Guidelines appraisal methodology may, in some cases, want to use a different parameter value for their particular initiative. Proponents may submit a BCA estimate using their own preferred values together with a second BCA using the Guidelines' values as a sensitivity test. Reasons for departing from the Guidelines' parameters and the justification for the preferred value should be detailed.

The following section provides details on the default values for environmental externalities (adjusted to 2005 dollars), assumptions behind each externality and basic steps for applying the values in an appraisal. To ensure the default values are applied under appropriate circumstances, the assumptions behind each value should be noted. Consult the references in this section for further information.

Air pollution

Recent studies identify a strong link between air pollution and increases in adverse health effects imposed on society as a result of transport, particularly in urban areas where population is highest.

Pollutants relevant to Australia, and provided in Volume 3, Appendix C, include:

- carbon monoxide (CO)
- oxides of nitrogen (NO_x)
- particulate matter (PM_{10}), and
- total hydrocarbons (THC).

Air pollution values are refined estimates identified in Pratt (2002). For an alternative to the air pollution parameters, other sources such as Austroads (2003c) and Austroads (2006) may be useful.

Assumptions:

- The total air pollution values in Volume 3, Appendix C were selected because they maximise use of Australian data in their calculations. They are higher than other sources because the values include larger health costs for each chemical type; for example, health costs associated with particulate matter are higher than other sources.
- The values in the Guidelines are calculated using $/tonne health costs from Watkiss (2002), emission factors from Cosgrove (2003), and ABS Survey of Motor Vehicle Use (SMVU) data. Bus values apply emission factors from Commonwealth of Australia (2003) for air pollution, and

Cosgrove (2003) for greenhouse gas emissions; however, these have not been adjusted for occupancy and caution should be applied when using these values.

- Cost per tonne values ($/tonne) and grams per vehicle-kilometre (g/km) exclude full life cycle analysis. The g/km incorporates fuel consumption.
- Watkiss (2002) reports $/tonne estimates through a bottom-up approach using results from the European Commission's ExternE project (1995–2005), and include health impacts, damage to buildings and damage to crops. ExternE excludes damage to ecosystems. The impacts excluded have a lower cost relative to human health effects. Values recommended by Watkiss (2002) are constrained to health impacts only.
- The ExternE methodology comprises four main steps:

  1. assessment of emissions
  2. assessment of air pollution concentrations
  3. assessment of impacts using dose-response functions, and
  4. assessment of economic value of impacts.

- The effects of pollutants attributed to transport on health are quantified using WTP and dose-response relationships, based on epidemiological studies that link pollution concentrations, or increments, to levels of health effects. The $/tonne costs include mortality and morbidity health impacts such as asthma.
- Cost per tonne values have been transferred to Australian values according to population. Population densities in Australia are compared against the site types reported in ExternE, based on a regression analysis.
- Due to lower, more dispersed populations (urban areas are spread over larger areas) than in Europe, secondary air pollution values are considerably lower in Australia (Watkiss 2002). Additionally, because population is concentrated along narrow areas (the coasts) and is very sparse elsewhere, the effects of any regional scale pollution are much lower than in Europe.
- For the purpose of the Guidelines, a population-weighted average of the $/tonne costs for Australian cities is calculated to obtain national $/tonne costs for urban locations, using Watkiss (2002).
- Air pollution values are estimated as being close to zero in rural locations. In order to assess these non-urban emissions, more consideration of the locations of roads, vehicle flows and population is required. Although vehicles travelling on inter-urban routes between cities have some air pollution costs, these are much lower than for urban areas. Modelling research may be required for further analysis.
- Rural values in the Guidelines are approximately zero or are one per cent of the urban value. Rural bus and light truck values are assumed to be almost zero, as most vehicle-kilometres are in urban areas (Austroads 2006).
- The calculation methodology adopted in BTE (1999a), for heavy trucks only, is applied to rail.

Steps for application

Quantifiable impacts:

- Establish the environment (specific air pollutants) and vehicle compositions.
- Users may apply $/tonne values to models, where it is necessary to fully develop an appraisal.

Valuation:

- Apply cents/vkm or cents/ntkm values.
- Multiply vehicle-kilometres obtained by cents/vkm or cents/ntkm for the Base and Project Cases.
- Incorporate totals into BCA.

Greenhouse gas emissions

Valuation of greenhouse gas emissions is speculative at this point in time. Under a national program to achieve a target reduction in greenhouse gases, it is unclear how much of the burden is attributed to transport. In the current absence of an international market value for carbon dioxide equivalent emissions, the Australian Greenhouse Office (AGO) suggests using, on an illustrative basis, the values contained in its 1999 publication, *Issuing the Permits*. The discussion paper suggests a permit price range of \$10 to \$50 a tonne carbon dioxide equivalent (CO_2-e). The lower value of \$10/tonne CO_2-e is consistent with the upper bound of the cost to government of abatement purchased under Round 1 of the Greenhouse Gas Abatement Program (GGAP). The abatement purchased under GGAP relates to the period 2008–2012, which aligns with the first commitment period under the Kyoto Protocol (Commonwealth of Australia 2003). Because there are no definitive estimates of the cost of greenhouse gas emissions, a lower cost of \$10/tonne CO_2-e is recommended in the Guidelines.

Assumptions:

- CO_2-e is a measure used for the emissions of CO_2, CH_4 and N_2O.
- In the Guidelines, greenhouse gas values are calculated using the AGO \$/tonne costs, emission factors from Cosgrove (2003), which contain a petrol/diesel fleet mix average, and SMVU data.
- CO_2-e emission rates assume a global warming potential of one for CO_2, 21 for CH_4 and 310 for N_2O.
- As the greenhouse effect is global in nature, the values are the same for urban and rural locations.
- The calculation methodology adopted in BTE (1999a), for heavy trucks only, was applied to rail.

Steps for application:

As for air pollution.

Alternatively, users can consider a range of \$/tonne costs, as a result of uncertainties in valuation and absence of Australian data. A sensitivity analysis using a low and a a high range (e.g. A\$10/tonne to A\$50/tonne CO_2-e) may be undertaken if further analysis is required.

Noise pollution

Analysis of noise pollution requires the use of a logarithmic scale of decibels (dB), and a frequency sensitivity is included by applying an 'A-weighting' scale (dB(A)).

The noise depreciation index (NDI) is a widely used parameter to link road traffic noise to a monetary value using hedonic pricing. The NDI gives an estimate of the depreciation (%) in-house value for a unit (1 dB) increase in noise level above the threshold level selected (measured as daily equivalent noise levels (Leq). In Australia, a NDI of 0.5 per cent of property value per dB(A) is typically applied for noise levels in excess of a threshold level of 50dB(A)–55dB(A) (Austroads 2003c).

Other approaches to value the cost of noise include estimation of control costs, health costs and WTP for reduced disturbance (Austroads 2003c).

Noise estimates recommended in the Guidelines are based on WTP and are derived from Austroads (2003c) and Austroads (2006).

Assumptions:

- Values are based on the Infras/IWW (2000) study methodology, which estimates the total noise costs.
- These values were selected for the Guidelines because they represent the sum of a WTP component and include a health cost component.
- The WTP component is expressed as a share of per capita income, and the health cost component considers the health effects of noise exposure such as disturbance and other stress reactions and health risks.
- Infras/IWW (2000) suggests that approximately 60 per cent of the total costs are associated with the WTP for noise reduction and 40 per cent are associated with health effects (Austroads 2003c).
- Austroads (2003c) notes that key factors affecting the unit costs include exposure levels of the population (urban/rural density and day–night population movements), and variations in per capita income that affect WTP. See Austroads (2003c) for further information.

- Values are also dependent on vehicle speed, proximity to the road, gradient of the road, surface and the noise–generation function of each vehicle type (tyre noise, engine type, maintenance). All else being equal, roads with high-speed traffic generate more noise than roads with low-speed traffic, and roads close to houses cause more disturbance than roads some distance away. The effect of lower speeds is partially offset by the proximity of houses. See Pratt (2002) for further information.
- Rural heavy vehicles values are estimated as 10 per cent of the urban value. As noise is mostly an urban issue, influenced by population and vkt, rural noise unit costs for passenger and LCVs are set to zero. It is recommended that for rural townships, the urban value should be used.
- For rail, the estimates were obtained from Laird (2005).

Steps for application:

Quantifiable impacts:

- Establish the environment (urban or rural), proportions of vehicles and vehicle types travelled.
- Determine the vehicle kilometres of traffic breaking noise thresholds by vehicle type.

Valuation:

- Apply cents/vkm or cents/ntkm values.
- Multiply vehicle-kilometres obtained by cents/vkm or cents/ntkm for the Base and Project Cases.
- Incorporate values into the BCA.

Water pollution

Estimates for water pollution include values of organic waste and persistent toxicants (road run-off from vehicles such as engine oil leakage and disposal, road surface, particulate matter and other air pollutants from exhausts, tyre degradation). Values are disaggregated by cents/vkm for passenger vehicles and cents/ntkm for freight transport. Costs are unlikely to reflect all damage costs from transport alone, and depend on rainfall intensity, drainage path length, type of road and type of system.

Assumptions:

- Methods to value water pollution impacts resulting from road transport include dose-response, contingent valuation and WTP. Estimates provided in the Guidelines are based on mitigation costs, which value transport-related impacts by estimating social costs of installing mitigation devices (i.e. vegetation, sedimentation tanks, combined catchment and treatment of stormwater run-off) over entire road networks or on a per vehicle-kilometre basis. See Pratt (2002) for further detail.
- Rural values are estimated as one-tenth of the urban figure for passenger vehicles, and one-hundredth of the urban figure for freight. Notably, values are so small for heavy freight vehicles that they are documented as zero in the table.
- Rail parameters are estimated by scaling the Infras/IWW (2000) nature value for heavy vehicles and rail freight.

Steps for application:

- Estimate potential environmental impact of run-off from road vehicles.
- Estimate the degree of rainfall intensity, mitigation devices, type of road and drainage path length.
- Estimate vehicle-kilometres.
- Multiply by cents/vkm or cents/ntkm value for the Base and Project Cases.
- Quantify on an initiative-by-initiative basis, because the initiative may be site-specific.

Nature and landscape

Estimates are based on Austroads (2006). Austroads refers to the Infras/IWW (2000) study, which includes effects such as loss of natural areas, ecological impacts (to land, water and biodiversity) and reductions in the quality of the landscape. This methodology is based on the costs of repair and compensation measures (Austroads, 2003c). See Austroads (2003c) for further detail on the calculation methodology.

Assumptions:

- Although nature and landscape values are difficult to quantify and are highly specific to the initiative being appraised, the values recommended in the Guidelines provide a rough guide.
- Austroads (2006) notes that as most infrastructure is in urban areas, rural values should be set at 30 per cent of urban values. Austroads (2006) calculates LCV rural values as one per cent of LCV urban values. See Austroads (2006) for further information.
- Rail parameters are estimated by scaling the Infras/IWW (2000) nature value for heavy vehicles and rail freight.

Steps for application:

- Determine the area of land impacted (e.g. direct and indirect land).
- Estimate vehicle-kilometres.
- Multiply by cents/vkm or cents/ntkm value for the Base and Project Cases.
- Quantify on an initiative-by-initiative basis, because the impact may be site-specific.

Urban separation

Estimates for passenger and freight vehicles are sourced from Austroads (2003c) and Austroads (2006). Austroads refers to the Infras/IWW (2000) study, which assesses the constraints to mobility of pedestrians as a technique to value urban separation. See Austroads (2006) and Infras/IWWW (2000) for further detail.

Assumptions:

- Although urban separation values are difficult to quantify and are highly specific to the initiative being appraised, the values recommended in the Guidelines provide a rough guide.
- The Infras/IWW (2000) study applies typical average waiting times that depend on traffic volume and road type (Austroads 2003c).
- Rail parameters are estimated by scaling the Infras/IWW (2000) urban value for heavy vehicles and rail freight.

Steps for application:

- Determine the constraints to mobility of pedestrians.
- Estimate vehicle-kilometres.
- Multiply by cents/vkm or cents/ntkm value for the Base and Project Cases.
- Quantify on an initiative-by initiative basis, because the impact may be site-specific.

2.9.2 Estimation of externalities specific to the initiative

Ways to value externalities in specific cases include[18]:

- directly using market prices
  - value of reduced output of producers less the resource cost of inputs saved (for example, if pollution reduces crop yields)
  - loss of earnings of people affected, also called the human capital approach (for example, pollution affecting people's health reducing their capacity to work)

18 The techniques described here are for cases where the externality is negative and therefore imposes costs. Externalities may be positive and give rise to benefits. This section is based on Perkins (1994, Chapter 11) and Boardman et al. (1996, Chapters 10 and 11).

    - opportunity cost, that is, the difference between the NPV of the initiative being appraised and the next best option that does not give rise to the externality
    - mitigating, preventative or replacement expenditures, that is, estimating the expenditures incurred to mitigate or prevent the effects of the externality (for example, costs of insulation against noise, medical costs of treating health problems caused by pollution), or to replace damaged or lost assets
- indirectly using surrogate prices
    - hedonic pricing, that is, using changes in land values to estimate how people value externalities (for example, to estimate noise costs)
    - travel costs, that is, using costs of travel (vehicle operation, fares, time, accommodation, admission prices) to estimate the value of recreational sites such as national parks, and
- asking people through 'contingent valuation' surveys. In its simplest form, people might be asked to state their maximum WTP to avoid the externality being costed. More sophisticated methods involve asking consumers to rank feasible combinations of quantities of the externality and monetary payments, or asking consumers to respond 'yes' or 'no' to payments at different levels to avoid the externality. The main problem with contingent valuation methods is the hypothetical nature of the questions being asked. Careful design of the questions and training of interviewers is required.

2.9.3 Calculation of benefits or costs

No additions to the corresponding section in Volume 3 are made here.

2.10 Discounting and decision criteria

2.10.1 The discount rate

When investing capital in an initiative, a large quantity of resources are forgone to establish the initiative in return for a stream of benefits spread over a number of years. Present consumption is forgone in return for future consumption. For individuals, a dollar today is worth more than a dollar in the future for two reasons:

- uncertainty about the future, and
- there is an expectation that society and individuals will be better off in the future than they are now (Perkins 1994, p. 53).

Discounting is the reverse of adding interest—it reduces the money value of future sums to reflect the preference for (or trade-off between) present benefits over future benefits. It enables future benefits and costs to be appraised in terms of a common Reference (Base) Year, and enables different time profiles of benefits and costs to be compared on the same basis.

A higher discount rate makes it more difficult for initiatives to pass a BCA. The size of the discount rate can affect initiative rankings. A higher discount rate favours initiatives where benefits are realised earlier in the lives of initiatives, and vice versa.

Choosing a discount rate is problematic. In a BCA, the aim is to measure the opportunity cost of resources used. Capital invested can come from three sources, and benefits can be returned to these three sources. Each source has its own discount rate:

1. deferred private consumption
2. forgone or delayed private sector investment, and
3. net borrowing from overseas sources.

In a tax-free world, the discount rates for the three sources of funds are identical. Taxes drive wedges between them. Taxes on interest reduce the rate for deferred private consumption. Corporate income taxes raise the rate for forgone private sector investment. The theoretically correct approach is to discount at the rate for one of the sources, and then apply shadow-price adjustment factors to those parts of costs coming from, and benefits going to, the other two sources. However, the information requirements of such an approach make it impractical.

In an open economy, where the country has a good credit rating, governments and firms can borrow (or lend) in international capital markets without affecting the interest rates paid (or received) over the relevant range. Since funds invested can ultimately be sourced to overseas borrowing, the international cost of capital to the country appears to be the appropriate discount rate. However, the benefits from an initiative can still be consumed or re-invested. The data required to estimate amounts consumed and re-invested, and the discount rate for private consumption, make this approach also impractical.

In the absence of a better solution, BTE (1999b, p. 78) concludes that the most appropriate discount rate to use for BCA is the government bond rate. Because the government will not default on loans, the bond rate provides a ready measure of the cost of capital free of any risk premium. The nominal bond rate needs to be adjusted for inflation to obtain the real rate. Using the real yield on indexed bonds (see the Reserve Bank of Australia website, www.rba.gov.au) avoids the need to adjust for inflation.

The private sector practice of adding a risk premium to the discount rate is not appropriate for evaluation of public sector initiatives, as discussed in Section 2.11.

To ensure consistency of appraisal within a jurisdiction, it is important that transport and roads agencies adopt a single rate. Treasuries and finance departments usually nominate a discount rate to use for BCAs.

2.10.2 Timing of benefits and costs

No additions to the corresponding section in Volume 3 are made here.

2.10.3 Net present value (NPV)

A positive NPV means that implementing the initiative represents an improvement in economic efficiency, compared with the Base Case.

The NPV is used to compare mutually exclusive options for the same initiative. Options may include different solutions to the same challenge, different routes for a new road or railway and different scales and implementation times for the same initiative.

2.10.4 Benefit–cost ratio (BCR)

The BCR measure is used to rank initiatives where there is a budget constraint. In a budget-constrained situation, net economic benefits to society are maximised by undertaking initiatives in descending order of BCR ranking, until the budget is exhausted. Provided the BCA accounts for all benefits and costs, that is, there are no non-monetised benefits or costs, this maximises the economic efficiency objective, subject to the budget constraint.

BCR definition

The definition of BCR provided in Volume 3 places the present value of infrastructure operating costs (ongoing costs) in the numerator as a negative benefit. This differs from BCR definitions in many BCA manuals where infrastructure operating costs are added to investment costs (or capital costs) in the denominator.

Perkins (1994) defines BCR as having operating costs in the *denominator* and uses the term 'net benefit investment ratio' (NBIR) to refer to the ratio with operating costs in the *numerator*. According to Perkins, the BCR with operating costs in the denominator cannot be used to rank initiatives. Its only use is as an indicator of whether or not the NPV is positive. NBIR should be used to rank initiatives because only the investment costs of new initiatives are paid from the budget, not the associated operating costs. If operating costs are paid from the same budget as capital initiatives, the NBIR is still the appropriate measure for ranking initiatives, because the operating costs arising from *new* initiatives come from *future* budgets, not the currently allocated budget. The issue of *future* operating costs competing for scarce funds out of *future* budgets is addressed in detail below. The Guidelines use the BCR term for NBIR because BCR is well-known and is often defined as if it were the NBIR.

Why the BCR is used to rank initiatives

The following demonstrates why the BCR must be used to rank initiatives and why infrastructure operating costs should be included in the numerator.

For an individual initiative i

$$NPV_i = PVB_i - PVOC_i - PVIC_i$$

where

PVB_i = present value of benefits

$PVOC_i$ = present value of infrastructure operating costs, and

$PVIC_i$ = present value of investment costs.

In a budget-constrained situation, choose the combination of initiatives that achieves the maximum value of

$$\sum_i NPV_i = \sum_i (PVB_i - PVOC_i - PVIC_i) \quad \text{subject to} \quad \sum_i PVIC_i \leq F$$

where F is the budget constraint, that is, the total amount of funds available to invest in capital initiatives *in the current period*. If operating costs are paid from the same budget as investment costs, the amount F is the residual after paying system operating costs for the current period. All the costs that comprise *PVOC* for initiatives under consideration are paid for out of budgets in future periods, not the current period. Hence only *PVIC* appears in the constraint.

To solve the maximisation problem, exclude all initiatives with negative NPVs, because they cannot feature in the optimal solution. If the sum of the investment costs of the remaining initiatives is less than the budget constraint, that is, the constraint is non-binding, $\sum_i PVIC_i < F$, then there is no need to rank initiatives. The optimal solution is simply to implement all initiatives with positive NPVs.

If the sum of the investment costs exceeds available funds, then the optimal solution could be found by identifying all possible combinations of initiative for which $\sum_i PVIC_i = F$ (assuming that all the investment costs for all possible combinations of initiatives exactly sum to F). The expression to maximise can be rewritten as

$$\sum_i NPV_i = \sum_i (PVB_i - PVOC_i) - \sum_i PVIC_i = \sum_i (PVB_i - PVOC_i) - F$$

Since F is common to all solutions, the expression to be maximised is reduced to

$$\sum_i (PVB_i - PVOC_i)$$

For each initiative, dividing $PVB_i - PVOC_i$ by $PVIC_i$ gives the value of benefits less the operating costs the initiative contributes per dollar of investment cost. The optimal combination of initiatives can be found by selecting initiatives in descending order of $(PVB_i - PVOC_i)/PVIC_i$, that is, the BCR with operating costs in the numerator, until the budget constraint is exhausted.

It follows that only costs paid from the current budget should be included in the denominator. Costs incurred by society during the construction phase of an initiative that are not paid from the government's budget, such as externalities arising from noise, dust and traffic disruption, should be counted as negative benefits in the numerator.

If only part of an initiative's funding comes from the government's budget and the remainder is sourced from the private sector, see Section 2.18.

Lumpy initiatives

Ranking initiatives by BCR is not guaranteed to give the best economic efficiency outcome where initiatives are large relative to the size of the budget constraint and, after funding the last initiative that can be afforded, there are still funds left over. For example, say, after funding higher-BCR initiatives, $100 million is left in the budget and the next three initiatives in the BCR in order or merit are:

- A: cost $50 million BCR 4.0
- B: cost $100 million BCR 3.5
- C: cost $50 million BCR 2.0.

After funding Initiative A, the remaining $50 million is insufficient to pay for Initiative B. Initiative C, however, could be included. The total benefits from implementing Initiatives A and C together are ($50m×4.0) + ($50m×2.0) = $300 million. However, if the remaining $100 million is used to fund Initiative B instead, the total benefits are $100m × 3.5 = $350 million. Therefore, different combinations of initiatives should be tested to find the optimal combination.

Cut-off BCR over time

If a budget-constrained government agency is able to shift funds through time by borrowing or lending at the discount rate, the cut-off BCR or marginal BCR (i.e. the BCR of the last initiative implemented) should be kept uniform over time to ensure the most efficient allocation of resources.

To demonstrate this, say, the marginal BCR this year is 2.0 and next year it is 3.0. If $1 of capital spending is shifted from this year's budget to next year's budget, society forgoes $2 in benefit. The $1 could be invested elsewhere for the year. So it is worth *$1 × (1+r)* next year, where *r* is the discount rate. The benefit from investing in next year's initiatives is *$3 × (1+r)*, which is worth $3 when discounted back to this year. The net gain to society is $1 in benefit in present value terms. As more funds are shifted from this year to next year, the marginal BCR this year rises and the marginal BCR next year falls. When they are equal, no further gains can be made by shifting funds through time.

The implication for governments is that budgets for infrastructure spending should be expanded in times when there is a strong demand for funds expressed in greater numbers of initiatives with high BCRs; budgets should be contracted at times when demand is weak, with a view to maintaining a fairly constant cut-off BCR over the long-term.[19]

Treatment of infrastructure operating costs

The question of whether to place operating costs in the numerator or the denominator leads to the issue of operating costs of current capital initiatives competing for funds with capital initiatives out of future budgets. This is usually the case for road agencies. New capital initiatives add only a small fraction to the total maintenance costs of a road or rail network, and only well into the future when periodic maintenance is due. So it is generally safe to ignore impacts of current capital initiatives on future maintenance budgets.

19 This conclusion is based purely on micro-economic resource allocation considerations. There are also macro-economic considerations.

In the event that an assessment is necessary, an optimisation problem could be set up with assumptions about the size of future budgets and the benefits and costs of future initiatives. A simple approach is to assume values for cut-off BCRs in future years. Maintenance costs for all initiatives ranked in the current period could be multiplied by the cut-off BCR in the year in which they are incurred. Assuming a constant value for μ over time, the BCR definition would then become $BCR = \frac{PVB - \mu PVOC}{PVIC}$ where μ is the assumed future cut-off or marginal BCR.

For example, say, an initiative implemented today creates a maintenance need of $1 million in 10 years time, funded out of the same budget as capital initiatives. Assuming that the cut-off BCR in 10 years time is 3.0, spending $1 million on maintenance in 10 years time precludes the opportunity to invest $1 million in capital initiatives with a BCR of 3.0. Hence, the opportunity cost of the maintenance commitment generated by the current initiative is $3 million in forgone benefits in 10 years time.

Provided μ stays constant over a long period of time (this is the economically efficient policy), placing operating costs in the denominator will not give the wrong decision. This is because if the BCR equals μ, then the BCR is the same under either definition

$$\frac{PVB - \mu PVOC}{PVIC} = \mu$$

$$PVB - \mu PVOC = \mu PVIC$$

$$PVB = \mu (PVOC + PVIC)$$

$$\frac{PVB}{PVOC + PVIC} = \mu$$

It follows that if the BCR exceeds μ under one definition, it will also exceed μ under the alternative definition. Hence, as long as the BCR is above μ under both definitions and μ is constant over time, there should be no wrong decisions. Rankings, however, may be distorted as Table 2.4 shows. Initiatives A and B have the same level of investment costs. With operating costs in the denominator, Initiative B is ranked above A, even though it has a lower NPV.

Table 2.4: Example comparing BCR definitions

($ millions)

Initiative	PVB	PVOC	PVIC	NPV μ = 2.0	BCR: OC in numerator μ = 2.0	BCR: OC in denominator
A	400	50	100	200	3.0	2.667
B	300	10	100	180	2.8	2.727

Notes: PVB = present value of benefits.
PVOC = present value of operating costs.
PVIC = present value of investment costs.
BCR: OC in numerator = benefit–cost ratio with operating costs subtracted from the numerator.
BCR: OC in denominator = benefit–cost ratio with operating costs added to the denominator.
μ = long-term cut-off BCR, the factor by which operating costs are multiplied to reflect the opportunity cost of funds in the future.

Placing operating costs in the denominator can lead to wrong decisions if μ changes over time, which is quite possible because initiatives can have very long lives. Considering the numerical example in Table 2.4, say, there is $100 million left in the budget for the current period and the choice is between Initiative A and Initiative B. When the BCR is defined with operating costs in the numerator, Initiative A is chosen—the correct decision. When the BCR is defined with operating costs in the denominator, Initiative B is chosen—the wrong decision. The cut-off BCR, μ, has changed from 3.0 in the current period to 2.0 in future periods when maintenance costs are incurred. If the current cut-off BCR of 3.0 continued into the future, the decision to select

Initiative B is correct, as Table 2.5 demonstrates. The BCR with operating costs in the denominator is guaranteed to produce correct decisions only if the cut-off BCR stays constant over time.

Table 2.5: Example comparing BCR definitions

($ millions)

Initiative	PVB	PVOC	PVIC	NPV μ = 3.0	BCR: OC in numerator μ = 3.0	BCR: OC in denominator
A	400	50	100	150	2.5	2.667
B	300	10	100	170	2.7	2.727

Note: See notes to Table 2.4 for explanation of column headings.

The numbers in Tables 2.4 and 2.5 are contrived to demonstrate the point. In practice, operating costs for road and rail infrastructure initiatives tend to be very small in relation to benefits and investment costs, so placing operating costs in the numerator or the denominator rarely affects initiative rankings or investment decisions. An exception might be sealing a gravel road, which leads to a high saving in maintenance costs relative to the level of investment costs. For non-infrastructure initiatives, such as intelligent transport system solutions, the relativities between benefits, operating costs and investment costs could be very different. In these cases, the possibility of assuming a value of μ above one might be considered if operating costs are paid for from the same budget as capital costs in the future.

The level of μ over time depends on the balance between the demand for infrastructure spending and the supply of available funds. μ rises if funding for transport infrastructure fails to keep pace with increasing demand, changes in the locations of population and economic activity and replacement needs of existing infrastructure.

The Guidelines recommend defining BCR with operating costs in the numerator for two reasons:

1. As shown above, in some circumstances, placing operating costs in the denominator can lead to wrong decisions.
2. With operating costs in the numerator, there is no ambiguity about what is a cost and what is a negative benefit (or what is a benefit and what is a negative cost)—all costs paid for from the current budget-constrained funds belong in the denominator; all other benefits and costs belong in the numerator. With operating costs in the denominator, ambiguities can arise, as the following examples show:
   - The cost of maintaining a length of newly constructed rail track is paid for by the track operator out of revenues, not out of future government budgets. With the BCR defined with operating costs in the denominator, maintenance costs belong in the denominator as do maintenance costs for a road initiative. However, the denominator then overstates the impact on the government's budget of selecting the initiative, suggesting that the infrastructure operating costs should be placed in the numerator.
   - Infrastructure initiatives that replace ageing infrastructure can save on future maintenance costs. Upgrading from bitumen to a concrete pavement saves on maintenance costs. Therefore, there can be operating benefits. With the BCR defined with operating costs in the denominator, it is not clear whether maintenance benefits should be treated as a benefit in the numerator or a negative operating cost in the denominator.

BCR and mutually exclusive initiative options

BCRs should never be used to choose between mutually exclusive options for the same initiative, because they remove the effects of different initiative scales. In the example given in Table 2.6, society is better off by $200 million by investing in Option B compared with $150 million by investing in Option A. So Option B should be selected. However, Option A has a higher BCR.

Table 2.6: Example showing misuse of BCR

($ millions)

OPTION IDENTIFICATION	PVB	PVIC	NPV	BCR
A	250	100	150	2.5
B	400	200	200	2

2.10.5 Incremental BCR

The IBCR is the ratio of the present value of the additional benefit to the present value of the additional cost when going from one size or standard of initiative to the next. Selection between mutually exclusive options can be undertaken using the incremental BCR (IBCR) defined as

$$IBCR = \frac{PV(B_2 - OC_2) - PV(B_1 - OC_1)}{PV(IC_2) - PV(IC_1)}$$

where the subscripts represent Options 1 and 2. Operating costs may be grouped with investment costs in the denominator—this has no effect on the decision, provided it is consistent.

For the example in Table 2.6, the IBCR is (400–250) / (200–100) = 1.5 > 1.0. So Option 2 is preferred. If there was a third option with benefits of $450 million and costs of $300 million, its IBCR compared with Option 2 is (450–400) / (300–200) = 0.5 < 1.0, indicating that Option 3 is less desirable than Option 2. This can be verified by comparing the NPV for Option 3 of $150 million with the NPV for Option 2 of $200 million.

In the absence of a budget constraint, the size or standard of the initiative should be increased until the IBCR falls below one. With a budget constraint, the IBCR should not fall below the cut-off BCR applied generally.

2.10.6 Internal rate of return

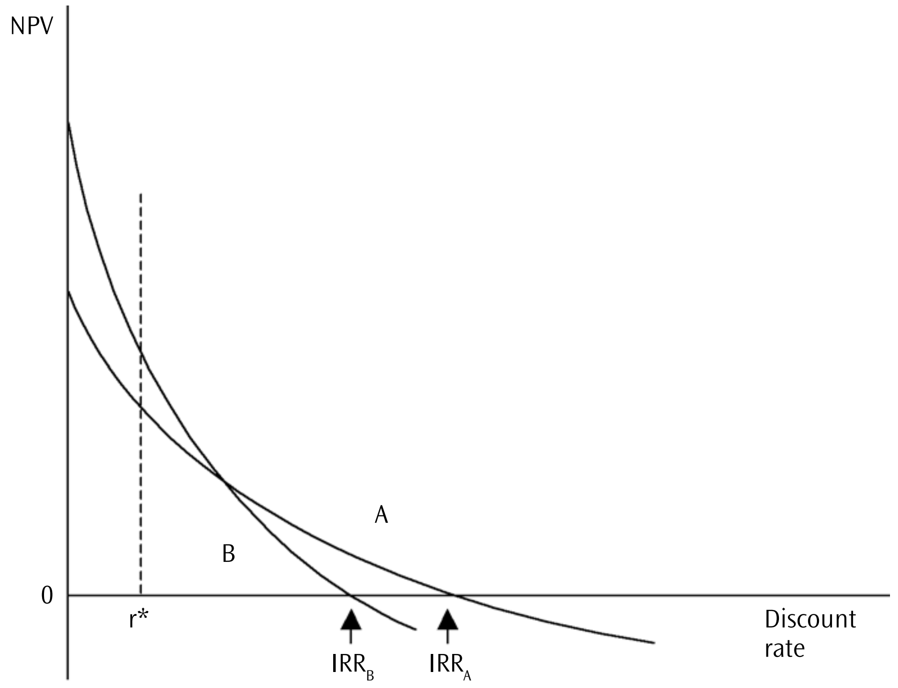
The internal rate of return (IRR) is defined as the value of the discount rate where the NPV equals zero. It represents the minimum discount rate at which the initiative is viable in economic terms. For example, if an initiative has an IRR of 8 per cent, it has a positive NPV at discount rates below 8 per cent. The initiative is rejected (NPV<0) at discount rates above 8 per cent.

There is no formula for the IRR; it is found by iteration. Spreadsheet packages such as Excel and Lotus 123 have functions to find the IRR.

The IRR can be used the same way as the NPV to indicate whether or not an initiative is of overall benefit to society as a whole. It provides an indication of the worth of an initiative, without the need to be specific about the discount rate.

The IRR should never be used to rank initiatives or to choose between mutually exclusive options as this would involve comparing initiatives using different discount rates. Figure 2.24 illustrates this by showing plots of NPVs against the discount rate for two hypothetical initiatives, A and B. Initiative A has a higher IRR than Initiative B. If the discount rate is r^*, ranking the initiatives by IRR leads to the wrong decision.

Figure 2.24: Demonstration of why initiatives should not be ranked by IRR



IRRs cannot be calculated for initiatives where the stream of net benefits does not change sign. For example, a BCA of increasing the mass limits for heavy vehicles has no initial investment costs.

If the stream of net benefits changes sign more than once (for example, a construction phase followed by a period of positive net benefits, followed by a period of negative net benefits), it is possible to have multiple IRRs. The IRR criterion then should not be used.

2.10.7 First-year rate of return

The first-year rate of return (FYRR) is the level of benefits minus operating costs in the first year of operation of the initiative discounted to year zero, divided by the present value of investment costs

$$FYRR = \frac{B_{t_f}}{(1+r)^{t_f}} \Big/ \sum_{t=0}^{t_f-1} \frac{IC_1}{(1+r)^t}$$

where t_f is the first year of operation of the initiative.

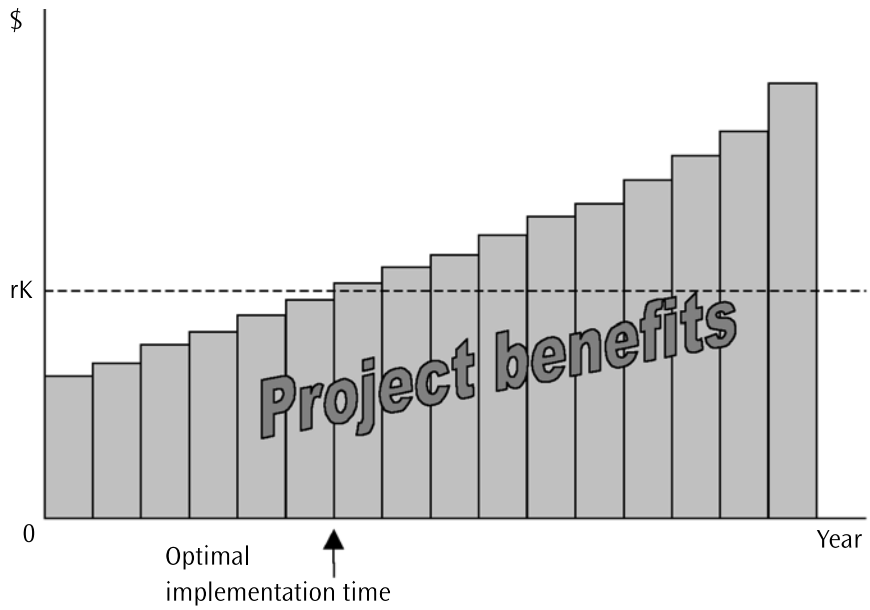
The FYRR can indicate whether an initiative's optimal implementation time is in the past or future, and whether deferral is warranted. Provided the assumptions underlying the criterion are met, the optimal implementation time is the first year in which the FYRR is greater than the discount rate.

Figure 2.25 shows the basis of the criterion assuming that all investment costs are spent in year zero and the initiative starts operation then, ($t_f = 0$). The block heights represent net benefits in each year the initiative is in place. Benefits increase over time due to growth in demand. The dashed line is set at the discount rate, multiplied by the capital cost (rK). Optimal implementation time occurs in the first year when $B_0 = rK$. For years prior to the optimal time, when $B_o < rK$, delaying the initiative by one year results in a loss to society of the benefits for that year. However, the capital costs of the initiative can be invested elsewhere (or not borrowed from overseas, saving on interest), saving society the time value of the funds for one year, rK. Society is better off by $rK - B_0$ by delaying the initiative. For years after the optimal time, when $B_0 > rK$, the value of forgone benefits by delaying the initiative by one year exceeds the time value of the capital costs. Society is worse off by $B_0 - rK$, so delaying the initiative imposes net costs on society.

If an initiative passes the FYRR test, $B_0 > rK$, its optimal implementation time lies in the past. If it fails, that is, $B_0 < rK$, its optimal implementation time is in the future and deferring the initiative should be considered.

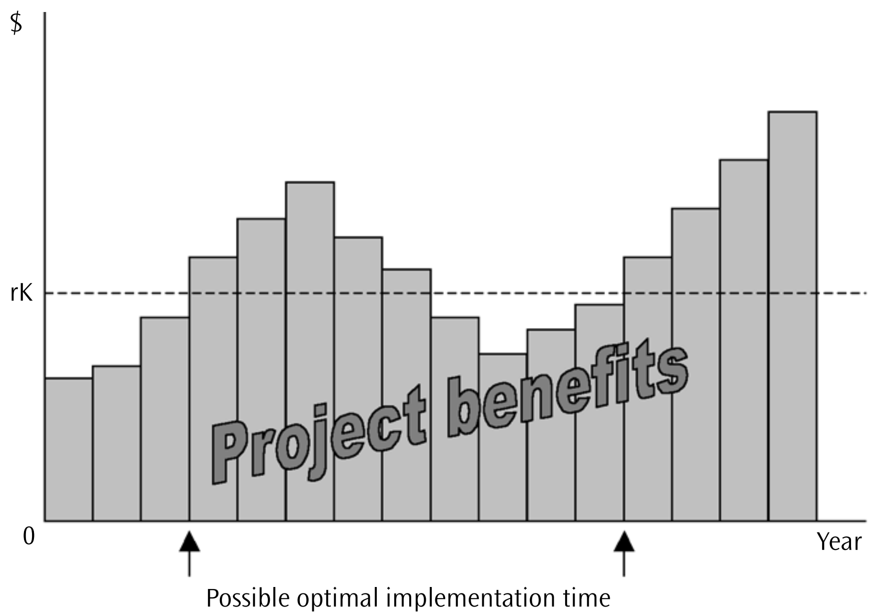
When the initiative takes time to construct, *f* years, the benefit forgone due to a one-year delay in commencement occurs *f* years into the future, while the saving in the time value of the capital invested ($r \times$ the present value of the investment costs) is in today's dollars. This is why first-year benefits are discounted by t_f years to the present in the FYRR formula.

Figure 2.25: Optimal initiative implementation time



The FYRR criterion is based on an assumption that benefits do not fall. In Figure 2.26, due to a dip in initiative benefits, there are two possible optimal times. The only way to determine the true optimal time is to treat the two optimal times as mutually exclusive options, that is, to compare the NPVs with the initiative implemented at each of the two times when $B_0 = rK$.

Figure 2.26: Possible optimal initiative implementation times



When comparing different implementation times, the NPV calculations must be made from the same year of analysis. For example, if the current year is the year of analysis, implementation in Year 5 would cause construction costs to be discounted by five years, first-year benefits by six years and so on. Implementation in Year 10 would cause construction costs to be discounted by 10 years, first-year benefits by 11 years and so on.

Another important assumption underlying the FYRR criterion is that changing the year of implementation does not alter benefits in any years. For example, regardless of whether the initiative is completed in 2010 or 2015, net benefits in 2020 are the same. Delaying construction might also delay major maintenance expenses. However, these are incurred well into the future so that the effects of changing their time in an initiative's NPV is minimal. Small violations of this assumption are acceptable.

Where maintenance costs occur at fixed numbers of years after implementation, that is, delay of the implementation by (say) five years delays all maintenance costs by five years, the present value of the maintenance costs should be added onto the capital cost to calculate the FYRR. The initiative's delay leads to additional gains to society by delaying future maintenance costs. The size of the gain is given by the discount rate multiplied by the present value of the future maintenance costs. However, there may be offsetting additional maintenance costs in the Base Case if the initiative is delayed.

Where the effects of delaying the implementation time are more pronounced, or more complex, the NPV should be recalculated for each implementation year to find the year with the maximum NPV for a given year of analysis. Section 2.11.10 continues this point, discussing situations where deferring an initiative creates costs or alters risks.

It is recommended that all initiatives are subjected to the FYRR test, with the result reported in the initiative's Business Case.

For transport infrastructure initiatives evaluated over long periods of time with high demand growth, levels of congestion can become very large in the Base Case. Forecast benefits in the latter years of the initiative's life can be extremely large, leading to high BCRs. However, the implication that the initiative is urgently needed may be quite misleading.

To illustrate, say, a two-lane road has an average annual daily traffic (AADT) level of 9000 vehicles per day in 2006 and this level is forecast to grow at 3 per cent per annum. Using the FYRR criterion, the optimal implementation time for duplicating the road occurs when traffic grows to 10 000 vehicles per day in 2010. By 2034, the AADT has passed 20 000 vehicles per day. In the Base Case, there are 20 000 vehicles per day on a two-lane road, causing vehicle speeds to be reduced well below free-speed levels. When road-user costs at these speeds are compared with road-user costs in the Project Case, large benefits ensue. Even though heavily discounted, large benefits can significantly inflate BCRs.

2.10.8 Discounting for financial analysis

No additions to the corresponding section in Volume 3 are made here.

2.11 Risk and uncertainty[20]

Defining risk and uncertainty

All benefits and costs that go into a BCA are forecasts of the future. Risk arises from the possibility that a forecast will prove to be wrong. According to the Penguin Dictionary of Economics, risk is defined as, 'A state in which the number of possible future events exceeds the number of events that will actually occur, and some measure of probability can be attached to them' (Bannock et al. 2003, p. 338). A distinction is sometimes drawn between 'risk' and 'uncertainty'. 'Risk' occurs where the probability distribution is known, and 'uncertainty' where it is not. For the purposes of initiative appraisal, this distinction is irrelevant because analysis of uncertainty requires a probability distribution to be specified. In this context, the two terms may be used interchangeably.

An alternative approach in the literature is to define 'uncertainty' as imperfect knowledge about the future, and 'risk' as uncertain consequences. For example, the weather tomorrow is uncertain, but, without consequences, there is no risk present. Risk only exists when the weather has the potential to cause some economic loss or gain, physical damage or injury, or delay (Austroads 2002, p. 3).

The main sources of risk for investment initiatives are:

- construction costs differing from expected costs because of changes in input costs or unforeseen technical factors
- operating costs differing from expected costs because of changes in input costs or unforeseen technical factors
- unrealised demand forecasts, a risk that increases the further into the future projections are made
- environmental impacts that differ from the expected or are unforeseen, and
- network effects, where an asset is part of the network (for example, an individual length of road or rail track) and decisions made elsewhere in the network impact the initiative in question.

A distinction can be drawn between 'downside risk' and 'pure risk'. There is a good deal of evidence that ex-ante evaluations of investment initiatives tend to be over-optimistic compared with ex-post performance (Flyvbjerg et al. 2002; 2003; 2006). People tend not to consider what can go wrong, causing assessments to be biased in favour of the initiative. Also, where probability distributions are skewed, people choose the modal value of the variable (the highest point of the probability distribution) rather than the mean. Consequently, the forecasts employed in evaluations tend to be more favourable than expected values (the mean of the probability distribution). The probability of a below-forecast outcome is then greater than for an above-forecast outcome. For construction and operating costs, the tendency is to underestimate, and for demand, overestimate. The difference between an initiative biased on the optimistic side and the expected value is termed 'downside risk'.

Downside risk is part of a broader problem called 'optimism bias', which includes organisation of decision-making processes and strategic behaviour that favour implementation of large, capital-intensive initiatives (UK Department for Transport 2004). The Danish sociologist, Bent Flyvbjerg has written extensively on the subject (see, for example, Flyvbjerg et al. 2003). Flyvbjerg et al. (2002) found that of 258 transport initiatives, the actual capital cost of road and rail initiatives was, on average, between 20 and 45 per cent more than estimated. In a later study of 210 transport initiatives, the same authors (2006) found that actual patronage for rail initiatives was 49 per cent of forecast patronage.

If downside risk is eliminated from projections, variation about the expected value remains, and is called pure risk. With an unbiased forecast, there is an equal probability of a better or a worse outcome than the forecast. Pure risk can be further divided into random variation, which can be eliminated by diversification, and risk correlated with the general level of economic activity that cannot be diversified away.

20 The approach to risk advocated in this section is based on BTRE (2005), which provides a more detailed discussion.

Risk premium in the discount rate

One approach to incorporating risk into appraisals is to adjust for risk by adding a premium to the discount rate. Adding a few percentage points to the discount rate reduces the present value of returns from the initiative.

The risk premium approach is valid for financial analysis where the weighted average cost of capital includes the risk premium. For BCAs, however, it is not recommended. It can distort ranking of initiatives.

In situations where initiatives are ranked to allocate program funding, it is more appropriate to discount at the long-term government bond rate, that is, the risk-free rate.

Downside risk

The addition of a risk premium to the discount rate reduces the present value of future benefits and costs. The proportional reduction rises each year into the future as the power in the denominator of the present value formula grows. The risk premium properly adjusts for downside risk only if the extent of net benefit overestimation rises each year in line with the proportional reduction. This would occur only by coincidence. It should not occur where demand is estimated by econometric modelling. The typical statistical procedures used to make projections of demand growth are designed to minimise bias when applied correctly, and so should not be subject to downside risk.

A risk premium does little to correct for the risk of underestimation of construction costs because construction costs occur in the first few years. Failure of net benefits to reach expectations once the initiative is complete may appear as soon as the initiative commences, rather than build up gradually over time in line with the adjustments made by a risk premium. Examples include higher than expected operating costs and, where an initiative faces competition, a failure to realise the expected market share.

The risk premium approach breaks down completely where there are negative net returns in some years of an initiative's life. An example is a nuclear power station, with a decommissioning cost at the end of its life. Raising the discount rate reduces the significance of future costs, making the initiative appear *more* attractive.

In conclusion, raising the discount rate to incorporate a risk premium is a poor way to correct for downside risk in a BCA. The way to ensure that projections are free of downside risk is discussed below.

For financial analyses, as noted in Section 2.10.8 in Volume 3, the discount rate to use is the weighted average cost of capital. Lenders include a risk premium in the interest rate they charge. The risk premium in the interest rate compensates for the losses made when borrowers default on loans.[21] The risk to which the premium applies is the downside risk that the borrower will go bankrupt, not the risk of an individual initiative underperforming. The revenues and costs in financial analyses should still be expected values, free of downside risk.

Pure risk

In share market terminology, pure risk can be broken down into 'diversifiable' and 'non-diversifiable' risk. The former (also called 'unsystematic' or 'idiosyncratic' risk) is risk that is not correlated with the share market as a whole and can be eliminated by holding a diversified portfolio. Losses on some shares are offset by gains on others, so the overall performance of the portfolio is smoothed. Diversifiable risk can be eliminated by diversification. 'Undiversifiable' risk

21 Say a lender has a portfolio of $100 million of loans and each year, on average, loses one per cent of their capital ($1 million) as a result of borrowers defaulting on their loans. The risk-free interest rate is 5 per cent. If the lender charged all customers a 6 per cent interest rate, the expected rate received would be 5 per cent, the risk-free rate, after taking account of loan defaults.

(also called 'systematic' or 'market' risk) cannot be diversified away and occurs when share prices are correlated with the market as a whole.

In the capital asset pricing model, the level of undiversifiable risk associated with a share determines the return above the risk-free return demanded by the market—that is, the risk premium. The higher the degree of correlation between the return for an individual share and the return for the market as a whole, the greater the risk premium is for that share. A share that has no correlation with the market as a whole is only required to earn the risk-free rate of return, that is, zero risk premium. Diversifiable risk has no effect on the risk premium.

The same reasoning can be applied to public sector investment initiatives. Governments invest in a large number of initiatives and their benefits and costs are spread over many individuals. Consequently, when individual initiatives over or underperform, the gains or losses are spread over many individuals, each of whom is only a little better or worse off than expected. Furthermore, the welfare of each individual is affected by many initiatives. Just as for a share portfolio, the initiatives that do worse than expected are more or less offset by initiatives that do better. So some of the risk associated with public sector initiatives can be classed as diversifiable. Since individuals are not selecting a portfolio of public sector initiatives in the same way they select a share portfolio, some of the diversifiable risk remains, but in most cases, it is likely to be small enough to be ignored for practical purposes.

The undiversifiable component of risk in public sector initiatives is the risk arising from benefits being correlated with movements in the economy as a whole. To the extent that the benefits from an initiative are correlated with individual consumption, there are grounds for making a negative adjustment to an initiative's benefits. However, a risk premium approach is correct only under a very restrictive set of assumptions. The assumptions are likely to be approximately met for an investor in shares, but are unrealistic for most public sector investment initiatives.

Furthermore, when plausible estimates of the variability of benefits and their correlation with the level of overall economic activity are substituted into the consumption capital asset pricing model formula, the percentage downward adjustment of benefits for undiversifiable risk is very small in relation to the margins for error in BCAs (BTRE 2005).

Pure risk can, therefore, usually be ignored for BCAs. The exception is where the initiative has a large effect on the welfare of affected individuals so the risk is not diversified away by other public sector initiatives. Examples include initiatives that supply essential services to isolated communities where the success or failure of the initiative can make a large difference to the communities' welfare. BTRE (2005) shows how to adjust benefits and costs for risk in these situations.

The remainder of this section is mostly concerned with treatment of downside risk in appraisals.

2.11.1 Sensitivity analysis

Sensitivity analysis is a simple way to analyse the uncertainty surrounding BCA results, but it is a limited tool. In its most basic form, sensitivity analysis involves changing one variable at a time by a standard percentage, say, plus 10 per cent followed by minus 10 per cent, or by an absolute amount, to gauge how much a NPV changes. If the NPV changes by only a small amount (e.g. ±10 per cent change causes ±2 per cent change in NPV), the implication is that the uncertainly surrounding the variable is not very important and is not critical to decision-making. Conversely, if the affect on NPV is large in percentage terms, the robustness of the BCA can be called into question. It may be worthwhile to expend more resources to obtain a better estimate of the variable, though this will not reduce the risk arising from inherent volatility of the variable.

The variations used for sensitivity tests should be chosen considering the range of plausible values that a variable can take. For example, the Austroads Guide to Project Evaluation (2005a, p. 27) recommends ±50 per cent for generated traffic and ±10 to ±20 per cent for total traffic volume—there is far more uncertainty attached to estimates of generated traffic than to estimates of traffic volume. Hence, when selecting the variables and the amounts of variation for sensitivity tests, judgments about the probability distributions for those variables need to be made.

The amounts by which variables are changed do not have to be symmetrical. Even with the best intentions (unless the probability analysis described below is applied), downside risk may be present. In the Austroads Guide (2005a), the percentage *increases* applied to construction costs are greater than the percentage *decreases*. Also, some variables may be constrained on one side. For example, car occupancy cannot take on values below one.

Table 2.7 shows the Austroads recommended sensitivity ranges for road initiatives. This list is not exhaustive. Other possible sensitivity tests may be appropriate depending on the initiative; for example, where exchange rate movements affect costs, or where implementation of another initiative affects benefits. Some of the values in Table 2.7 can be transferred to rail initiatives.

Table 2.7: Sensitivity variable and ranges recommended by Austroads

VARIABLE	SUGGESTED MINIMUM VALUE	SUGGESTED MAXIMUM VALUE
Capital items[a]		
Concept estimate	–20% of estimate	+20% to 35% of estimate [b]
Detailed costing	–15% of estimate	+15% to 25% of estimate [b]
Final costing	–10% of estimate	+10% to 20% of estimate [b]
Network operation	–10% of estimate	+10% of estimate
Traffic		
Total traffic volume (AADT)	–10% to –20% of estimate	+10% to +20% of estimate
Proportion heavy vehicles	–5 percentage points	+5 percentage points
Average car occupancy	–0.3 from estimate	+0.3 from estimate
Normal traffic growth rate	–2% pa (absolute) from the forecast rate	+2% pa (absolute) to the forecast rate
Traffic generated by specific (uncertain) developments	Zero	As forecast
Traffic diverted or generated by the initiative	–50% of estimate	+50% of estimate
Traffic speed changes	–25% of estimated change in speed	+25% of estimated change in speed
Changes in crash rates	–50% of estimated change	+50% of estimated change

a. The appropriate range for capital costs depends on the detail of investigations, designs and costing. The concept estimate relates to initial pre-feasibility or sketch-planning estimates. The final costing relates to estimates after the final design stage.

b. The range of values relates to different types of initiative. Costing for more routine initiatives (road shape correction, resealing) are generally more accurate than those for larger initiatives (new construction).

Source: Austroads 1996, p. 28; 2005a, p. 27.

Limitations of sensitivity testing include:

- choosing the variables to test and the ranges over which to test involves subjective judgment
- it can easily become a mechanistic exercise that does not encourage genuine exploration of risks or lead to minimisation of downside risk

- only one variable is changed at a time so interactions between variables that reduce or magnify changes in NPV are not observed[22]
- interpretation of the results is subjective, and
- there is no aggregated statistic that sums up the results of a series of sensitivity tests in a single number.

Spreadsheets are ideally suited for conducting sensitivity tests. It is recommended that the list of key variables to be tested be grouped together in an easily accessed part of the spreadsheet (e.g. upper left corner). If proponents submit the spreadsheets containing BCA calculations, government assessors can perform their own sensitivity tests. Some relationships between variables are obtained by computer models outside the spreadsheet, so sensitivity tests cannot be undertaken using the spreadsheet alone. Examples of these variables are operating costs of cars, trucks or trains, estimated using complex computer models that are not easily linked to a spreadsheet. In such cases, the proponents and government assessors can agree on the sensitivity tests to be undertaken.

The Guidelines recommend a large number of parameter values. The standard advice throughout the Guidelines is that proponents are free to depart from these parameters, providing a justification is given, together with a sensitivity test using the recommended parameter value.

For financial analyses, the same sensitivity tests may be carried out as for the BCA.

2.11.2 Risk analysis

The discussion at the start of this section identified two broad types of risk: downside risk and pure risk. BTRE (2005, p. 19) makes the following recommendations:

- Minimise downside risk by ensuring that the benefits and costs are expected values. The state-contingent approach is recommended.
- The required adjustments for pure risk are usually so small that, for practical purposes, it can be ignored. The exception is where the initiative has a large effect on the welfare of individuals; in which case, the recommended approach is to estimate a 'certainty equivalent' value of the benefits.

Downside risk arises from a failure to consider what can go wrong. So if downside risk is to be eliminated, a complete assessment of the possibilities of what can go wrong is needed. The state-contingent approach provides a framework for this. It provides a process that asks a complete set of 'what if?' questions. Sections 2.11.2 to 2.11.9, taken from BTRE (2005), describe how to implement the state-contingent approach. The Austroads Guide to Project Evaluation (2005a) and Austroads (2002) are also useful references.

Concepts used in state-contingent analysis of risk

Outcome

An outcome is the result of a situation involving uncertainty. For example, an initiative failing for technical reasons is an outcome. Absence of failure for technical reasons is another outcome. For forecast demand, an AADT of 11 841.4 vehicles per day is an outcome. Fail/not fail is a discrete variable. An AADT value is a continuous variable.

Event

An event is any collection of outcomes. For example, partial failure and complete failure on technical grounds can be grouped together and defined as an event, 'partial failure or worse'. For a continuous variable, an AADT in the range >8000 to ≤12 000 can be defined as an event.

22 It is possible to undertake two-way sensitivity tests. Results can be presented in a table with changes in one variable as column headings and changes in the other as row headings. However, there are a large number of possible pairs of variables, and only a select few can be presented in this way.

Event space

An *event space* (also known in statistics as a sample space or possibility space) is the set of all the possible events. For technical success of the initiative, the event space might be (fail, partial fail, not fail) or (partial failure or worse, not fail). Outcomes or events may be distinguished by time, for example (fail in Year 1, fail in Year 2, fail in Year 3, …, fail in Year 20, not fail).

As the list of outcomes in an event space is exhaustive, the probabilities of all events in an event space should sum to one. For example, if the event space for the technical success or failure of a piece of infrastructure is (fail, not fail), the probabilities could be (0.01, 0.99).

For a continuous variable such as AADT, the event space could range from zero to infinity. The possibilities for partitioning the event space for a continuous variable into events are limitless. For AADTs, examples are {≤10 000, >10 000} and {≤8000, >8000 to ≤12 000, >12 000}. A continuous variable has a probability distribution associated with it. The probability associated with each event in a continuous event space is measured as the area under the probability distribution between the event boundaries.

State of nature

A *state of nature* is a collection of events selected from different event spaces. For example, the initiative not failing and AADT being greater than 12 000, is a state of nature.

State space

A *state space* is the set of all possible states of nature. If there are n event spaces identified by the subscripts $i=1, 2, 3, \ldots, n$, and each event space contains mi outcomes, the maximum number of possible states of nature will be $m1 \times m2 \times m3 \times \ldots \times mn$. The reason this is a maximum, and not a total, is that some events make the events in other event spaces redundant. For example, if the initiative is a road tunnel and it fails, the different possible AADT outcomes are irrelevant. If one or more event spaces are treated as continuous variables without partitioning into discrete events, the number of states of nature becomes infinite.

States of nature may be identified using an event tree; an example is shown in Figure 2.27.

The probability associated with a particular state of nature is the product of the probabilities of its constituent events. The probabilities of all the possible states of nature must sum to one.

2.11.3 Level of detail

No additions to the corresponding section in Volume 3 are made here.

2.11.4 Sources of risk

The first step in implementing the state-contingent approach is to identify the possible events that will impact on benefits and costs. Events may include equipment breakdowns, adverse weather, technical difficulties, unanticipated environmental or planning requirements, industrial disputes, the population level and so on. Checklists based on experience, examining similar current or previous initiatives, holding a brainstorming session or compiling historical information can assist this process. The Environment Impact Statement should identify environmental risks. Judgment is needed to make decisions about excluding events with small probabilities or small impacts on costs or benefits.

2.11.5 Values of risky variables

No additions to the corresponding section in Volume 3 are made here.

2.11.6 *Probabilities of events*

Probabilities are attached to each event. Analysis of historical data (for example, on flood heights and frequencies, or accidents) or engineering models (for example, of bridge design) may provide objective means to assign probabilities. For most cases, subjective judgment is necessary. The probability attached to each state of nature is calculated by multiplying together the probabilities of the constituent events.

2.11.7 *States of nature and probabilities*

Each state of nature is associated with a unique stream of year-by-year benefits and costs, and these have to be estimated. A BCA is always a comparison between two states of the world: a Base Case and a Project Case. A benefit or cost is the difference between the forecast level of a variable in the Base Case and its forecast level in the Project Case. Either, or both, the Base Case and the Project Case may be different in different states of nature, An example of a situation in which the Base Case differs is where flooding leads to rapid deterioration of an existing road in the Base Case. The benefit from replacing it with a new road, which is less vulnerable to flood damage, is therefore greater in states of nature where floods occur, because the Base Case will be worse.

2.11.8 *Expected values of BCA results*

Expected value of NPV

The final step is to use the probabilities to calculate the expected value of each benefit and cost for each year, and to discount them at the risk-free rate. All benefits and costs for a state of nature should be multiplied by the probability for that state and the results summed. Alternatively, the figures could be discounted first to obtain the NPV under each state of nature, with the expected NPVs derived by multiplying by the probabilities and summing.

Expected values of other BCA results

When calculating expected values of BCRs or IRRs, multiplying by probabilities before, or after, discounting is *not* a matter of choice. Multiplying all benefits and costs of an initiative by a probability has no effect on the BCR or the IRR. These measures must be calculated for each state of nature *first*, and only then can the expected values be derived.

The FYRR criterion can be applied in the presence of risk using the expected values of the capital cost and first-year benefit provided:

- expected values of net benefits in subsequent years do not decline at any stage over the life of the initiative, and
- the probabilities of the states of nature, and the benefits and costs under the various states of nature, do not change as the implementation time changes—this possibility is explored in Section 2.11.10.

If either or both of these conditions do not hold, the expected NPV must be calculated for each year the initiative could be implemented to find the implementation time with the maximum expected NPV. The year of analysis should be the same for all calculations. For example, if the current year is chosen as the year of analysis and the initiative could be constructed in less than one year, implementation in five years time would cause construction costs to be discounted by five years, first-year benefits by six years and so on. Implementation in Year 10 would cause construction costs to be discounted by 10 years, first-year benefits by 11 years and so on.

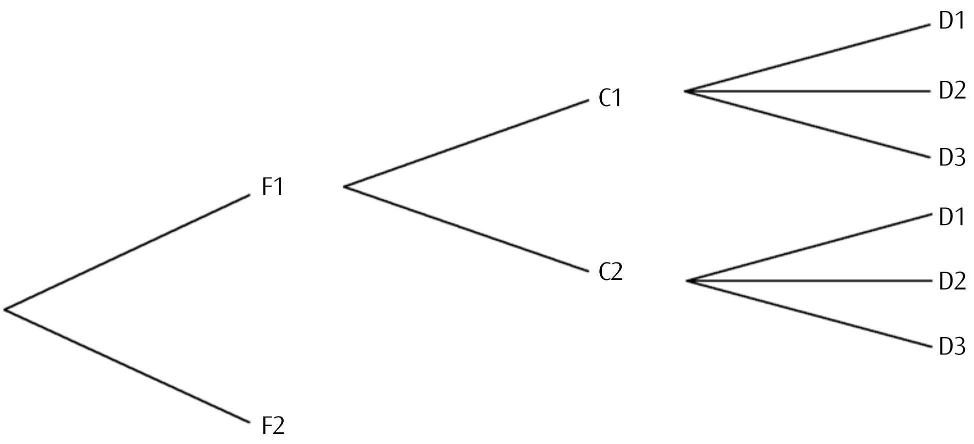
A simple numerical example

Say, there are three event spaces for a transport infrastructure initiative, comprised of the following events:

- the initiative may proceed normally (F1) or fail due to unexpected technical difficulties and be abandoned (F2)
- annual operating costs may be $10 million (C1) or $20 million (C2), or
- annual growth in demand may be 2 per cent (D1), 4 per cent (D2) or zero (D3).

There are 2×2×3 = 12 combinations, but only seven states of nature, because failure of the initiative (F2) renders the other event spaces irrelevant. These combinations could be mapped in an event tree as in Figure 2.27. Events that render other events redundant should be placed in front of, or above, the events they make redundant.

Figure 2.27: An event tree



Source: BTRE 2005, p. 22.

The same tree is set out in tabular form in Table 2.8, with probabilities attached.

Table 2.8: Event tree in tabular form

TECHNICAL FAILURE		ANNUAL OPERATING COSTS		DEMAND GROWTH		PROBABILITY OF STATE OF NATURE
Occurrence	Probability	Occurrence	Probability	Occurrence	Probability	
F1	0.8	C1	0.6	D1	0.6	0.288
				D2	0.2	0.096
				D3	0.2	0.096
		C2	0.4	D1	0.6	0.192
				D2	0.2	0.064
				D3	0.2	0.064
F2	0.2					0.200
Total						1.000

Source: BTRE 2005, p. 22.

Other assumptions are:

- the construction cost is $100 million spread evenly over two years
- the life of the initiative is eight years after completion of construction
- benefits are $40 million in the first year and grow at the same rate as demand, and
- the discount rate is 5 per cent.

Table 2.9 shows calculation of the expected NPV in two ways:

- taking the expectation of the NPVs for all states of nature, and
- taking the expectation of annual net benefits for each year and then discounting them.

The same result, $44 million, is obtained both ways. For the BCR, the expected value is 1.4, obtained by taking the expectation of the BCRs for each state of nature. The expected IRR is problematic because the IRR for the F2 state of nature of negative 200 per cent—loss of the entire capital with no return—distorts the result. The IRR measure cannot be relied on when the pattern of benefits and costs departs from the conventional pattern of early costs during the investment phase, followed by a continuous stream of positive benefits. So the expected IRR is not available in this case. The expected FYRR is 21 per cent = $21m/$98m.

Table 2.9: Calculation of expected NPV and BCR

($ millions)

	C1D1	C1D2	C1D3	C2D1	C2D2	C2D3	F2	EXPECTED VALUES
Year	Annual net benefits							Mean
0	-50	-50	-50	-50	-50	-50	-50	-50
1	-50	-50	-50	-50	-50	-50	-50	-50
2	30	30	30	20	20	20	0	21
3	31	32	30	21	22	20	0	21
4	32	33	30	22	23	20	0	22
5	32	35	30	22	25	20	0	23
6	33	37	30	23	27	20	0	23
7	34	39	30	24	29	20	0	24
8	35	41	30	25	31	20	0	25
9	36	43	30	26	33	20	0	26
Probability	0.288	0.096	0.096	0.192	0.064	0.064	0.200	
NPV	104	122	87	42	60	25	-98	44
NPV × prob	29.9	11.7	8.4	8.1	3.8	1.6	-19.5	44
PV benefits	201	219	185	140	158	123	0	
PV costs	98	98	98	98	98	98	98	
BCR	2.1	2.2	1.9	1.4	1.6	1.3	0.0	
BCR × prob	0.59	0.22	0.18	0.27	0.10	0.08	0.00	1.4
IRR	23.3%	25.3%	21.3%	13.2%	15.8%	10.4%	-200.0%	

Source BTRE 2005, p. 24.

Financial analysis

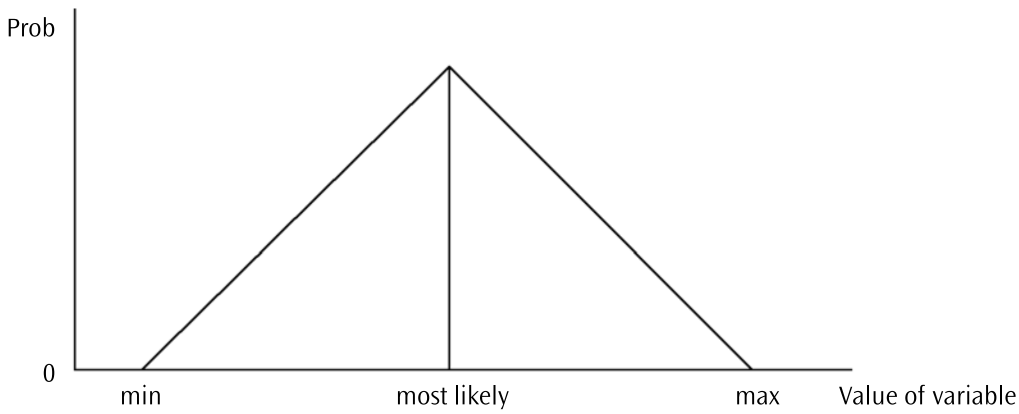
Although the discount rate used for financial analysis (the weighted average cost of capital) includes a risk premium, there is still a need to minimise downside risk. The risk premium in the discount rate compensates lenders and shareholders for bearing the risk that the firm will go bankrupt. It is not a mechanism to offset bias in financial calculations. The set of states of the nature and probabilities developed for the BCA is largely transferable to the financial analysis, the differences being:

- exclusion of event spaces that concern benefits and costs excluded from the financial analysis, for example, externalities
- addition of event spaces that are specific to the financial analysis, for example, a court ruling on a taxation matter, and
- replacement of social with private benefits and costs, for example, replacing user benefits with revenues.

2.11.9 Computer software

Where the number of states of nature or the number of uncertain variables is large, the combinations of input values can become extremely large as well. Computer programs such as @RISK, which links with widely used spreadsheet programs such as Excel and Lotus 123, can facilitate the process. The user can specify probability distributions for continuous event spaces. The triangular distribution, shown in Figure 2.28, is one of simplest forms of probability distributions, obtained by specifying three values: the minimum, the most likely and the maximum value. The 'Trigen' distribution in @RISK requires the user to specify a further two points: the bottom **x** per cent and the top **x** per cent. This ensures that a reasonable proportion of sample points is drawn towards the ends of the distribution.

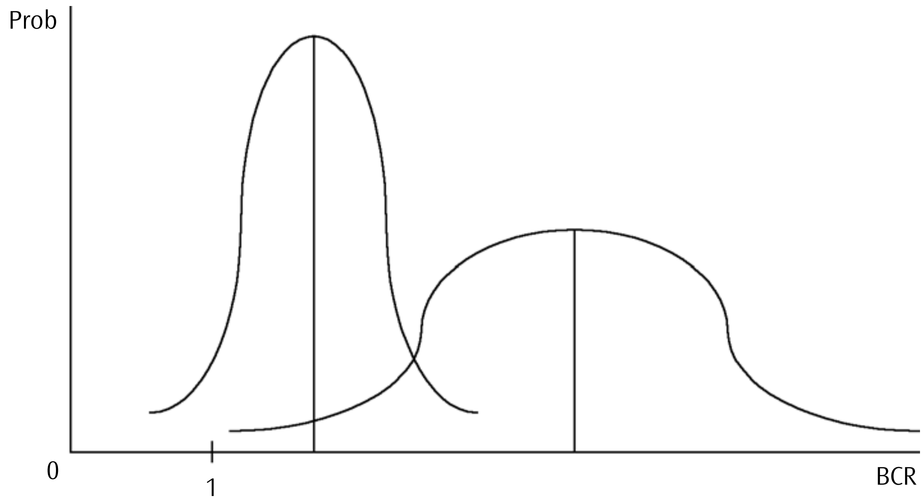
Figure 2.28: Triangular probability distribution as defined in @RISK



Programs such as @RISK use sampling procedures with the number of iterations specified by the user. In a typical iteration, the program draws a random sample value for each uncertain variable, with the sampling dictated by the associated probability distributions. The set of sample values are then inserted into the BCA spreadsheet to find the NPV, BCR and other required results. By repeating the process a large number of times, probability distributions of BCA results are constructed, such as those illustrated in Figure 2.29.

Estimates of the variances of results are also created, together with expected values. However, given that pure risk should be ignored, variances are of no relevance to decisions about whether initiatives should proceed or for ranking initiatives. In Figure 2.29, the initiative with the higher BCR is still preferred even though it has a greater variance. The purpose of risk assessment using the state-contingent approach is to minimise downside risk, not to estimate variances.

Figure 2.29: Probability distribution of benefit–cost ratios



For more discussion on using @RISK in the road appraisal context, including a worked example, see Austroads (2002).

Implementation in practice

The SMT and rapid BCA template require proponents to address a series of questions about risks. Some of these questions relate to risks that the initiative will be delayed. Major cost side (e.g. excess costs) and benefit side (e.g. where benefits are not realised) risks also need to be specified. For the SMT and rapid BCA, it is not necessary to conduct a state-contingent assessment. For detailed BCAs, a comprehensive risk analysis is likely to be warranted and may require use of @RISK.

A risk analysis can be undertaken at a range of levels of detail. The larger the initiative, the greater the level of detail expected. The level of detail may be negotiated between the proponent and the government agency assessing the proposal.

At the highest level of detail, three probability distributions could be specified: benefits, operating costs and investment costs. The next level might distinguish between the major benefits and costs, for example, the level of demand, user cost savings, routine maintenance costs, timing and costs of major rehabilitations and replacements, planning and design costs and construction costs of major stages of the initiative. For a highly detailed risk analysis, separate probability distributions might be specified for some of the following:

- benefits—population projections, income projections, elasticities, user cost savings, generated demand, diverted demand and accident cost savings
- operating costs—hourly rates for labour, hours of labour required for different types of maintenance task, material costs by type of material and frequency of breakdowns, and
- investment costs—planning and design costs, hourly rates for labour, hours of labour required for different tasks, unit costs for particular inputs, quantities of inputs consumed, effects of labour disputes, effects of weather, other causes of delays and unanticipated technical difficulties.

For externalities, the unit values should be central values. Environmental risks should be assessed in the Environmental Impact Statement and, by attaching probabilities to them, can be brought across into the risk analysis for the BCA.

Care needs to be taken to avoid double-counting; for example, if different values of demand elasticity are considered, there is no need to consider generated and diverted demands.

Although the purpose of the analysis is to minimise downside risk, it has to encompass better-than-expected as well as worse-than-expected outcomes. Minimising downside risk does not mean finding the most pessimistic result for the BCA; it means locating the result that is the mean of the probability distribution. Upside and downside risks both need to be assessed in order to estimate the mean.

For each variable, a decision is required about whether to treat it as continuous or discrete. If the different values of the variable relate to clearly defined scenarios, a discrete variable should be used, even where the variable is intrinsically continuous. For example, although population is continuous, if ABS projections are provided as high, medium and low projections, the variable should be treated as discrete.

Wherever possible, the analysis should be based on identifying circumstances (states of nature) in which the variable will take on different values. It can defeat the purpose of the analysis to assume that the initial estimate of the variable is the central value and to construct a symmetrical probability distribution around it, whether discrete or continuous.

There will, of course, be many cases where a symmetrical distribution is justified. Examples include instances where the variable is estimated using statistical or econometric methods that produce unbiased results when correctly applied. It is still worthwhile including these in the risk assessment because, due to non-linearities, the effect on the final BCA results may not be symmetrical.

Certainty equivalent approach

The reasons pure risk can be ignored, in most practical situations, are provided at the beginning of this section. The exception, where the initiative has a large effect on the welfare of affected individuals, is likely to be rare. It is most likely to occur for essential services to isolated communities. An example is an initiative that provides all-weather access along a road that is the only link into a town and the success of the initiative makes a large difference to the livelihoods of residents (for example, ±20 per cent or more of their annual incomes). In this case, the effects of success, or failure, of the initiative on consumption levels are not diversified away by offsetting effects from other initiatives. Furthermore, since the potential changes to incomes are large, residents are likely to have a significant WTP to reduce the risk.

In such situations, the correct approach is to estimate a 'certainty equivalent'. The certainty equivalent of a risky benefit received by an individual is the money amount received with certainty so that the individual is indifferent between the risky benefit and the certain amount. For example, an individual might be indifferent between a 50:50 chance of receiving $10 000 (expected value $5000) and a certain $4000. The latter is the certainty equivalent.

The procedure for estimating a certainty equivalent is described in detail in BTRE (2005). It is not covered in the Guidelines because it is rarely needed. The approach involves assuming a utility function from which to estimate the expected level of utility, and then to convert the expected utility level back into dollars.

2.11.10 Risk management strategies

Risk management can be defined as the process of assessing exposure to risk and determining how best to handle this exposure, with the aim of minimising risk and optimising the risk–benefit balance. Risk management is an ongoing process throughout the entire planning, construction and operation of the initiative. The conclusion that pure risk can reasonably be ignored in most situations has implications for risk management. Alternative risk management strategies can be compared using the state-contingent approach to find the strategy that yields the highest expected NPV. This simplifies comparisons between options having different levels of risks and costs. There is no need to make subjective judgements about the degree of risk to trade-off against net benefit.

In the numerical example in Table 2.8 demonstrating the state-contingent approach, there was a 20 per cent probability of technical failure of the initiative. In practice, such a large chance of technical failure is unacceptable for a major infrastructure initiative. Ways to reduce the probability of failure should be found, even though the cost of the initiative would increase. For example, a bridge could be built to have greater strength, or, if the initiative was a tunnel, more extensive geological studies could be undertaken to better understand the risks and to devise actions to reduce the probability of major technical difficulties during construction.

Risk management is usually thought of in terms of undesirable outcomes, both in terms of the probabilities and the costs imposed. However, it is possible for an initiative to be over-designed, that is, where it is preferable to accept an increased risk of an undesirable outcome in exchange for a cost saving.

The twin aims of risk management are risk minimisation and optimising the risk–benefit balance. Risk minimisation means taking any actions that reduce the probability or the cost of undesirable outcomes, where actions involve little or no cost. Risks also have to be minimised where required by law or by community expectations. Attaining the optimal risk–benefit balance requires determination of how much net benefit to sacrifice in order to reduce risk. In most cases, attaining the optimal risk–benefit balance is a matter of determining the option with the highest expected NPV.

Table 2.10 shows two options with different risk–benefit trade-offs. Spending an additional $50 million reduces the chance of technical failure from 20 per cent to five per cent and generates a net gain in expected NPV of $25 million after reducing the NPVs for both states of nature by $50 million. The gain comes because of the change in probabilities.

Table 2.10: Comparisons of risk options using expected NPV: changing probabilities

($ millions)

	NPV	PROB	PROB × NPV	EXPECTED NPV
		Without extra spending to reduce risk		
Pass	$200	0.8	$160	
Fail	–$300	0.2	–$60	
				$100
		With an extra $50m spent to reduce risk		
Pass	$150	0.95	$142.5	
Fail	–$350	0.05	–$17.5	
				$125

Another option might leave the probabilities unchanged, but reduce the cost of failure. This is the case for measures taken to lessen damage in the event of a disaster. In Table 2.11, spending an additional $50 million reduces the NPV of failure to zero, so that the NPV in this state of nature becomes –$50 million after allowing for the additional spending. The net benefit is an increase of $10 million in the expected NPV. The net gain from spending to reduce the cost of failure is given by

(probability of failure × present value of benefit) – present value of cost.

This amount has to be positive for the investment to be worthwhile. The benefit, that is, the reduction in cost to society in the event of failure, is only realised in the event of failure and so is multiplied by the probability. The cost is incurred in all states of nature and is multiplied by unity. In terms of the example in Table 2.11, the net benefit is $10 million = 0.2 × $300 million – $50 million.

Table 2.11: Comparison of risk options using expected NPV: reducing costs of failure

($ millions)

NPV		PROB	PROB × NPV	EXPECTED NPV
Without extra spending to reduce cost of failure				
Pass	$200	0.8	$160	
Fail	–$300	0.2	–$60	
				$100
With an extra $50m spent to cost of failure				
Pass	$150	0.8	$120	
Fail	–$50	0.2	–$10	
				$110

Risk–benefit trade-offs should be considered in cases where the certainty equivalent is required, that is, where the particular risks could be associated with large changes in the welfare of a small number of individuals. In these rare situations, the objective is to maximise the certainty equivalent of the NPV.

Dividing the initiative into stages can also be a risk management strategy. It can reduce risks by deferring commitment of part of the total investment until uncertain developments have played out, and by 'testing the waters' before committing to the entire investment. Staged options can be evaluated in the same way as other options with differing risks, by estimating expected NPVs.

Deferring the initiative

Deferring the initiative is one of the many strategies available for managing risk.

Probabilities of states of nature can change with implementation in situations where the costs or benefits of an initiative are affected by an uncertain one-off event in the near future. There may be the option to defer the initiative until after the outcome of the one-off event is known and the uncertainty is removed. The possibility of a negative outcome for the initiative is thereby avoided. A wait-and-see policy might be worthwhile when benefits or costs are highly dependent on a future decision by government to approve, or by the private sector to undertake, a significant residential or industrial development. Other uncertain future events that could be worth waiting for include decisions to proceed with other initiatives in a network (network risk) and possible significant changes in the overall level of economic activity or in exchange rates.

The wait-and-see option is preferable if it is associated with a higher expected NPV than the 'build-now' option, calculated from a common year of analysis. For example, say, forecast traffic along a road will be higher if a residential or industrial development proceeds. Without the additional traffic caused by the development, the initiative has a negative NPV. If it is uncertain whether or not the new development will occur, a probability could be assigned to the development, say, 50 per cent, creating a branching of the event tree. Under the wait-and-see approach, the implementation decision would not be made until it was known with certainty whether the development proceeds. The possibility of implementing an initiative with a negative NPV is therefore avoided altogether.

The way to analyse this situation is to first estimate the expected NPV in the usual way for the build-now option. It is necessary to assume a time at which the decision about the development will be known—say, five years into the future. There is a 50:50 chance of high benefits (positive NPV) with the development and low benefits (negative NPV) without it; the high benefits would commence sometime after the decision is made. The expected NPV of the build-now option is

0.5 × NPV with the development + 0.5 × NPV without the development.

A second expected NPV calculation should be made assuming the initiative is implemented in five years time and the development takes place with certainty. Deferring the initiative by five years saves capital costs via discounting, and eliminates the 50 per cent chance of a negative NPV should the development not proceed. Benefits accruing from the initiative during the first five years, however, are lost. From the standpoint of today, the expected NPV of the deferment option is

0.5 × NPV with deferral and with the development + 0.5 × zero.

Comparison of the two expected NPVs shows whether the deferral option is better than the build-now option.

It should be noted that the NPV calculations all have to be made from the same year of analysis.

The deferral option may impose some additional costs not incurred under the build-now option. For example, if it is required to maintain the option, the land may have to be purchased now instead of in five years time. A cheaper alternative might be to pay the owner an amount in exchange for an option to purchase land in five years time. The amount of the payment would not enter into the BCA because it is a transfer, but any costs from the restricted use of the land, because of the risk of resumption in five years time, should be counted. Costs of keeping the deferral option open should be built into the calculation, so the expected NPV of the deferment option becomes

0.5 × NPV with deferral and with the development + 0.5 × zero + option costs.

Like the 'real options' approach used in finance, a monetary value can be derived for the option of deferral. Assuming deferral is preferable in the absence of any option cost, the maximum acceptable option cost is the amount by which the expected NPV, with deferral, exceeds the expected NPV by building now. This could be compared with the actual option cost to decide whether deferral is worthwhile.

2.12 Adjusted benefit–cost analysis

2.12.1 Use of adjusted BCA

The BCA methodology aims to maximise the economic efficiency objective. It recognises a number of other objectives such as safety and environment, but only in so far as they are consistent with economic efficiency. The equity objective is not taken into account. The adjusted BCA methodology is a formal and transparent way to re-weight or incorporate non-efficiency objectives.

It is a hybrid of multi-criteria analysis and BCA that retains the dollar measurement of BCA. Adjusted BCA is not an essential component of the Guidelines appraisal framework. It is included in the Guidelines as an option should governments wish to introduce into the process a more formal way of ranking initiatives according to predetermined weights for objectives.

Both rapid and detailed BCAs may be converted into adjusted BCAs.

A simple example of an adjusted BCA is the economic evaluation software packages developed by some road agencies that produce an additional set of BCA results, with car passenger time accorded a zero value. By highlighting the initiatives that are particularly beneficial for freight and those that are less beneficial for freight, the analysis can provide useful information for a government that wants to place a high priority on freight transport relative to passenger transport.

With the exception of impacts of a one-off nature and the distributional multiplier, the same specified values and weights should be employed for all initiatives or options assessed in the same period by a jurisdiction. The weights used to calculate the distributional multiplier (explained below) should be the same. It would defeat the purpose of the analysis to have weights that vary between initiatives or options.

The exception is a subjectively determined monetary value assigned to a one-off benefit or cost, specific to the initiative being appraised and that is impossible to value in money terms. Even so, some degree of consistency could be achieved by selecting monetary values with reference to previous decisions.

A weight or a nominated value for a benefit or cost used in an adjusted BCA has to represent the value judgments of the government. Clearly, ministers, political advisers and senior officials have to own any process to reveal their government's preferred weights. Various government policy and strategy documents and statements may provide broad guidance as to the objectives and impacts that might be weighted, and the sizes of the weights.

In selecting weights, the government needs to be aware of the practical consequences for ranking of initiatives. Sets of weights and nominated values can be tested on appraisals undertaken in the recent past to gauge the effects. Ideally, the weights and values should be revised in each period in the light of the experience in previous periods, and to reflect any changes in the government's revealed objectives and priorities.

The implicit weights preferred by the government could be derived for an existing set of initiatives by undertaking an iterative process. By varying the weights until the resulting rankings closely match the desired rankings, in turn, the value judgments of the government will be reflected. Although there is a loss of transparency because weights do not have to be nominated beforehand, this approach provides consistency because all initiatives are subject to the same set of weights. In addition, extraneous considerations are not allowed to intrude.

Arguments for and against adjusted BCA

BCA aims to be as comprehensive as possible in identifying and measuring benefits and costs. Impacts of initiatives that are not priced by markets are, as far as practical, converted into monetary amounts. Considerable research has been undertaken into valuing environment and safety impacts. The values attached to these impacts are intended to represent the WTP of consumers. Hence, BCA takes account of safety and environmental objectives, but only in so far as they are consistent with economic efficiency.

There are three reasons why BCA, by itself, might be considered too narrow to be the sole basis for prioritising investment initiatives:

1. BCA values benefits and costs on the basis of preferences of individuals (WTP), which the government may, in some cases, regard as inappropriate.
2. BCA omits certain impacts because they cannot be expressed in monetary terms.
3. BCA ignores social equity, as it takes no account of how the benefits and costs are distributed among members of society.

Actual prioritisation of initiatives involves subjective consideration of all relevant factors of which the BCA is only one, albeit an important one. Systems of multi-criteria analysis that use quantitative procedures to rank initiatives are an attempt to improve subjective decision-making by:

- ensuring that all relevant factors are considered and that irrelevant factors are excluded
- increasing transparency by requiring decision-makers to be explicit about the objectives, weights attached to objectives and assessments of initiatives against objectives, and
- providing a degree of consistency in decision-making because the same procedure and set of weights are applied to each initiative, and can be used again in subsequent rounds of decision-making.

It is common to combine BCA and multi-criteria analysis by treating the result of the BCA as one of the criteria in the multi-criteria analysis.

In common with multi-criteria analysis, the adjusted BCA approach allows subjective judgments to be incorporated in a quantitative manner, improving transparency, rigour and consistency across decisions.

Adjusted BCA has three advantages over quantitative forms of multi-criteria analysis that involve attaching scores to impacts and weights to objectives:

1. Retention of the monetary measuring rod of BCA should add to transparency. For example, a subjectively set value of $30 a tonne multiplied by 10 000 tonnes of greenhouse gas emissions is far more meaningful than a score of 50 on a zero to 100 scale multiplied by a weight of 0.1.
2. By keeping within the BCA framework, adjusted BCA protects against some of the pitfalls of poor application of multi-criteria analysis such as double-counting, insufficient disaggregation of impacts and failure to account for non-linear dose-response relationships.
3. Via the discounting mechanism, adjusted BCA is able to distinguish in a rigorous manner between impacts occurring at different times.

The main argument against adjusted BCA is that it 'distorts' the results of BCAs in such a way that it can give less economically efficient initiatives precedence over more efficient initiatives. It could lead to some highly wasteful initiatives being implemented. However, this is the desired result if it reflects government directions. As a safeguard, adjusted BCA results should never be reported separately from the corresponding unadjusted BCA. This ensures that the potential efficiency losses from decisions based on adjusted BCA results are transparent.

2.12.2 Adjustments

There are four categories of adjustments that can be made:

1. parameter replacement
2. weighting
3. inserting values for omitted benefits and costs, and
4. distributional weighting.

2.12.3 Parameter replacement

If the government wants to place higher values on, for example, safety or environmental concerns, it can arbitrarily set a higher unit value; for example, a higher value for the statistical value of life. To tilt the investment program in favour of providing benefits to freight and business travel, a value of zero could be attributed to time savings to private motorists.

2.12.4 Weighting

The same effect can be achieved by applying weights to particular benefits and costs: greater than one to give the impact greater weight, and less than one to give the impact less weight. For example, if the government wants to tilt the investment program in favour of the environment, environmental benefits and costs might be multiplied by a factor of, say, 1.5. To favour freight, benefits accruing to freight could be multiplied by a factor greater than one, or benefits to private cars multiplied by a factor less than one.

2.12.5 Inserting values for omitted benefits and costs

For impacts excluded from BCA on the grounds that it is too difficult to arrive at satisfactory estimates of consumers' valuations (intangibles), the government could simply assign a monetary amount. For example, the New Zealand *Project Evaluation Manual* (Transfund 1997) nominates NZ$30 per tonne of carbon dioxide as the cost of greenhouse gas emissions. The value should be expressed as a total amount where the impact cannot be quantified in physical units; for example, the aesthetic effects of an initiative.

For impacts where no guidance is available, the Transfund manual recommends adjusting the BCR and back-calculating to obtain the implied dollar value placed on the impact. As a safeguard, the manual recommends that the total value of intangible benefits not exceed 10 per cent of the benefits included in the tangible BCR.

2.12.6 Distributional weighting

BCA adds together benefits and costs regardless of to whom they accrue. Yet all initiatives have distribution impacts and there will inevitably be gainers and losers. In the transport context, equity tends to be seen more in terms of geographical location than levels of income and wealth. See Section 1.2 for a discussion of equity in transport planning.

An approach to addressing equity objectives suggested in the BCA literature is to weight benefits and costs according to the affected group. Hence, benefits and costs accruing to individuals in less populated areas could be given a higher weight than those accruing to individuals in more populated areas. This would introduce a systematic bias in favour of initiatives in regional areas.

A major difficulty with the weighting approach is that it can be difficult to estimate how the benefits and costs from initiatives are distributed. Depending on levels of competition, benefits to industries may be passed down the production chain and may ultimately be spread widely, but thinly, throughout the whole economy. For example, an initiative that improves international competitiveness will cause the exchange rate to be stronger than it would otherwise be, making imports cheaper for all Australians. An initiative that benefits an industry in one region could harm competing producers in other regions. Benefits from improved infrastructure accrue to both local traffic and through traffic having disparate origins and destinations, spreading benefits widely.

A simple and practical approach is for the analyst to make a subjective judgment about how the benefits are distributed among a small number of broadly defined groups. For example, it might be judged that the split of benefits for a particular initiative is 50 per cent to regional areas, 30 per cent to urban areas and 20 per cent nationwide. If the weights adopted for the three groups were 1.5, 0.5 and 1.0 respectively, then the benefits from this initiative would be multiplied by a factor of 1.1 (= 50 per cent × 1.5 + 30 per cent × 0.5 + 20 per cent × 1.0). Thus, initiatives with a higher proportion of benefits accruing to regional areas as compared with urban areas could be advantaged in a consistent and transparent manner.

2.12.7 Distributional multiplier

No additions to the corresponding section in Volume 3 are made here.

2.12.8 Calculate adjusted NPV and BCR and report results

No additions to the corresponding section in Volume 3 are made here.

2.13 Application to road initiatives

2.13.1 Models and assumptions

No additions to the corresponding section in Volume 3 are made here.

2.14 Application to rail initiatives

2.14.1 Models and assumptions

No additions to the corresponding section in Volume 3 are made here.

2.14.2 Pricing and assumptions

No additions to the corresponding section in Volume 3 are made here.

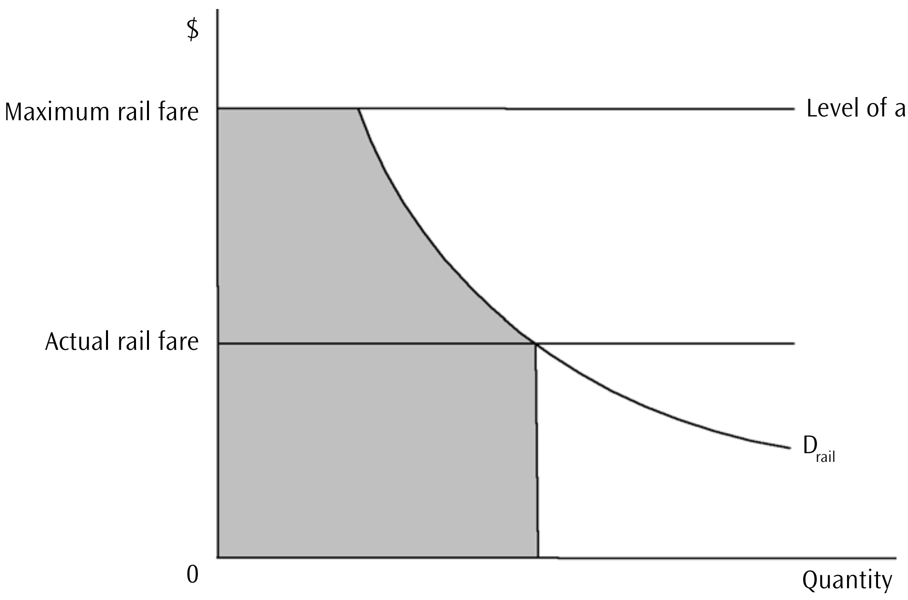
2.15 Application to new services where traffic is diverted

For a totally new transport service, the benefit is the entire WTP area under the demand curve, less social costs. Measurement of the benefit requires knowledge of the demand curve over a substantial part of its length—information that is extremely difficult to obtain. In a mature economy with a well-developed transport network, most new services will already have one or more fairly close substitutes. For example, if the initiative is a new railway line, usually road links already exist joining the centres to be served by the new service. In the road network, construction of a new link draws traffic from longer, or more congested, existing routes. The price and quality attributes of sources of diverted traffic for the new service can provide information about the value diverted users place on the new service.

For demand *generated* by the new service (as distinguished from demand that is *diverted*), prices and qualities of existing substitute services can provide information about the value of the service to users, because the existing services are not sufficiently valuable to new users to stimulate the demand generated by the new service. Figure 2.30 provides a simple example. For a BCA of a proposed fast passenger train service between Sydney and Melbourne, the estimated demand function (D_{rail}) is asymptotic to the price axis. It is considered that no-one will pay more than the airfare for the same journey, because the air alternative is faster. The demand function is truncated above the airfare and the WTP benefit estimated as the shaded area in Figure 2.30.

For further discussion about estimating benefits of generated demand where there is an alternative service, see Harvey (2002). This reference also provides more detail on the methodology put forward in the corresponding section in Volume 3.

Figure 2.30: Benefit from a new fast train passenger service



2.16 Flood immunity projects

Initiatives that improve flood immunity of road or rail infrastructure reduce the frequency and duration of closures due to floods, and the damage to the infrastructure. Benefits accrue to transport users, producers and consumers dependent on transport, and to road and rail infrastructure providers. There may be some negative side-effects if the initiatives cause floodwaters to spread more or permit increased travel by heavy vehicles on saturated pavements.

Benefits discussed here are reductions in:

- waiting time
- diversion to alternative routes
- returning to the point of origin
- decisions not to travel or to delay travel
- use of air transport
- loss of perishable goods
- lost production
- demurrage of ships
- size of inventories
- wash away of transport infrastructure
- other short-term damage to transport infrastructure, and
- long-term damage to transport infrastructure.

The list is not exhaustive. In deciding whether to add in other benefits, the usual cautions about double-counting apply. For example, if the cost of car users' time is valued at the standard rates per hour for work and non-work travel, then it is double-counting to include losses to tourist operators and other businesses at destinations.

Primary information requirements for appraisal of initiatives that improve flood immunity of road or rail infrastructure are estimates of:

- number of times the infrastructure is closed per year and the average duration of closure in days or hours, and
- for each day or hour of closure of the road or rail line, the volume of traffic or trains with information about their characteristics e.g. vehicles (proportions of cars and trucks by type), train characteristics (numbers and types of engines and wagons, tonnages, passenger numbers), freight carried (quantities and types).

Most of the material in this section is written to apply to road initiatives. However, all the benefits and costs discussed can apply equally to initiatives that improve flood immunity of rail infrastructure, though opportunities for 'diversion to alternative routes' and 'returning to the point of origin' are rare for rail initiatives. The last section addresses aspects that apply to only rail initiatives.

Road closure could occur at different flood heights for different vehicle types since trucks can pass through deeper water than cars. The Base Case (and possibly the Project Case) may have to be split into two or more components applying over different ranges of flood heights. If a risk analysis is carried out, as proposed in Section 2.11, alternative scenarios with different flood heights with associated probabilities could be examined.

Where road closure is regular and expected, as is the case in some far northern areas of Australia during the wet season, residents and businesses are usually prepared for it, mitigating many of the costs that would otherwise be imposed. In such cases, it is important to use seasonal traffic flow data because many road users will have decided not to travel during the wet season.

Initiatives that improve flood immunity may be interrelated in the sense that a link, corridor or network could be cut at a number of points simultaneously. Depending on the spread of the rain leading up to the flood, a number of rivers in a district or region could flood at once. Closures could move from one crossing to another where a road runs parallel to a river and the flood moves downstream. The duration of closure of the link, taken as whole, is from the beginning of the first individual closure until the end of the last. Section 2.1.11 discusses interrelationships between initiatives. Where initiatives are this closely related, it may be necessary to appraise them as a single combined initiative instead of, or in addition to, appraisal as individual initiatives.[23]

2.16.1 Waiting

The cost per hour or day of waiting could be estimated using the standard values of time (see Austroads 2005c and 2006). For cars, only daylight hours should be counted on the grounds that relatively fewer cars travel at night. There may also be some accommodation costs. For freight vehicles, the cost of crew time is included in the Austroads hourly costs. Separate time values for the freight carried are provided by Austroads on a per vehicle basis, which can be used for general non-bulk freight. Lower values may be appropriate for non-perishable bulk freight because it normally has a relatively low value-to-weight ratio.

For perishable freight and livestock, information should be collected on the value of the freight and the consequences of different delay times e.g. loss of value and costs incurred to preserve the freight.

For each hour of road closure, assumptions need to be made about the proportions of traffic that will:

- wait
- divert to alternative routes, where alternatives exist
- turn back, and
- not travel or wait at their points of origin.

These estimates are required for the Base Case and possibly for the Project Case if there remains a reasonable probability of closure when the initiative is implemented.

Other basic information requirements include:

- AADT, vehicle composition and hourly volume distribution
- the cost to travel by alternative route(s) by vehicle type, and
- the cost per hour or day of waiting by vehicle type.

Waiting traffic accumulates. Say, the maximum waiting time was 24 hours. Traffic arriving in the first hour would wait an average of 23.5 hours, traffic in the second hour, 22.5 hours and so. Unless the hourly volume distribution is close to uniform, it cannot be assumed that the average waiting time is half the maximum. If, on the hand, the road is closed for a matter of days, the number of arrivals per day should fall as more travellers decide not to travel.

2.16.2 Diversion to alternative routes

In some situations, alternative routes are available. The availability of alternative routes could vary with flood height. Floods above a certain height may close off the alternative route, causing another alternative route to be used or eliminating the possibility of diversion altogether. This may necessitate splitting the Base Case (and possibly the Project Case) into two or more components, applied over different ranges of flood heights.

23 Appraisal of potential interrelated flood immunity projects together as a rapid appraisal could help to determine a target flood immunity level for a corridor, as part of the economic assessment discussed in Section 1.4 in Volume 5.

In deciding whether to wait or divert to a longer alternative route, travellers compare the generalised costs. The longer the expected wait time, the more costs travellers will be willing to incur by diverting. If it is assumed that the media provides travellers with reasonably accurate forecasts of waiting time, given the costs of waiting and the costs of using an alternative route, the maximum number of hours and days waiting can be found by dividing the additional cost of taking the alternative route, over and above the cost via the primary route, by the cost per hour or day of waiting. Say, for a given class of vehicle, the alternative route costs an additional $600 to use, while the cost per day waiting is $300. A traveller will choose to use the alternative route only if the waiting time is greater than two days. Most traffic arriving two or more days before the road reopens will divert to the alternative route, and most traffic arriving within two days of reopening will wait. A simple approach is to assume that *all* traffic within each vehicle class arriving before the critical time for that class diverts and *all* traffic arriving after waits. A more sophisticated approach is to model the decision for each time period (say, hourly) using a logit curve with the wait/divert proportion a function of the cost difference (see the discussions on logit curves in Section 2.4.5 in Volume 5, Part 2 and in Volume 5, Part 3).

2.16.3 Returning to the point of origin

Where no viable alternative route is available, some private cars and freight vehicles may return to their point of origin. To estimate the costs involved requires knowledge or assumptions about the origins of the vehicles affected. The decision about whether a vehicle returns to its point of origin could be modelled by comparing the generalised cost of waiting with the generalised cost of returning to the origin and subsequently returning to the point where the road is cut after the road has reopened. The cost is the combined generalised cost of the two trips, less the cost of the waiting time saved. A logit curve could be used to estimate the proportion of traffic turning back given the waiting time, since the proportion of vehicles returning would increase with the expected waiting time.

Traffic *en route* before the road is closed may turn back before it reaches the point of closure. The generalised cost of continuing the journey and waiting can be compared with the generalised cost of returning to the origin and recommencing the trip later.

2.16.4 Decisions not to travel or to delay travel

The proportion of road users deciding not to travel is likely to rise with the duration of the road closure. A longer road closure gives more potential travellers time to reverse their decision to travel.

Where no viable alternative route is available, traffic due to depart from its point of origin after the closure occurs (and possibly some time before, if the closure is anticipated), and which arrives at the closure point prior to reopening, would delay departure to avoid waiting, or not travel at all. Say, the journey from the point of origin to the break in the road is 48 hours and closure occurs at midday on 20 February, reopening at midday on 25 February. On the assumption that the media provides accurate forecasts of closure and reopening times, then the traffic due to depart between midday 20 February and midday 23 February has the option of waiting at the point of origin, where waiting costs are likely to be lower, or not travelling at all. There may be alternative work for the trucks and drivers. Perishable freight may be able to be stored. Motorists may be able to delay their holiday plans. To allow for waiting at points of origin, information is required on the origins of the traffic and its characteristics to estimate costs.

A decision not to travel imposes costs. The cost on each individual affected must range between zero and the lower of the generalised cost they would incur by waiting or diverting. In many cases, the trip would be put off to another time or an alternative destination substituted with little cost imposed on the traveller. A practical approach is to assume a uniform distribution of costs between zero and the lower of the costs of waiting and diverting. The mid-point is taken to represent the average cost of not travelling.

An exception is perishable crops that must be picked and transported to markets within a short period around the time of ripening. Mangoes are an example of this type of crop. Estimation of costs of lost production is discussed below.

2.16.5 Use of air transport

Airlifts may be required for medical evacuations and to bring in essential supplies if stocks run out. Research into past floods could provide information on typical lengths of time before an airlift is necessary and the frequencies and costs of the airlifts. For medical evacuations, it is important to distinguish between emergencies for which a medical evacuation by air is necessary regardless of the road closure, and those of a less urgent nature that can be attributed to the closure. For airlifts of supplies, the costs of the supplies themselves should be omitted because the supplies are required in the absence of the road closure.

2.16.6 Lost production

Agricultural produce still on the farm, or being processed, can lose value or perish altogether. If there is a loss of value because harvesting or delivery is delayed, the cost is the reduction in market price. If the produce is lost altogether, the cost is the market price plus any disposal costs, less all resource costs saved by not selling the goods (e.g. harvesting, processing, packaging, transport).

The same principle can be applied for lost production by manufacturing, mining and service industries. The cost to society is the value of the output lost, less all resource costs saved by not producing the output and delivering it to customers.

Note that this applies only to genuinely lost production. If the harvesting of a crop or provision of a service is delayed, there may be very little cost imposed. Overtime payment to labour could be the only cost.

If individuals are unable to travel to work, the cost is earnings before income and payroll tax, less travel and other presentation costs saved. The value of the leisure time gained by the individual should be deducted as well, but is normally too difficult to value.

2.16.7 Demurrage of ships

If deferral of delivery of mining or agricultural commodities to ports leads to ships waiting, the demurrage charge can be used as a measure of the cost.

2.16.8 Size of inventories

Households and businesses in remote areas subject to periodic isolation due to flooding need to hold higher levels of stores of food, household supplies and inventories. A reduction in the frequencies and durations of periods of isolation will reduce the sizes of stores that need to be held. Given the difficulty of monetising this benefit and its relatively small size, it can reasonably be left as a non-monetised benefit.

2.16.9 Wash away of transport infrastructure

Flood waters may wash away infrastructure such as pavements, culverts and bridges. Rushing water can also undermine the foundations of bridges and embankments. Lower-level floods may not cause any damage. Hence, an estimate has to be made of the threshold flood height at which a washout may occur. Benefits from initiatives that reduce washouts will be realised only for floods above the threshold height. There may be more than one threshold for greater levels of damage associated with higher floods. When estimating benefits from damage avoided, remember that replacement of damaged transport infrastructure is likely to be more expensive in remote areas.

Road closure can be extended beyond the time when the flood level drops while the road is being repaired. Increased closure time will affect the value of the time-dependent benefits of flood alleviation initiatives discussed above.

2.16.10 Other short-term damage to transport infrastructure

Pavements are particularly susceptible to damage if used by heavy vehicles straight after being inundated. Potholing and rutting may occur after inundation, requiring repairs. Drains may need to be cleared. Roadside furniture (signs, guardrails) may have been damaged or washed away. Earth and rocks may need to be moved from roads.

2.16.11 Long-term damage to transport infrastructure

Heavy traffic on an inundated pavement will damage sub-grades, causing time for pavement rehabilitation or reconstruction to be brought forward. The cost of pavement damage (benefit of flood alleviation) can be estimated by projecting forward maintenance, rehabilitation and reconstruction costs with and without the flood alleviation initiatives and taking the difference in the present values at the discount rate being used for the BCA (see Section 2.17 for a discussion of road maintenance costs).

In the Base Case, there may be damage costs for pavements on routes to which traffic diverts.

2.16.12 Negative side-effects

Construction of causeways to raise the level of roads can impede flows of floodwaters causing the flood to spread over a greater area. There could be costs from damage to property and other transport infrastructure, and loss of agricultural production. Valuation of the damage to transport infrastructure and lost agricultural production has already been addressed in this section, but in the case of negative side-effects the damage and losses are greater in the Project Case.

Closure of a road due to flooding can *save* substantial repair costs on long lengths of road by keeping heavy vehicles off inundated or saturated pavements that are highly susceptible to damage. A flood immunity initiative that opens up damage-prone roads to heavy vehicles can generate costs (negative benefits), which need to be deducted from the Project Case benefits.

2.16.13 Effects of differences in duration of road closure

Most of the costs of flooding mentioned above rise at a less than proportionate rate with flood duration. The longer the waiting time, the more transport users have opportunities to chose less expensive options such as using alternative routes, turning around and not travelling. Some costs, such as loss of production and use of air transport, could rise more than proportionately.

To the extent that the daily cost is different for each day of road closure, an assumption that all flood closures are of average duration could lead to significant errors in estimating benefits. If this is the case, a probability distribution of flood durations should be assumed. The distribution is skewed in the direction of below average durations.

2.16.14 Rail initiatives

When a railway line is blocked by floodwaters or washed away, there are rarely alternative routes or opportunities for trains to return to their point of origin. Passengers on delayed trains might be carried by bus around the blocked track to a train on the other side of the flood water or carried by bus for the rest of their journey. This saves on waiting time and some resource costs for the remainder of the train journey, but there are additional costs of accommodation and bus transport.

Where a large part of rail freight has to pass through a single terminal at its destination, there could be further delays if the terminal becomes congested.

2.17 Maintenance initiatives

2.17.1 Capital versus maintenance expenditures

The distinction between capital and maintenance expenditures is not always clear. Capital expenditures create new infrastructure or raise the standard of existing infrastructure above its initial standard. Maintenance expenditures either restore an existing asset that has deteriorated, part or all of the way up to its initial standard, or slow the rate of future deterioration. The distinction becomes blurred when an asset has deteriorated to a very low level and undergoes a major rehabilitation or complete replacement. If the rehabilitated asset, or new asset, is of the same standard or capacity, the expenditure is categorised as maintenance under this definition. In practice, the accounting convention is to class it as capital.

This section of the Guidelines sets out a methodology for evaluating maintenance initiatives for a seamless transition to the methodology for evaluating capital initiatives.

2.17.2 Rail maintenance and renewals

Undesirable incentives, which could lead to under-investment and below-optimum operating conditions, may result from governments funding rail maintenance and renewals.

An infrastructure operator with commercial objectives has no incentive to spend on maintenance if the government agrees to pay for it. There are many situations where capital expenditures and other operating expenses are substitutable for maintenance costs.

Firstly, there is the capital cost/maintenance trade-off for the infrastructure itself— higher quality infrastructure has lower maintenance costs and, conversely, lower quality infrastructure has higher maintenance costs.

Secondly, there are trade-offs between track maintenance and rollingstock operating and maintenance costs. Optimising the track–wheel interface involves complex engineering and economic considerations. Wheels have to be ground to maintain a profile that minimises damage to tracks. Tracks need grinding to ensure optimal contact between wheels and tracks. Failure to do so can be damaging to both wheels and tracks. Higher train speeds place greater strain on tracks and the strain will be greater if the lateral alignment of tracks is not maintained. With government funding of rail maintenance, there are incentives for trains to be operated in ways that maximise profitability without taking account of damage to tracks, and for tracks to be over-maintained to save on maintenance costs for rollingstock and allow higher train speeds.

Where tracks reach the end of their economic lives, the replacement infrastructure usually has a significant capital component. For example, timber sleepers are usually replaced with concrete sleepers and heavier rail weights are used. Section 2.17.7 in Volume 3 discusses how to appraise initiatives that combine capital and maintenance elements.

Government funding of rail capital initiatives can create incentives to over-invest in infrastructure quality to save on maintenance costs. The government agency providing funding for capital initiatives should be aware of this during the identification and appraisal phases of initiatives.

2.17.3 Road maintenance

The types of road maintenance works required varies with the type of road: unsealed, sealed or concrete. This section of the Guidelines concentrates on sealed pavements because they account for most of the length of main roads in Australia. Concrete pavements are a low maintenance option, requiring some routine maintenance and occasional roughening and resealing of joints. They are long-lived, but, being a relatively recent phenomenon, it is uncertain just how long-lived.

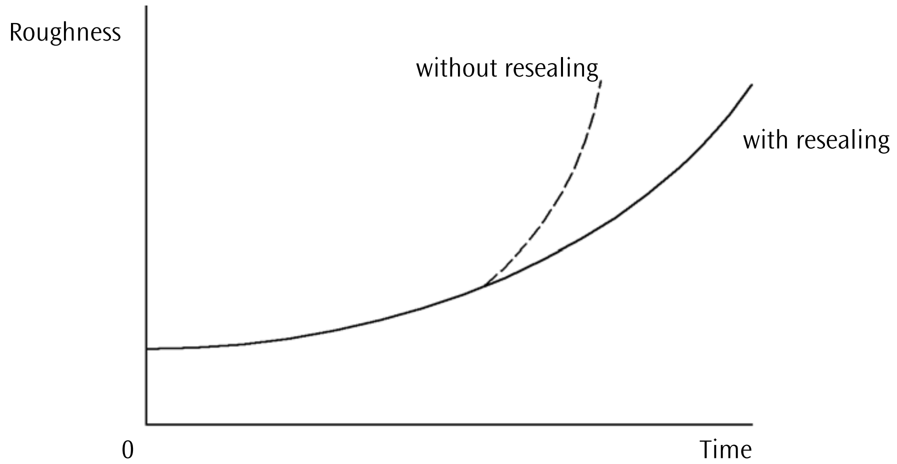
Road maintenance falls broadly in two categories:

1. routine—cutting grass beside roads; repairing and replacing signs, guideposts and barriers; keeping drains clear; repairing potholes and patching, and
2. periodic—resealing and rehabilitation.

Resealing for flexible pavements involves spraying bitumen over the pavement surface and spreading a thin layer of crushed rock. The purpose is to seal up cracks to prevent water from entering the sub-grade material and weakening the pavement. It does not significantly reduce road roughness. For asphaltic concrete pavements, resealing involves applying an asphalt overlay. Resealing is required about every seven to 15 years, depending on the climate.

The rate of deterioration depends on the pavement strength, the standard of maintenance (reseals and patching), weathering and damage from vehicles. Figure 2.31 shows road roughness over time, with and without resealing. Failing to reseal when required can considerably shorten pavement life.

Figure 2.31: Effect of resealing in pavement deterioration



Rehabilitation can involve:

- applying a thick asphalt overlay
- resheeting—removing the seal, putting down a layer of granular material such as gravel or crushed rock, and resealing, or
- stabilisation—removing the seal, milling the existing pavement material, possibly adding extra material, adding cement, recompacting and resealing.

Complete replacement of the pavement can be regarded as maintenance to the extent that it restores the pavement to its original condition. In this section of the Guidelines, the term rehabilitation is considered to include reconstruction.

Rehabilitation results in a considerable improvement in the roughness and extension of pavement life. Figure 2.32 shows how roughness can change over time given the effects of rehabilitations, with a full reconstruction occurring after two rehabilitations. These cycles can be 15 to 40 years long, so the length of time covered by Figure 2.32 could be well over 100 years. The choice of treatment depends on factors such as the cost of materials, the structural strength of the existing pavement compared with the strength required to meet future needs, constraints on raising the level of the road's surface and limitations on the length of time the pavement can be closed to traffic. Some treatments, for example, a thin asphalt overlay, will not bring the roughness all the way down to the level of a new pavement.

Figure 2.32: Pavement cycles

Rougher pavements lead to greater costs being incurred by road users. Increased roughness may reduce free-speed (the speed vehicles travel in the absence of any congestion), which leads to increased time costs. Increased roughness directly raises vehicle operating costs by imposing greater wear and tear on vehicles. Pavement deterioration also takes the form of cracking and rutting. Cracking and rutting do not affect vehicle operating costs, but rutting is a safety concern because water accumulates in the ruts in wet weather reducing skid resistance.

2.17.4 Cost minimisation for periodic maintenance

Economic assessment of maintenance initiatives requires a different approach from capital initiatives. For capital initiatives, the Base Case can usually be clearly identified as continued use of the existing infrastructure. For maintenance initiatives, the Base Case is not easy to identify because the alternative to a particular maintenance treatment under consideration is not to do-nothing, but a range of possible maintenance treatments undertaken at different times. A maintenance decision made now affects the timings of future maintenance treatments, shifting future road-user and road-agency costs in time. In contrast, a capital initiative usually generates a steady stream of benefits.

The usual way to approach economic assessment of road maintenance is cost minimisation, which obviates the need to specify a Base Case. From any number of alternative combinations of maintenance treatments and timings of treatments, the combination that minimises the discounted present value of social costs should be found. The term social costs here comprises road-user costs as a function of road condition, costs to the road agency, delays to road users while maintenance works are being carried out and costs of externalities and accidents.

Minimising social costs is equivalent to maximising net benefits relative to any given Base Case, *provided demand is perfectly inelastic*. If demand changes as a result of an initiative, total costs are increased because of the higher numbers of users. However, this is more than offset by the benefits received by the additional users. The cost minimisation framework treats the increased costs from additional users as a negative factor, with no way of recognising the offsetting benefits. Fortunately, the assumption of perfectly inelastic demand for maintenance initiatives is usually realistic. Variations in road roughness over the applicable range do not lead to very large differences in costs for individual road users, and, during the course of any one trip, a road user is likely to pass over a large number of different segments of pavement at different stages of their lifecycles. Hence, rehabilitation of any one segment of pavement, in isolation, is unlikely to alter total trip costs sufficiently to produce any significant amounts of diverted or generated demand. To sum up, for cost minimisation, the volume and composition of traffic must be exogenous. It may, however, change over time.

Estimating road-user costs for each year requires a pavement deterioration relationship and a road-user cost relationship. The pavement deterioration relationship provides an estimate of roughness as a function of variables such as pavement age, usage (cumulative equivalent standard axle loads), climate (Thornwaithe index (Martin 1996) or environmental coefficient), pavement strength (modified structural number) and cracking. Variables entering into the road-user cost relationship include roughness, vehicle numbers (AADT), traffic composition (percentages of trucks of different types) and road characteristics (number of lanes, lane widths, shoulder widths, gradient, alignment).

Pavement deterioration relationships are not very accurate. Site-specific factors that cannot be included in a general relationship, such as the quality of materials and workmanship, the effects of extreme wet weather events and overloaded trucks, can impact pavement deterioration. A shortage of good-quality time-series data on pavement deterioration under varying conditions in Australia is also an issue. In practice, data on pavement strengths is scarce because it is not as easy to measure as roughness. Where deterioration algorithms are considered untrustworthy, expert judgment is the best substitute.

If there is only one type of rehabilitation treatment with a single implementation cost, the optimal rehabilitation times can be found by using a simple algorithm to find the minimum point, since there will be a unique minimum. BTRE's Road Infrastructure Assessment Model (RIAM), used for strategic-level analysis, operates on this basis. Where there are a variety of possible treatments with different costs, different initial reductions in road roughness and different impacts on pavement strength, there are potentially a very large number of local optimum points. ARRB Group's Pavement Life Cycle Cost and PLATO models use a 'genetic algorithm'. A genetic algorithm tests large numbers of potential solutions and 'interbreeds' the better solutions to produce improved solutions. There is no guarantee it will find the best of all possible solutions, but it will find a very good solution. Better solutions are found the longer the algorithm is allowed to run. The drawback is that run times are long, so the models cannot be used to individually analyse all segments in a large road network database. For analysis of a network or corridor, segments need to be grouped together into bins with similar age, traffic and pavement strength characteristics.

2.17.5 Model calibration and running

In setting up the optimisation problem, a deterioration curve has to be fitted that passes through the roughness level of the pavement at the time it was last measured. Using this deterioration curve, roughness is then projected forward to the starting year of the analysis.

A technical problem with economic models of pavement deterioration is determining the number of years to project the modelling forward. Flexible pavements in dry parts of Australia can last for well over 50 years before rehabilitation or reconstruction is required. If the final year of analysis is set too early, the model may increase the time intervals between rehabilitations pushing the time of the last rehabilitation just beyond the final year, distorting the result.

One solution is to extend the number of years of analysis far into the future, say, 100 years, by which time, discounting has reduced the effects of whether or not the last rehabilitation falls within the analysis period to a negligible amount. The number of years of analysis required depends on the discount rate and the length of the time interval between rehabilitations. It may be necessary to constrain traffic after a number of years so that volume–capacity ratios do not reach unrealistically high levels. Usually, maintenance models only consider maintenance and do not allow for expansion of capacity to accommodate traffic growth.

Another solution is to make the number of years of analysis infinite by stopping traffic growth after a period of time so that the rehabilitation cycles become uniform in length. For each of these cycles, the rehabilitation cost is incurred at the start and road-user costs rise each year over the length of the cycle as the pavement roughens. The present value of road-agency and road-user costs over a single cycle of *n* years is

$$a = \text{rehabilitation cost} + \sum_{t=1}^{n} \frac{ruc_t}{(1+r)^t}$$

where ruc_t is the road-user cost in year t of the cycle. If these uniform cycles start in year T^*, the year of the first rehabilitation after traffic growth has stopped, and continue on in perpetuity, the present value discounted to year zero is

$$\frac{a}{(1+r)^{T^*}[1-(1+r)^{-n}]}.$$

This amount must be added to the present value of road-agency and road-user costs for all years prior to year T^* to obtain the full cost that has to be minimised.

A further possible problem with cost minimising maintenance models is extreme solutions. If the cost of a particular rehabilitation treatment is low relative to the benefits in terms of reducing roughness, the optimal solution may be to apply the treatment as often as possible, for example, every year if the model works in one-year intervals. This is a 'corner solution'. The opposite extreme solution may occur for low trafficked pavements where the total benefits of roughness reduction to users are small in relation to rehabilitation costs. Road roughness may then be permitted to reach very high levels before rehabilitation is economically worthwhile, extrapolating the deterioration and road-user cost relationships beyond the ranges over which they are valid.

Extreme solutions can be avoided by setting upper and lower bounds on either roughness levels or cycle lengths. Upper bounds on roughness may also be required for equity or meeting community expectations about minimum road standards.

Budget constraints

Budget constraints are usually imposed on road-agency costs in maintenance models. If routine maintenance is treated as fixed, the residual amount left in the budget, after paying for routine maintenance, is available for periodic maintenance. Since periodic maintenance involves consideration of expenditures over long periods of time, the problem is far more analytically difficult than for capital initiatives where the problem is to allocate funds out of a budget for a single year or for a few consecutive years. Assumptions have to be made about the availability of funds for many years into the future. The budget constraint could be set either as a series of amounts, one for each year, or as a single present value of future available funds. For a single segment of road pavement considered in isolation, only the latter assumption makes sense.

When maintenance standards are determined exogenously, the required budget becomes endogenous. As part of the strategic planning processes, the road agency may determine a set of standards consistent with the budget constraint. For individual pavements, it is then a question of achieving the required standard at the least life cycle cost for the road agency.

2.17.6 Marginal BCR

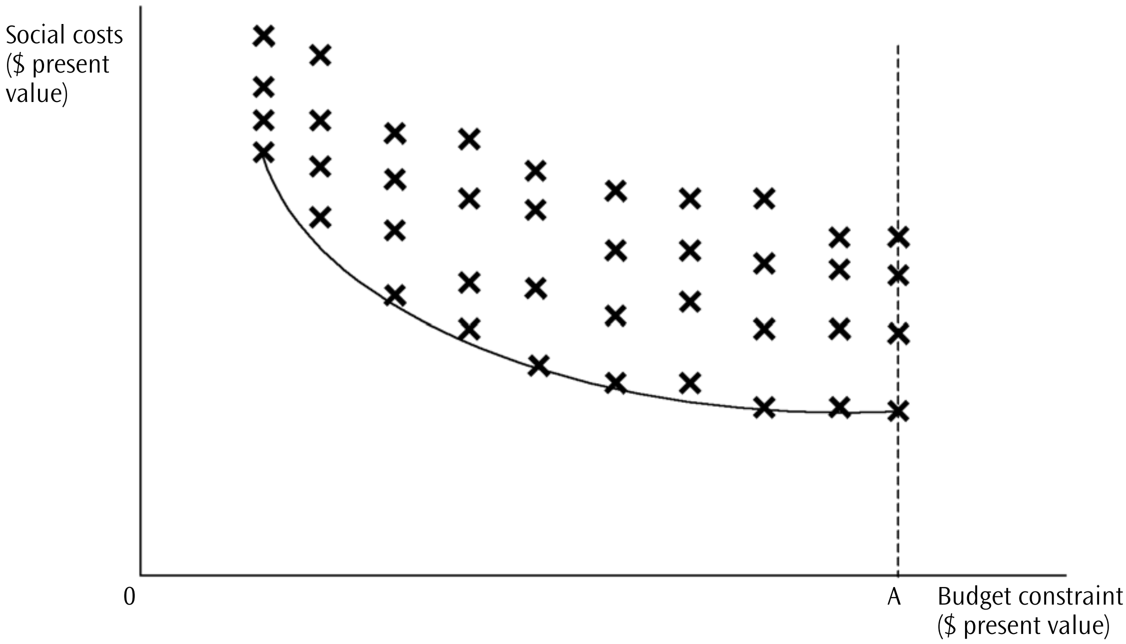
The life cycle cost minimisation approach does not produce a BCR that enables comparison between the relative values of maintenance and capital initiatives within a budget constraint. A Base Case has to be specified to estimate a BCR. With a vast number of possible combinations of maintenance treatments and timings over the life of a pavement, the value of the BCR depends on which Base Case is selected.

Section 1.5 in Volume 5 proposes a method for assessing whether the relative split of total funds for a road system devoted to capital works and maintenance is optimal, in economic terms. Where the present value of total social costs (mostly road-agency and road-user costs) of an entire network is minimised subject to a budget constraint, the marginal BCR can be found by reducing the budget constraint, say, by $1 million, and rerunning the model to obtain the new present value of total social costs. Provided that the budget constraint is binding, the new present value of the total system costs should be higher. The marginal BCR is found by dividing the difference in the present values by the change in the budget constraint. If the marginal BCR for maintenance is greater than the cut-off or marginal BCR for capital works, then it is economically efficient to shift funds from capital spending to maintenance, and vice versa.

A similar approach could be adopted for the maintenance schedule of a single homogeneous segment of pavement. An important difference is that for a large number of pavements considered together (i.e. a whole network), the budget constraint could be set *either* as a series of year-by-year amounts or as a single present value. For a single length of pavement, the budget constraint can be set only as a discounted present value because there is zero spending on periodic maintenance for most years, interspersed with an occasional year of high spending when a treatment is undertaken.

Figure 2.33 shows how the results might appear if a series of discounted present values of total social costs (including road-agency costs) were estimated with different budget constraints expressed as present values of road-agency spending and plotted on a chart. For each budget constraint, only the maintenance schedule with the minimum present value of social costs is of interest. Joining the minimum values together, a curve should be obtained. Higher amounts of spending should lead to lower social costs until the budget constraint ceases to be binding and there is unconstrained cost minimisation. At this point (*A* in Figure 2.33), the curve becomes flat. Analysis with budget constraints beyond this point has no effects because the budget constraint is non-binding—the same minimum cost solution will be derived with funds left over.

Figure 2.33: Estimation of marginal BCR for maintenance

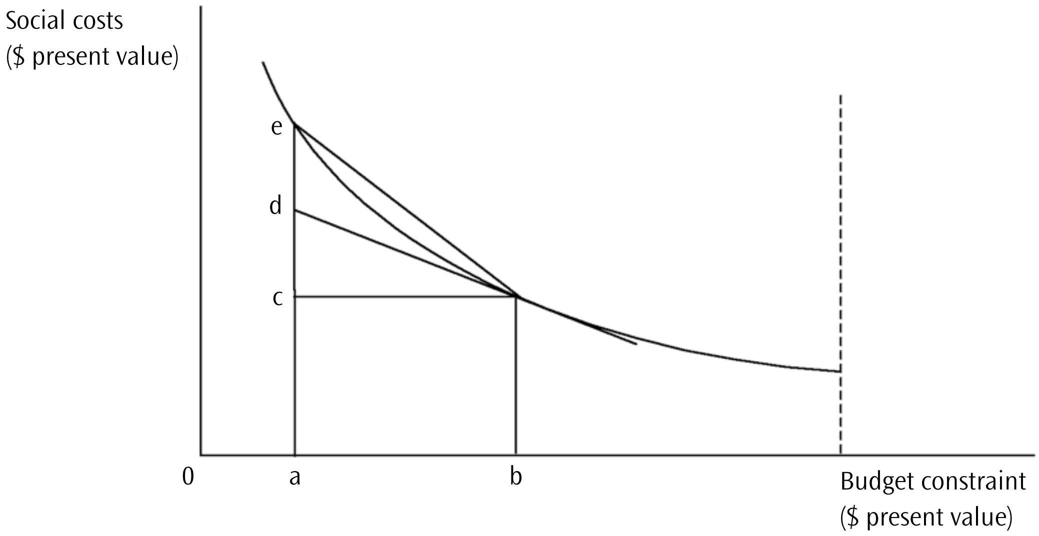


The marginal BCR for any given budget constraint is the slope of the curve in Figure 2.33 at the budget level, multiplied by minus one to make it positive, plus one. For example, say, adding an additional $1 million to the present value of the maintenance budget reduces the present value of total social costs by $3 million. The slope of the curve in Figure 2.33 at this point is –3.0. For simplicity, assuming away any other costs, the present value of road-user costs would have to fall by $4 million to create a reduction of $3 million in the present value of total costs. The marginal benefit–cost ratio is therefore 4.0. At the point where the budget constraint just ceases to be binding and the curve becomes flat (slope of zero) (budget of *A* in Figure 2.33), the marginal BCR becomes one. See the corresponding section in Volume 3 for the formula and a more detailed numerical example.

With an economically optimal allocation of maintenance funds across a network, the marginal BCR is the same for all pavement segments. The marginal BCR for maintenance defined in this way can be compared with the BCRs of capital initiatives. As shown in Section 2.10.4 in Volume 5, with an economically optimal allocation for spending over time, the marginal BCR is constant over time.

When estimating a marginal BCR for road maintenance, it is important that the points sampled in the region of the desired marginal BCR are not too far apart. Figure 2.34 shows how an exaggerated value can be obtained for a budget constraint *b*, by comparing social costs at *b* with social costs at a much lower budget constraint *a*. The marginal BCR is estimated as *ce/ab* + 1, which is significantly larger than the correct value of *cd/ab* + 1.[24] A related point to note is that a marginal BCR for maintenance is valid only over a limited range of changes to the budget constraint.

Figure 2.34: Effects on marginal BCR: estimate of size of change in budget constraint



This methodology could be implemented starting with a budget constraint for a single homogeneous road segment. However, a better allocation of funds occurs if there is flexibility to shift funds between road segments in different parts of the network. A way of gaining this flexibility, while staying within a budget constraint for the whole network, is to determine the marginal BCR associated with the given budget constraint via the process described in Section 1.5. Hence, the marginal BCR becomes the starting point instead of the budget constraint for an individual segment. By undertaking an analysis as described in Figure 2.33 for an individual pavement segment, the set of maintenance treatments and budget constraints consistent with the desired marginal BCR could be obtained. Doing this for all segments in the network produces an optimal allocation of maintenance funds within the budget constraint for the whole network.

2.17.7 Combination capital and maintenance initiatives

No additions to the corresponding section in Volume 3 are made here.

24 Say, ce/ab is 1.0 and cd/ab is 0.5. Then the marginal BCRs will be 2.0 and 1.5 respectively.

2.18 Toll roads and leveraging

2.18.1 Economic evaluation of private sector proposals

Compared with the option of full public provision, private sector participation can affect the economic attractiveness of an initiative in four ways:

1. Direct effect of a toll on benefits. Imposition of a toll or other additional charges is likely to alter the price paid by users, which affects the demand for infrastructure use. This in turn affects benefits and costs of the initiative. The net effect could be either positive or negative (see Section 2.18.2).
2. Better management by the private sector. The private sector may be able to undertake certain tasks better than the public sector (lower social costs). For example, the private sector may be better at maintaining control over costs. Ideally, the arrangement with the private sector would transfer risks to the party best able to manage them. Hence, if the private sector is better at avoiding construction cost overruns, appropriate risk transfer produces the incentive structure most likely to maximise overall benefits to society. In particular, the private sector should be rewarded for cost containment, but should also bear the costs of overruns. Expected values of investment and operating costs could therefore differ, depending on the form and level of private sector participation. A public sector comparator (see Section 2.19) is a good source of data for a BCA that aims to take account of the effects of private sector participation.
3. Easing the budget constraint. Private sector funding can free up scarce government funds for other uses, including other infrastructure initiatives. If the additional funds made available from leveraging are spent on initiatives with BCRs greater than one, there is an additional net economic benefit to society (see Section 2.18.3).
4. Diversion of government funds away from more attractive initiatives. If the initiative with private sector participation has a BCR below the government's cut-off rate, the government's contribution may come at the expense of more economically attractive initiatives. However, provided the BCR is above one, the initiative still yields positive net economic benefits. These two opposite effects can be compared to determine if the initiative is worth supporting (see Section 2.18.4).

2.18.2 Direct effect of a toll on benefits

Pricing assumptions affect levels of demand, as noted in Section 2.4.7. A toll on a road adds to the private generalised cost of using the road. For an initiative involving construction of the new toll road, the toll reduces the amount of generated traffic and traffic diverted to the new road from alternative routes. In uncongested conditions, a toll will unambiguously reduce benefits because the toll deters some traffic from reaping the benefit of using spare capacity on the toll road. In congested conditions, the toll can act as a congestion charge and could generate either benefits or costs, depending on how closely the toll approximates the optimal second-best congestion charge. The optimal charge is 'second-best' because congestion on alternative routes is not charged. The toll, therefore, leads to a gain in economic efficiency for the toll road by raising the private cost on the toll road closer to the marginal social cost, but causes an offsetting loss in efficiency by distorting the allocation of vehicles between the toll road and alternative routes. How these two effects balance out is case-specific. A toll is also likely to differ from the second-best optimal congestion charge because tolls usually remain unchanged throughout the day, while the optimal congestion charge fluctuates with demand.

In a congested urban area, in which travellers choose between the faster toll road and slower alternative routes, the benefits from the toll road could be underestimated if the traditional assumption, that there is a uniform value of travel time savings (VTTS) for all traffic with each vehicle class, is made. In practice, there is a distribution of VTTS with work travellers having higher values than non-work travellers, and VTTS for cars rising with vehicle occupancy. Freight vehicles have different values per passenger car unit (PCU), depending on the nature of the freight carried.

Where there is a distribution of VTTS, a toll on the faster route can split the traffic stream so that road users with high values of time choose the toll road and users with low values of time choose alternative, slower routes through the network.

Say, a road user incurs $c_2+\pi$ in generalised costs on the toll road, where c_2 is the cost of vehicle operation and time and π is the toll. The lowest generalised cost achievable for alternative routes is c_1. The toll deters traffic from the toll road, reducing congestion on the toll road and ensuring that $c_1>c_2$. By switching from the alternative route to the toll road, a vehicle saves $c_1–c_2$ but the driver pays π. The driver will stay on the alternative route if $c_1–c_2<\pi$, otherwise he or she will switch to the toll road. If generalised cost is assumed to be a linear function of time (expressing vehicle operating costs as an amount per period of time and adding this to the VTTS), the cost difference between the two routes (excluding the toll) can be expressed as $c_1–c_2=v(t_1–t_2)$, where v is the value of time for a vehicle, comprising both driver/passenger costs and vehicle operating costs, t_1 is the time taken to travel between two points via the alternative route and t_2 is the time taken via the toll road. All vehicles with a value of time below a critical value $v^*= \pi/(t_1–t_2)$ will use the alternative route and all vehicles with a value of time above v* will use the toll road.[25]

If a uniform value of time for all road users in each vehicle class is assumed, benefits in time savings for users of the toll road could be significantly underestimated.

To evaluate a toll road allowing for differences in values of time, a frequency distribution of traffic by value of time must be developed. Ideally, separate frequency distributions would be specified for different vehicle types. Within the cars category, work and non-work travel and vehicles with different occupancies could be treated separately. In Harvey (2005), the frequency distributions for the different vehicle categories are assumed to be normal. However, Hensher and Goodwin (2003) point out that empirical evidence suggests that VTTS distributions are skewed with a disproportionately large number of individuals having a relatively low VTTS. The frequency distributions of values of time for different vehicle categories can be combined to produce a value of time distribution for all traffic between a given origin–destination pair. The critical value of time, v*, can be identified providing an estimate of the split of traffic using the toll road and alternative routes. A separate benefit calculation is required for traffic in each value of time bin within the distribution.

2.18.3 Initiative easing budget constraint

Private sector participation can affect the economic merit of an initiative through freeing up scarce government funds for other uses, including other infrastructure initiatives. If economic efficiency is the sole objective, and the total value of economically warranted initiatives exceeds the budget, initiatives should be ranked in descending order of BCR, until the budget is exhausted (see Section 2.10.4). If the additional funds made available from leveraging are spent on initiatives with BCRs greater than one, there is a net economic benefit to society. Say, the initiative that just missed out has a BCR of 2.5 and costs \$100 million. A private sector contribution of \$100 million to undertake a more highly ranked initiative would free up government funds to undertake this lower ranked initiative, generating a net gain to society of \$150 million = \$100 million × 2.5 – \$100 million.

It is assumed that the more highly ranked initiative is sufficiently attractive so that it is undertaken with or without leveraging. The public policy question is whether or not to leverage. If tolling increases the BCR, then the government has the option of levying a toll without inviting private sector participation and can set the toll to promote the economic efficiency objective. With leveraging, the toll will be set at a level negotiated with the private sector partner and there may be cost savings to take into account. If the net present value is smaller with leveraging, due to net negative economic efficiency effects from tolling, leveraging may still be preferable depending on a comparison with the benefit from easing the budget constraint.

25 This discussion is based on Harvey (2005). See also Hensher and Goodwin (2003).

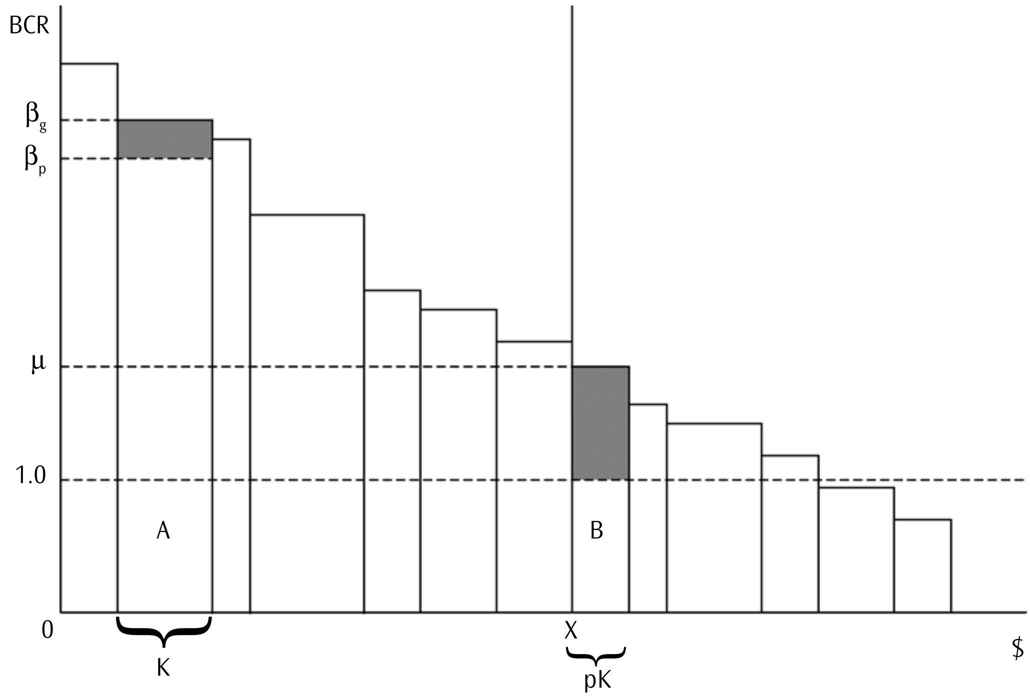
In the numerical example above, say, the highly ranked initiative costs $200 million with a BCR of 4.0 (total benefit $800 million) without leveraging (and tolling) and 3.5 (total benefit $700 million) with leveraging. Leveraging makes society worse off by $100 million. However, leveraging is the better option when the gain of $150 million from easing the budget constraint is taken into account.

The decision can be made using the 'BCR of leveraging'. For the BCR of leveraging to be applicable, the initiative has to have a high enough BCR and meet other objectives so that it is attractive enough to proceed, regardless of whether there is private sector participation. A decision has to be made about whether or not to allow leveraging to occur.

In Figure 2.35, initiative capital costs are shown on the horizontal axis and BCRs on the vertical axis. With total funds of *0X* available, all initiatives with BCRs above μ are implemented. μ could be termed the marginal BCR. It is assumed that $\mu > 1$.

Say that leveraging is possible for Initiative *A* and its BCR is β_g without tolling. The BCR is reduced to β_p where there is private sector participation and a toll is imposed. The private sector would fund proportion *p* of the capital cost *K*, leaving the government to fund *1-p*. The size of the budget available for spending on transport initiatives has effectively been increased by *pK*. With these extra funds, Initiative *B* with a BCR of μ, can now be implemented. For practical purposes μ could be either the BCR for a specific initiative or a general cut-off BCR.

Figure 2.35: Benefit and cost of leveraging



If Initiative *A* is leveraged, the net gain to society compared with full government financing is:

- the benefits from spending the additional funds made available on Initiative *B:* μpK
- the loss of benefit from *pK* not being available elsewhere in the economy, assumed to equal *pK*, and
- the loss of benefit caused by tolling: $(\beta_g - \beta_p)K$.

Adding these: $\mu pK - pK - (\beta_g - \beta_p)K$.

So leveraged financing is preferable if:

$$p(\mu - 1) - (\beta_g - \beta_p) > 0$$

or if, what might be called the BCR of leveraging, exceeds one; that is:

$$\frac{p(\mu - 1)}{\beta_g - \beta_p} > 1$$

In terms of Figure 2.35, leveraging is preferred if:

- the area of the shaded rectangle on the right, which is the gain from implementing the marginal initiative $(\mu - 1)pK$ = \$150 million = (2.5–1) × 0.5 × \$200 million for the numerical example

is greater than

- the area of the shaded rectangle on the left, the loss from leveraging $(\beta_g - \beta_p)K$ = \$100 million = (4.0–3.5) × \$200 million.

In the case of the numerical example, the BCR of leveraging is:

$$\frac{p(\mu - 1)}{\beta_g - \beta_p} = \frac{0.5 \times (2.5 - 1)}{4.0 - 3.5} = 1.5 > 1$$

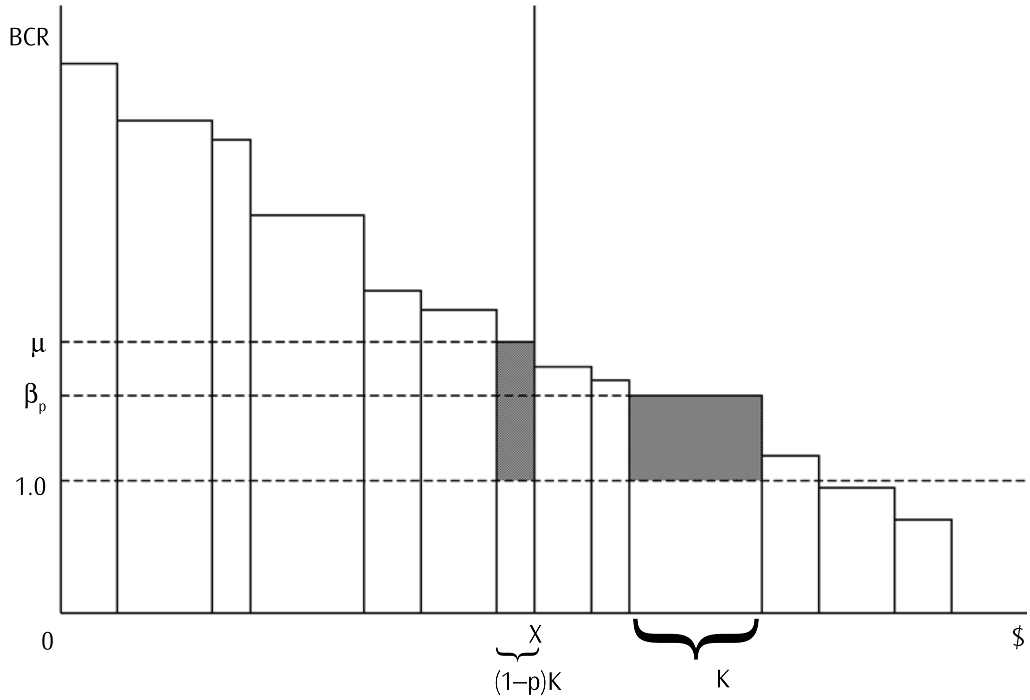
2.18.4 Initiative diverting government funds from more attractive initiatives

An initiative with a BCR that is too low to attract funding, but above 1.0, is still of overall benefit to society. Say the government and the private sector agreed to jointly fund a \$100 million initiative having a BCR of 2.0. The net gain to society from the initiative is \$100 million = \$100 million × 2.0 – \$100 million. However, the government contribution is at the expense of other initiatives with higher BCRs. Say the initiative forgone has a BCR of 2.5 and the government contribution is \$50 million. Then society is worse off by \$75 million = \$50 million × 2.5 – \$50 million. The net result is a gain to society of \$25 million = \$100 million – \$75 million. The gain will vary with the share the government contributes, ranging from zero for a two-thirds government contribution to \$100 million for no contribution.

The government's decision can be made using the 'leverage-adjusted BCR'. For the leverage-adjusted BCR concept to be relevant, the initiative has to have a BCR greater than one, but too low for the initiative to be implemented given funding constraints. A further prerequisite is that the private sector will not implement the initiative of its own accord.

In Figure 2.36, Initiative *A* has a BCR below the cut-off point, μ, but still greater than one. If it is implemented with government funding, there is a net loss to society because it would crowd out other initiatives with higher BCRs. However, with leveraging, there are circumstances where it is more economically efficient to implement Initiative *A*.

Figure 2.36: Benefit and cost of implementing a low BCR leveraged initiative



Using the same notation as above, the net gain to society from implementing a leveraged initiative, compared with not implementing it, is:

- the benefit from implementing the initiative—βpK (βg is irrelevant in this case because there is no question of implementing the initiative without leveraging)
- the loss of benefit from pK not being available elsewhere in the economy, assumed to equal pK, and
- the loss from not having the government contribution available to spend on the marginal initiative $B - \mu(1-p)K$.

Adding these: $\beta_p K - pK - \mu(1-p)K$.

So it is preferable that the initiative proceeds, even though it displaces investments with higher BCRs, if

$$\beta_p - p - \mu(1-p) > 0$$

or if $\frac{\beta_p - p}{1-p} > \mu$.

The value of $\frac{\beta_p - p}{1-p}$ might be termed the leverage-adjusted BCR and can be used for prioritising initiatives.

Even though the leveraged initiative has a lower BCR than the initiative with the marginal BCR that it displaces, ($\beta_p < \mu$), the size of the leveraged initiative is larger, so that, in absolute terms, the gain to society from implementing the leveraged initiative is greater than that for the marginal initiative.

In terms of Figure 2.36, leveraging is economically desirable for the initiative to proceed if:

- the area of shaded rectangle on the right, that is, the gain from implementing the leveraged initiative $(\beta_p - 1)K$ = \$100 million = (2.0–1) × \$100 million

is greater than

- the area of the shaded rectangle on the left, that is, the loss from forgoing the marginal initiative $(\mu - 1)(1 - p)K$ = \$75 million = (2.5–1) × (1–0.5) × \$100 million.

In the case of the numerical example, the leverage-adjusted BCR is

$$\frac{\beta_p - p}{1-p} = \frac{2.0 - 0.5}{1 - 0.5} = 3.0 > \mu = 2.5$$

2.19 Public sector comparator

2.19.1 Public sector comparator (PSC) model

More detail is available on the public sector comparator (PSC) model from internal guidelines developed by jurisdictions that use the model in documents on Internet sites such as Partnerships Victoria (2001; 2003a), Industry Canada (2003) and HM Treasury (1999).

2.19.2 Reference initiative

No additions to the corresponding section in Volume 3 are made here.

2.19.3 Discount rate

In the corresponding section in Volume 3, it states that the discount rate should be the weighted average cost of capital to the government. This may be the bond rate, but it could be higher if the government relies on higher-cost sources of funds from private sector participation in other initiatives.

Section 2.11 in Volume 5 discusses two ways to adjust for risk in appraisal of initiatives: direct adjustment of benefits, revenues and costs via the state-contingent approach, and adding a risk premium to the discount rate. The latter is only valid if the assumptions of the consumption capital asset pricing model hold, which is unlikely to be the case for BCAs. For a financial analysis, the weighted average cost of capital (WACC) includes a risk premium, but it is not adjusted specifically for the initiative in question.

There is disagreement about which approach to take for the PSC. HM Treasury (1999, Section 3.3.8) recommends that the discount rate is the WACC and that cash flows are adjusted for risk. This is the approach recommended in Volume 3. Partnerships Victoria (2003b) requires higher discount rates to be used for initiatives with greater amounts of systematic risk and that the discount rate is adjusted to reflect the level of risk retained by the government. When all risks are retained by the government, the discount rate is estimated from the capital asset pricing model (CAPM). As more risks are transferred to the private sector, the discount rate will fall and reach the risk-free rate once the government has transferred all risks. The Guidelines do not recommend this approach due to concerns about:

- the ability to accurately estimate appropriate adjustments to make to the discount rate as risks are transferred
- whether a government should be concerned with systematic risk for its own funds in the same way as a private individual[26]
- inclusion of the stock market equity premium in the discount rate, and
- the assumptions of the consumption CAPM not holding; see BTRE (2005) for an explanation of the equity premium and the assumptions of the CAPM.

Note that being a financial analysis, the PSC is undertaken in nominal terms, so a nominal discount rate should be employed.

2.19.4 to 2.19.9 Public sector comparator model

No additions to the corresponding section in Volume 3 are made here.

26 Risk adversity depends on an assumption of diminishing marginal utility of money. The consumption CAPM is based on an assumption of a utility function that reflects this, increasing at a diminishing rate. This seems reasonable for individuals but is questionable for governments.

2.20 Post-completion evaluation

The strategic planning framework set out in Volume 2 and the appraisal methodology for individual initiatives set out in Figure 1.1 in Volume 3 finish with a review phase. The reviews put in place a formal mechanism to gather feedback to support learning from the past and to uncover potential problems at an early stage. The review can focus on any level of the planning, appraisal and implementation processes—strategies, investment programs or specific initiatives. This section focuses on evaluation of the appraisal stage for individual initiatives.[27]

The methodology is tailored to fit the appraisal process outlined in Volume 3. Post-completion evaluation (PCE) reports on initiatives appraised using the Guidelines methodology should provide valuable feedback for improving future editions of the Guidelines.

The aim of PCE is to assess ex-post, the efficiency and effectiveness of investment decisions and the management of their implementation. It involves comparing actual actions or outcomes with planned actions or outcomes to show the extent to which the stated objectives and expected results were achieved.

PCE supports the goal of continuous improvement in the capital investment process. More precisely, reasons for conducting the evaluation include to:

- improve appraisal methods and practice
- improve the current procedures for making and implementing decisions
- improve the accuracy of forecasts of costs, benefits, traffic demand and infrastructure condition
- create incentives for procedures, decisions and plans to be properly implemented since PCE increases the probability that failure to follow procedures or plans and bad decisions are noticed[28], and
- identify any corrective actions that need to be taken to adjust for deviations from forecast in the performance of the initiative.

2.20.1 Subject of post-completion evaluation

The various types of PCE discussed here apply to individual initiatives. They can also be applied to programs or groups of related initiatives; for example, all the initiatives in a corridor, initiatives of a particular type or initiatives for a particular mode.

PCEs can be applied to the *process* that leads to the outcomes, or to the *outcomes* themselves. Table 2.12 sets out stages of development of an initiative consistent with the approach in the Guidelines (see Figure 1.1 in Volume 3) together with possible reviews that might be undertaken regarding processes and outcomes. Process reviews look at how outcomes were achieved. Outcome reviews involve comparing actual with predicted outcomes.

27 See Austroads (2005d) for an alternative discussion of post-completion evaluation.

28 This also applies to forecasts in BCAs because PCE can expose proponents of initiatives who consistently underestimate costs or overestimate benefits by large margins.

Table 2.12: Types of review of an initiative

PROCESS REVIEW	OUTCOME REVIEW	
	Actual outcomes	Sources of predicted outcomes
Identification, consideration of options	Initiatives identified	Government strategy documents, stakeholder views of options
Appraisal	BCA, SMT, other analyses, Business Case	Guidelines, knowledge of correct methodologies, Austroads publications, government strategy documents
Recommendation	Recommendation for or against	Business Case
Planning and design, budget development	Scope, costs, risks	Appraisal
Management and implementation	Scope, costs, problems	Appraisal, plans and designs, budget
Operation	Performance	Appraisal

2.20.2 Standards of comparison

Figure 2.37 illustrates the three types of PCE discussed in this section.

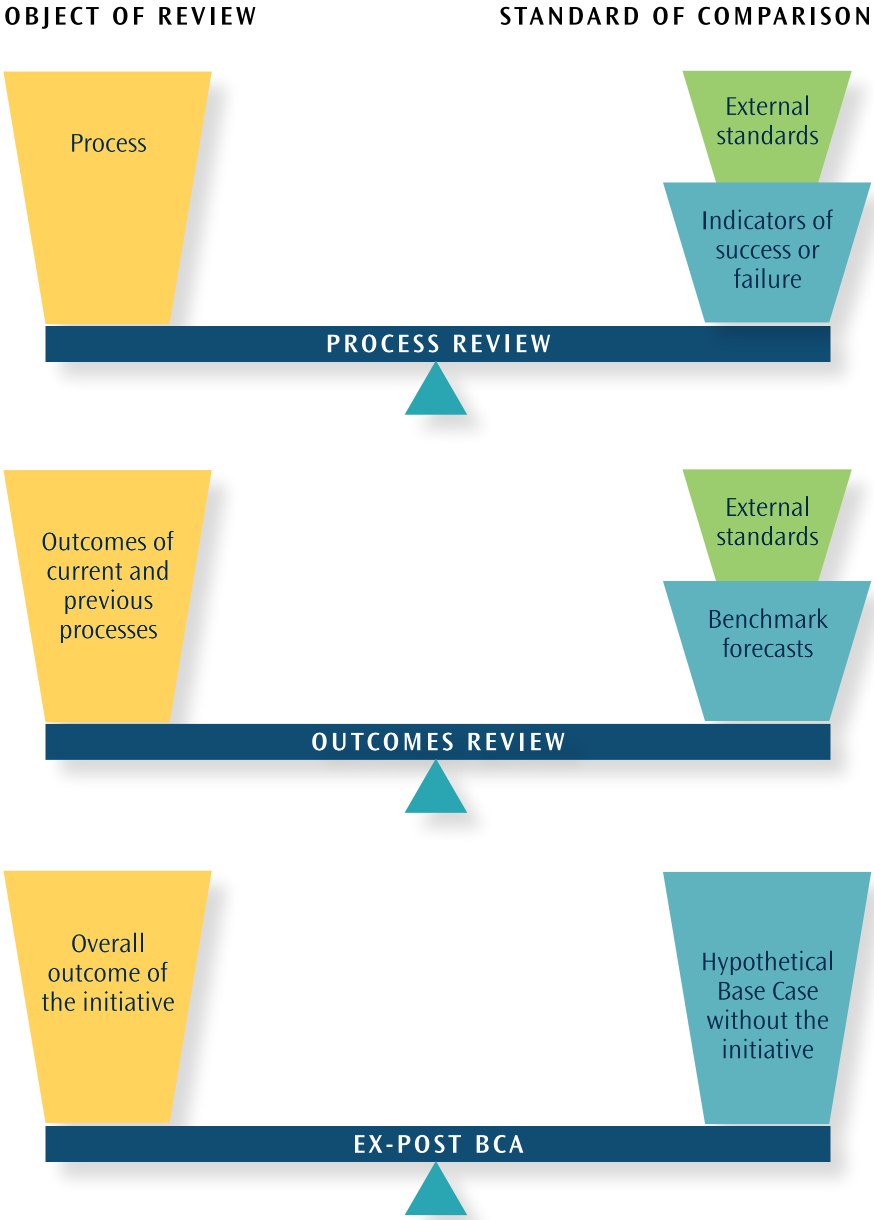
For a process review, the effectiveness of the process is assessed against external standards about desirable attributes of processes. Another basis on which to assess a process is the indicators emerging from outcome reviews; in particular, successes and failures in meeting benchmarks or forecasts.

For an outcome review, the standards of comparison are external standards such as desirable attributes of outcomes (e.g. correct BCA methodology) and benchmarks or forecasts established during a previous process.

The two reviews should be carried out simultaneously for a single stage so that any process-related reasons for successes or failures can be explored. The outcome review at one stage may point to process issues in earlier stages.

The third type of review, ex-post BCA, is discussed in Section 2.18.6.

Figure 2.37: Types of post-completion evaluation



The decision made at ministerial level does not appear in the table. Decisions by ministers are based on a wide range of factors, including value judgments, which are presumed to represent the view of the community as a whole. It is not appropriate for a PCE to question these judgements. The initiative recommendation is similarly based on a wide range of considerations and involves subjective judgments. It too should not be questioned by a PCE. The initiative recommendation stage is included in Table 2.12 because it is legitimate to ask how well the Business Case and the underlying appraisal process served the needs of those responsible for developing recommendations.

The following set of questions, structured in line with Table 2.12, is intended to be indicative rather than exhaustive.

Identification of the initiative, consideration of options

Process review:

- Did the initiative have clear objectives?
- Were government strategies adequately formed and communicated?
- How were the initiatives and options identified?
- Was there a thorough exploration of options, including non-infrastructure and multi-modal solutions?
- Was there significant justification for why an option was discarded?
- Was good use made of stakeholder input e.g. key informant interviews, focus group discussions, community and group interviews and surveys?
- Was potential co-funding adequately sought e.g. other agencies, the private sector, beneficiaries?
- Was adequate information or data acquired to assess whether an option was worth pursuing in more detail?

Outcome review:

- How well did the objectives of the initiative and options selected for analysis align with government strategies?
- Does the initiative address problems or issues that need to be addressed?
- Was the initiative adequately scoped e.g. utilisation of the service provided by the option; how the community, environment, economy and safety is impacted; the current usage of assets or patterns of behaviour?
- Was the initiative adequately described in terms of location, nature of the capital works, estimated benefits and costs, initiative development and level of support for the initiative?
- Do the options selected to take forward to the analysis stage appear feasible in terms of resource requirements?
- Is the selection of options to take forward to the analysis stage soundly based, as evidenced by early identification of highly desirable options and elimination of undesirable options?

Appraisal

Process review:

- Was the appraisal process properly followed, with all the necessary analyses undertaken?
- How were the assumptions determined?
- Did decisions to accept, reject or alter the scope of initiatives or options follow on from the results of the analyses undertaken?
- Were independent reviews (within the department or jurisdiction) of the SMT, rapid BCA, detailed BCA and other analyses undertaken?
- Has the process brought to light suggestions for improving the Guidelines, including the SMT template?

Outcome review:

- Were all the questions raised in the SMT template adequately addressed by the proponent, using the best available information at the time, and applying sound judgment, including identifying risks and options?
- Has it been demonstrated that the initiative is likely to have the greatest net return compared to other options?
- Have optimal timing issues been adequately investigated?

- Was the initiative well defined e.g. adequate identification of issues the initiative addresses or services it will support and for whom, identification of capital works and benefits and definition of what the initiative will deliver?
- Were the rapid and detailed BCAs methodologically correct, and carried out in accordance with the Guidelines? The list of points to check here could be quite long. A few important ones are that the analysis:
  - did not omit significant benefits or costs
  - did not count benefits or costs that should not be included, or double-counts any benefits or costs
  - used sound methods, data and assumptions in making forecasts of costs, demands and benefits
  - used parameter values specified by the Guidelines and, when alternative values have been used, justifications have been provided, reasonable alternative values have been used, and a sensitivity test has been provided using the Guidelines parameters, and
  - included a proper risk analysis, including assignment of reasonable probabilities.
- If adjusted BCAs were undertaken, has the methodology been correctly implemented? Were the assumptions about the distribution of benefits reasonable? Were the values placed on weights sensible and likely to produce outcomes in line with government objectives? Were the results of adjusted BCAs presented alongside unadjusted BCA results?
- Were the other analyses (financial, Environmental Impact Statement, etc.) undertaken using the correct methodology and reasonable assumptions?
- Were the different types of analysis (SMT, BCA, financial, Environmental Impact Statement, etc.) consistent with each other (scope, parameters, assumptions, forecasts, etc.)?
- Was the SMT template completed adequately e.g. answers to questions fully completed; calculation of economic estimates such as BCR, NPV, IRR; identification of quantifiable and unquantifiable environmental impacts; estimation of the percentages of total benefits and costs and identification of the gainers and losers from the initiative?
- Was the Business Case complete? Did it present an accurate and balanced summary of the various analyses undertaken in regard to the proposal? Are the points for and against the initiative consistent with the results of the analyses undertaken?

Recommendation

Process review:

- How were initiatives prioritised?
- How well does the Business Case support the process of prioritising initiatives? Is there missing information or superfluous information? Is the level of detail right? Can the presentation be improved?

Outcome review:

- Does the recommendation accord with the Business Case? If not, what other factors influenced the recommendation? (The purpose of these questions is not to be critical of the recommendation, but to elicit how well the Business Case and the appraisal process in general have supplied the information required to make the recommendation.)

Planning and design, budget development

Process review:

- Were responsibilities for planning clearly established?
- How well did the contractors undertaking planning and design tasks perform?
- How effective was contract surveillance?

- Were changes in scope based on a careful review of options, and then correctly treated and approved?
- Were contingency plans made (risk management)?
- Has a review of the impact on the community been undertaken?
- How were assumptions about unit costs and resource requirements determined?

Outcome review:

- How did cost estimates compare with those in the appraisal stage analyses and what are the reasons for differences (e.g. changes in scope, different assumptions for unit costs and resource inputs)?
- Were forecast inflation, interest rates and exchange rates based on realistic assumptions?
- How do the risks allowed for compare with those identified at the appraisal stage?
- Have all the requirements (legal, environmental, safety and stakeholder consultation) identified in the appraisal stage been addressed?

Initiative management and implementation

Process review:

- How well did contractors perform?
- How effective was contract surveillance?
- Was a cost accounting and classification system designed appropriate to the initiative's methods for budgeting, assessing, analysing and reporting costs?
- Were implementation plans established for specific activities with predetermined resource constraints?
- Were the most cost effective budgeting methods used on the basis of available information?
- Was information on actual expenditure available on a timely basis?
- Was a performance review system put in place for internal services, operations, equipment?
- Were progress reports available on a timely basis?
- Were changes in scope correctly treated and approved?

Outcome review:

- How do construction costs compare with budgeted costs, and with costs estimated during the appraisal stage and what are the reasons for differences (e.g. changes in scope, differences in unit costs and resource inputs)?
- To what extent were the actual circumstances of the initiative anticipated by the risk assessment process?
- Has the initiative had any unexpected consequences?
- To what extent, if any, did community members complain about side-effects from construction?

Operation

Process review:

- Is the initiative being used for the purpose specified?
- Is the initiative being operated as planned?

Outcome review:

- How do demand levels, benefits, revenues, operating costs, accident rates and other annual indicators of performance differ from forecasts made in the appraisal stage?
- Was there any impact on modal shares?

- To what extent has the initiative achieved its objectives? (This may include comparisons between pre- and post-implementation performance indicators such as travel times, accident rates, increased market share for rail).
- Could technology have played a greater role in contributing to desired outcomes?
- To what extent were the actual circumstances of the initiative anticipated by the risk assessment process?
- How successful were the risk management measures undertaken?
- Has the initiative had any unexpected consequences, good or bad?

Benchmarks

Evaluations of outcomes can be streamlined if, during the appraisal and design phases, benchmarks are set with times to check whether the benchmarks have been achieved. For example, construction costs estimated during the appraisal stage, and re-estimated during planning and design, can be compared with actual costs following completion. The traffic-level forecast from the BCA for, say, two years after completion might be set as a benchmark for subsequent checking. The benchmarks need to be clearly defined to ensure valid comparisons.

Templates for tables of benchmarks can be established. One table (or set of tables) could cover the construction phase and another the post-completion phase. Total construction costs are the starting point, which can be broken down into components such as sections or stages of the initiative, and, within each component, direct and indirect costs for materials and labour by type could be considered. The detailed BCA establishes benchmarks for the design stage, which in turn establishes benchmarks for the construction stage via preparation of the budget for construction.

If benchmarks are established for unit costs and for physical quantities of inputs, concepts from cost accounting might be employed to identify variances due to changes in unit costs (price variances), changes in design or scope, and efficiencies in use of labour and materials.

Timing of construction can give rise to further benchmarks.

In the operation phase, operating costs, demand levels, revenues and benefits can be compared with forecasts made in the BCA and financial analysis. Environmental impacts may be compared with forecasts in the Environmental Impact Statement. The timing of comparisons during the operation phase requires careful consideration. For some benchmarks, a time interval too soon after completion of the initiative is not sufficient to establish levels and trends with adequate certainty. The trade-off is that, as time passes, lessons about the reliability of forecasting practices employed in the past become less relevant to the present time.

The earliest time at which a comparison between an actual and a forecast value can be undertaken depends on the variable in question. In the case of a 2.5 per cent forecast traffic growth rate over 30 years, PCE two years after completion of the initiative is unlikely to be very informative because of the influence of short-term fluctuations. The trend can be gauged only by looking at traffic levels for a series of years over a long period. It is worthwhile to compare forecast with actual traffic levels in such a case within a few years of completion of the initiative to check whether demand is affected by unforeseen developments.

For diverted and generated traffic, a PCE after two or three years is a useful exercise because traffic is expected to respond immediately after completion of the initiative along the lines forecast by short-run elasticities. After several years, the results should be moving towards levels predicted by long-run elasticities.

For safety benefits, a longer time frame is required because the initiative is expected to reduce the *probability* of events that occur infrequently. The case study at the end of Volume 5, Part 2 notes a related problem, 'regression to the mean'.

For infrastructure operating costs, components that are broadly constant from year to year can usefully be assessed after completion as soon as normal operating conditions are established.

Examples are costs of running a new road toll collection system or railway signalling system. The same applies to savings in train operating costs or improvements in reliability expected to result from an initiative.

A need for rehabilitation of new infrastructure may not be required for decades after completion, so rehabilitation time could not be used as a benchmark. An exception is a technical failure that brings forward rehabilitation by a significant amount. This can be assessed in the short-term by checking that deterioration rates and current maintenance expenses are within normal bounds.

In summary, PCEs undertaken within a few years of completion of the initiative should use variables to which the initiative is expected to make a difference in the short-term as benchmarks. Variables that can only be assessed over a longer period because they relate to long-term trends or infrequent occurrences, are not suitable as short-term benchmarks, but should be checked for substantial deviations from forecasts due to unforeseen circumstances.

Benchmark levels set during the appraisal period should have times for review attached with these considerations in mind.

2.20.3 Depth and frequency of post-completion evaluation

PCEs can be carried out with a wide range of intensity. Prudent use of resources suggests that all initiatives should be subject to some basic level of PCE, with more detailed levels of PCE reserved for select groups of initiatives. Setting benchmark levels during the appraisal and (for construction) design phases, can facilitate cost-effective screening of large numbers of initiatives for construction performance immediately after completion, and for operating performance at set time intervals after completion. Further investigation, including reviews of processes, is warranted where such screenings find significant variances, in order to explain the causes.

Good candidates for detailed PCEs include initiatives that are:

- large
- appear to have gone poorly or exceptionally well
- of a type that recurs
- especially risky, including pilot initiatives for testing innovations
- strategically important
- long-term (interim PCEs, say annually)
- staged (PCEs of stages), or
- programs involving a series of smaller initiatives (for example, the Australian Government's Black Spot Program).[29]

RTA (1999) suggests that, as a broad guide, roughly one in 10 major initiatives should be subjected to a full PCE, although all major initiatives should be subjected to some form of review in terms of assumptions versus reality.

Where certain types of initiatives are not of a sufficient scale to require an individual PCE, the owner of the initiative should undertake an ex-post evaluation of a representative initiative at least once every five years.

2.20.4 Evaluators

The PCE should not be undertaken by the same personnel responsible for the process or the work being evaluated. They should, of course, be consulted when their expertise and knowledge is required. They should be briefed comprehensively on the outcomes and on factors contributing to differences between the expected and realised outcomes.

29 This list is based on Austroads (2005d, p. 4).

2.20.5 Assessment and reporting

RTA (1999) proposes the following steps for carrying out a PCE:

- establish the purpose of the evaluation
- define the problem or focus of attention
- establish methods of evaluation
- decide on the suitable method of measurement
- establish criteria
- decide on the intensity of the evaluation
- decide on the type of information required
- collect data
- analyse data in terms of the criteria and the focus of attention
- decide on the form and language of the presentation, and
- produce a final report.

Where process, or outcomes, is assessed as either significantly good or significantly poor, the reasons need to be determined. Consequences of internal management and planning processes should be separated from impacts of external factors. The aim should always to be to identify lessons for the future, not to allocate blame for mistakes, or to state what should have been done in the past with the benefit of hindsight. Where external factors have an adverse impact, the issue for consideration is whether actions can, or should, be taken in the future to mitigate these factors.

If a PCE includes a recommendation for a major process change, the recommendation should be made only in broad terms, suggesting a direction for change. A sample size of one does not provide a strong basis on which to recommend detailed changes to processes. It is not within the scope of a PCE to consider arguments for and against particular changes.

Reports should be made available to those responsible for:

- proposing and developing the initiative
- evaluating the stages
- managing or implementing those stages for current and future initiatives, and
- managing or implementing current and future PCEs.

2.20.6 Ex-post BCAs

Ex-post BCAs attempt to answer the question: 'With hindsight, how strong was the economic justification for the initiative?' They differ from the process and outcome PCEs discussed elsewhere in this section in several ways:

- Ex-post BCAs are concerned with the overall outcome of the initiative, not individual components.
- Like conventional BCAs, ex-post BCAs assess how well the initiative met the economic efficiency objective. However, an ex-post BCA may be accompanied by a broader examination of impacts of the initiative; for example, impacts on the region, the environment and other parts of the transport network.
- The standard of comparison for ex-post BCAs is not forecasts made in the past or externally established standards, but a hypothetical Base Case in which the initiative was never implemented. The Base Case is not the state of the world *before* the initiative commenced, although pre-implementation information is used for estimating Base Case characteristics.
- Ex-post BCAs involve analytical effort to estimate benefits and costs. Some elements are already known with complete certainty because they have occurred in the past. Planning, design and construction costs, infrastructure operating costs, demand levels, accident rates and physical

> levels of externalities are known with certainty up to the present time for the Project Case of an ex-post BCA. With the exception of planning, design and construction costs, these all have to be estimated for the Base Case, and projected into the future for both the Base and Project Cases.

Care is needed in specifying the Base Case. Assumption of a do-nothing or status quo Base Case can exaggerate benefits. The Base Case needs to include forecast maintenance costs. If the existing asset is expected to wear out and need replacing within the life of the initiative (or some years after), this issue should be incorporated in the ex-post BCA, as for a normal BCA.

Demand growth assumed for the Base Case is usually the same as for the Project Case. The difference in demand levels between the Base and Project Cases is generated and diverted traffic, which is likely grow at the same rate as the base demand. A good reason is needed for assuming a higher growth rate for the Project Case than for the Base Case because it implies that traffic generated or diverted by the initiative is growing faster than Base Case demand.

Case study

The following case study summarises the methodology employed in BTCE (1995), *Evaluation of the Black Spot Program*, an ex-post BCA of the Black Spot Program, pointing out some of the difficulties.

A 'black spot' is defined as 'a site, road length or area with a history of casualty crashes' (BTCE 1995, p. 5). The Australian Government's Black Spot Program provides funds to state and territory governments to implement 'treatments' for identified black spots. Treatments include new traffic lights, modifications to existing traffic lights, roundabouts, intersection channelisation and provision of medians. After the program was operating for a few years, the BTCE was asked to make an economic evaluation of the program. A sample of 254 projects was selected out of a total of 3176 projects approved during the first three years of the program's operation.

The frequencies of each crash type that occurred in the period between completion of the treatment and the evaluation provided a basis for estimating Project Case crash costs. For the Base Case, the relevant crash costs were those that would have occurred *in the absence* of the initiative, not *before* the initiative. The crash frequency before the initiative was not simply assumed to continue in the Base Case; first, because of the general trend in road crashes and, second, because of the regression-to-mean effect.

There may be a general upward or downward trend in crash frequencies due to factors that are not site-specific such as changes in driver attitudes, law enforcement or car safety. Pre-initiative crash frequencies had to be adjusted for general trends to obtain the Base Case frequencies.

The particular sites selected for treatment came to the attention of governments because they had recent high crash frequencies compared to all other sites. If crash frequencies during a given period of time behave according to some probability distribution centred on a mean, it is likely that many of the sites with the highest crash rates over the period are going through a phase of above-average crash rates. The regression-to-mean effect implies that without treatment of the site, crash frequencies in subsequent periods are statistically more likely to be closer to the average and therefore lower.

As BTCE (1995, p. 76) explains:

> Black spots are usually selected for treatment because they have experienced a large number of crashes in a recent period. Due to the statistical randomness associated with the occurrence of crashes, a site with a high number of crashes in a given period is likely to have a lower number in the subsequent period, even without any treatment. Some part of the observed 'benefits' of a treatment could therefore be illusory.

Due to inadequate historical data on treated sites, there was nothing that could be done about the regression-to-mean problem for the 1995 study. However, for a more recent Black Spot Program evaluation, BTE (2001), statistical methods were used to correct for the effect.

References

Australian Bureau of Statistics (ABS) 2000, *Survey of Motor Vehicle Use,Australia,* Cat No. 9208.0, ABS, Canberra

Australian Greenhouse Office 1999, *National Emissions Trading: Issuing the Permits,* Discussion Paper No. 2, Commonwealth of Australia, Canberra.

Australian Greenhouse Office 2004, *Australian Methodology for the Estimation of Greenhouse Gas Emissions and Sinks 2002, Energy (Transport)*, National Greenhouse Gas Inventory Committee.

Austroads 1996, *Benefit-Cost Analysis Manual*, Austroads, Sydney.

Austroads 2002, *Economic Evaluation of Road Investment Proposals: Risk Analysis*, Austroads, Sydney. Available at http://www.Austroads.com.au.

Austroads 2003a, *Economic Evaluation of Road Investment Proposals: Improved estimation of hourly traffic volume distributions*, AP-R219, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2003b, *Economic Evaluation of Road Investment Proposals: Valuing Travel Time Savings for Freight*, AP-R230, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2003c, *Valuing Environmental and Other Externalities,* Austroads Report AP-R229/03. Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2003d, *Forecasting Inter-Regional Freight Transport from Regional Development*, AP-R226/03, Austroads, Sydney.

Austroads 2005a, *Guide to Project Evaluation—Part 2: Project Evaluation Methodology*, AGPE02/05, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2005b, *Guide to Project Evaluation—Part 3: Models and Procedures*, AGPE03/05, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2005c, *Guide to Project Evaluation—Part 4: Project Evaluation Data*, AGPE04/05, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2005d, *Guide to Project Evaluation—Part 7: Post-Completion Evaluation*, AGPE07/05, Austroads, Sydney. Available at http://www.austroads.com.au.

Austroads 2006 , *Update of RUC Unit Values to June 2005*, Austroads Technical Report, AP-T70/06, Austroads, Sydney. Available at http://www.austroads.com.au

Bannock, G., Baxter, R.E. and Davis, E. 2003, *Penguin Dictionary of Economics*, 7th edition, London.

Bates, J., Polak, J., Jones, P. and Cook, A. 2001, 'The valuation of reliability for personal travel', *Transportation Research, Part E, 37(2),* pp. 191–229.

Bennathan, E., and Walters, A. A. 1969, *The Economics of Ocean Freight Rates,* Praeger: New York.

Boardman, A. E., Greenberg, D. H., Vining, A. R. and Weimer, D. L. 1996, *Cost–Benefit Analysis: Concepts and Practice*, Prentice Hall, Upper Saddle River, NJ.

Bray, D. and Tisato, P. 1997, 'Broadening the debate on road pricing', Paper presented at the 21st *Australasian Transport Research Forum*, Adelaide, September.

Bray, D. 2005, 'Improving economic evaluation of urban transport projects in Australia', Paper presented at the 28th *Australasian Transport Research Forum*, Sydney, September.

British Department for Transport 2004, *Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document*, Report prepared by B. Flyvbjerg in association with COWI. Available at http://www.dft.gov.uk/stellent/groups/dft_localtrans/documents/page/dft_localtrans_029645-01.hcsp (website of the UK Department of Transport).

BTCE 1995, *Evaluation of the Black Spot Program*, Report 90, AGPS, Canberra.

BTCE 1996, *Valuing Transport Safety in Australia*, Working Paper 26, Bureau of Transport and Communications Economics, Canberra.

BTE 1999a, *Competitive Neutrality Between Road and Rail*, Working Paper 40, BTE, Canberra.

BTE 1999b, *Facts and Furphies in Benefit–Cost Analysis: Transport,* Report 100, BTE, Canberra.

BTE 2000, *Road Crash Costs in Australia,* Report 102, BTE, Canberra. Available at **http://www.btre.gov.au**.

BTE 2001, *The Black Spot Program 1996–2002: An Evaluation of the First Three Years*, Report 104, BTRE, Canberra.

BTRE 2005, *Risk in Cost-Benefit Analysis*, Report 110, BTRE, Canberra.

BTRE 2007 forthcoming, *OZPASS: BTRE/TRA Non-Urban Passenger Travel Projection Model—User Guide: Version 1.0,* BTRE, Canberra.

Button, K. J. 1993, *Transport Economics*, 2nd edition, Edward Elgar, Cheltenham, UK.

Commonwealth of Australia 2003, *Appropriateness of 350 Million Litre Biofuels Target*, Report to the Australian Government, Department of Industry, Tourism and Resources, Canberra, December.

Cosgrove 2003, *Urban Pollutant Emissions from Motor Vehicles: Australian Trends to 2020*, BTRE Consultancy Report for Environment Australia. Available at **http://www.btre.gov.au/ea_btre.pdf**.

European Commission 1995–2005, *ExternE: Externalities of Energy, A Research Project of the European Commission*, Series of reports available at http://www.externe.info/.

Flyvbjerg, B., Skramris Holm, M. K. and Buhl, S. L. 2002, 'Underestimating costs in public works projects—Error or lie?', *Journal of the American Planning Association, 63(3)*, pp. 279–95.

Flyvbjerg, B., Bruzelius, N. and Rothengatter, W. 2003, *Megaprojects and Risk: An Anatomy of Ambition*, Cambridge University Press, Cambridge.

Flyvbjerg, B., Skramris Holm, M. K. and Buhl, S. L. 2006, 'Inaccuracy in traffic forecasts', *Transport Reviews, 26(1),* January, pp. 1–24.

Harberger, A. C. 1972, *Project Evaluation: Collected Papers*, Macmillan, London.

Harvey, M. 1996, 'Interactions between investments in road networks' in *Road-96, Conference proceedings*, Christchurch, New Zealand, vol. 6, September 1996, pp. 217–32.

Harvey, M. 2002, 'Measuring the benefits of new transport services', Paper presented at the 25th *Australian Transport Research Forum*, Canberra.

Harvey, M. 2005, 'The economic efficiency of lane differentiation policies', Paper presented at the 28th *Australasian Transport Research Forum*, Sydney.

Hensher, D.A. and Goodwin, P. 2003, *Using Values of Travel Time Savings for Toll Roads: Avoiding Some Common Errors*, Working Paper ITS-WP-03-12, Institute of Transport Studies, Sydney.

HM Treasury 1999, *How to Construct and Public Sector Comparator*, Treasury Taskforce: Private Finance, Technical Note No. 5. Available at **http://www.hm-treasury.gov.uk/media/9D527/PPP_TTF_Technote5.pdf**
#search='treasury%20taskforce%20public%20sector%20comparator.

Industry Canada 2003, *The Public Sector Comparator: a Canadian Best Practices Guide.* Available at http://strategis.ic.gc.ca/epic/internet/inpupr-bdpr.nsf/en/h_qz01557e. html.

Infras/IWW 2000, *External Costs of Transport—Accidents, Environmental and Congestion Costs in Western Europe,* Zurich/Karlsruhe.

Johansson, P. 1987, *The Economic Theory and Measurement of Environmental Benefits*, Cambridge University Press, Cambridge.

Laird, P. 2005, 'Revised land freight external costs in Australia', Paper presented at the 28th *Australasian Transport Research Forum*, Sydney, September.

Mackie, P. J., Jara-Díaz, S. and Fowkes, A. S. 2001, 'The value of travel time savings in evaluation', *Transportation Research, Part E, 37,* pp. 91–106.

Martin, T. C. 1996, *A Review of Existing Pavement Performance Relationships,* Research Report No. ARR 282, ARRB Transport Research, Vermont South, Victoria.

McIntosh, P. T. and Quarmby D. A. 1972, 'Generalised costs and the estimation of movement costs and benefits in transport planning', *Highway Research Record*, 383, pp. 11–26.

Neuberger, H. 1971, 'User benefit in the evaluation of transport and land use plans', *Journal of Transport Economics and Policy*, 5 January, pp. 52–75.

Partnerships Victoria 2001, *Guidance material: Public Sector Comparator: Technical Note*. Available at **http://www.partnerships.vic.gov.au**.

Partnerships Victoria 2003a, *Guidance material: Public Sector Comparator: Supplementary Technical Note*. Available at **http://www.partnerships.vic.gov.au**.

Partnerships Victoria 2003b, *Use of Discount Rates in the Partnerships Victoria Process*, Technical Note, July.

Perkins, F. 1994, *Practical Cost Benefit Analysis*, Macmillan, South Melbourne.

Pratt, C. 2002, 'Estimation and valuation of environmental and social externalities for the transport sector'. Paper presented at the *Australasian Transport Research Forum (ATRF) Conference*, Canberra, October.

Pressman, I. 1970, 'A mathematical formulation of the peak-load pricing problem', *Bell Journal of Economics and Management Science, Autumn, 1(2),* pp. 304–26.

Roads and Traffic Authority 1999, *Economic Analysis Manual*, Roads and Traffic Authority of New South Wales, Sydney.

Sugden, R. and Williams, A. 1978, *The Principles of Practical Cost–Benefit Analysis*, Oxford University Press, Oxford.

Transfund 1997, *Project Evaluation Manual*, Transfund New Zealand, Wellington.

Transfund 2004, *Transfund's Project Evaluation Manual, Amendment No. 8, Manual Number: PFM2, Effective 1 October 2004*, Transfund New Zealand, Wellington.

UK Department for Transport 2006, *Transport Analysis Guideance—webTAG*. Available at **http://www.webtag.org.uk**.

Watkiss, P. 2002, *Fuel Taxation Inquiry: The Air Pollution Costs of Transport in Australia*, A Report for the Commonwealth of Australia, March.

3

Rail Demand Forecasting and Cost Modelling: Demand Forecasting and Modelling

UNIVERSITY OF SOUTH AUSTRALIA

Professor Michael A P Taylor

Report to the Guidelines Assessment Methodology Working Group

July 2004

UNIVERSITY OF SOUTH AUSTRALIA

TSC Transport Systems Centre

3.1 Introduction

This paper provides information on demand forecasting and modelling in a multi-modal transport system context, on the basis of a summary of existing knowledge and an assessment of accepted good practices. There is an underlying emphasis on freight transport systems, although much of the good practice relates to, and is derived from, passenger travel, which historically has been the subject of most of the research, development and application in transport modelling. This paper attempts to develop common concepts for both passengers and freight, and to highlight those areas of analysis where differences may be expected.

The paper begins with an overview of transport modelling, indicating how modelling is performed in the context of a supply-demand equilibrium, and that the outputs of a transport model may then be used in analysis of transport impacts including economic, financial, social and environmental impacts.

Sections 3.3–3.6 consider market segmentation for passenger and freight transport, the importance of price and service quality variables in influencing decisions about transport choices and therefore demand for transport services, methods for the measurement of service quality, and the assessment of levels of customer service through the use of relevant performance indicators. A discussion of demand elasticities is then provided. This discussion ties these sections together. Demand elasticity analysis provides a powerful method for estimating effects on demand for a given transport service or mode or of changes to that service or mode or of changes in an alternative service or mode.

Section 3.7 introduces discrete choice models, which are widely used to estimate transport behaviour and market shares, including modal split. These models, based on the principle of utility maximisation by consumers, provide powerful analytical tools that may be used on policy formulation and evaluation, as they identify critical factors influencing consumer choices.

The use of statistical models and data for estimating models and their parameters, including discrete choice models and time series analysis, is introduced in Sections 3.8 and 3.9. These models may be used for the prediction of future demands. Section 3.10 provides summary information on available computer software for use in model development and parameter estimation.

Sections 3.11–3.13 consider issues in data collection, including a general methodology for surveys, an introduction to survey techniques, and the statistical methods for estimating desirable survey sample sizes.

The final section of this paper is a bibliography, providing cited references and further reading for each section of the Guidelines.

This paper should be read in conjunction with the paper, 'Estimation of impacts of rail infrastructure improvements', which is included as Part 4 of this Volume. It includes information on representative costs in rail systems and features of rail operation that are relevant to demand forecasting and modelling.

3.2 Overview of transport modelling

Modelling provides a powerful tool for the planning and design of transport systems. While practitioners may be able to develop and assess simple systems using a mental picture of a proposed system and by applying manual methods of analysis, this approach becomes less useful when networks and their associated management and control systems become more complex. Mathematical, or computer-based, models capable of representing the behaviour of real world systems, offer a valuable alternative in these cases. These models can take care of data management as well as the required calculations, and can present the results of an analysis so that the analyst can concentrate on devising alternative scenarios, interpreting the results, and making decisions. Modelling can serve the following functions:

- *Testing ideas and theories* without the need to install costly prototype systems or disrupt existing operations. Model-based testing also provides a mechanism for indicating the relative influence of different factors and for identifying cause and effect relationships.
- *Prioritising options* for which models may be used to test, evaluate and rank alternatives, and to suggest the most promising options for further, detailed study.
- *Minimising risk* as models can be used to undertake sensitivity analysis of a given system in which there is uncertainty about the values of some of the parameters or when different operating conditions may be imposed.
- *Explaining ideas* as a good model package will provide a valuable tool for developing understanding of a complex logistics system and will offer visualisations of the system at work and of the likely consequences of alternative scenarios and management strategies.

Modelling can have a key role at all stages of the planning and implementation cycle—strategic, tactical and operational. At the strategic planning level, models can assist with long-term resource planning for the design and upgrading of networks and infrastructure systems. In tactical planning, models can assist in 'fine tuning' or incremental adjustments to systems resource needs and overall operation, for instance in anticipating and responding to changing market conditions. Models for operational planning can assist with the day-by-day allocation of resources and the management and control of daily or real-time operational procedures e.g. scheduling daily vehicle fleet operations and crew rosters.

Modelling can be applied to individual components of a logistics system or to the system as a whole. For example, there are specific techniques for analysing a given component of a logistics system, such as inventory models. When modelling complex logistics systems, there are different approaches that may be adopted. The nature and form of the model will depend on the particular approach adopted. Following the work of D'Este (2001), we can identify three major approaches to (freight) transport systems modelling:

- optimisation or 'prescriptive' modelling
- simulation or 'descriptive' modelling, and
- network modelling.

Optimisation is a mathematical modelling approach designed to find the best course of action subject to a prevailing set of conditions or 'constraints'. The constraints indicate the bounds of feasibility that apply. Thus the optimisation approach may be used to allocate limited resources to specific activities to achieve the best outcome in terms of (say) maximum profit or of minimum operating cost. It may be termed 'prescriptive' because the approach prescribes a desired outcome for a system as defined by the constraints. The optimisation model then finds a set of system settings that lead to that outcome.

Simulation modelling takes a different approach. While optimisation seeks the best solution or decision, simulation mimics the behaviour of the system. It then allows for the testing and evaluation of alternatives using a 'what if' approach, in which the analyst changes some component of the system and runs the model to compare performance between the original version and the modified version. Simulation models use a combination of mathematics and logical relationships to represent the factors and interactions within a logistics system. A useful way to view a simulation model is as a 'laboratory' version of the real world system. These models are thus descriptive. Young, Taylor and Gipps (1989) discussed the process and steps in the development of a simulation model.

The third general approach to transport systems modelling is the network approach. This has long been the preferred approach for modelling flows in passenger transport systems, and there are many established model packages in this field (e.g. see Ortuzar and Willumsen, 1994). In many instances, network models can be converted to an equivalent set of equations in an optimisation model format, so there is some commonality between these two approaches. Hensher and Button (2000) provides a useful description of network modelling concepts, theories and approaches.

The strength of the network modelling approach is that it is well suited to modelling transport systems and can efficiently represent and optimise a large, complex transport network. In addition, the formulation of the logistics system as a network model is relatively easy to comprehend and visualise because there is a direct correspondence between model components (nodes and links) and real world features. However, as for the other modelling approaches, a network model is still a highly idealised representation of the real world logistics system. In particular, the standard algorithms applied in network models are not particularly good at accounting for the discontinuities and granularity (or 'lumpiness') that characterise the logistics systems that underpin freight transport. In logistics systems, for example, the product is likely to be expressed in terms of individual items while consignments are in units such as shipping containers or cartons, which are then consolidated for transport and supplied in units such as truck or rail car or ship loads. There can thus be a mismatch between the units of 'flow' used at different points in a logistics chain. Network models generally consider flows as continuous, such as the number of passengers using a public transport system or the numbers of cars on a road. While these are still discrete flow units the continuous flow analogy is a useful approximation for them. This is much less true for logistics systems. D'Este (1996) provides a full critique of the network modelling approach.

Modelling the performance of transport systems requires an appreciation of the balance or equilibrium between demand and supply. The schematic diagram of Figure 3.1 shows the overall modelling framework. The framework defines three component models and their data and information needs. This particular framework is taken from Tanaguchi, Thompson, Yamada and Van Duin (2001), and is itself based on the generic transport systems modelling framework proposed by Taylor, Bonsall and Young (2000). The three component models are:

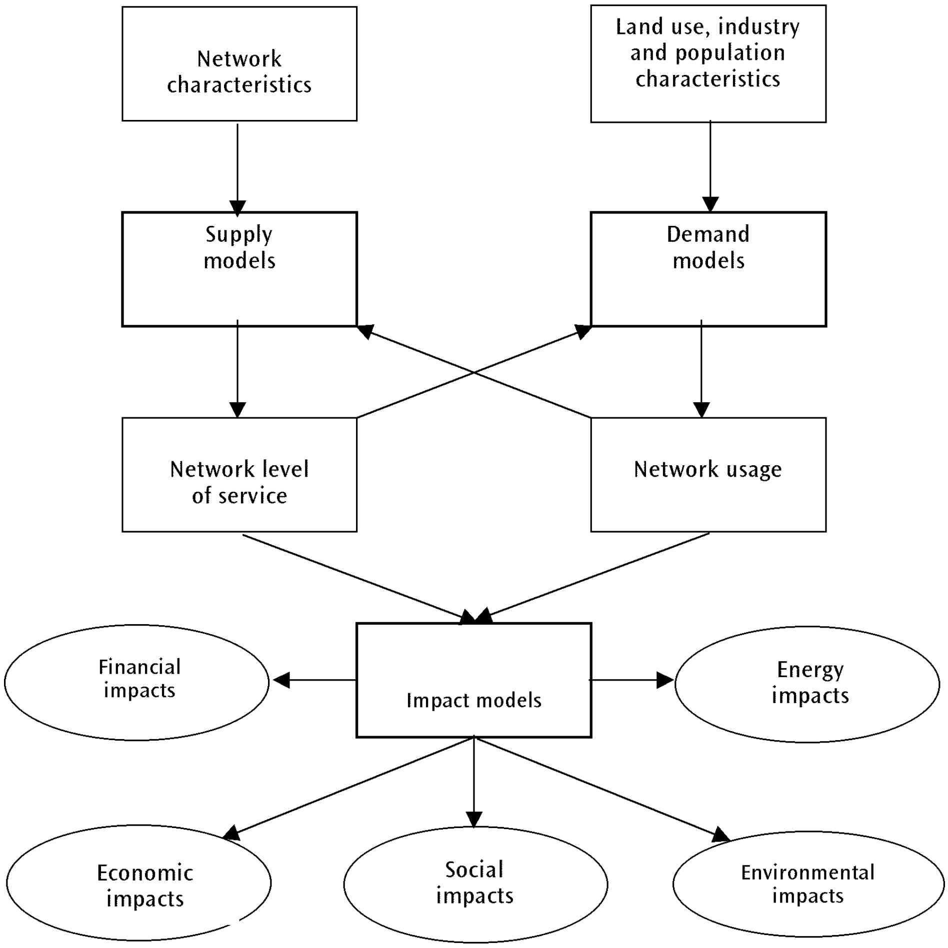
1. a *supply model* that uses the network characteristics (link and node characteristics) and an estimate of network usage (e.g. link flows) to estimate the network level of service (e.g. link travel times)
2. a *demand model* that uses the population and industry characteristics of the study region (e.g. land use distribution) and the estimate of network level of service to estimate the network usage, and
3. a set of *impact models* that may use the outputs of the combined demand-supply modelling subsystem to estimate impacts in terms of financial costs, economic impacts, social impacts, environmental impacts and energy impacts.

The supply model and the demand model operate as a combined model subsystem, iterating until a balance between network usage and level of service is obtained. This balance is a vital part of a transport model—the demands for different transport services and modes determine the operating state of the transport system and its components, while the absolute and relative performance levels of the components and the system set the conditions that determine the demands. In other words, the inputs to the demand submodel are the outputs of the supply submodel, and vice versa. This iterative modelling subsystem within the overall modelling framework is highlighted in Figure 3.1.

The impacts are generally those affecting other systems related to but not part of the transport system. These may be seen as externalities.

Jones, Simpson and Alchin (2001) described the use of a large-scale travel demand model (the Sydney Transport Model developed and maintained by the New South Wales Transport and Population Data Centre) of supply-demand equilibrium in the strategic planning of a railway system.

Figure 3.1: Schematic diagram of an overall framework for transport systems modelling



Source: Taniguchi, Thompson, Yamada and Van Duin (2001)

3.3 Market segmentation

The starting point for analysis of travel demand is to note that travel is almost always a derived demand. Travel occurs and goods are shipped because people want to undertake specific activities at different locations in an area, or because goods and commodities are required at different spatial locations to where they were produced or are stored. Thus the transport activity only occurs because of some other underlying demand.

Demand characteristics (price, income and cross elasticities; sensitivity to time; damage to freight; comfort for passengers; growth rates) and transport costs (packaging, type of service demanded) will vary for different segments of the market. In estimating quantities of freight that could switch between road and rail as a result of changes in relative prices or service qualities, it is useful to segment the market into types of freight that are being transported and their origins and destinations because rail can generally compete effectively only for long-distance freight that is not time sensitive.

Passenger trips may be categorised in many ways, including factors such as trip purpose, trip frequency, trip timing, trip distance and spatial separation of origin and destination (O–D), and travel mode used. Further, the socio-economic characteristics of individual travellers and the households to which they belong are also important determinants in predicting the travel behaviour of those individuals (see Section 3.7). The breakdown shown in Table 3.1 can be considered as a broad brush categorisation of passenger travel.

Table 3.1: Classification of passenger travel in terms of trip purpose, trip frequency, trip timing, spatial separation and transport mode

PURPOSE	FREQUENCY	TRIP TIMING	TRAVEL DISTANCE	TRANSPORT MODE
Work Education Shopping Personal business Work related Social Recreational	Regular Infrequent Occasional 'One-off'	Peak period Business hours Off-peak Late night Weekday Weekend	Local To city centre Inter-suburb Regional Inter-city Inter-state International Origin–destination* *within a locality* *separated, good access separated, difficult access* *through*	Private car: *driver* *passenger* Public transport: *rail* *bus* *tram/LRT* *taxi* *ferry* *aircraft* *ship* Non-motorised: *walk* *bicycle*

*As far as origins and destinations are concerned, the major distinction will be between O–D pairs that are in close proximity (i.e. a trip from origin to destination is completely local in nature and therefore could be made easily on foot, or might never be made by train because the station accessed from the origin is the same as that used to access the destination), within a locality, and those where the origin and destination are separated so that mechanised modes of transport such as rail that could be used for them, are separated. See Figure 3.2 for a pictorial representation of 'within a locality' and 'separated' origin-destination movements. There is also a question of the ease of access to (origin) or from (destination) rail or bus services for a given trip. Good access implies an ability to conveniently walk from home to the station or stop or from the station or stop to the destination. Difficult access means that some intermediate transport modes (e.g. car passenger as a 'kiss and ride' user, car driver as a 'park and ride' user, bicycle or local feeder bus) would be required to reach a line haul public transport service. The degree of access is a major factor influencing travellers' choices of transport modes.

Knowledge of spatial patterns of travel demand is used in transport planning for network and service design. A common method of describing travel demand in a region is through the use of origin-destination (O–D) matrices, which are tables of trip or commodity movements between the various O–D pairs that exist in a study region (Taylor, Bonsall and Young 2000, pp.114–116). Consider the schematic map of such a region as shown in Figure 3.2. The study region is identified by a cordon line drawn around it. Travel movements across the cordon line indicate trips made to and from the region. These are external origins and destinations. Observations on the cordon line can be used to assess the numbers and patterns of these through trips. Internal or local trip movements may also exist, for trips in which either or both of the origin and destination of the trip are located inside the study area. Further information on through trips (e.g. routes chosen for the segments of those trips inside the study area) and on local trips can be gathered by defining screenlines inside the study area, and then making observations of vehicle movements at the screenlines. Traffic management studies are often concerned with the split of proportions of through traffic and local traffic (the local trips) in the study area.

Figure 3.2: Origins and destinations and trip movements in an identified study region

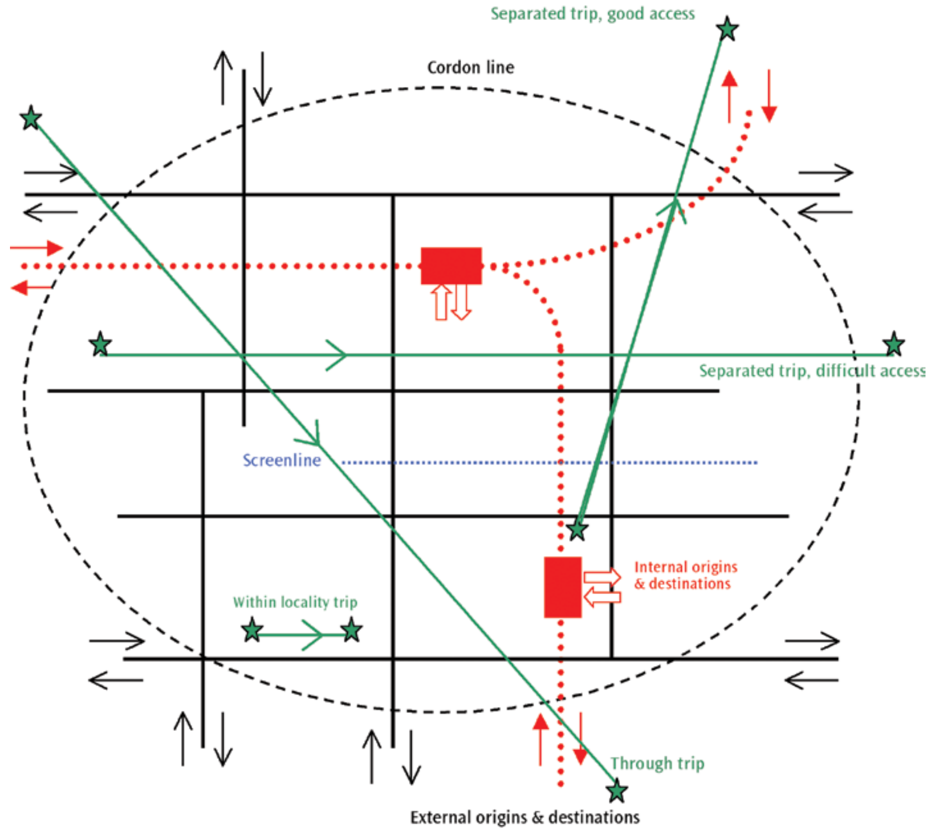


Figure 3.2 shows different O–D configurations. These include through trips and local trips (those that have at least one trip end—origin or destination—inside the study area). Local trips may be further subdivided into the categories of within locality, separated with good access and separated with difficult access, as indicated in the figure.

A typical structure for an O–D table is given in Table 3.2. Travel movements may be expressed in units of vehicles, passengers or commodity flows.

Table 3.2: Typical structure of an origin–destination matrix of travel movements

ORIGIN		DESTINATION	
		External	Internal
		1 … M_1	1 … M_2
External	1 .. N_1	Through trips	Local trips, destination in study region (separated trip, good/difficult access)
Internal	1 .. N_2	Local trips, origin in study region (separated trip, good/difficult access)	Local trips (within locality)

Similarly to passenger travel (see Table 3.1), freight trips may also be categorised using a number of dimensions. Building on the system suggested by D'Este (2000) for urban freight, the following key factors can be considered: commodity, load type, vehicle type and function, transport distance and transport mode. In addition, trip frequency and trip timing are also important, as for passenger travel. Commodities include agricultural products and food, raw materials including minerals and ores, finished manufactured goods, fuel, building and construction materials, and waste products. Table 3.3 shows a broad brush categorisation of freight trips in terms of these factors.

Table 3.3: Classification of passenger travel in terms of commodity, load type, vehicle type, transport distance and transport mode

COMMODITY	LOAD TYPE	VEHICLE TYPE	TRANSPORT DISTANCE	TRANSPORT MODE
Agricultural products and food	Full truck load	Light commercial vehicle	Local	Road vehicle
Raw materials	Partial truck load	Rigid truck	Inter-suburb	Rail
Finished goods	Individual consignment	Articulated truck	Regional	Pipeline
Fuels	Container	Truck/trailer	Inter-city	Air
Building and construction materials	Bulk liquid	Road train	Inter-state	Waterborne
Waste products	Bulk solid	Flat bed rail wagon	International	
		Hopper rail wagon		
		Tanker rail wagon		
		Double stack rail wagon		
		Ship		
		Aircraft		

Freight flows are also described using O–D matrices. Raimond, Peachman and Akers (1999) illustrated the use of O–D matrices for describing freight movements in a large urban region (Sydney).

Temporal distributions of travel demand are also important. Travel demand varies over the hours of the day, the days of the week and the weeks of the year. Cyclic and seasonal patterns can be ascertained to describe these patterns of variation and these patterns used to predict the demands for transport services. Temporal variations of travel demand are considered in Section 3.9.

3.4 Measurement of price and service quality variables

Service quality provides an overall measure of the convenience and attractiveness of a given service or operation. It will include the total time taken for the trip or operation to be completed for a given service user, and possibly the components of that total time (e.g. access time, waiting time, riding time, stopped time, etc.), as well as service frequency, reliability, comfort and cleanliness for passengers, freedom from damage or pilfering for freight, safety and security. Time taken, service frequency and service reliability are perhaps the most important factors affecting service quality. They are also the easiest of these factors to measure and assess.

Time taken is perhaps the basic factor to be considered for most travel. In addition to 'up-front' money costs incurred in undertaking travel and transport activities (such as fares and charges), the major component of overall cost of transport is travel time. Associated with travel time is delay. A number of travel time and delay components and parameters may be defined. Given that the overall travel time for a given trip is T, then this time can be said to be composed of a free flow component of travel time (T_0) and an additional component (D) which can be defined as the total or 'system' delay incurred in the journey, i.e.

$$T = T_0 + D \quad (4.1)$$

This basic relationship suggests some important concepts with regard to travel time and delay. Firstly, some finite time (the free flow time, corresponding to the minimum time required to traverse the network for the specified trip given that there is no hold up or delay incurred e.g. because of other traffic using the network) will always be required to complete the travel. Secondly, while delay is often made up of small and separate amounts incurred at bottlenecks, hold-ups due to other vehicles using the system, waiting for connecting services or standing in queues, we can sum these components together to form a total delay time. In some situations, there may be need to isolate individual components of delay, for instance in analysis of public transport rider-ship it is common to differentiate between time spent waiting at a stop for the transit vehicle and time spent in the vehicle (even when the vehicle is stationary). This is done because travellers are seen to place different weights on waiting time at a stop compared to in-vehicle time (a ratio of 2:1 is often assumed) and such weights would be reflected in the utility function associated with the given transport mode in discrete choice modelling (see Section 3.7). The question of the ease of access was also noted in Section 3.3.

An alternative split of travel time sometimes used in traffic engineering is between running time (T_r, the time that a vehicle is in motion) and stopped time or 'stopped delay' (D_s, the time the vehicle is at rest over the journey). Then

$$T = T_r + D_S \qquad (4.2)$$

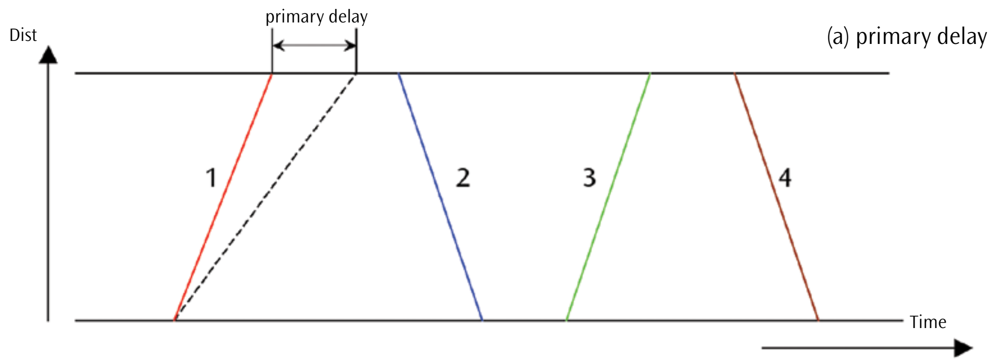
This relation applies to air and rail transport just as it does to road transport. There is no direct comparison between the components of equations (4.1) and (4.2), except that we can say that $T_r \geq T_0$ and $D_s \leq D$. This implies that stopped delay is a lower bound on total delay time. This is logical because the braking and acceleration behaviour that a vehicle will engage in over its journey must add additional travel time to the journey (compared to the free flow time) even if the vehicle never came to rest in that journey. Stopped time may be easier to measure than total delay time because a definite criterion can be established to ascertain when the vehicle is stationary (i.e. has zero speed, or a speed less than a set threshold such a 2.0 km/h). Nevertheless, delay is more than just stopped time.

Delays may also be divided into recurrent delay and non-recurrent or incident-based delay. Recurrent delay reflects regular variations in travel demand over time e.g. over the hours of the day or days of the week. These delays reflect the general differences in travel times between peak and off-peak periods. It is generally assumed that travellers and network users have some reasonable knowledge of these recurrent variations in travel times and plan their trips using this knowledge. Incident-based delay on the other hand is unpredictable. It occurs as a result of random or at least unforeseen events on the network, such as breakdowns, accidents or even maintenance works that may occur at any location and differ widely from day to day. Network users cannot often escape these sources of delays, although modern Intelligent Transport Systems technologies may be able to provide users with some advance information about their occurrence in real time. Non-recurrent delay is consequently weighted more highly than recurrent delay in determining traveller behaviour.

In railway systems, delays are also defined in terms of primary delay and secondary delay components. Primary delay is delay time accruing to a train due to its own performance, when it suffers travel time in excess of the scheduled travel time because of abnormally long technical operations (e.g. locomotive change) or commercial operations (e.g. delays in loading goods or passengers, or in attaching or detaching wagons). Other sources of primary delay would include mechanical failure and environmental conditions (e.g. adverse weather or line blockages from external causes). Secondary delay occurs to a train whose progress is interrupted because of delays experienced by other trains. If there is congestion the primary delay generates repeated secondary delays (for other trains). A train is said to be operating effectively if it incurs no secondary delay.

These delay components can be seen in Figure 3.3, which presents two simple train path diagrams. A train path diagram is a plot of the position of a train along a rail line over time. This plot shows the trajectory of the train. The slope of the trajectory line indicates the travel speed of the train at a given point on its journey—a horizontal line segment on the trajectory means that the train was stationary at that point. These diagrams are used in train scheduling, amongst other things.

Figure 3.3: Simple train path diagrams showing primary and secondary delays to individual trains



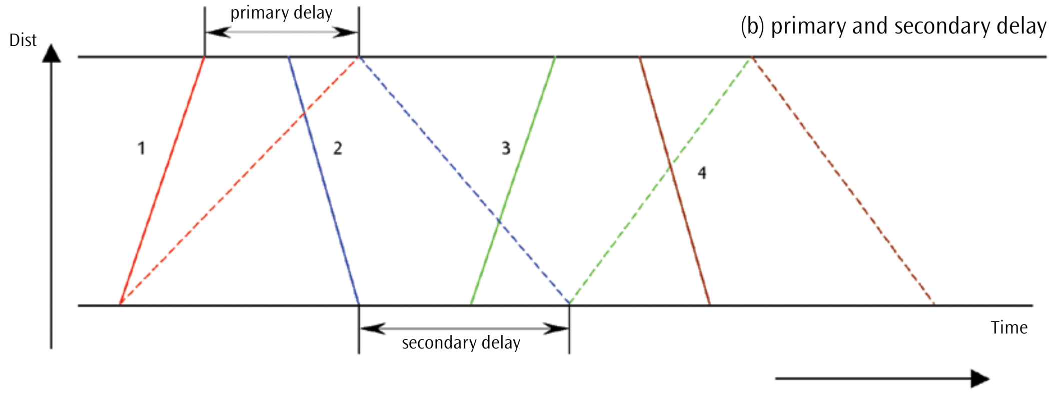


Figure 3.3 (a) shows a set of trains using a single line track and travelling in opposite directions (indicated by the positive or negative slopes of the trajectory lines). In this case, train 1 experiences some hold up, which is a primary delay to that train. None of the other trains experience any disruption to their progress. Figure 3.3 (b) shows a situation in which the primary delay experienced by train 1 causes a secondary delay to train 2 (which cannot enter the line section until train 1 has exited it), and in fact the secondary delays cascade as trains 3 and 4 also encounter secondary delays.

It is important to note that the travel times of all individual vehicles making the same journey or travelling on some part of a network will also vary, so that travel times (and delays) are best studied through the use of a statistical distribution of individual times. Travel time and delay distributions generally exhibit positive skew i.e. there will be a tendency for a few individual times to be much larger than the mean travel time ($\overline{T}$), and there will not be the same tendency for corresponding very short times to be observed. Thus distributions such as the Gamma distribution or the log-normal distribution are often used to represent travel times and delays. A useful measure of the variation in travel times is the coefficient of variation, defined as the ratio of the standard deviation of the distribution of times to the mean time. This dimensionless parameter is also used as an indicator of the reliability of travel time on a network. Other related measures of system performance may be obtained from these distributions of travel time and delay. These are percentiles from the distributions, such as the percentage of individual travel times exceeding some set limit (either an absolute travel time, say the percentage of travel times exceeding 90 minutes, or a relative travel time, say the percentage of travel times more than three times the average travel time, or a direct percentile value such as the travel time exceeded by ten per cent of travellers—this would be the 90th percentile travel time).

One problem with the use of actual travel times is that it is difficult to compare these between different parts of a network, e.g. between different links when one link is much longer than the other. One way of being able to make comparisons is to use a dimensionless congestion index e.g.

$$CI = \frac{T - T_0}{T_0} = \frac{D}{T_0} \qquad (4.3)$$

This parameter is the ratio of the total delay to the free flow travel time, and it can be compared between different network components, transport modes and transport services. The use of this parameter and other parametric measures of congestion is described in Taylor, Woolley and Zito (2000).

Distributions of congestion indices can be developed from sets of observed travel times and delays using equation (4.3). These distributions will also tend to have positive skew, and may be subject to the same analysis as the raw distributions (travel times and delays).

Service frequency indicates the rate at which services are provided over a given time period. The reciprocal of service frequency is headway, which is the average time between successive services. Low service frequencies mean long headways and therefore longer delays—having just missed one service there is then a wait of one headway until the next vehicle arrives for the next service. High service frequency implies more flexibility for users, and therefore more choices and better service quality. Service frequency would be measured as the average number of services per unit time (e.g. the average number of buses per hour on a route). Consideration may also be given to the variability in service frequency, especially in congested networks. This could be done by measuring both the mean and the standard deviation of headways (and hence service frequencies) and the actual performance compared to planned schedules and timetables.

Bell and Iida (2003) provide examples of the study of travel time variability, network reliability and the impacts of different components of delay. Travel time reliability is a factor of considerable importance for freight deliveries, especially when the commodity being transported is time-sensitive (e.g. its quality, and therefore value, may deteriorate over time so that delivery must be made within a certain time after despatch—live seafood is an obvious example but there are many others) or in just-in-time supply chain systems where a downstream process requires regular, periodic supply of raw materials. In transport systems where travel time reliability is poor (e.g. under heavy congestion) travel time reliability may deteriorate. One measure of travel time reliability is the probability of reaching a given destination within a specified time interval—in practice, for deliveries that are repeated, this measure can be approximated by empirical observation of the range of arrival times and the proportion of times that fall into a selected interval. The conservative correction to variability may be to allow much longer scheduled travel times, but early arrival may be as worrisome as late arrival in just-in-time operations, further accentuating the importance of travel time reliability.

Building on the discussion above, the following points may be used in estimating the value of service quality improvements to customers:

- Customers can be assumed to put money values on quality attributes, so that these attributes can be used in an analysis in the same way that prices and costs to customers can be handled. There is also scope for customers to trade-off quality attributes against price, because of customers' attribution of money values to service quality variables.
- Service quality attributes and perceived equivalent money values will have statistical distributions. Any distribution has a mean value which is a first estimate of the value of that distribution to the population that it represents.
- Where projects improve service quality, knowledge of these values is necessary to estimate effects on market share and total sales, benefits for benefit-cost analysis, and revenues for financial analysis.
- Subsequent sections of this paper (e.g. Sections 3.6, 3.7, 3.9, 3.11 and 3.12) discuss the use of statistical models and survey techniques to help estimate the relevant values.

3.5 Demand elasticities

Analysis using demand elasticities is based on the assumption of a direct relationship between the change in a policy-dependent variable and the corresponding change of a particular transport choice. The elasticity of demand, with respect to a given parameter (explanatory variable), may be seen as the percentage change in quantity demanded resulting from a one per cent change in the value of the parameter[1]. Both the magnitude of the percentage change and the sense of that change (positive or negative) are important. Direct elasticity refers to the change of demand for one transport service (or mode) in terms of the change in a parameter affecting that service (or mode). Cross elasticity refers to the change of demand for a service or mode resulting from the change in a parameter affecting a different (competing) mode or service. Thus, for example, the change in patronage on a rail service as a result of the increase in the fare charged for that service could be estimated using the direct demand elasticity for rail with respect to fare. The increase to an alternative mode (e.g. bus) resulting from passengers switching modes could be estimated using the cross elasticity of demand for bus with respect to rail fare.

Similarly, the elasticity of demand with respect to income measures the proportionate change in demand resulting from a proportionate change in the income of the consumer. As consumers have more income, so their choices sets may expand or change.

Elasticity models are used to estimate these effects. If the value of the parameter is *P* and the demand (number of trips, amount of freight, etc.) is Q, then the elasticity η is given by

$$\eta = \frac{\Delta Q}{Q} \Big/ \frac{\Delta P}{P} = \frac{P}{Q}\frac{\partial Q}{\partial P} \qquad (5.1)$$

The partial derivative $\frac{\partial Q}{\partial P}$ is used to indicate that all of the other explanatory variables (parameters) that can affect demand are assumed to be held constant. This derivative is the slope of the demand curve that relates Q and *P*. The slope of the curve may vary along its length so that the value of η as defined by equation (5.1) is only valid for the particular point at which the slope is measured. It should strictly be called the 'point' elasticity.

Elasticity values can be derived for many parameters, such as fares, costs and charges, travel times, service reliability, service frequency or service quality (e.g. passenger comfort or possibility of damage to or loss of commodities). Provided that the relative change in the value of each of these parameters is small, the overall change in demand can be estimated as a simple sum of the individual change estimates

$$\frac{\Delta Q}{Q} = \eta_A \frac{\Delta P_A}{P_A} + \eta_B \frac{\Delta P_B}{P_B} + \eta_C \frac{\Delta P_C}{P_C} + \ldots \qquad (5.2)$$

for parameters P_A, P_B, P_C, etc.. Models of this type have been used for many years by transport operators. They provide a quick estimation of the likely effects of proposed service changes, based on previous experience, as long as the service changes are incremental in nature.

1 More formally, the elasticity of demand with respect to a particular parameter is defined as the ratio of the proportionate change in demand to the proportionate change in the relevant parameter.

Point elasticity as defined by equation (5.1) is rarely available for practical applications, because knowledge of the mathematical shape of the demand curve would be required to determine it—and this is seldom the case. Rather, most elasticity values used in practice are arc elasticities, estimated by considering measured differences in demand at different values of a given parameter. If the demand Q_1 corresponds to value P_1 and demand Q_2 corresponds to value P_2 of the parameter, then the arc elasticity $\bar{\eta}$ is given by

$$\bar{\eta} = \frac{(Q_2 - Q_1)}{\frac{1}{2}(Q_2 + Q_1)} \Bigg/ \frac{(P_2 - P_1)}{\frac{1}{2}(P_2 + P_1)} = \frac{\Delta Q}{\Delta P}\frac{\bar{P}}{\bar{Q}} \qquad (5.3)$$

where $\bar{P}$ and $\bar{Q}$ are the respective mean values.

The demand to use a particular mode or service can be affected by changes in travel parameters applying to competing modes as well as changes in the parameters of the given mode. The effect on demand for use of mode H of changes in a parameter P_G on mode G is given by the cross elasticity

$$\eta_H^G = \frac{\Delta Q_H}{Q_H} \Bigg/ \frac{\Delta P_G}{P_G} \qquad (5.4)$$

Elasticity values are generally estimated from time series data. For instance, transport operators may collect information on patronage or commodity movements over time, and this can be used to suggest the effects on demand stemming from changes in transport costs or other service parameters (for both their own operations and those of their competitors).

Elasticity values for passenger and freight transport and by transport mode may be found in reported studies such as Oum, Waters and Yong (1992) and Luk and Hepburn (1993). A substantial database of elasticity values has been compiled by the Bureau of Transport and Regional Economics and is available at the BTRE website www.btre.gov.au under databases and products. A selection of typical values is provided as Table 3.4. Taplin, Hensher and Smith (1999) provided a discussion on the conceptual and theoretical requirements on the estimation of elasticity values.

Certain conventions need to be followed when interpreting published values of elasticities. In the case where η denotes a demand elasticity with respect to parameters such as cost or trip time the value of the elasticity is negative—an increase in price of a commodity or service would normally be expected to lead to a decrease in the demand for it, but common practice among economists is to quote only the magnitude (modulus) of the value i.e. as a positive number, with the implicit assumption that the value is actually a negative number. Cross elasticity values are generally positive—an increase in the charges for use of a competing service would be expected to result in an increased use of the other service.

Table 3.4: Selection of freight transport elasticity values

QUANTITY	VARIABLE	COUNTRY	ELASTICITY VALUE		BTRE DATABASE REFERENCE	ORIGINAL REFERENCE
			Short run	Long run		
Road freight	Relative price road/rail	Australia	-0.37	–	ID 21, Table 1B13	Luk and Hepburn (1993)
	Road price		-0.33	–		
Rail freight	Diesel fuel price	Australia	-0.55	–	ID 11, Table 1D11	BTCE (1991)
Rail freight	Rail price	Canada	-0.64	–	ID 13, Table 1D39	Oum (1989)
	Road price		0.84	–		
Road freight	Rail price		0.06	–		
	Road price		-0.04	–		
All freight	Road price	Australia	-0.39	-0.80	ID 15, Table 1D41	Luk and Hepburn (1993)
Road freight	Road price	Australia	-0.12	–	ID 60, Table 2B04	Luck and Martin (1988)
Rail freight	Rail price	USA	-1.06	–	ID 74, Table 2C07	Hsing (1994)
	Road price		1.21	–		
	Income		0.48	–		
Rail freight	Service frequency	Australia	0.18	–	ID 165, Table 4C07	ARRDO (1982)
Rail freight	Rail price	Australia	-0.10	–	ID 460, Table 9C01	BTCE (1990)
	Income		0.59	–		
	Road price		0.12	–		
Road Freight	Income	Australia	1.93	–	ID 462, Table 9B15	BTCE (1990)
Road freight	Road price	Australia	–	-0.19	ID 479, Table 9D20	NIEIR (1990)

Source: BTRE (2004)

The demand elasticity approach to estimating changes in freight flows on rail from infrastructure and service improvements was used by Booz, Allen and Hamilton in its recent audit of the ARTC rail network (BAH, 2001). This study used three factors relating to a rail corridor—transit time, reliability and service availability—to estimate future demands following infrastructure and service improvements. Table 3.5 shows the elasticity values used in the network audit study.

Table 3.5: Service elasticity values used by Booz, Allen & Hamilton in their audit of the ARTC rail network

	PRICE	TRANSIT TIME	RELIABILITY	SERVICE AVAILABILITY
Long-haul corridor	-1.1	-0.3	0.6	0.4
Short-haul corridor	-1.1	-0.4	0.6	0.5

Source: BAH (2001, p.17, Table 4.1)

Parameters with elasticity values that have magnitudes less than one are said to be inelastic. This implies that a one per cent change in the parameter value has a relative effect of less than one per cent on demand. Parameters for most transport systems have been found to have such values.

A distinction needs to be drawn between the elasticity of demand for transport as a whole (e.g. road and rail combined) and segments of the market (e.g. road and rail considered separately between a given origin-destination pair). The elasticity of demand for freight transport as a whole (η_{FT}) depends on the elasticities of demand (η_d) and supply (η_s) for the product being carried and the fraction of the demand price that is spent on transport (f). Bennathan and Walters (1969, pp.109-112) have shown that this overall transport elasticity is given by

$$\eta_{FT} = f\left[\frac{\eta_s \eta_d}{\eta_s - (1-f)\eta_d}\right] \qquad (5.5)$$

where the product is supplied under competitive conditions. For a product available in infinitely elastic supply, the formula reduces to $f\eta_d$. For a segment of the market where customers have the option of switching to another mode or form of transport, the elasticity will be greater, and may well be highly elastic for some narrowly defined market segments (e.g. for a road corridor where users can chose between alternative routes, or for a particular truck type where customers have the option of switching to other truck types).

Elasticities also need to be considered in terms of time span. Most elasticities estimated using time series data are short run elasticities. The range of possible adjustments to changes is greater in the long run than in the short term. For instance, an increase in fuel prices may have small effects immediately (the demand would be said to be 'inelastic') because travellers and operators have little scope to change behaviour initially—they will still be using the same vehicles and engaged in the same travel—but in the long run they may make substantial changes (e.g. use of new, more fuel efficient vehicles, changes in destination requiring short trip lengths or changes of mode) so that the eventual effects are more pronounced. Thus there is a need to consider both short run and long run elasticity values and to use care in the selection of appropriate values in an analysis.

Elasticity analysis and elasticity values can be useful in project evaluation. They may be used to forecast market responses to price and service changes (in terms of both generated and diverted traffic), from which estimates of consumer surplus gains can be made. They may also be used to estimate appropriate prices to charge to obtain maximum efficiency or maximum profit.

3.6 Performance indicators for customer service

Customer service performance for 'fee for service' transport modes, including both freight and passenger services, may be assessed in terms of a number of performance indicators. In general terms these indicators may be seen as assessing either the efficiency of the service (perhaps as a utilisation of available capacity), the effectiveness of the service (as a indication of how well it meshes with customer needs) and user satisfaction (who well the service is received or perceived by actual and potential customers). There are a large number of potential indicators, and some typical formulations would include:

- *Measures of occupancy or utilisation*, such as the ratio of tonne-km carried to vehicle-km or capacity-km of service provided, or the ratio of passenger-km to seat-km.
- *Percentage of on-time arrivals*, assessed as the percentage of vehicles or scheduled services that arrive at a designated destination within a pre-set time window e.g. ($-\Delta t_1, +\Delta t_2$) of a scheduled arrival time t_A—this means that to be 'on time' the actual arrival time t of the vehicle should fall in the time interval $t_A - \Delta t_1 \leq t \leq t_A + \Delta t_2$. It is normal for on-time running to include considerations of early arrival as well as late arrival. This is certainly important in scheduled public transport services where adherence to a set schedule will be an important factor in engendering passenger confidence in the service. This will also be important for freight deliveries operating in supply chains under just-in-time principles where early arrival may be as unwelcome as late arrival (Taylor, 2001). It is also common for the time window to be asymmetric about the scheduled arrival time t_A, as indicated in the above inequality.

- Percentage of vehicles running late (or running early) is an alternative presentation of compliance with on-time service. Again set time differentials are used to indicate if a service is early, late or on time.
- Statistical distributions of vehicle and passenger travel times and delays on a daily or weekly basis, from which parameters such as mean value, standard deviation, coefficient of variation and percentile values (see Section 3.4) may be extracted. These statistical parameters may then be monitored over time to provide ongoing performance.
- Adherence to schedules, especially frequency of service. In a high frequency service regime such as public transport services in an inner city area, scheduled arrival time may be less important than actual frequency of service (alternatively measured as the headway between successive transit vehicles). The regularity of headway (frequency) may then be the relevant performance issue and the operator's concern may be to ensure regularity—for instance light rail transit (LRT) services sharing road space with general traffic may encounter hold ups due to traffic congestion. Subsequent LRT vehicles then become bunched behind a leader and are unable to pass that vehicle. This can lead to an excessive headway for the lead vehicle followed by a rapid succession of the subsequent LRT vehicles, which provides poor service to customers. The average and standard deviation of actual headways between the vehicles and comparisons with service design headway would be a performance indicator.
- Customer satisfaction indices, a variety of measures of customer satisfaction, based on the quality of the service and perceptions of this quality may also be considered. These would include the extent of service availability on demand, safety and security, vehicle cleanliness, and staff attitudes and helpfulness to customers. Factors or variables reflecting these service characteristics may be measured using attitudinal surveys of customers.

Many of these indicators can be applied to different parts of the operation of a given transport system, such as by route or by commodity type carried. The computation of performance indicators at these more disaggregate levels can provide vital diagnostic advice to service operators and system managers about overall system performance and trends and variations in performance, much as 'quality control' systems are employed in manufacturing industry. Lee (1989) explored this issue more fully for the case of urban public transport systems, but his concepts and generalised models for diagnosis of systems operation are equally applicable to other transport services. Nash (2000) considered performance indicators for use in railway planning in the United Kingdom.

Tanaguchi, Thompson, Yamada and Van Duin (2001) and Bell and Iida (2003) discuss the measurement and application of parameters reflecting service reliability, including service appraisal on the basis of late or early running and the use of penalty functions that reflect the extent of problems associated with these. Lam and Bell (2003) indicate how the service parameters ban be used in models of transport services.

Anzelark, Esterman and Homburg (2002) describe a monitoring system applied to an urban public transport system to gauge the level of customer satisfaction with services.

3.7 Discrete choice models

Discrete choice models have been employed widely in travel demand analysis since the 1970s, with the most common application being in the choice of travel mode. Modal split is the relative proportion of travellers or shippers using one particular mode compared to the other available modes. Most models of modal choice use a 'utility' function representation of the attributes of the different modes and of the travellers or shippers as a main set of independent (explanatory) variables. The utility function is usually a weighted sum of the modal and personal attributes considered (e.g. travel time and reliability, travel cost, service frequency and socio-economic characteristics). The simplest choice models consider only two alternatives (e.g. mode *A* and mode *B*) and are known as binary choice models. In general terms a binary choice model can be expressed as

$$\frac{p_A}{p_B} = F(U_A, U_B)$$

where p_A and p_B are the probabilities of choosing modes *A* and *B*, U_A and U_B are the utility functions for modes *A* and *B*, and $F(U_A, U_B)$ is some suitable function. The models are often expressed for one of the modes only, for example, as

$$p_A = f(U_A, U_B)$$

and they can be extended to include more than two alternatives in the choice set. The discrete choice models are often termed 'behavioural' because they can represent causality in that they can be derived from a theory that explicitly maps out the decision-making processes of the individual taking the decision. The theoretical basis is usually that of utility maximisation. It assumes that the utility an individual ascribes to an alternative is defined by a utility function in which the attributes of the alternative and characteristics of the individual are determining factors. The choice of a particular alternative is made on the basis of comparing the levels of utility derivable from each of the available alternatives.

Of necessity, the models estimate the probability that an individual, in a given situation, will choose a particular alternative rather than the definite selection of a preferred alternative. Assume that individual *i* can choose one alternative *r* from a set of *K* available alternatives and that the utility of alternative *r* is for that individual is given by U_{ri}. Then alternative *r* will be chosen if

$$U_{ri} \geq U_{ki} \quad \text{for all} \quad r \neq k \in K.$$

Given that there will almost always be some uncertainty concerning the specification of the utility function—because of measurement errors, omission of unobserved attributes and other specification errors—utility functions have a random component. Thus it is only possible to determine the probability that a given alternative will be chosen. We can represent the utility function U_{ri} as

$$U_{ri} = V_{ri} + \varepsilon_{ri}$$

where V_{ri} is the deterministic part of the utility function and ε_{ri} is the random part. Then the probability that individual *i* will select alternative *r* can be written as

$$p_{ri} = \Pr\{U_{ri} \geq U_{ki}\} = \Pr\{V_{ri} + \varepsilon_{ri} \geq V_{ki} + \varepsilon_{ki}\} = \Pr\{V_{ri} - V_{ki} \geq \varepsilon_{ki} - \varepsilon_{ri}\} \text{ for all } k \in K \mid k \neq r.$$

Specific mathematical forms of the choice model then emerge depending on the assumptions adopted about the form of the joint distribution of the random errors $\varepsilon_{ki} - \varepsilon_{ri}$. If this distribution is assumed to be the normal distribution, then the choice probability model is the probit model. Unfortunately, this model is mathematically intractable. As a result, the practice is to assume that the distribution follows Weibull distribution, which approximates the normal distribution to some degree. The advantage of this assumption is that the resultant choice model is the multinomial logit model, which is mathematically tractable. The functional form of the multinomial logit model is

$$p_{ri} = \frac{\exp(U_{ri})}{\sum_{k \in K} \exp(U_{ki})} \qquad (7.1)$$

The binomial form of this model—using the earlier notation of alternatives *A* and *B*— is

$$p_A = \frac{\exp(U_A)}{\exp(U_A) + \exp(U_B)} = \frac{1}{1 + \exp(U_B - U_A)}$$

This function is such that if U_A and U_B are equal then the probability of choosing each of the two alternatives is 0.5, while if $U_A > U_B$ then the probability of choosing *A* is greater than that of choosing *B*. The utility functions are generally weighted linear functions of the attributes e.g.

$$U_{ri} = \alpha_r + \sum_{j=1}^{J} \beta_{rj} X_{rji} + \sum_{l=1}^{L} \gamma_l Y_{li} \qquad (7.2)$$

where α_r is an alternative-specific constant, the $\{\beta_{rj}, j = 1, \ldots, J\}$ are constant coefficients for the attributes (e.g. service variables) $\{X_j\}$ of the alternative and the $\{\gamma_l, l = 1, \ldots, L\}$ are constant coefficients for the attributes (e.g. socio-economic characteristics) of the individual decision-maker *i*.

The coefficients of the utility function may be used to provide further information. For example, if a utility function takes the form

$$U = \alpha + \beta_1 \times price + \beta_2 \times time + \beta_3 \times reliability \qquad (7.3)$$

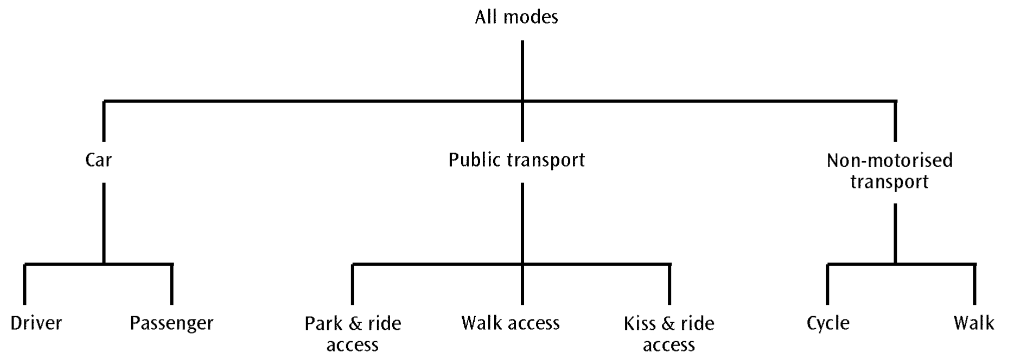
then by dividing both sides of equation (7.3) by β_1 yields the money value of time $(\beta_2\ \beta_1)$ and the money value of reliability (β_3/β_1).

Coefficients in the utility functions are generally estimated from observed data sets using maximum likelihood techniques (Section 3.8) implemented in software packages such as LIMDEP (see Section 3.10).

While the multinomial logit model of equation (7.1) is a powerful tool for understanding travel choices, it has some significant limitations. The most important of these is its reliance on the 'Axiom of the Independence of Irrelevant Alternatives (IIA)' (Luce, 1959), which states that '*if a set of alternative choices exists, then the relative probability of choice among any two alternatives is unaffected by the removal (or addition) of any set of other alternatives*'. This means that the ratio p_A/p_B is independent of the other alternatives available in the choice set. This property is the basis of the multinomial logit model. Unfortunately, while the model is attractive and easy to use, it really applies only in rather special circumstances and its use in practice can lead to certain anomalies (e.g. the 'red bus-blue bus' anomaly[2]). The general solution is to use nested logit models that present a hierarchy of choices in which the decision-maker usually has to choose between no more than two alternatives at any point in the nested structure. Binomial logit models are generally used at each of the decision points. An example is shown in Figure 3.4, which is the nested logit model for mode choice applied in the Western Australian Department for Planning and Infrastructure's STEM (Strategic Transport Evaluation Model) model for travel demand analysis in the Perth metropolitan area.

2 Consider the case where travellers can choose between two modes—say car and bus—to make a given trip, and further assume that the utility values for both of these modes are equal. Then there are probabilities of 0.5 for the use of both modes. Now assume that the public transport operator paints half of the buses red and the other half blue. There are now three modes apparent: car, red bus and blue bus. All modes still have the same utility values, so the ratio between any pair of modes is unity, and the overall modal split is (according to the multinomial logit model) one third car, one third blue bus and one third red bus. The proportion of travellers using cars has then decreased from 0.5 to 0.33, but nothing has actually changed except the colours of the buses. This is an illogical result and is due to the fact that the two bus modes are actually variations of the same choice alternative for the travellers—they are not truly independent alternatives.

Figure 3.4: Structure of the nested logit modal choice model in the stem model of the multimodal transport system in metropolitan Perth



Source: Western Australian Department for Planning and Infrastructure.

This model incorporates three broad modes of travel: car, public transport and non-motorised. The car mode is split into car driver and car passenger. Public transport is split into three separate elemental modes, depending on the mode of access taken to use the transit services (note that this model does not distinguish between rail or bus services). Non-motorised transport is split into bicycle and pedestrian modes. Douglas, Franzmann and Frost (2003) describe the development of similar discrete choice models for modal choice in Brisbane.

The use of multinomial logit models in freight transport studies is illustrated in Wigan, Rockliffe, Thoresen and Tsolakis (1998). This study used the models to estimate the value of time spent in transit and reliability of arrival time for both longhaul and metropolitan freight.

More complicated, but more generally applicable, choice models are also available e.g. the mixed logit models (see Louviere, Hensher and Swait 2000; Train, Revelt and Ruud 2004). While the standard logit model assumes that the utility function coefficients are the same for the entire population (e.g. everyone has the same values of time and of other quality attributes), the mixed logit model allows the analyst to make the more realistic assumption that the coefficients vary across the population according to some distribution (e.g. uniform, normal, log-normal). The mixed logit model relaxes the assumption of IIA, and allows correlation in the unobserved components of utility between alternatives. The model is, however, much more data intensive and computationally exhaustive.

It must be noted that the reliance on the 'independence of irrelevant alternatives' axiom does not invalidate the multinomial logit model. It is a perfectly acceptable model as long as care is taken to ensure that its basic assumptions are not violated in a given application.

One advantage of the multinomial logit model is that it is possible to derive point elasticity values from it. Considering the choice model defined by equations (7.1) and (7.2), it can be shown that the elasticity of the probability of choosing mode *A* for individual *i* with respect to changes in X_{rji} (the *jth* independent variable of alternative *r*) is given by

$$\eta^{r}_{Aji} = \left[\delta_{Ar} - p_{ri}\right]\beta_{rj} X_{rji}$$

where δ_{Ar} is a delta function defined as

$\delta_{Ar} = 0$ if $A \neq r$ (cross elasticity)

$\delta_{Ar} = 1$ if $A = r$ (direct elasticity).

This result indicates that the direct elasticity for alternative *A* depends only on the attributes of that alternative, while for the cross elasticities only the attributes of the other alternative *r* enter the equation.

Properly established discrete choice models are powerful decision-support tools for both policy analysis and project evaluation. At a basic level, the models can be used to estimate changes in market shares across modes or for alternative routes if a project is undertaken that will reduce the price or improve one or more of the service quality attributes for one mode or route. The models may also be used to help a transport operator determine the price to charge that will maximise profits. At a more complex level, the models can be used to estimate the welfare gains to customers from an improvement in a service quality attribute, which is important for inclusion in benefit-cost analysis.

For further information on the development, estimation and application of discrete choice models, the determination of elasticity values from discrete choice models, and the application of the models in project evaluation, see Ortuzar and Willumsen (1994); Taplin, Hensher and Smith (1999); Hensher and Button (2000) and Louviere, Hensher and Swait (2000).

3.8 Estimation techniques

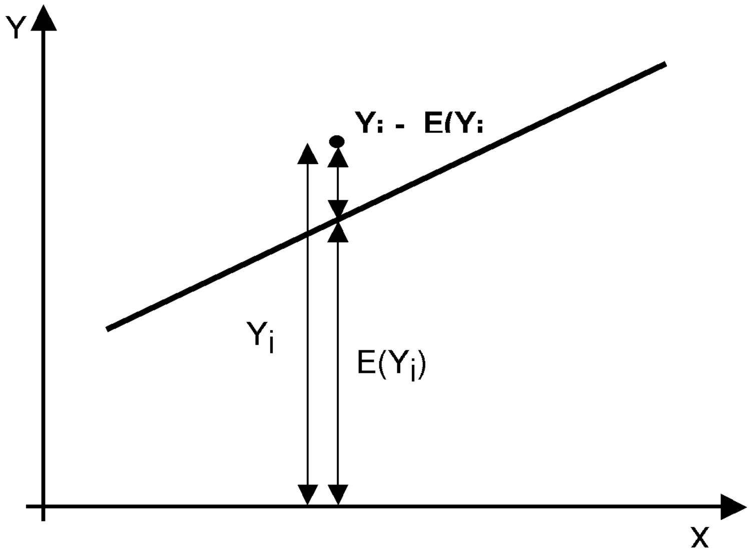
Frequently, analysts may wish to develop relationships between the factors contained in their data. These relationships are statistical models and provide a useful, but simplified, view of the process being studied. This simplicity is limited by the need to provide a reasonable representation of the process. Statistical models use sample data to develop a mathematical relationship between factors, often of the form

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \ldots + b_m X_{im} + \varepsilon_i \qquad (8.1)$$

where the Y and X_j terms represent the variables considered in the model, the subscript i represents the i^{th} observation, the b_0, b_1, ... b_m terms represent the model parameters, and the ε_i is an error term. The variables can also be divided into dependent and independent variables. Dependent variables (Y) are the variables to be explained by the model. The right-hand side of equation (8.1), less the error term, is the expected value $E[Y_i]$ of Y_i given a set of observed values X_{ij}, $j = 1 \ldots m$. The independent (explanatory) variables (X_j) are the variables providing the explanation.

Regression is a general statistical tool through which the analyst can develop a relationship between dependent and independent variables. It can be used to develop descriptive relationships or investigate causal relationships. The first step in developing a relationship is the definition of what constitutes a good fit. The answer is usually a fit that leaves a small total error. One error type is shown in Figure 3.5. It is defined as the vertical distance from the observed data point Y_i to the corresponding fitted data point $E(Y_i)$. Mathematically this is the difference $(Y_i - E(Y_i))$. This error is positive when the data point is above the line and negative when it is below the line. Several methods for minimising this error can be suggested. These include minimising the error term itself, minimising the absolute value of the error term $|Y_i - E(Y_i)|$, and minimising the square of the error term $(Y_i - E(Y_i))2$. The last approach is the most commonly used technique, and leads to what is known as 'least squares' parameter estimation.

Figure 3.5: Model errors comparing observed and estimated data



The least squares method is based on the following mathematical analysis. To fit a linear relationship

$$E(Y_i) = a + b X_i \qquad (8.2)$$

to a data set $\{X_i, Y_i, i = 1, \ldots, n\}$, form the 'Residual Sum of Squares' (RSS) given by

$$RSS = \sum_{i=1}^{n} (Y_i - a - bX_i)^2 \qquad (8.3)$$

and then select the values of the parameters a and b that minimise the value of RSS. It can be shown that these best values of *a* and *b* are

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{Y}\,\overline{X}}{\sum_{i=1}^{n} X_i^2 - N\overline{X}^2} \quad \text{and} \quad a = \overline{Y} - b\overline{X}$$

The degree of fit of the estimated relationship can be tested statistically. One common measure of the goodness of fit is the correlation coefficient (*r*) or, more commonly, the coefficient of determination (r^2). The coefficient of determination for the simple regression relation (equation (8.2)) is given by

$$r^2 = \frac{\left[\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

and this coefficient is the proportion of the overall variance in the independent data $\{Y_i\}$ explained by the regression relationship ($0 \leq r^2 \leq 1$).

This method extends immediately to multivariate relationships of the form of equation (8.1), to give what is known as multiple linear regression. While multiple regression is a straightforward, powerful and widely-used model estimation technique, care is needed in the development of any multiple regression model intended for practical applications, and generally the advice of a statistician will be helpful. For example, the choice of the set of independent variables (the 'X_{ij}', $j = 1, \ldots, m$) needs to be considered carefully. There are three broad issues concerning the independent variables that may affect the structure and development, and hence usefulness of a multiple regression model:

- multicollinearity
- heteroskedasticity, and
- autocorrelation.

Multicollinearity occurs when (as is often the case) some of the 'independent' variables are highly correlated with each other—the statistical theory on which multiple regression is based assumes them to be independent of each other. Thus care is needed in selecting a subset of the independent variables for inclusion in the multiple regression relationship. The usual approach is to enter or delete independent variables one by one in some pre-established manner. Some of the approaches used are forward inclusion, backward elimination and stepwise solution. With *forward inclusion*, independent variables are entered one-by-one only if they meet certain statistical criteria. The order of inclusion is determined by the respective contribution of each variable to the explained variance. In *backward elimination*, variables are eliminated one-by-one from a regression equation that initially contains all variables. With a *stepwise solution*, forward inclusion is combined with the deletion of variables that no longer meet the pre-established criteria at each successive step.

Heteroskedasticity is also a common phenomenon. It relates to the variation of the absolute errors in measurement of an independent variable with the actual magnitude of that variable. Statistical theory for regression analysis assumes that the variance of the error term is the same for all values of the independent variable. This is called homoskedasticity. There are many instances where this may not be the case. For instance, in considering the masses of consignment loads to be transported, an error in measurement of 1 kg would be insignificant for an item of machinery of mass (say) 400 kg, but would be very significant for a parcel of mass 5 kg. For standard regression analysis, the consequences of heteroskedasticity in an independent variable are to underestimate the standard error of the estimated (dependent) variable (Y) and to falsely raise the apparent significance of independent variables (X_j). It does not affect the estimated values of the model coefficients, but does reduce the efficiency of the model as an estimating tool. One solution is to use weighted least squares analysis, in which each observation of an independent variable is adjusted for the expected size of its error term. Another possibility is to transform an independent variable (e.g. by using its logarithm, log(X)). This approach is useful for variables where there is a great range of sizes in the observations (e.g. the distribution of household incomes in a region).

Autocorrelation occurs when successive observations of a given variable are highly correlated with each other. This phenomenon frequently arises in data observations taken over time. The problem for multiple regression analysis is again the underlying assumption that the observed values of the independent variables are independent of each other. In fact, autocorrelation is most important in its own right and there is a special field of statistical analysis devoted to 'time series' analysis. This is discussed in Section 3.9. Autocorrelation reflects cyclical behaviour in a variable over time, such as seasonal variations in (say) agricultural production, travel demand or road crashes in a region. Knowledge of these cycles is important in understanding the patterns of demand for transport services, among other things.

There are a number of statistical tests and procedures to identify and cope with these potential problems in regression modelling.

As suggested by equation (8.1), multiple regression analysis is based on the assumption that there is a linear relationship between the dependent and the independent variables. When the relationships are non-linear, it may still be possible to use standard multiple regression—by transforming the relevant variables. For example, if there is a multiplicative relationship

$$Y = kX_1X_2 \tag{8.4}$$

then this can be transformed into a linear relationship by taking logarithms of both sides of equation (8.4), i.e.

$$\log(Y) = \log K + \log X_1 + \log X_2$$

which is then a linear model.

Likewise, if there is a quadratic relationship between a dependent and an independent variable, e.g.

$$Y = a + b_1X_1 + b_2X_2^2$$

then the variable $Z = X_2^2$ may be used instead, as $Y = a + b_1X_1 + b_2Z$ is a linear relation that can be treated using the standard regression approach.

It may be noted that when logarithms are taken of both the dependent and independent variables (e.g. as for equation (8.4)) then the regression coefficient for the independent variable is an elasticity estimate (Section 3.5).

Chapter 17 of Taylor, Bonsall andYoung (2000, pp.392–410) provides a more detailed discussion of regression modelling techniques. See also Fahrmeir and Tutz (1994).

Multiple linear regression is a technique often used for estimating model parameters. However, it requires that assumptions be made about the distribution of errors—strictly that the independent variables all follow normal (Gaussian) distributions. An alternative technique for estimating the parameters of a model that does not have this limitation is maximum likelihood estimation. It is a standard statistical technique employed for estimating the parameters of statistical distributions. The method requires the analyst to develop a likelihood function which represents the combined probability a given data set comes from a particular distribution with specified parameter values. A simple illustration should convey the idea. Consider a random variable X with probability density function $f(X|\theta)$ where θ is a parameter defining the distribution. A data set of observations $\{X_i, i = 1, \ldots, n\}$ has been collected. Then the likelihood function for this data set with parameter θ is the product of the probabilities of observing the set of data observations, which is

$$L(\theta) = f(x_1|\theta)f(x_2|\theta)\ldots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{8.5}$$

The parameter θ may then be estimated by finding the value of the parameter that maximises the value of $L(\theta)$. This is usually done by taking the logarithms of both sides of equation (8.5), which yields the equation

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log(f(x_i|\theta)) \tag{8.6}$$

which is a linear sum of terms and can be optimised using standard methods. The maximum likelihood technique can be used, for instance, to determine the parameters of statistical distributions (e.g. normal distribution or Poisson distribution) fitted to data sets or, as shown in Taylor, Bonsall and Young (2000, pp.414–5), as an alternative method to determine the coefficients in a multiple linear regression relationship. The technique forms the basis of the parameter estimation procedures used to determine the coefficients in logit models of discrete choice processes (e.g. see Louviere, Hensher and Swait, 2000).

A further method for model parameter estimation is 'entropy maximisation' or the use of the mathematical theory of information (Shannon 1948, Van Zuylen and Willumsen, 1980). This method is used, for instance, in the estimation of origin-destination matrices from observed link counts (see Taylor, Bonsall and Young (2000, pp.116–20)).

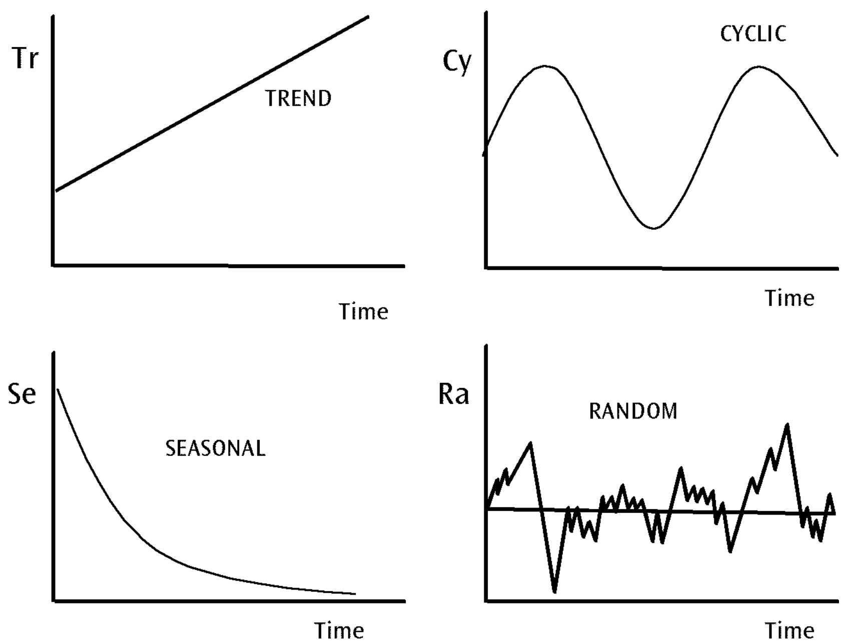
3.9 Projection of demand growth

Many transport problems are a combination of deterministic and stochastic processes that may vary over time. Time series analysis is the statistical technique used to study such problems. From a stochastic point of view, traffic flow on a particular link of a network could be regarded as consisting of four components: trend (*Tr*), seasonal (*Se*), cyclic (*Cy*), and random (*Ra*) components. The trend component may result from long- term growth in traffic. Seasonal variation may result from different flows at different times of the year. Cyclic components can result from long term economic changes. The random component may result from short-term variations in traffic flow. When there is an additive relationship between the components, they can be combined as in equation (9.1) and Figure 3.6

$$X(t) = Tr(t) + Se(t) + Cy(t) + Ra \tag{9.1}$$

The loads on a transport network reflect the time-dependent variations in social, economic, industrial, agricultural and recreational activities in the area it serves, as well as long-term trends in the levels of those activities. For instance, traffic data—such as hourly or daily traffic volumes—that are indicative of these loads, must be recorded as time dependent data. Freight flows in a region may vary over time (e.g. weeks and months of the year) as an outcome of the seasonal factors in production and also in demand. The distinguishing feature of time-dependent data is that they come from processes that are undergoing continual change. If we are to understand the data, we must extract the components of change that are involved and separate the random effects from the trends and cycles that influence the data. Stationary processes (those whose parameters are stable over time) offer the possibility of repeating observations in order to uncover the degree of variability existing in the data. Time series data do not permit repetition of observations, for data collected at one point in time will, of necessity, differ from those collected at other times.

Figure 3.6: Components of a time series



Methods for the analysis of time series data, including the use of moving averages and autocorrelation coefficients, are outlined in Taylor, Bonsall and Young (2000, pp.416–22) and described more fully in text books such as Chatfield (1984).

Time series data are important information sources for transport analysis. They include data on system performance and impacts over time, such as freight flows, passenger movements or road fatalities, as well as economic performance and activity data (e.g. quarterly GNP statistics and fuel sales) and socio-economic data such as population and employment data sets. Gargett and Perry (1998), Amoako (2002) and Sutcliffe (2002) provide recent examples of the use of time series data in analysis of freight movements in Australia.

3.10 Software for model estimation

The methods described in Sections 3.8 and 3.9 are commonly applied in many areas of quantitative analysis. There are many commercial software packages available that can be employed for data analysis, estimation of model parameters and assessment of the validity and goodness of fit of statistical models.

The tasks involved in data analysis include exploratory data analysis, descriptive statistics, statistical inference and comparisons of parameter estimates, analysis of variance, distribution fitting, goodness of fit tests, multiple regression, factor analysis and time series analysis.

The Excel spreadsheet software in the Microsoft Office software suite has considerable capability for standard statistical analysis, including descriptive statistics, regression analysis and more besides.

More substantial statistical analysis is best undertaken using specialist, but commercially available, software.

The SPSS package (SPSS Inc., 2004) is perhaps the most widely known statistical analysis software for general applications. SPSS has been in use (in various versions) since the 1960s and has an enormous international user base and historical track record. SPSS includes a substantial array of the available statistical techniques with a unified database system and high quality graphical outputs available.

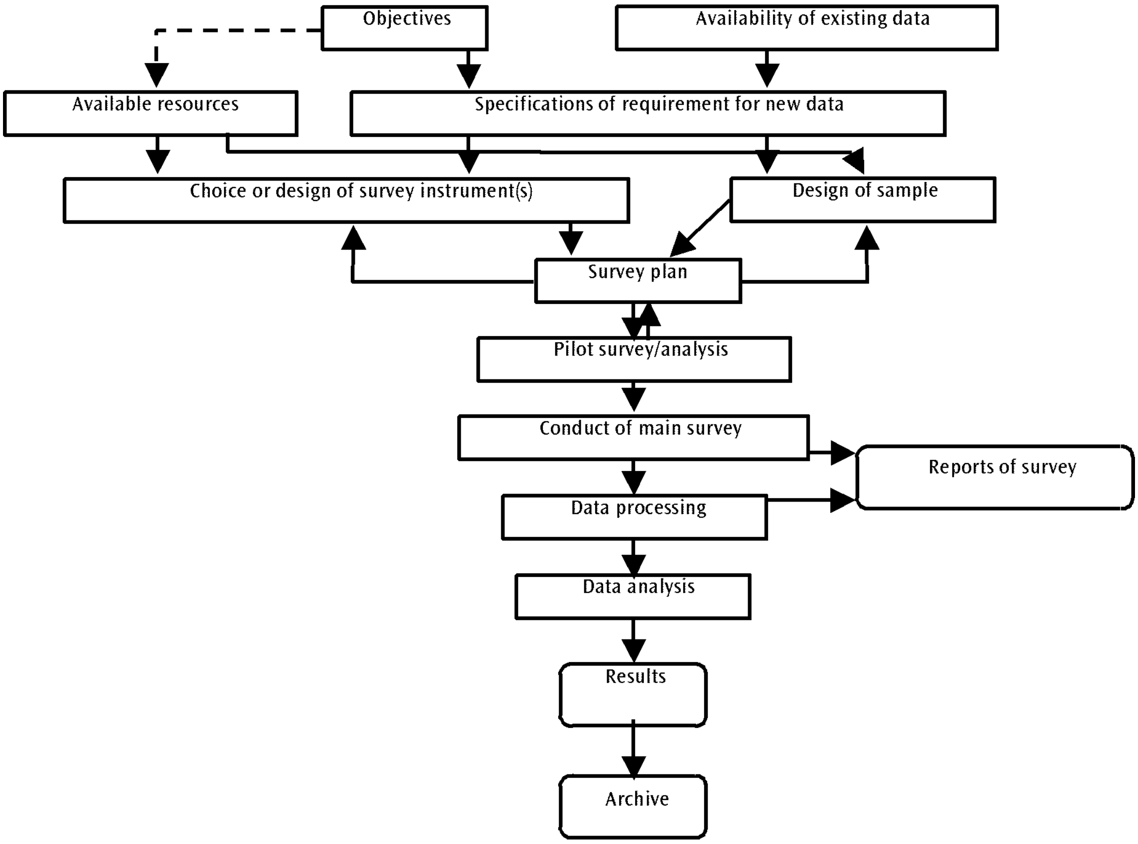
There are other statistical analysis packages available, such as S-PLUS (Venables and Ripley 1994, Insightful Inc. 2004), which offer similar capabilities to SPSS and occasionally some other useful features. For example, for specialist applications S-PLUS offers programmers a capability to develop and include their own procedures within its analytical shell.

Specialist fields such as multivariate statistical modelling using generalised linear models (e.g. Fahrmeir and Tutz 1994) are of importance in transport data analysis and may require specialist software. For example, the estimation of parameters for discrete choice models (as described in Section 3.7) is widely undertaken using the LIMDEP software package (Econometric Software Inc, 2004).

3.11 Survey methodology and data requirements

Models and analytical procedures need data. The data need to be relevant, current and accurate if useful results are to be gained from modelling and analysis. Data collection is expensive, time consuming and not always straightforward, so care is needed in the planning, design and conduct of surveys. Without this attention, resources—time, people and money—can easily be wasted for little gain. High quality and relevant data are essential for analysis and serve to support policy formulation and decision-making. Poor quality or inappropriate data are to the detriment of informed decision-making. One useful way to approach data collection is to view the survey process from the systems perspective. Figure 3.7 provides one such process model. This figure represents a transport survey data collection as a process starting with the specification of objectives of the survey and running through to the archiving of results. Note the existence within Figure 3.7 of various feedback loops indicating that survey design is not a purely sequential process; for example, analysts must be prepared to modify their survey instruments and sample frames in the light of the outcome of the pilot survey.

Figure 3.7: A systems process model for transport data surveys



Source: Taylor, Bonsall and Young (2000, p.138)

This process model identifies a number of steps and stages in the collection and analysis of data. These steps may be grouped into three broad stages.

1. *Preliminary planning*, in which the purpose and specific objectives of the survey are identified, specifications of the requirements for new data are determined (in the light of existing data sources) and resources available or required for the survey are identified.
2. *Survey planning and design*, in which the appropriate survey instrument is selected and the sample design (including target population, sampling frame, sampling method and sample size) is undertaken, leading to a survey plan and the conduct of a pilot survey to test all aspects of the plan and to ensure that it works and provides the required data and that they are compatible with the proposed analysis. This is an iterative stage, in which pilot survey outcomes may lead to revisions in the survey plan. Good, and successful, surveys necessarily pay significant attention to getting this stage right.
3. *Survey conduct*, in which the full survey is undertaken, data extracted and analysed, study reports prepared and databases archived for future reference.

This process is fully explained in Taylor, Bonsall and Young (2000, pp.137–145).

3.12 Survey techniques

Surveys are used to obtain data, which are then used to estimate model parameters for predicting behaviour of transport users, to make demand forecasts and to estimate the economic and financial values of projects. Most transport data surveys are sample surveys, in that only a small fraction of the overall population (e.g. of travellers, vehicles, network links or customers) is surveyed to provide data that are then extrapolated to provide descriptions of the total population. Data are either collected at a few locations taken to represent transport activity, travel movement and traffic flow across the study area, or a sample of individual travellers, customers or operators is surveyed because it is infeasible, impractical or uneconomic to survey the entire population. This means that survey data often need expansion from the sample to represent the full population. As discussed in Sections 3.11 and 3.13, care in survey organisation and attention to detail are needed to ensure that the survey data can properly represent their parent population.

There are two broad approaches to data collection:

- observational (passive) surveys in which surveyors (human or mechanical) record the occurrence (and often time of occurrence) of specified transport events or phenomena, such as the passage of vehicles past a point on the road, the arrival of trucks at a warehouse, or the number of passengers exiting from a railway platform in a specified time interval; and
- interview (active) surveys, in which the surveyors make contact with the individual travellers, customers or decision makers to seek information directly from them, such as the freight movements surveys undertaken by the Australian Bureau of Statistics (ABS) in the mid 1990s (Maitland and Higgins 1999).

The information gathered in active surveys can be much richer than that available from passive surveys:

1. Observations are limited in scope to the direct area under study, for instance the arrival of a vehicle at a cordon line indicates the point at which the vehicle entered or left the study area, but provides little information on the actual origin or the ultimate destination of the trip, nor the frequency with which the vehicle makes that trip or the purpose for which it is made. An active interview or questionnaire survey could obtain this additional information.
2. Observational surveys are limited to study of actual behaviour at the study site. They provide information on 'revealed demand', the actual behaviour that is occurring under the environmental conditions pertaining to the study area at the time of the survey. Revealed demand is the observed use of an area or facility. Environmental states such as traffic congestion or lack of parking and seasonal conditions (including time of day) may restrict the ability of some individuals to access the specific site or facility, or to choose to use an alternative (e.g. another destination) that they would not do if the opportunity arose. This phenomenon is known as 'latent demand' and its extent cannot be gauged using observational surveys. An active survey method could seek to determine the existence and extent of latent demand, especially if the survey is designed and applied to include 'non-users' of a facility or service as well as the users[3].

The passive surveys attempt to make no interference with the normal operation of the survey site and not to disturb the behaviour of the individuals under observation. The active surveys cannot avoid some interference and may provide disturbances, which could even affect the behaviour of the respondents. Great care is need in the survey design for both observational and active surveys to ensure that any interference is minimised and that significant bias is not introduced into the survey results because of the manner in which the data were collected.

3 For example, on-board surveys conducted on bus, train or tram are often used to collect data on public transport users, but could not indicate much about those travellers who are potential users of public transport, but are currently using some other mode. This is one reason for the use of home interviews in general travel surveys. Likewise, shippers of freight may be better surveyed at their company locations.

Active ('interview') surveys may be conducted in three alternative ways:

- direct personal interviews
- remote interviews, generally conducted by telephone, but also possible over the internet, and
- questionnaire surveys.

Personal interviews may be conducted in a variety of locations. Interviews in people's homes have been widely used for collecting detailed data on the travel behaviour of households and individual. Most metropolitan areas and other large cities have databases of personal travel conducted using 'household interviews'. Household interview surveys were conducted in Adelaide in 1999 and Perth in 2002–03, among other cities, while Sydney has a rolling cycle of home interview surveys running continuously. Freight and commercial vehicle surveys are also conducted using interviews at company offices (e.g. Raimond, Peachman and Akers, 1999) where they may involve both staff and customers. Interviews may also be conducted outside the home or business: for instance at shopping centres or recreation facilities, at railway stations and airport terminals, on board (public transport) vehicles, or by the roadside (many truck surveys have been conducted using the latter). Direct interviews can be used to collect detailed data about businesses, households and individuals and their travel behaviour and freight systems usage, and about other traits such as attitudes and perceptions. In some cases, direct interviews may be conducted in a laboratory setting as well as 'in the field'. Stated preference studies are often undertaken in a laboratory where specialised resources can be used (e.g. to create a simulated environment for the respondent to be immersed in the situational context behind the survey). Hensher, Brotchie and Gunn (1989) described a methodology for surveys of rail passengers using active survey techniques. Hensher and Golob (1998) described an interview survey of shippers and freight forwarders conducted in Sydney in 1996.

One problem with the direct interview is that it can take a considerable period of time to complete, which may cause significant inconvenience to the (volunteer) interviewee. A second problem is that the interviewer must make the direct contact with the survey respondent. This may involve considerable time spent travelling by the interviewer, to visit the respondents and the need for multiple 'call backs' if the respondent is not 'at home'.

One solution to the first of these problems is the use of questionnaire surveys, often to be returned through the post ('mail back') to the surveyor at some future time. A case in point would be roadside or in-vehicle surveys. These can be attempted by direct interview but individual travellers may be delayed and inconvenienced in the process, or may reach their destination before the conclusion of the interview. It may then be more reasonable to distribute a written questionnaire to the travellers, to be completed and returned after the journey is finished. The questionnaire can contain a number of questions similar to those posed in an interview, but there are many limitations that will be in place. For example, the question must be clear and unambiguous as it stands, without the opportunity for an interviewer to offer an explanation as is the case in interview surveys. Fewer questions can be posed as excessively long questionnaires will reduce the number of completed responses. There is also the possibility for the respondent to offer false or misleading answers, that an interviewer would recognise as such but are much harder to detect on a written form. The major problem with questionnaire surveys of this type, however, is the likelihood of a low response rate. While response rates of the order of 10–20 per cent may be acceptable in some areas of (say) general market research on consumer goods, there is a considerable body of transport research that suggests that much higher levels of response are required—response rates of 80 per cent or better may be necessary to properly gauge true levels of travel activity in a population. Richardson, Ampt and Meyburg (1995) provide a full discussion of this issue, as well as detailed advice on the conduct of both interview and questionnaire surveys.

The third alternative is the use of telephone (or internet-based) surveys. These are similar to the direct interview except that the interview is conducted remotely over a telecommunications network, with telephone interviews usually using random dial-up access. The advantages of this survey technique are cost and convenience. The interviewer stays in the one place (a call centre) and can make contact with a large number of respondents in a short period of time. Data entry can also be automated, as responses are directly entered into a computer database during the interview. The disadvantages of the technique include the relatively short length of the interview that is normally possible over the telephone and perhaps a growing 'consumer resistance' to the telephone interview given that the method is widely used in general social and market research surveys and in direct marketing. Richardson, Ampt and Meyburg (1995), among many other transport survey researchers, maintain that good quality data on travel behaviour (at least of a quality commensurate with that obtainable from face-to-face interviews and questionnaires) cannot be collected using telephone interviews.

Internet-based surveys are, in some ways more, like observational surveys in that generally respondents to the surveys will find the survey website of their own volition, rather than through an active encouragement by the surveyors. This technique is being used, for instance, in studies on driver route choice, where detailed information is required that is quite difficult to obtain through more conventional survey approaches (see Abdel-Aty, 2003). However, bias in sampling is quite likely to be an issue in these surveys and the field is as yet relatively unexplored.

In terms of costs, the direct interview surveys are generally the most expensive, followed by questionnaire surveys (especially with regard to the number of valid completed questionnaires returned) and then telephone surveys. Observational surveys can be quite cheap, at least in terms of elemental costs such as hourly paid wage rates or if automatic data loggers can be employed (e.g. as in automatic vehicle counts), but large scale observational surveys such as vehicle number plate surveys that require large number of observers can prove to be very expensive, and sometimes not particularly efficient in terms of the collection of usable, quality data.

A further distinction in interview surveys should be drawn between the collection of revealed preference data (what people are seen to do or record that they have done) and stated preference data (what people say they would do in different circumstances, such as when faced with changes in transport fares, services, or costs). By and large the household travel surveys concentrate on revealed preference data. They record historical data on travel behaviour, that then form a snapshot of travel activity in an area at one point in time. These data provide little information on how people might change their behaviour in response to new transport policies, or to changing travel environments, or to the availability of new modes or services. Stated preference data may be used for these purposes and stated preference experimental methods provide powerful tools in this regard. At the same time, there are considerable problems in ensuring that stated preference information is valid and reliable. Louviere, Hensher and Swait (2000) provide a full coverage of this survey methodology and its use. It should be noted that stated preference methods are a rich source of information for the development of discrete choice models of travel behaviour.

For further reading on transport survey methods see Taylor, Bonsall and Young (2000), Louviere, Hensher and Swait (2000) and Richardson, Ampt and Meyburg (1995).

3.13 Sample size estimation

As indicated in Section 3.12, most transport data surveys are sample surveys. Sampling is usually necessary because it is too expensive to survey all members of the population (e.g. to obtain travel diaries from all inhabitants of a metropolitan area), or it is physically impossible to do so (e.g. to test the roadworthiness of all vehicles), or because the survey testing process is destructive (e.g. to determine the strength of railway sleepers). Almost all transport surveys involve observing some members of a target population in order to infer something about the characteristics of that population. In this sense they are statistical sampling surveys. The effectiveness of the survey is dependent on choosing an appropriate sample. Sample design is thus a fundamental part of the overall survey process, as described in Section 3.11. It includes the following elements:

- definition of target population
- definition of sampling unit

- selection of sampling frame
- choice of sample method
- consideration of likely sampling errors and biases, and
- determination of sample size.

Two main methods exist for selecting samples from a target population. These are judgment sampling and random sampling. In random sampling, all members of the target population have a chance of being selected in the sample, whereas judgment sampling uses personal knowledge, expertise and opinion to identify sample members. Judgment samples have a certain convenience. They may have a particular role, by way of 'case studies' of particular phenomena or behaviours. The difficulty is that judgment samples have no statistical meaning; they cannot represent the target population. Statistical techniques cannot be applied to them to produce useful results, for they are almost certainly biased. There is a particular role for judgment sampling in exploratory or pilot surveys where the intention is to examine the possible extremes of outcomes with minimal resources. However, in order to go beyond such an exploration, the investigator cannot attempt to select 'typical' members or exclude 'atypical' members of a population, or to seek sampling by convenience or desire (choosing sample members on the basis of ease or pleasure of observation). Rather, a random sampling scheme should be sought, to ensure that the sample taken is statistically representative. Random samples may be taken by one of four basic methods (Cochran, 1977): simple random sampling, systematic sampling, stratified random sampling and cluster sampling. Taylor, Bonsall and Young (2000, pp.155–58) describe each of these sampling methods and their applications, as does Richardson, Ampt and Meyburg (1995).

Simple random sampling allows each possible sample to have an equal probability of being chosen, and each unit in the target population has an equal probability of being included in any one sample. Sampling may be either 'with replacement' (i.e. any member may be selected more than once in any sample draw) or 'without replacement' (i.e. after selection in one sample, that unit is removed from the sampling frame for the remainder of the draw for that sample). Selection of the sample is by way of a 'raffle'. A number is assigned to each unit in the sampling frame, and repeated random draws are made until a sample of desired size is obtained. For most applications, the use of a table of random digits is the most convenient means of drawing a random sample. Most statistical text books contain such tables. The methods of statistical inferences applied to sample data analysis are predicated on the basis that a sample is chosen by simple random sampling. Data collected using other sampling methods need to be analysed using known techniques that include corrections to approximate simple random samples.

There is always a possibility that a sample may not adequately reflect the nature of the parent population. Random fluctuations ('errors'), which are inherent in the sampling process, are not serious because they can be quantified and allow for using statistical methods[4]. However, if due to poor experimental design or survey execution, there is a systematic pattern to the errors, this will introduce bias into the data and, unless it can be detected, it will distort the analysis. A principal objective of statistical theory is to infer valid conclusions about a population from unbiased sample data, bearing in mind the inherent variability introduced by sampling. Bias and sampling errors are the two, quite different, sources of error in experimental observations. As described in Richardson, Ampt and Meyburg (1995, pp.96–101), bias (or systematic error) needs to be removed from sample data before statistical analysis can be attempted, for statistical theory treats all errors as sampling errors.

A distribution of all the possible means of samples drawn from a target population is known as a sampling distribution. It can be partially described by its mean and standard deviation. The standard deviation of the sampling distribution is known as the standard error. It takes account of

4 Noting that minimisation of experimental errors is of course important in improving the precision of parameter estimates based on survey data.

the anticipated amount of random variation inherent in the sampling process and can therefore be use to determine the precision of a given estimate of a population parameter from the sample.

Surveys for specific investigations usually attempt to provide data for the estimation of particular population parameters, or to test statistical hypotheses about a population. In either case the size of the sample selected will be an important element, and the reliability of the estimate will increase as sample size increases. On the other hand, the cost of gathering the data will also increase with increased sample size, and this is an important consideration in sample design. A trade-off may occur, and the additional returns from an increase in sample size will need to be evaluated against the additional costs incurred. If the target population may be taken as infinite, then the standard error (s_X) of variable X is given by

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad (13.1)$$

where s is the estimated standard deviation of the population and n is the sample size, assuming that the sampling distribution is normal. Even when the sampling distribution is not normal, this method may still apply because of the Central Limits Theorem which states that the mean of n random variables from the same distribution will, in the limit as n approaches infinity, have a normal distribution even if the parent distribution is not normal. The standard deviation of the mean is inversely proportional to $\sqrt{n}$. The implication of equation (13.1) is that as sample size increases, standard error decreases in proportion to the square root of n. Here is an important result. The extra precision of a larger sample should be traded off against the cost of collecting that amount of data. To double the precision of an estimate will require the collection of four times as much data.

Similar results are found for other statistical parameters. For instance, the standard error (s_p) of a proportion p (e.g. a measure such as 'the proportion of households owning one vehicle') is given by

$$s_p = \sqrt{\frac{p(1-p)}{n}} \qquad (13.2)$$

The practical application of these results requires some prior knowledge of the population, for example a prior estimate of the sample standard deviation (s) in the case of the mean value of variable X, or an initial estimate of the proportion p. The results of previous surveys, or the pilot survey, may be used to provide this knowledge.

3.14 Bibliography

The bibliography is arranged in order of the sections of Part 3. The citations under each heading include the references given in the text and some other works that are of relevance and interest to the subject matter covered in the specific sections.

3.14.1 Overview of transport modelling

D'Este, G M (1996), An event based approach to modelling intermodal freight systems. *International Journal of Physical Distribution and Logistics Management 26*, pp.4–15.

D'Este, G M (2001), Freight and logistics modelling. *In* Brewer, A M, Button, K J and Hensher, D A (eds). *Handbook of Logistics and Supply Chain Management.* (Pergamon-Elsevier: Oxford), pp.521–34.

Hensher, D A and Button, K (eds) (2000), *Handbook of Transport Modelling*. (Pergamon-Elsevier: Oxford).

Jones, M, Simpson, B and Alchin, S (2001), Strategic models for new railway planning – a New South Wales perspective. *Papers of the 24th Australasian Transport Research Forum 24,* Hobart, April, CD-ROM (Department of Infrastructure, Energy and Resources: Hobart).

Ortuzar, J de D and Willumsen, L G (1994), *Modelling Transport* (2nd ed). (John Wiley and Sons: New York).

Taniguchi, E, Thompson, R G, Yamada, T and Van Duin, R (2001), *City Logistics: Network Modelling and Intelligent Transport Systems*. (Elsevier: Oxford).

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.107–24.

Young, W, Taylor, M A P and Gipps, P G (1989), *Microcomputers in Traffic Engineering*. (Research Studies Press: Taunton).

3.14.2 Market segmentation

D'Este, G M (2000), Urban freight movement modelling. In Hensher, D A and Button, K (eds) *Handbook of Transport Modelling*. (Pergamon-Elsevier: Oxford), pp.539–52.

Raimond, T, Peachman, J and Akers, G (1999), Estimating small area commercial transport movements. *Papers of the Australasian Transport Research Forum 23* (1), pp.493–506.

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.114–20, 276–77.

3.14.3 Measurement of price and service quality variables

Anzelark, A, Estermna, A and Homburg, A (2002), Beyond customer satisfaction: measuring service provision. *Papers of the 25th Australasian Transport Research Forum 25,* paper no 44. CD-ROM. (Bureau of Transport and Regional Economics: Canberra).

Bell, M G H and Iida, Y (eds) (2003), *The Network Reliability of Transport.* (Elsevier-Pergamon: Oxford).

Ortuzar, J de D and Willumsen, L G (1994), *Modelling Transport* (2nd ed). (John Wiley and Sons: New York).

Tanaguchi. E, Thompson, R G, Yamada, T and Van Duin, R (2001), *City Logistics: Network Modelling and Intelligent Transport Systems.* (Pergamon-Elsevier: Oxford).

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.207–8.

Taylor, M A P, Woolley, J E and Zito, R (2000), Integration of the global positioning system and geographical information systems for traffic congestion studies. *Transportation Research C 8C* (1-6), pp.257–85.

3.14.4 Demand elasticities

ARRDO (1982), The use of demand elasticities in road planning. ARRDO Report no 82/2. (Australian Railways Research and Development Organisation: Melbourne).

BAH (2001), *Appendix A—evaluation methodology: interstate rail network audit*. Report to ARTC. Booz Allen and Hamilton, Sydney.

Bennathan, E and Walters, A A (1969), *The Economics of Ocean Freight Rates*. (Praeger: New York).

BTCE (1990), Freight flows in Australian transport corridors. Occasional paper 98. (Bureau of Transport and Communications Economics: Canberra).

BTCE (1991), Demand forecasts and demand elasticities for Australian transport fuel. Occasional Paper 103. (Bureau of Transport and Communications Economics: Canberra).

BTRE (2004), Transport elasticities database. http://dynamic.dotars.gov.au/btre/tedbindex.cfm. Bureau of Transport and Regional Economics, Canberra [last accessed 06/04/04].

Hsing, Y (1994), Estimating the impact of deregulation on the elasticity of the demand for railroad services. *International Journal of Transport Economics 21* (3), pp.301–12.

Luck, D P and Martin, I J (1988), Review of road cost recovery. BTE Occasional Paper 90. (Bureau of Transport Economics: Canberra).

Luk, J Y K and Hepburn, S (1993), New review of Australian travel demand elasticities. Research Report ARR249. Australian Road Research Board, Melbourne.

NIEIR (1990). Implication of adopting CO2 emission reduction targets in transport sector. A study for the Bureau of Transport and Communications Economics by the National Institute of Economic and Industry Research. (Bureau of Transport and Communications Economics: Canberra).

Oum, T (1989), Alternative demand models and their elasticity values. *Journal of Transport Economics and Policy 23* (2), pp.163–87.

Oum, T, Waters, W G and Yong, J (1992), Concept of price elasticities of transport demand and recent empirical results. *Journal of Transport Economics and Policy 26* (2), pp.139–54.

Taplin, J H E, Hensher, D A and Smith, B (1999), Preserving the symmetry of estimated commuter travel elasticities. *Transportation Research B 33B* (3), pp.215–332.

3.14.5 Performance indicators for customer service

Anzelark, A, Esterman, A and Homburg, A (2002), Beyond customer satisfaction: measuring service provision. *Papers of the 25th Australasian Transport Research Forum 25*, paper no 44, Canberra, October. CD-ROM. (Bureau of Transport and Regional Economics: Canberra).

Bell, M G H and Iida, Y (eds) (2003), *The Network Reliability of Transport*. (Pergamon-Elsevier: Oxford).

Lee, D H (1989), Transit cost and performance measurement. *Transport Reviews 9* (2), pp.147–70.

Lam, W H K and Bell, M G H (eds) (2003), *Advanced Modeling for Transit Operations and Service Planning*. (Pergamon-Elsevier: Oxford).

Nash, C (2000), Modelling performance: rail. In Hensher, D A and Button, K (eds). *Handbook of Transport Modelling*. (Pergamon-Elsevier: Oxford), pp.565–75.

Tanaguchi. E, Thompson, R G, Yamada, T and Van Duin, R (2001), *City Logistics: Network Modelling and Intelligent Transport Systems*. (Pergamon-Elsevier: Oxford).

Taylor, M A P, Woolley, J E and Zito, R (2000), Integration of the global positioning system and geographical information systems for traffic congestion studies. *Transportation Research C 8C* (1-6), pp.257–85.

Taylor, S Y (2001), Just-in-time. In Brewer, A M, Button, K J and Hensher, D A (eds). *Handbook of Logistics and Supply-Chain Management*. (Pergamon-Elsevier: Oxford), pp.213–24.

Vuchic (1981), *Urban Public Transportation: Systems and Technology*. (Prentice-Hall: Englewood Cliffs).

3.14.6 Discrete choice models

Douglas, N J, Franzmann, L J and Frost, T W (2003), The estimation of demand parameters for primary public transport service in Brisbane. *Papers of the 26th Australasian Transport Research Forum 26*, paper no 42, Wellington, October. CD-ROM (NZ Institute of Highway Technology: New Plymouth, New Zealand).

Hensher, D A and Button, K (eds) (2000), *Handbook of Transport Modelling*. (Pergamon-Elsevier: Oxford).

Louviere, J J, Hensher, D A and Swait, J D (2000), *Stated Choice Methods: Analysis and Application*. (Cambridge University Press: Cambridge).

Luce, R D (1959), *Individual Choice Behaviour*. (John Wiley and Sons: New York). Ortuzar, J de D and Willumsen, L G (1994). *Modelling Transport* (2nd ed). (John Wiley and Sons: New York).

Taplin, J H E, Hensher, D A and Smith, B (1999), Preserving the symmetry of estimated commuter travel elasticities. *Transportation Research B 33B* (3), pp.215–32.

Train, K, Revelt, D and Ruud, P (2004), Mixed logit estimation routine for panel data. http://elsa.berkeley.edu/Software/abstracts/train0296.html [last accessed 08/07/04].

Wigan, M R, Rockliffe, N, Thoresen, T and Tsolakis, D (1998), Valuing longhaul and metropolitan freight travel time and reliability. *Papers of the 22nd Australasian Transport Research Forum 22* (1), pp.29–37.

3.14.7 Estimation techniques

Econometric Software Inc (2004), LIMDEP and NLOGIT software. www.limdep.com [last accessed 06/04/04].

Fahrmeir, L and Tutz, G (1994), *Multivariate Statistical Modelling based on Generalized Linear Models*. (Springer-Verlag: New York).

Insightful Inc (2004), S-PLUS and other software tools. www.s-plus.com [last accessed 06/04/04]

Louviere, J J, Hensher, D A and Swait, J D (2000), *Stated Choice Methods: Analysis and Application*. (Cambridge University Press: Cambridge).

SPSS Inc (2004), Statistical package for the social sciences. www.spss.com [last accessed 06/04/04]

Shannon, C E (1948), A mathematical theory of communication. *Bell Systems Technical Journal 27*, pp.379–423 and 623–56.

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.116–20, 387–416.

Van Zuylen, J H and Willumsen, L G (1980), The most likely trip matrix estimated from traffic counts. *Transportation Research B 14B* (4), pp.281–93.

3.14.8 Projection of demand growth

Amoako, J (2002), Forecasting Australia's international container trade. *Papers of the 25th Australasian Transport Research Forum 25*, paper no 60, Canberra, October. CD-ROM. (Bureau of Transport and Regional Economics: Canberra).

Chatfield, C (1984), *The Analysis of Time Series Data: An Introduction* (3rd ed). (Chapman and Hall: London).

Gargett, D and Perry, R (1998), Interstate non-bulk freight. *Papers of the 22nd Australasian Transport Research Forum 22* (1), pp.19–28.

Ortuzar, J de D and Willumsen, L G (1995), *Modelling Transport* (2nd ed). (John Wiley and Sons: New York).

Sutcliffe, P (2002), Features of seasonality and its effects on road freight transport task. *Papers of the 25th Australasian Transport Research Forum 25*, paper no 30, Canberra, October. CD-ROM. (Bureau of Transport and Regional Economics: Canberra).

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.416–22.

3.14.9 Software for model estimation

Econometric Software Inc (2004), LIMDEP and NLOGIT software. www.limdep.com [last accessed 06/04/04].

Fahrmeir, L and Tutz, G (1994), *Multivariate Statistical Modelling based on Generalized Linear Models*. (Springer-Verlag: New York).

Insightful Inc (2004), S-PLUS and other software tools. www.s-plus.com [last accessed 06/04/04].

SPSS Inc (2004), Statistical package for the social sciences. www.spss.com [last accessed 06/04/04].

Venables, W N and Ripley, B D (1994)., *Modern Applied Statistics with S-PLUS*. (Springer-Verlag: New York).

3.15.10 Survey methodology and data requirements

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.137–45, 147–162, 291–302.

Richardson, A J, Ampt, E S and Meyburg, A N (1995), *Survey Methods for Transport Planning*. (Eucalyptus Press: Melbourne).

3.15.11 Survey techniques

Abdel-Aty, M A (2003), Hybrid distribution and response techniques for an origin-destination travel survey. *ITE Journal 73* (2), pp.22–27.

Hensher, D A, Brotchie, J F and Gunn, H (1989), A methodology for investigating the passenger demand for high-speed rail. *Papers of the 14th Australasian Transport Research Forum 14* (1), pp.459–76.

Hensher, D A and Golob, T (1998), Searching for policy priorities in the formulation of a freight transport strategy: an analysis of freight industry attitudes towards policy initiatives. *Papers of the 22nd Australasian Transport Research Forum 22* (1), pp.1–17.

Louviere, J J, Hensher, D A and Swait, J D (2000), *Stated Choice Methods: Analysis and Application*. (Cambridge University Press: Cambridge).

Maitland, E and Higgins, R (1999), Recent Australian Bureau of Statistics initiatives in the development of reliable measures of the freight task in Australia. *Papers of the 23rd Australasian Transport Research Forum 23* (1), pp.319–32.

Raimond, T, Peachman, J and Akers, G (1999), Estimating small area commercial transport movements. *Papers of the Australasian Transport Research Forum 23* (1), pp.493–506.

Richardson, A J, Ampt, E S and Meyburg, A N (1995), *Survey Methods for Transport Planning*. (Eucalyptus Press: Melbourne).

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.137–62, 291–302.

3.15.12 Sample size estimation

Cochran, W G (1977), *Sampling Techniques* (3rd ed). (John Wiley and Sons: New York).

Richardson, A J, Ampt, E S and Meyburg, A N (1995). *Survey Methods for Transport Planning*. (Eucalyptus Press: Melbourne).

Taylor, M A P, Bonsall, P W and Young, W (2000), *Understanding Traffic Systems: Data, Presentation and Analysis* (2nd ed). (Ashgate Publishing Co: Aldershot), pp.159–62.

4

Rail Demand Forecasting and Cost Modelling: Estimation of Impacts of Rail Infrastructure Improvements

Brian Hollins
Arduus Pty Ltd

Dr Peter Pudney and Professor Phillip Howlett
Centre for Industrial and Applied Mathematics
University of South Australia and

Professor Michael A P Taylor
Transport Systems Centre
University of South Australia

Report to the Guidelines Assessment Methodology Working Group

July 2004

UNIVERSITY OF SOUTH AUSTRALIA

TSC
Transport Systems Centre

4.1 Introduction

This paper provides information on methods for estimating the impacts of rail projects on costs, service qualities and performance indicators, based on a summary of existing knowledge and an assessment of accepted good practices. This paper also includes information on typical parameter values for modelling and analysis.

The purpose of this paper is to provide an overview of approaches to determining the impacts of rail infrastructure improvements. The paper examines the following:

- the needs for impact estimation and requirements specific to rail transport (Section 4.2)
- rail costs and benefits (Sections 4.3 and 4.4), including below-rail costs, above-rail costs, other costs and avoidable costs
- rail infrastructure projects, programs and strategies (Section 4.5)
- estimation and simulation techniques and contemporary approaches to rail modelling (Section 4.6), and
- performance analysis and measurement (Section 4.7).

The final section of this paper is a bibliography, providing cited references and further reading for each section of the paper.

The paper also contains set of parameter values and representative externality values for use in studies of rail infrastructure improvements and impact analyses.

This paper (or Part 4) should be read in conjunction with the paper 'Demand forecasting and modelling' (or Part 3). Part 3 includes information on methods for assessing transport service performance, modelling demand for transport services, and estimating future demands. It also describes survey methods and techniques for the collection of data for modelling and performance assessment.

4.2 Needs for impact estimation

Impact estimation is often developed within a sequential decision-making framework, where the proposed infrastructure improvement represents the last step towards an assessment of the extent to which pre-determined transport objectives are achieved. Usually the estimation process is focused on the satisfaction of a forecast level of future demand or on the contribution of the infrastructure proposal to a reduction in net operating costs.

At the core of the estimation of the impacts of rail infrastructure improvements, there is likely to be a myriad of modelling techniques, ranging from the simulation of train operations, to scheduling and possession planning, to capacity determination and possibly to the evaluation of maintenance practices, for example. Recent emphasis on the development of links between various models has seen the evolution of suites of rail simulation and modelling tools, although it would seem that the overarching architecture has not yet advanced to the point where the trade-offs between 'above rail' and 'below rail' actions are estimated as the norm. 'Below rail' refers to the activities such as train control and the infrastructure such as track, structures, catenary and signals necessary to sustain train operations. 'Above rail' refers to train operating activities.

In part, this is due to the complex characteristics of railways. In the context of rail performance measurement, Nash (2000) identified the following three characteristics:

1. *Multiplicity of outputs*: not only in terms of passenger and freight movement but also encompassing the multiplicity of origin–destination sets and cost structures, the different densities and handling requirements of freight commodities and the temporal effects on capacity of demand in peak periods.

2. *Complexity of the production process*: particularly rail technology where a train service comprises an array of locomotives, wagons, track, signalling, terminals and the like; a complexity amplified by shared costs and economies of scale, scope and traffic density[5]
3. *Operating environment and government intervention*: exert influence on rail performance through the spatial attributes of each corridor and the extent of regulation and public sector participation in employment and investment decisions.

While it could be argued that roads and vehicle traffic also display these three characteristics, railways are different in the sense that operations are underpinned by a fundamental premise—the safe separation of trains—that provides for almost total control of entry onto the network and ordered movement within the network.

This premise and the three main rail performance measurement characteristics help explain the complexity inherent in the estimation of impacts of rail infrastructure improvements. One means of simplifying the estimation process is to categorise and integrate the objectives of rail infrastructure improvements, possibly using efficiency and effectiveness groupings, where, in broad terms, efficiency considers train operations and track maintenance, and effectiveness considers the contribution of improved rail to transport outcomes.

4.3 Rail costs

An understanding of rail costs is essential in the estimation of impacts of proposed infrastructure improvements. Rail costs can be categorised as below rail, above rail or other.

4.3.1 Below rail costs

Rail infrastructure maintenance represents the principal below rail cost. The extent of rail infrastructure maintenance will be determined by the interaction between track usage, the normal life expectancy of each class of below rail asset, the terrain (grade and soil) traversed by the rail corridor and climatic conditions.

In assessing the Queensland Rail (QR) undertaking, the Queensland Competition Authority (QCA, 2000) identified that most 'below rail' assets are long-lived, with long-term maintenance cycles as such changes in costs that will not be apparent until the maintenance cycle has been completed. Forecast costs can also be influenced by management decisions on track standards based on the authorised levels of usage, including for example, the imposition of speed and tonnage restrictions as a means of extending maintenance intervention intervals.

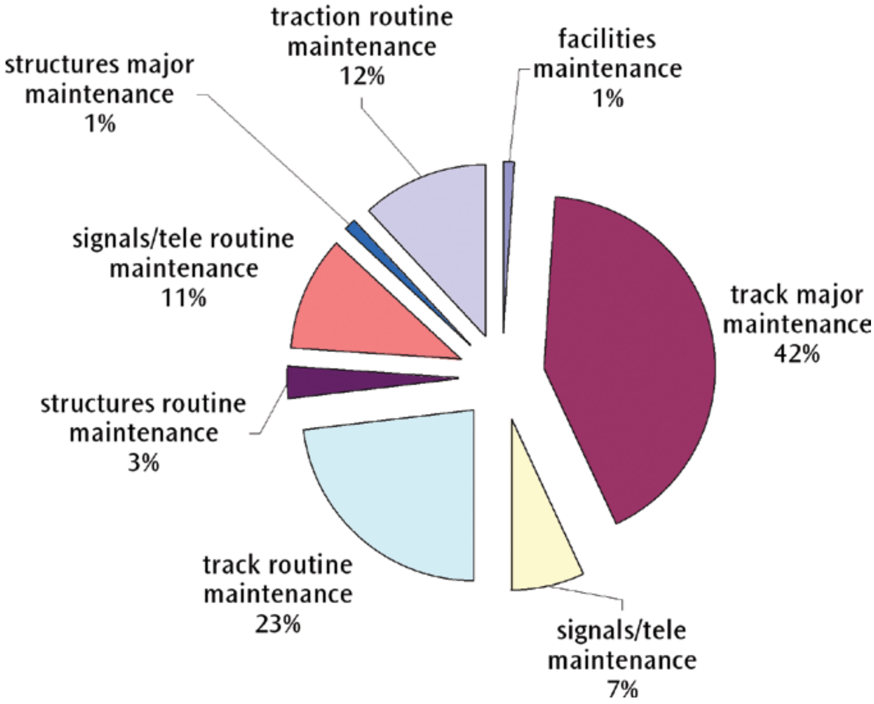
Notwithstanding the variations in expected life for each class of assets and the effects of usage and climate on life expectancy, it is normal for railway managers to prepare long-term maintenance plans. The pie-chart in Figure 4.1 depicts the principal elements of rail infrastructure and identifies the proportions of a ten-year budget assigned for routine and major maintenance of each element.

5 Shared costs cannot be attributed to a particular form of traffic or a specific train service. Shared costs comprise common costs and joint costs. Shared costs are common costs if the services that share these costs are not jointly determined (i.e. one service does not unavoidably create the other). A shared cost is joint when the provision of one service requires the output of another service.

In a railway, economies of:

- scale are generally related to the size of the network
- scope relate to the extent to which the railway can handle a diverse range of traffic types, and
- traffic density concerns the extent to which unit costs can be reduced as traffic density is increased (unless the route is already congested).

Figure 4.1: Infrastructure maintenance elements. The chart shows percentage of total maintenance budget, given a ten year planning horizon



Source: QCA (2000, p.25)

In terms of asset classes, the QCA work indicated that track typically comprises 65 per cent of the total rail maintenance budget (QCA 2000, p.25). Signals and telecommunications represent the next most significant budgetary expenditure.

On older rail corridors, annual expenditure on structures can fluctuate substantially, depending not only on the frequency and severity of floods but also on whether the structures and the track approaches to each structure have been designed for flood immunity or maintained to a flood proof only standard.

In most railways, the maintenance strategy will distinguish between routine and major maintenance activities. Routine maintenance activities usually include inspections[6], fencing, formation and drainage maintenance, turnout and joint maintenance, curve lubricator maintenance, the construction of fire breaks and weed and vegetation control[7]. Except for weed and vegetation control, local gangs undertake most routine maintenance activities, generally with the minimum of mechanised assistance.

Major maintenance activities include:

- Formation rehabilitation.
- Ballast cleaning, encompasses the removal of accumulated fines, coal dust and broken ballast and the insertion, where necessary, of new ballast to restore the drainage and elasticity properties of the track structure. On average, QR annually spends the same amount on ballast as it spends on concrete sleepers (Martin 2001, p.2).

6 In QR, the frequency of track audit inspections varies between ten and 14 months. Notwithstanding the level of sophistication of track condition monitoring and analysis, walking inspections are still considered mandatory. For example on timber sleepered track, these inspections cover five per cent of the total track km in the audited area, with 2 x 100m sleeper tests completed in each kilometre walked (Dennis 2001, p. 4 and 9).

7 *Weed control* is usually undertaken on a district basis by a specialist person, travelling in a road vehicle. *Vegetation control* is often sub-contracted, particularly for the clearance of overhanging vegetation.

- Tamping (or re-surfacing), which involves the packing of ballast around the sleepers to provide for continued track elasticity, where the track structure remains in place after a train has passed over it.
- Rail grinding, profiling of the rail head to provide for the most optimal wheel-rail interface. The grinding process also removes wheel burns and corrugations from the rail.
- Occasionally, re-sleepering or re-railing in those areas where problems are recurrent. Re-railing involves the replacement of rails but not sleepers, particularly on curves. Re-railing or re-sleepering of complete corridors is generally regarded as track renewal.

Specialist mechanised gangs undertake the major maintenance activities. In the context of the type of maintenance, major maintenance generally represents 40 per cent to 50 per cent of the total rail maintenance budget (QCA, 2000).

Whilst maintenance regimes will vary according to traffic volumes, narrow gauge maintenance (as per the QR network) is more intensive than standard gauge maintenance regimes. The Australian Rail Track Corporation (ARTC) has provided some information on maintenance expenses on the standard gauge Defined Interstate Rail Network (DIRN). ARTC (2001a, pp.9–10) indicates that infrastructure maintenance, which includes routine maintenance and variable and fixed major periodic maintenance, consumes about 73 per cent of ARTC's total operating expenditure. Of the infrastructure maintenance expenditure, about 64 per cent is spent on routine maintenance. ARTC's unit costs for track maintenance average around $10,100 per track km across the DIRN. Unit costs for individual segments may vary depending on density of traffic and terrain. A better unit of measure is then perhaps gross tonne kilometre (GTK). ARTC's average infrastructure maintenance unit cost (including signals and communications maintenance) is around $1.70/'000 GTK.

Train control, which also includes transit management and data processing, is around six per cent of ARTC's operating expenditure. ARTC (2001a, p.10) indicates that the unit cost of train control was about $292 per thousand train-km in 2001.

System management and administration (including IT, property management, security, accounting, insurance, strategic management and executive costs) is about 12.5 per cent of ARTC's operating expenditure. Maintenance contract management (administration, accounting, project management and maintenance planning) accounts for about five per cent, and operations and safety management another 2.5 per cent.

Irrespective of track gauge, asset life expiry on lightly trafficked lines is likely to have the most significant impact on track standards and will drive the maintenance program. As with many such costs, railway maintenance costs have both fixed and variable components. The variable cost may be quite significant. For example, rail renewal costs on a track carrying 2.0 mgtpa represent only three per cent of the maintenance budget. In contrast, on a track with 20 mgtpa, rail renewal is 20 per cent of the maintenance budget (QCA 2000, Table 2.1, p.5).

Where track usage is the primary driver of maintenance, the scope and nature of the maintenance intervention will encompass a linked series of major maintenance activities including, for example, ballast cleaning, tamping and rail grinding. In effect the aggregation of activities represents a major maintenance intervention strategy, which requires the development and approval of a long-term, integrated maintenance budget. Nonetheless, there is a point at which assets expire through the interaction of age and usage and major maintenance or maintenance intervention strategies are no longer appropriate. At such times track renewal is undertaken, through re-sleepering, re-railing or re-laying[8].

8 *Re-laying* of the track structure encompasses replacement of track, sleepers, fasteners and ballast and may need to be preceded by formation restoration work.

While there is some inter-dependency between major maintenance and track renewal, it is likely that maintenance will be costed separately and renewal costs will be treated as capital investment, particularly if the renewal encompasses an element of track upgrading (QCA 2000, p.21).

In terms of below rail costs, it is reasonable to conclude that:

- on lightly trafficked routes, the proportion of variable maintenance costs is relatively low
- the cost of maintenance per unit of output steadily reduces as tonnages increase
- a reduction in axle load results in a corresponding reduction in maintenance costs
- other than on tight curves where rail head lateral wear will occur due to centripetal forces for all travel speeds, a decrease in train speed would normally result in a reduction in maintenance costs. This is because maximum permissible vertical wear is an inverse function of maximum train speed (Profillidis 2000, pp.105–16), and
- in time, uncertainty about the availability of funds for track renewal (capitalised maintenance) will induce the preparation of alternative or sub-optimal maintenance strategies and operational responses.

4.3.2 Above rail costs

Modelling of above rail inter-modal freight operations normally includes estimation of the following costs:

- capital cost and economic life of locomotives and wagons
- locomotive and wagon service, repair and maintenance
- crewing (labour)
- load damage, and
- fuel and lubricants.

This modelling may also provide for the cost of:

- train utilisation
- crew training, of which a significant outlay is likely to be for a driver training simulator, and
- refuelling and maintenance facilities.

In comparing the aggregation of these costs for particular train services, it is apparent that total above rail costs are influenced substantially by the nature of operations, including the spatial characteristics of each rail corridor. On a cost per net tonne-kilometre ($/NTK) basis, the lowest above rail costs in Australia are incurred by bulk ore train services. It is understood that such best practice costs are in the order of $0.01/NTK. The nature of inter-modal freight train services is such that above rail costs as low as bulk ore operations could not be achieved.

Generally, it can be expected that the capital cost of a new diesel powered locomotive is likely to be in the range $4.5m to $6m. For example, in March 2003, EDI Rail (2003) signed a $61m contract with QR for 11 new 4000 class narrow gauge locomotives. These locomotives have a rated power output of 2460 kW (3300 hp), giving as a rough guide, $1.68m/'000 hp as the capital cost for locomotives.

While a diesel-powered unit is more expensive than electric-powered, a diesel-powered locomotive offers relatively greater flexibility in that the only restrictions on its deployment within the rail network will be gauge and axle loading. Most inter-modal freight train services in Australia are diesel-hauled.

The purchase price of a typical container-carrying wagon is approximately $150,000, noting that container-carrying wagons vary in terms of gauge and container capacity and may also be linked in five packs. The number of container slots on a wagon and the double-stacking capability determines wagon capacity. At present, the kinematic envelope of a corridor dictates the feasibility of double stacking operations on the DIRN. Double stacking of containers has yet to be proven feasible on a narrow gauge railway.

Given that a significant proportion of the major assemblies of locomotives and rollingstock are imported, exchange rate fluctuations will affect capital costs. Depending on the practices of individual operators, the up-front capital cost of locomotives and wagons could also be influenced if purchase contracts include provision for major rebuilding or half-life re-fits. A half-life re-fit of a locomotive could cost one-third of the purchase price of the locomotive and around ten per cent to 15 per cent of the original price of a wagon. In modelling above rail costs, allowance should also be made for a fleet rolling stock reserve and for the salvage value of locomotives and possibly wagons.

While usage is most likely to determine the timing of rebuilding or re-fitting, the availability of, and need to adopt, new technology may also impact timing (and of course cost). Nonetheless it could normally be expected that the economic life of a locomotive will be 20 to 25 years and a wagon 25 to 30 years.

The type and number of locomotives used on a particular train service can be determined from industry sources, observation, or possibly calculated on the basis of hp/tonne. Laird (2000, p.58) identified that a hp/tonne ratio of between 2.0 hp/tonne and 3.5 hp/tonne is usual for inter-modal freight train services in Australia.

Locomotive servicing, repair and maintenance requirements are determined by a combination of running and idling time, whereas distance travelled is the primary driver of wagon servicing needs. Generally, maintenance costs will be presented in the form of $ p.a. or $/km. The following indicative costs are suggested by industry sources as appropriate for modelling purposes:

- locomotive maintenance: $30–$55/h or $0.80–$1.50/km, depending on the type and age of the diesel electric locomotive (costs for an electric powered locomotive will be less), and
- wagon maintenance: $5000–$7000 p.a. or $45–$50/’000 wagon-km.

Some modelling techniques provide for the disaggregation of servicing and maintenance costs to a detailed component level. For example, the model used by Laird (2000) costs brake wear at $0.10 per wheel per stop, plus $0.025 cents for replacement.

The type of operation, relevant industry awards, enterprise bargaining arrangements and overheads will influence crew costs. As with rolling stock, an allowance should also be made for a fleet reserve, or relief crews. Two person crews continue to operate most inter-modal services and it is understood that the rates for these crews range between $100–$130 per train crew hour. Clearly track upgrades that result in a reduction in transit time have the potential to reduce overall train crew costs.

Occasionally a container and the freight in the container will be damaged. Anecdotal evidence suggests this is most likely to occur during lifting of the container onto and from the wagon. Most load damage would be sustained in an emergency-stopping situation, but fortunately such events are rare. Modelling undertaken for Queensland Transport by Laird (2000, 2002) includes a nominal allowance for damage of $2/train/day.

Fuel consumption is distance and time-related. The train parameters used by Laird, Michell and Adorni-Braccesi (2002) for a Superfreighter operating between Dynon and Chullora included:

- distance: 940 km
- transit time: 12 h
- locomotives: two 4000 hp locomotives
- trailing load: 2600 t
- fuel consumption: 13 200 L

This equates to a fuel consumption rate of 5.4 L/’000GTK. By comparison, industry sources indicate a narrow gauge coal train consumes 3.0 L/’000GTK. The fuel consumption, and other above rail costs, can be further verified from manufacturers’ data, trade publications and indices such as that produced by the Association of American Railroads (AAR).

4.3.3 *Other railway costs*

Consideration of railway costs also needs to include the costs associated with intermodal terminals and marshalling yards, the cost of capital, management, administration and corporate overhead costs and the costs passed on by safety and economic regulators. Other costs, which cannot be attributed to a specific train service, are assigned on a pro-rata basis across all services, using a measure such as $/'000GTK.

Intermodal terminals: Terminal costs will vary substantially depending on throughput rates and volume. The type of container lifting equipment, the length of sidings, the size and layout of the hardstand area and control mechanisms (including documentation) influences throughput rates. The BIE (1993) established that world best practice (WBP) for intermodal terminals and marshalling yards represented 20 per cent of total railway costs. Work undertaken by the BTE (1999) indicated that $0.0095/ntk could be attributed to rail terminal operations. The final report on the economic evaluation of the Darwin–Alice Springs Railway included a freight handling cost of $3/tonne, being the estimated benefit of eliminating the intermodal transfer of freight at Alice Springs (NT DoT&W 1999, p.5). More recently, industry sources have emphasised that handling costs are influenced by the spatial and temporal characteristics of each transport chain; however, as a rule of thumb the following rail costs represent best practice:

- $0.50/tonne for each loading or unloading of bulk grains or ore, where loadings and unloadings average less than 1.0 mtpa, and
- $15/empty TEU and $20/full TEU for containers.

Cost of capital: Regulators provide for the inclusion of capital costs (depreciation [a return of capital] and a return on assets) in the calculation of the revenue limits in a railway's access charges (ARTC 2001a, p.12). The value of these costs depends on the economic life and condition of assets, the Weighted Average Cost of Capital, the valuation and depreciation methodologies applied[9] and importantly, the accounting treatment of contributed assets and publicly funded infrastructure improvements. For below rail assets, the ARTC uses straight-line depreciation, with useful economic life as shown in Table 4.1.

Table 4.1: Economic life of ARTC assets

ASSET	USEFUL ECONOMIC LIFE (YEARS)
Bridges	40
Culverts	100
Signals and communications	10
Tunnels	50
Turnouts	12
Ballast	60
Rail	109
Sleepers	50

Source: ARTC (2003)

For the purposes of benefit–cost analysis, the more economically correct approach to spreading capital costs over the life of an asset is to convert the investment cost to an annuity at the discount rate used for the study. An annuity is the constant amount (S) received each year over a given number of years that discounts back to a given present value (PV). In this case PV is the initial investment value of the asset. S and PV are related by the following formula

9 The valuation methodology applied by Australian regulators to railways is Deprived Optimised Replacement Cost (DORC).

$$S = PV \frac{r}{1-(1+r)^{-n}}$$

where r is the discount rate and n is the number of years over which the annuity is taken (i.e. the economic life of the asset).

Management, administration and corporate overhead (MACO) costs: In the BIE (1993) work, corporate overheads represented 11 per cent of total costs in WBP railways. In its analysis of Class 1 railroads, the AAR (2002a) indicated that general and administrative costs represent, on average, 13 per cent of total operating costs. The ARTC (2001a) indicated that 17 per cent of its costs are incurred in contract management, system management and administration and insurance. When insurance is excluded, MACO costs, as a proportion of total ARTC costs, decreases to 12 per cent.

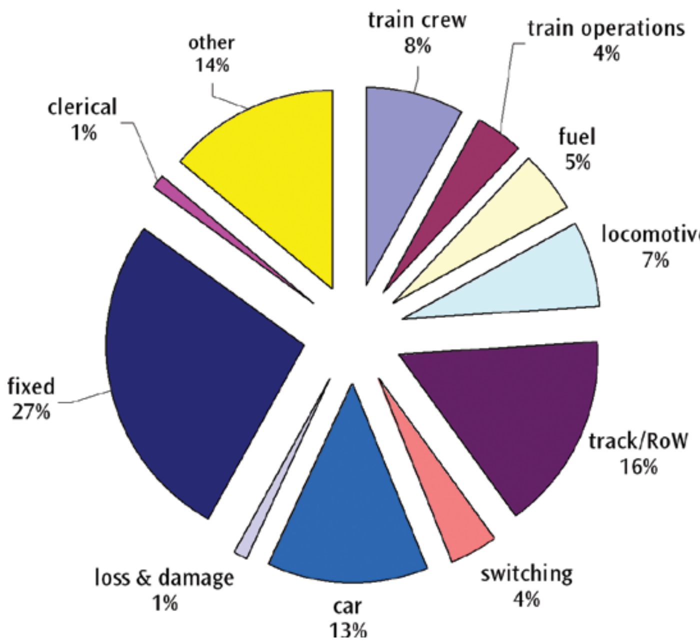
Costs passed on by safety and economic regulators: In Australia, statutory provisions relating to railways provide for the costs of safety and economic regulation to be borne by railways and through access charges, ultimately by end users. For example, in Queensland:

- Part 4A of the Transport Infrastructure (Rail) Regulation (Qld) provides for accredited railway operators and managers to pay an annual (safety accreditation) levy to the Queensland Government.
- Section 6.4(c) of the QR Access Undertaking provides for the recovery of fees levied by the QCA to be collected for the QCA by QR, through QR's access charges.

4.3.4 Total costs

Industry benchmarks, and the preceding analysis, suggest that above rail activities in Australia represent approximately one-third to one-half of total costs. Of course total inter-modal freight costs need to also include the costs of trucking the container to and form the inter-modal terminal. While it is critical that estimations account for individual operating parameters and differences in nomenclature between Australia and the US, the 'above' and 'below' allocations for US railways fall within the one-third to one-half range, as indicated in Figure 4.2. See also REEBIE Transportation Management Consultants (2004).

Figure 4.2: Typical allocations of rail costs in a US railway



Source: Reebie Transportation Management Consultants (2004)

A railway's determination of a freight rate may not always fully account for all current above rail and below rail costs. In the below rail context in Australia, this is seen in the negotiation of access prices between the access seeker and the railway, where the negotiation is guided by a regulated floor to ceiling price band or published reference tariffs (BTRE 2003, pp.75–77 and 79–84). A floor price represents the incremental, or avoidable cost, of providing access not including the costs of capital. A ceiling price is the total, or full, economic cost or stand-alone cost of access, including the costs of capital.

Typically, such variations are also seen in the determination of back-loading freight rates, where consideration is given to the cost avoided by not transporting the specific freight. For example, inter-modal trains on the North Coast Line in Queensland tend to carry more northbound freight than southbound. It could be expected that northbound freight rates fully reflect the costs of operating the train service. As the southbound train is in effect relocating to Brisbane, as a minimum, its freight rates need to cover the avoidable costs of moving the freight only. Such avoidable costs would include the costs associated with loading and unloading the freight and possibly the additional time, wear, tear and fuel consumed as a result of a heavier trailing load.

Further variations are seen in the productivity incentives offered to customers by freight train operators. For example, the NSW GIAC 2004 final draft report on grain logistics (GIAC, 2004) indicates that the structure of Pacific National rates includes productivity incentives for:

- outloading speed: $1.41 per tonne for improvement from <100 to >1000 tonnes per hour
- siding length: $0.51 per tonne for improvement from a siding of <20 wagons to a balloon loop
- shunting (if locomotive required): $0.89 per tonne for improvement in the outloading rate from <200 to >800 tonnes per hour, and
- silo operating hours: $1.55 per tonne for on demand operations.

A summary of values of rail cost parameters is given in Annex 1.

4.3.5 Avoidable costs

A railway operation could be characterised as a multi-product firm, since a single rail system usually carries both passengers and freight, provides services between a wide range of origins and destinations, and different types of freight have differing priority levels and handling characteristics. As a result, many components of costs are shared by different traffic tasks and there is no unambiguous way to allocate them. For example, how are track capital and fixed operating costs to be allocated between passenger and freight trains? For freight travelling between different origin–destination pairs that share a single length of track for part of their journey, should the track capital and fixed operating costs be allocated on the basis of gross tonne-kilometres, net tonne-kilometres, train kilometres, or wagon kilometres? The same applies to allocating system overheads across traffics. For accounting purposes, choices have to be made, but these inevitably involve an element of arbitrariness. The cost of carrying a particular traffic with all costs allocated using such arbitrary rules is called the fully distributed cost.

The avoidable cost concept can be useful in situations where there are significant shared costs. The avoidable cost of a given traffic task could be defined as the difference between total system cost with and without the traffic task. The task could be specified narrowly or broadly, and this will determine just which costs are avoidable, as the following examples show:

- In the example of backloaded freight on the Queensland North Coast line, the avoidable cost consists of the loading and unloading costs, additional operating time for trains, and the additional wear and tear and fuel consumed as a result of a heavier trailing load.
- For freight in a priority time slot on a major inter-capital route, the avoidable costs include, in addition to train operating costs and variable track maintenance costs, the additional revenue that could be earned by having the time slot available for other freight.
- For all freight and passengers using a branch line, the full cost of keeping the branch line open would be avoidable.

The avoidable cost provides a lower limit for setting prices if profitability is the objective. If the revenue earned from a traffic task was less than its avoidable cost, the railway operator would be better off by not undertaking the task at all. If all prices equalled avoidable costs, the enterprise as a whole would fail to cover total costs. So railway marketing managers would aim to set prices above avoidable costs so as to earn a contribution to the shared costs.

Avoidable costs should be used in financial and economic evaluations of proposals that involve adding or subtracting services or traffic.

Where trains and tracks are operated by different entities, the track-related part of financial avoidable cost to the train operator for a service or traffic will depend on the prices paid to the track operator. If the charge included a fixed component, this would not be part of the avoidable cost. For benefit–cost analysis purposes, resource costs are required, so access charges would be irrelevant. The avoidable costs would have to be estimated as the economic value of the resources saved by eliminating the service.

4.4 Rail benefits

Transport is a derived demand that can produce catalytic benefits for the economy, primarily through improvements in infrastructure that result in reduced transport costs. The catalyst for, and scope of, such benefits can range from the operational level to a large-scale or strategic level.

4.4.1 Operational level

At the below rail operational level, infrastructure improvements would normally result in lower track maintenance costs and reduced track possession time. Infrastructure improvements will also result in above rail benefits, including for example, lower fuel usage and decreased locomotive and rollingstock maintenance costs. Lower crewing and maintenance costs may also be achieved and, in time, there could be a net reduction in labour force size.

In the context of rail outputs, an improvement in track infrastructure should result in enhanced on-time performance. Reduced track possession time should equate to greater track capacity. The potential transport outcomes relate to lower rail transport prices and improved reliability and service availability.

It is possible to quantify these benefits using models calibrated with train and track performance data. Examples are contained in the *Rail Studies* completed by Laird (2000, 2002) for Queensland Transport.

In relation to time, Laird also examined the positive impact on crewing costs and the benefits inherent in the potential for increased train utilisation. Broader aspects of the relationship between time and cost concern the travel time savings accruing to passengers and freight from infrastructure improvements. The OXERA Environmental (2000) report includes a value of time for car drivers and rail passengers of £14 and £18 per person per hour respectively (in 1998/99 £ sterling)—see OXERA Environmental (2000, Table 2.2). For freight, de Jong (2000, p.559, Table 2) presents a range of values of time for freight transported by rail in the UK, ranging from $US0.09 to $US1.29 per tonne per hour (in 1999 $US).

In his examination of the value of travel time savings ($TTS), de Jong (2000, p.553) highlighted key differences between passenger ($PTTS) and freight ($FTTS), including the following:

- multiplicity of people involved in freight decision-making
- need to express $FTTS as $/hour, rather than $/minute (because of relatively longer distances and lower average speeds in freight transport)
- lack of uniformity in freight transport (for example in shipment size and value), and
- confidentiality of information.

In comparing road and rail improvement proposals, further work may be required on the valuation by mode of both $PTTS and $FTTS, given the disparities in the limited data available.

Ferreira and Camenzuli (1996, pp.184–185) indicate that the investment appraisal undertaken by railway operators is largely limited to financial evaluation, with no consideration of the benefits flowing to the community at large. Discussions with industry representatives in the preparation of this report confirmed these views, with assessments of net environmental and social benefits not historically included in modelling processes used by track owners and railway operators. Current modelling processes are discussed in Section 4.6. A wider consideration of modelling for travel demand forecasting and impact analysis is given in Part 3.

Improved safety, as a consequence of transport infrastructure investments, can provide substantial economic and social benefits. In the case of road transport, and given the major concerns about road safety held in the Australian community, improved safety for traffic operations and all road-users has been recognised for many years. The main issue is how to quantify these benefits. The standard approach is to estimate the reductions in the occurrence and severity of crashes, assuming that in reality, some crashes will continue to occur in the road system but different infrastructure investments and engineering measures and other countermeasures (e.g. in education and enforcement) will lead to improved traffic operating conditions under which fewer and/or less severe road crashes will be recorded. Research has demonstrated that certain countermeasures have positive safety outcomes for particular crash types. Careful statistical analysis is required to estimate the likely crash reduction effects of various measures—for example see BTE (2001). Further quantification of the economic value of crash reductions can then be made by the use of accepted crash costs data. Safety has long been an overriding concern in railway operations and similar analytical approaches may be employed to estimate the safety benefits of rail infrastructure investments, using the available databases on rail accident occurrence such as those of the Australian Transport Safety Bureau. (ATSB, 2004).

4.4.2 Strategic level

At the strategic level, infrastructure and service quality improvements would be expected to result in increased demand for rail and consequent better utilisation of the assets and improved productivity. A diversion of freight traffic from road to rail would constitute a substantial part of the increased demand for rail, which would then yield external benefits through reduced heavy goods vehicle usage of the road system, savings in fuel, improved road safety performance and reduced emissions of air and noise pollutants including greenhouse gases.

A recent major strategic level study is the interstate rail network audit conducted for the ARTC by Booz Allen Hamilton (BAH, 2001ab). This study included a performance evaluation of the national north–south rail corridors (Melbourne–Sydney, Sydney–Brisbane and Melbourne–Brisbane) and the east–west corridors (Melbourne–Adelaide, Melbourne–Perth and Sydney–Perth) on the basis of comparisons of different scenarios for rail infrastructure investment with the base case (present situation), from which an optimum level of investment could be derived. Besides the base case (BC), two investment scenarios based on upgrading the current DIRN were defined for the analysis:

- scenario S1, which specified minimum market improvements for each of the market segments represented by the three rail corridors (see above), with the minimum targets addressing rail's present critical service needs in each of the market segments, and
- scenario S2, which used 'stretch' targets and incorporated significant improvements in market characteristics. The S2 performance targets were based on providing a competitive service offering to road. The S2 investments included all of the S1 investments.

BAH (2001a) discusses the development of an optimal investment scenario (SO) based on the best mix of operational and investment measures from S1 and S2, which was analysed as the optimum investment scenario. In this scenario, a package of measures from S2 building on S1 was adopted, such that the net present value (NPV) of the scenario was maximised. In general this scenario did not include all of the measures proposed in S2. Three service characteristics assumed to be of major importance in improving rail market share were considered:

- transit time, defined as terminal-to-terminal travel time
- reliability, indicated by on time running performance (proportion of trains arriving within 15 minutes of the scheduled arrival time), and
- service availability, defined as the current proportion of the market for which rail was able to offer a broadly equivalent departure time as road.

In addition, the proposed inland route from Melbourne to Brisbane was included as an additional scenario for that corridor only.

Tables 4.2 and 4.3, taken from BAH (2001a), show the four performance scenarios for the three north–south market corridors and the three east–west market corridors respectively.

Table 4.2: Performance scenarios for north–south market corridors [Source: BAH (2001a)]

PERFORMANCE SCENARIOS	MELBOURNE–SYDNEY				SYDNEY–BRISBANE				MELBOURNE–BRISBANE				
	BC	S1	SO	S2	BC	S1	SO	S2	BC	S1	SO	S2	INLAND
Transit time (h)	13.5	11	10.5	9	21	19	17.5	16	36	32	29.5	27	27
Reliability (%)	55	75	75	95	50	75	75	95	45	80	75	95	95
Service availability (%)	50	70	75	80	25	50	60	70	60	85	85	90	90
Operational improvement		1500m train paths for all options -	-	-		1500m train paths for all options -	-	-		1500m train paths for all options -	-	-	1800m train paths DS
Capital costs ($M)		249	325	908		53	73	694		287	398	1614	1510
Market share (%)	11	19	20	26	19	27	30	36	21	32	35	39	54

Notes: All data taken from Exhibit 3.4 of BAH (2001a, p.11)
BC, S1, S2 and SO refer to the four general scenarios of Base Case, Scenario 1, Scenario 2 and Optimum Scenario.
For Melbourne–Brisbane, a fifth scenario for the proposed inland rail route was also included.
'Train path' refers to the maximum length of a train operating on the route. In some corridors specified train lengths have already been met.

Table 4.3: Performance scenarios for east–west market corridors [Source: BAH (2001a)]

PERFORMANCE SCENARIOS	MELBOURNE–ADELAIDE				MELBOURNE–PERTH				SYDNEY–PERTH			
	BC	S1	SO	S2	BC	S1	SO	S2	BC	S1	SO	S2
Transit time (h)	13	12	11.5	9	58	57	56	52	72	69	65	69
Reliability (%)	74	80	80	95	66	80	80	95	70	80	95	95
Service availability (%)	70	75	76	80	80	85	87	90	83	95	95	95
Operational Improvement	S1: no change from current S2: 1800m train paths for all options			DS	S1: no change from current S2: 1800m train paths for all options			DS	S1: no change from current S2: 1800m train paths for all options		DS	DS
Capital costs ($m)		113	133	810		27	52	626		8	78	37
Market share (%)	21	24	24	28	70	74	74	78	65	69	73	71

Notes: All data taken from Exhibit 3.5 of BAH (2001a, p.11)
BC, S1, S2 and SO refer to the four general scenarios of base case, Scenario 1, Scenario 2 and Optimum Scenario.
Train path' refers to the maximum length of a train operating on the route. In some corridors specified train lengths have already been met.

BAH (2001a, pp.13–19) discusses the outcomes of the study in terms of the performance of the different scenarios and the development of an investment plan based on the benefit–cost analysis. The investment required to achieve the S1 performance targets appeared to be well justified. The $337m overall network upgrade yielded a BCR of 3.7 and an NPV of $832m. All market corridors except Melbourne–Adelaide showed significant benefits and improvements to market share—this corridor was an exception because it had already been the subject of substantial recent investment. Most of the investment ($302m) was allocated to the north–south corridors, where the S1 improvements provided the largest benefits and potential improvements in rail freight volumes over the entire network. Under scenario S2, the investments required to achieve the performance targets generated notable improvements in modal share for rail. The overall result was a positive NPV of $323m with a marginal BCR of 1.2 over the base case. Table 4.4, taken from BAH(2001a, p.14), compares the results for S1 and S2 and shows that the additional capital expenditure generated a negative incremental return on the investment. This led to the development of the SO optimum scenario, involving investments beyond S1 but less than S2. Table 4.5 shows the overall optimised investment evaluation results. The resultant performance was shown in tables 4.2 and 4.3.

Table 4.4: S1 and S2 performance target evaluation results

	SCENARIO S1	SCENARIO S2	INCREMENTAL S1– S2
Undiscounted capital costs ($m)	337	2251	1914
Benefits ($m)	1138	2061	923
BCR	3.7	1.2	0.6

Source: BAH (2001a, p.14)

Table 4.5: Optimised investment evaluation results

RESULTS	SO INVESTMENT
Improvement in interstate intermodal market share from Base Case	38%
Undiscounted capital cost ($m)	507
BCR	3.2
Benefits ($m)	1453
NPV ($m)	994

Source: BAH (2001a, p.15)

BAH (2001a, pp.15–18) outlines a proposed investment plan for all corridors, which provides more details on the distribution of the investment and includes an investment ranking based on NPV.

4.4.3 Externalities

Usually benefit–cost analyses of rail projects include allowances for externalities such as noise, air pollution, greenhouse gas emissions and accidents. However, there appears to limited primary research into determination of the value of each externality. Users are recommended to refer to Appendix 2 of Volume 3 for rail environmental parameters, which have been derived from ARTC (2001a, 2001b), modified Infras/IWW (2000) and updated values from Pratt (2002). Additionally, a summary of accident and congestion values from a range of other sources are provided in Annex 2. See Section 2.8 of Volume 3—Estimate Safety Benefits for accidents, and 2.6 of Volume 3—Estimate User Benefits and 2.7 of Volume 3—Cross-Modal and Network Effects, for treatment of congestion in the context of these Guidelines.

4.5 Rail infrastructure improvement projects, programs and strategies

Section 2.1.3 of Volume 3 indicates:

> *'Where the impacts of a series of projects are closely interdependent, consider grouping the initiatives together, and then treating them as a single initiative'.*

In relation to railways, Ferreira and Camenzuli (1996, p.181 and 183) suggest that track investment must be structured into programs of projects because of the interdependence of functions and activities within a railway. They use the example of the potential impact of investment in locomotives, track, train control and terminal equipment in enhancing the reliability of train arrival times. They further identify the need to link projects into programs with a common objective or group of objectives. This issue was also discussed in BAH (2001a, pp.8–16).

An example of the practical application of the program approach to investment in rail infrastructure improvement is provided in the Rail Network Strategy for Queensland. The Strategy indicates that the Transport Services Contract (Rail Infrastructure) between Queensland Transport and QR includes funding of programmed works relating to specific types of work programs or designated asset classes such as the replacement of timber bridges or the refurbishment of turnouts and cross-overs (Queensland Government 2001, p.23).

Another example is seen in the simulation work completed by Laird (2000, 2002). In the smooth running study, Laird quantified the incremental costs that arise from slowing or stopping a standard gauge freight train and then returning it to normal speed. In the 'straight track' study he determined the benefits of curve easing on Queensland's North Coast Line, between Nambour and Townsville. Whilst both studies indicate that incremental savings in operating costs accrue on the completion of a project, it is apparent that the full range of benefits would not be realised unless the complete program of improvement projects in a corridor is undertaken. Only with both projects implemented would significant reductions in travel times be possible that would allow a more competitive rail service in this corridor.

Further guidance on an approach to infrastructure investment is provided in the Australian National Audit Office (ANAO, 1996). The asset management principles espoused in the handbook include (ANAO 1996, p.10):

- asset management decisions are to be integrated with strategic planning
- asset planning decisions are to be based on the assessment of alternatives, with consideration of life cycle benefits, costs and risks
- accountability is to be established for asset condition, use and performance.

In effect, the ANAO advocates the incorporation of asset strategies in operational and business plans. The ANAO indicates that such strategies encompass a four-step approach (ANAO 1996, p.16):

1. determination of asset needs, on the basis of forecast demand
2. evaluation of existing assets in terms of capacity to support demand, with emphasis on inventory identification and condition assessment
3. completion of a gap analysis, between existing assets and required assets
4. development of an asset strategy comprising acquisition, operation, maintenance and disposal plans.

It follows that this type of approach should be considered in the estimation of the impacts of rail infrastructure improvements. In these circumstances, the aim of the estimation process should be the maximisation of potential net benefits, through the identification of a grouping of possible projects that delivers the relatively greatest net benefits in the proposed program or strategy for a designated corridor or network.

In addition, there is a need to consider impacts on long term planning. This is of especial importance when considering changes in land use and development patterns that may result from major infrastructure development as discussed in Section 4.4.2. Hayashi and Roy (1996) provided a number of examples.

4.6 Estimation and simulation techniques

A variety of analytical techniques and modelling tools are available to assist in the estimation of rail systems' performance and impacts of infrastructure improvements. These tools begin with models to explain the movement of a single train along a track, and can then be extended to consider overall network performance.

In terms of the systems approach to transport modelling as outlined in Part 3 on Demand Forecasting and Modelling, the following models are all supply models as they predict the performance of components of a rail transport system under different levels of demand where the demands are supplied as exogenous inputs to the models.

None of the model packages surveyed were capable of estimating demand levels resulting from a given performance state (such models are described as demand models in Part 3, which also discusses the ensuing interaction between a demand model and a supply model to seek an equilibrium state for a transport system). One example of a full demand–supply equilibrium model applied to the study of an urban rail system is given by Jones, Simpson and Alchin (2001) for the Sydney region. A useful method for estimating incremental changes in demand for rail freight is the demand elasticity method described in Section 3.5 of Part 3. This method was used by Booz, Allen Hamilton (2001) in their audit of the ARTC network, with the results of that study summarised in Section 4.4.2.

4.6.1 Simulating the motion of a single train

The motion of a train along a track can be predicted using a mathematical model. Software tools based on these models can be used to answer questions such as:

- How long will it take to complete a journey?
- How much fuel or energy will be required?
- How fast will the train be travelling?

More detailed models can be used to predict, for example, traction motor currents or in-train forces.

The motion of a train along a track is determined by the forces acting on the train. The main forces are the traction and braking forces, running resistance, track curvature forces and track gradient forces. The motion of the train can be described by the equation of motion

$$m\frac{dv}{dt} = F(u,v) - R(v) - C(x,v) + G(x)$$

On the lefthand side of this equation, m is the mass of the train, v is the speed of the train, t is time, and dv/dt is the acceleration of the train.

The right-hand side of the equation is the sum of the forces acting on the train:

- $F(u,v)$ is the traction or braking force produced in response to control setting u when the train is travelling at speed v
- $R(v)$ is the resistive force at speed v due mainly to the rolling resistance of the wheels, bearing resistance and aerodynamic drag
- $C(x,v)$ is the force due to the curvature of the track when the front of the train is at position x and the train is travelling at speed v

- $G(x)$ is the force due to the gradient of the track when the front of the train is at x.

This is a relatively simple model. It does not model the details of the traction system, the dynamics of rotating parts, or delays in the air brake system. Such details are not necessarily required to give good predictions of train speeds and journey times.

Simulation of a train journey requires the following information about the train and track:

- maximum tractive effort of the locomotives as a function of speed
- service braking effort as a function of speed
- the total mass of the train, and the distribution of mass along the train
- the length of the train
- coefficients of running resistance, to allow the calculation of $R(v)$
- the maximum allowable train speed
- the gradient or altitude profile of the track
- the curvature of the track, and
- track speed limits.

There is also a need an algorithm to drive the train, that is to determine a sequence of control settings that will result in a speed profile that does not violate track or train speed limits. Many train simulators use a simple driving strategy that accelerates whenever the train speed is below the speed limit, maintains a constant speed when the train is at the speed limit, and brakes whenever necessary to avoid exceeding future speed limits. This driving strategy can be used to determine the minimum time required to complete a journey.

The mechanical work done by the train to overcome resistance, curvature and gradient forces can be calculated once the speed profile is known. A reasonably good estimation of fuel consumption can be made by multiplying the mechanical work done by a fuel efficiency factor. A better estimation of fuel consumption can be made by calculating the time spent in each control setting, multiplying each time by the corresponding the fuel flow rate, then summing.

In practice, train drivers do not drive using the minimum-time driving strategy. Some train simulators allow simple variations of the minimum-time driving strategy so that different driving styles can be modelled. Ultimately, however, the train simulator will calculate a single journey time and a single fuel consumption figure for a journey, whereas there will be significant variations in journey time and fuel consumption for real train journeys.

4.6.2 Fuel-efficient control of a single train

If the desired time for a journey is greater than the minimum possible time, then the control strategy can be chosen to give an 'optimal' journey. Typically, 'optimal' means a journey that minimises fuel consumption for a given journey time. As journey time increases, the control strategy can be adjusted so that the journey is still completed on time, but uses less fuel.

The problem of finding the control sequence that minimises the fuel used for a journey can be formulated and solved as an optimal control problem (Howlett 2000; Howlett and Pudney 1995, 2000).

Being able to calculate optimal journeys for a given journey time is often useful because:

- it is difficult to compare the fuel consumption of different types of locomotives hauling similar loads when the journey times are different; by calculating optimal journeys with a common journey time, a valid comparison can be made
- by calculating journeys with a range of journey times, it is possible to determine the effect of journey time on fuel consumption, and
- by calculating journeys with a common journey time, it is possible to determine whether it is more efficient to have one or two locomotives hauling a train—a higher power-to-mass ratio gives greater driving flexibility and, if the train is driven correctly, can give improved fuel consumption.

4.6.3 Interacting trains

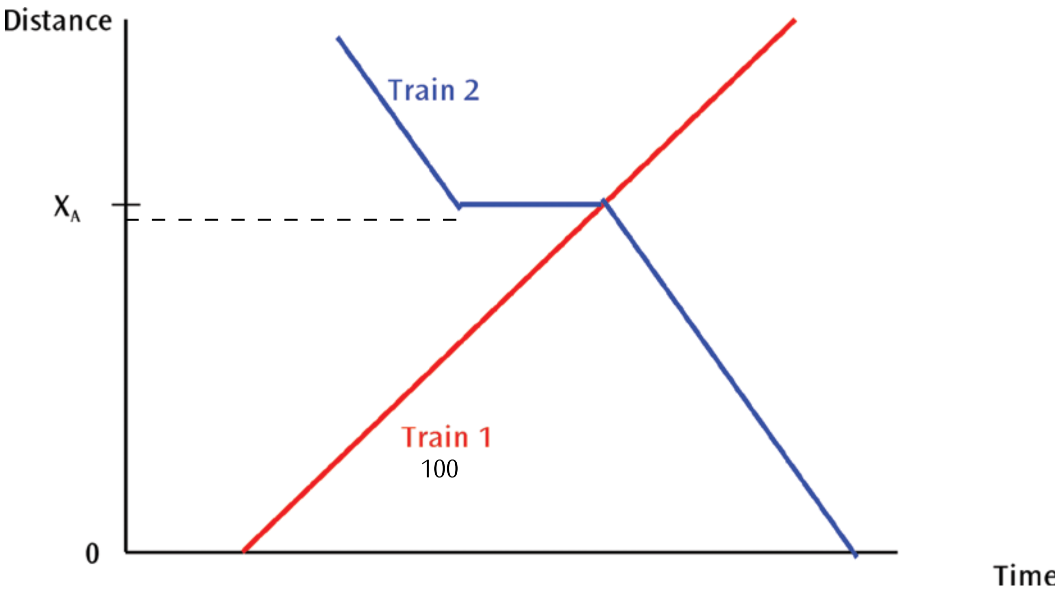
Trains interact. The effect of interaction is to increase the minimum possible journey times of individual trains, and to increase the fuel consumption required to meet a given journey time.

In the simplest case, a fast train travelling behind a slow train will be slowed to the speed of the train ahead by the signalling system. If the second train is repeatedly braking and accelerating in response to track signals then its fuel consumption will increase. Ideally, the second train should be paced so that it does not have to brake for signals, but in practise this is difficult to achieve.

Trains also interact at junctions and crossing loops. Frequently, a train will be driven to a junction or crossing loop and then wait for an opposing train movement. Ideally, trains would be driven efficiently to arrive at the interaction points at precisely the correct time, but once again this is difficult to achieve in practise. The cost is stop-start delays and increased fuel consumption.

Train path diagrams are used to show the progression of individual trains along a route, the interactions between trains, and the effects of those interactions on travel time and delay. Figure 4.3 shows a simple schematic train path diagram, involving two trains, travelling in opposite directions on a single track rail line. The train path diagram is a plot of the position of a train along a track at any point in time; it shows the trajectory of each train. The slope of the trajectory is the speed of the train at that point along the track. In Figure 4.3, Train 1 has priority along the track and therefore moves at a constant speed along the whole section, without delay. Train 2 progresses along the track until it reaches the passing loop at position X_A. It has to lay over in the loop until Train 1 has passed by. Then it can continue on its journey.

Figure 4.3: Simple schematic train path diagram showing two trains, travelling in opposite directions, on a single track rail line with a passing loop at position XA



The use of train path diagrams to analyse train delays was described in Part 3. Train path diagrams are widely used in determining train schedules. An example of a train path diagram for a busy rail line with mixed train traffic is shown in Annex 3.

A network of trains can be simulated, but the delays and fuel costs will depend on how the network is 'driven'. One approach is to drive each train as fast as possible, subject to the operation of the signalling system, to calculate the interaction times. The individual train journeys can then be recalculated to meet the target interaction times as efficiently as possible. This may result in a feasible timetable, but it is unlikely to be a good one. It may also result in deadlock, from which it becomes impossible to move each train forward to its destination. Different simulation dispatch procedures, used to determine when trains should wait and when they should move, may result in significantly different timetables, with significantly different journey times and fuel consumption.

Safe operation of trains is paramount in rail operations and a number of train control and safeworking systems have been developed over many years. The essence of a safeworking system is that one train cannot be placed in a situation where it could collide with a train ahead of it on the same track. This is normally accomplished by preventing the following train from entering a track section containing the leading train until that leading train has departed from the section. Only then can the second train enter the section. This applies most obviously on single track sections with trains travelling in opposite directions, where one train may have to rest in a passing loop until the other train has traversed the section. It also applies on double track systems for trains travelling in the same direction. Possible train control systems in use include 'Train Staff and Ticket', 'Electric Staff', 'Double Line Block', 'Section Authority', 'Centralised Train Control', 'Train Order Working' and 'Automatic and Track Control'. Many of these systems are very old and indeed originated in the nineteenth century, which is reflected in the technology with which they were implemented. Their basic operational principle was that a specific train took explicit control of a given track segment and no other train could enter that section until the first train had relinquished its control. A signal display visible to all trains indicated that control was in force. These systems are gradually being replaced by more modern automatic control systems that, as described by Vuchic (1981, pp. 536–48) offer improved track capacity and transit times. For example, in April 2000, ARTC replaced the century-old 'Electric Staff' (ES) control system on a seven km section of the Adelaide–Melbourne main line by a 'Centralised Train Control' (CTC) system, decreasing average travel time over the full route by 30 minutes (ARTC, 2000). Under ES control, a special instrument at each end of a section contains a number of Staffs[10], which are electrically interlocked so that only one Staff can be withdrawn at a time. Possession of the Staff is the authority to enter the single line section. Under CTC, all points and signals are controlled from a central location (VICSIG, 2004). Such systems require the continual monitoring of the locations of trains along a track from a central control centre (iQR, 2004).

4.6.4 Optimised timetables, congestion mapping and infrastructure planning

For optimum planning of train schedules there would be a need to automatically generate timetables that have trains interacting in a way that minimises the cost of lateness and fuel. Unfortunately, the difficulty of this problem increases rapidly with the number of track segments and the number of trains. Train scheduling techniques based on Lagrangean relaxation, network search techniques and probabilistic search techniques are described in the literature, but many of these do not work on realistic rail networks.

In collaborative research between a university and industry, UniSA and TMG International are developing software that will generate optimised timetables automatically. The work was originally funded by the Australian Research Council under its 'SPIRT' collaborative research program, and is continuing as part of the research program of the Cooperative Research Centre for Railway Engineering and Technologies (Rail CRC). The UniSA–TMG system uses a probabilistic search technique to compare thousands of different timetables, then selects the timetable that has the lowest total cost of lateness.

A fast, automated timetabling tool has many uses:

- *Better timetables*: Timetables are currently generated using manual scheduling methods, and it can take many weeks to generate a single feasible timetable. The automated tool can generate and evaluate hundreds of timetables every minute, giving the train planner the opportunity to investigate the effects of changes such as altered departure times or added trains.
- *Dynamic rescheduling*: The automated timetabling tool can be used in a control centre to calculate the best way to recover from operational disturbances.
- *Integrated scheduling*: Once train timetables can be generated automatically, train timetabling can be integrated with maintenance planning and crew rostering.
- *Congestion mapping*: An automated timetabling system can generate 'congestion maps' that show where and when the network is congested. The tool will also allow train planners to quickly investigate ways of relieving the congestion.

10 A 'Staff' is a special token used in safe working systems; possession of a Staff gives authority to enter a single line section of a rail track (see VICSIG, 2004).

- *Infrastructure planning*: An automated timetabling tool will allow network planners to investigate the effects of infrastructure changes on network operations.

Several of these issues are the subject of current research in the Rail CRC, which has a full research program 'Theme 3' on 'Optimal traffic control and scheduling' in the area of systems modelling of rail networks and services[11].

4.6.5 *Contemporary approaches to rail modelling*

There are a number of analytical and modelling tools currently used in the Australian rail industry, for investment appraisal and system performance management purposes. It should be noted that the Rail CRC has a substantial research program aimed at extending the capabilities of some of the existing tools and developing new tools.

An overview of a selection of the modelling tools currently used within the rail industry is shown in Table 4.5. As indicated at the start of Section 4.6, these models may all be classed as 'supply models' under the classification given in the modelling overview of Part 3. They all estimate the performance of some component(s) of a railway system under an assumed demand. None of the models cited in the table can be used to forecast that demand, although they provide necessary inputs to a 'demand model' as described in Part 3.

Table 4.6: Selection of contemporary modelling, simulation and operational tools

TOOL	REMARKS
MTRAIN	TMG provides[12]: • MTRAIN—for train performance and signal system simulation. • In 2000, Arup-TMG was commissioned by the ARTC to audit the Inland Railway proposal. Using MTRAIN, Arup-TMG constructed a geometric model of the Parkes to Brisbane and Newcastle to Brisbane corridors and simulated the operational performance of the corridors. [13]
CAPTURE/ PROVING	CAPTURE/PROVING—for timetable development, train flow monitoring and conflict detection.
RAIL//TRAIN	The RAIL//TRAIN tool—a single-train simulator based on the MTRAIN package, which was initially developed in about 1970 by the NSW Government Railways as a journey simulator for suburban trains. It calculates minimum-time journeys by default, but allows manual specification of the driving strategy. It also calculates journeys and signal interactions for a fleet of trains travelling in the same direction along a corridor. It has been used by most Australian railways.
RAIL//TABLE	RAIL//TABLE—a timetable development system that allows manual editing but automated checking of network timetables.
SKETC.H	SKETC.H—an infrastructure and timetable planning system that can be used to determine the scope for changing or reconfiguring infrastructure and developing different timetables.
TRNSYS	TRNSYS—for locomotive and train assignment and scheduling.
RAILSYS	RAILSYS—a suite of integrated tools for railway network planning, infrastructure maintenance, train performance, rollingstock scheduling and timetable development.
SCHEDULEMISER	ScheduleMiser—a tool for generating optimised train timetables that is currently being developed by TMG and UniSA as part of the Rail CRC program.

11 For details on the Rail CRC see http://www.railcrc.cqu.edu.au.

12 For details of the TMG tools see http://www.tmgint.com.au/pdf/Rail%20Business%20Units.pdf.

13 For details of the eight stages involved in the modelling, see Section 7.1 of ARTC 2001, Audit of the Inland Rail Proposal—Parkes to Brisbane, report prepared by Arup–TMG.

TOOL	REMARKS
FREIGHTMISER	The FreightMiser system is an in-cab advice system that helps the driver of a long-haul train stay on time and save fuel. The system is currently being developed by TMG and UniSA as part of the Rail CRC program. The FreightMiser system includes a journey optimisation tool that calculates energy-efficient control strategies for long-haul journeys.
IMPACT	Integrated Maintenance Planning and Capacity Tool (IMPACT) used by QR Network Access.
TMS and ICS	Pacific National uses the Train Management System (TMS) to assist in rail freight planning and operations and Intermodal Control System (ICS) for intermodal terminal operations. TMS and ICS are the core of the company's application package. TMS has three tiers, with client, business logic and database levels, using Oracle PL/SQL and Oracle 9i.[14]
OASIS	The Optimisation Alternatives Strategic Intermodal Scheduler (OASIS) is used to optimise the loading and unloading of containers between road and rail.
RAMS	Rail Access Management System (RAMS) used by ARTC, Westrail, Freight Australia, Tasrail, Freight Link, Australian Southern Railroad and QR (interstate network only). The system provides a comprehensive train information and integrated enquiry system for all trains operating across the ARTC network.[15]
PLANIMATE	Planimate from Interdynamics Pty Ltd is a discrete event simulation platform that has been applied to railway simulation problems. It is currently in use by QR as a general simulation modelling tool.
SimTrain	SimTrain is a train performance tool developed by Samron Pty Ltd. Key SimTrain data includes: › track file—detailing speeds and vertical and horizontal profiles, and › train file—detailing power, trailing load, train length, maximum speed, braking rates, fuel consumption, rolling resistance, etc. Laird (2000) indicates that the SimTrain model has been calibrated against field observations, deals with data at 10m intervals and can be operated in 'aggressive' or 'variable economy' driving modes. It can be used to determine run times, fuel usage, operational impacts of proposed infrastructure improvements or changed infrastructure conditions (e.g. speed restrictions), greenhouse emissions, etc.
DPM	The Dispatch Planning Model (DPM) developed by Berkely Simulation Software is a railway network simulator. It uses simplified train models and a priority-based dispatch algorithm to simulate the movement of trains on a rail network. DPM was used by National Rail briefly during the early 1990s.
NFG 2	NFG2 is a costing convention initially used by government-owned railways to determine cost transfers between states. It comprises 13 cost types: crew; fuel; shunting; corporate overheads; business overheads; renewable assets; non-renewable assets; handling of freight; handling of passengers; maintenance of locomotives, wagons, track, and signalling and communications.
ViziRail	ViziRail from the Intec Consulting Group is an integrated system for managing manually prepared network timetables.
Simu++	Simu++ is a simulation model for analysing railway operations on complete or partial networks, developed at the University of Hannover in Germany. It is currently used in NSW and Victoria.
TEM	The Train Energy Model (TEM) was developed by the Transportation Technology Centre Inc. for the Association of American Railroads. It is a single train simulator for long-haul trains, designed to calculate journey time and fuel use. It does not simulate fleets of trains. The locomotives, rolling stock and environment are modelled in more detail than RAIL//TABLE, but the results are generally similar.

14 For further information see www.ara.net.au/abdoc/PNTMS.pdf

15 For further information see http://www.artc.com.au/company/annual.htm http://www.arg.net.au/docs/ASR-RSD-019%20Part%2032.1.pdf.

Several of these models (including MTRAIN, Planimate, DPM and Simu++) are potentially suitable for assessing the impacts of rail upgrading and renewal projects on railway operating costs, trip times and reliability, for both individual projects and for whole investment programs.

A discrete event simulation model such as Planimate could be used to examine potential improvements in reliability from infrastructure improvements—this is also an area of current research in the Rail CRC research theme on 'optimal traffic control and scheduling' (see Section 4.6.4).

The ICS model has potential applications in estimating benefits from improvements to intermodal terminals.

All of the models in Table 4.5 are implemented as specialist software packages and may require specialised knowledge for their application. This review was unable to find any implementations of models in more general software platforms, such as spreadsheets. However, the Rail CRC's research program on traffic control and scheduling is concerned with the development of decision-support tools for rail network planning and a range of new models, including simple 'sketc.h planning' tools, are being developed in this. Details on the work and research outputs of this CRC are available at www.railcrc.cqu.edu.au.

The major limitation of all the models is that they are supply-side models. They can estimate how some part of a rail system will perform under a given level of demand, but they cannot, by themselves, be used to predict the likely level of demand. As indicated at the start of this section, one approach to modelling demand is to use the economic elasticity approach, as described in Section 3.5 of Part 3 and as applied in the recent ARTC rail network audit (BAH, 2001b).

4.7 Performance analysis and measurement

Generally, the analysis and measurement of rail performance has tended to focus on total business or network performance. For example:

- Nash (2000) suggested measures for comparing individual railways, including partial productivity, total factor productivity and data envelopment.
- Oum et al, (1998) provided a comprehensive survey of methodologies for measuring railway efficiency and productivity, including a selection of partial indicators of rail productivity and a summary of productivity and efficiency estimates undertaken in the US, Canada and Europe, using econometric models.

This section outlines:

- an approach to the analysis of corridor performance
- measurements of rail infrastructure performance 'efficiency' and 'effectiveness'
- national and state approaches to performance reporting, and
- other issues that may arise in an analysis of railway performance.

4.7.1 Corridor performance analysis

When considering the impacts of proposed rail infrastructure improvements, it may be feasible to adapt data envelopment analysis[16] to assess the forecast performance of the improved rail corridor relative to a pre-determined, similar WBP corridor. However, such an approach is likely to encounter the difficulties inherent in making comparisons. In particular the availability of robust data, the accurate identification of outputs and inputs and the extent of aggregation may be problematic.

16 The theory of information asymmetry was espoused initially by George Akerlof and developed further by Mirrless and Vickrey in the early to mid-1990s and subsequently, in relation to markets, by Akerlof, Spence and Stiglitz. In 1970 George Akerlof wrote '*The Market for Lemons: Quality Uncertainty and the Market Mechanism*, a study of the effects of information asymmetry on the second-hand market for used cars. In 1996 James Mirrless and William Vickrey were awarded the Bank of Sweden prize in Economic Sciences in Memory of Alfred Nobel for their contributions to the theory of incentives under asymmetric information. The 2001 Prize was awarded to George Akerlof, A. Michael Spence and Joseph Stiglitz for their analyses of markets with asymmetric information.

A more viable approach would be to compare current corridor performance against predicted performance following the infrastructure improvements. A comparison of this nature reviews the relationship between the attributes underpinning rail's market success and the current and forecast future attributes of the rail corridor. If there is a reasonable degree of substitution possible between road and rail, then price, transit time, reliability and service availability will be the primary factors in determining rail's market success. Analysis could include the use of benchmarks set with reference to the cost and service quality attributes that rail may need to achieve in order to maintain or improve its competitive position vis-à-vis road.

The corridor attributes to be considered should include average speed, axle loads, train length and corridor capacity and utilisation. Other attributes impacting on rail's performance include for example, the influence of traffic density and traffic mix on reliability and service availability, and rail's proportion of total 'door-to-door' transit time. In the context of intermodal terminal performance on the corridor, rail's market penetration will also be influenced by road access to rail terminals and the extent of short loading (i.e. containers booked on a train but not carried).

4.7.2 'Efficiency' and 'effectiveness' indicators

In terms of 'efficiency', the indicators for a rail network could include:

- labour productivity (net tonne-km [ntk]/employee, ntk/locomotive driver, ntk/maintenance employee), and
- capital productivity (ntk/locomotive, ntk/wagon, ntk/track km, ntk/track machine).

However, in the context of the impact of proposed infrastructure improvements to a corridor, it may be more appropriate to examine the current and forecast level of capital utilisation, possibly in terms of train paths used/train paths available and trailing gtkms/ntks (as a proxy for loading/train).

Providing sufficient information is available, 'effectiveness' indicators for a corridor include:

- modal share, including the potential for traffic diverted from road as a result of rail infrastructure or performance improvements
- safety (incident and accident rates, fatality rates, temporal and spatial incident and accident trends, accident severity and costs)
- ecological sustainability (energy use, train productivity, greenhouse gas and particulate emissions, noise), and
- other avoided costs (road capacity and/or maintenance, provision of parking space, etc.).

4.7.3 National and state approaches to reporting rail performance

In August 2002 the Australian Transport Council (ATC) endorsed revised performance targets for the DIRN, subject to the resolution of track leasing arrangements between NSW and the ARTC. The agreed performance targets for intermodal trains relate to on-time reliability, transit time, train length and double- stacking capability for the eight corridors comprising the DIRN.

The Bureau of Transport and Regional Economics (BTRE) and the Australasian Railway Association (ARA) are collaborating on the provision of relevant performance data for monitoring and reporting against the ATC targets.

In Queensland, the Transport Services Contract (Rail Infrastructure) between Queensland Transport and QR provides for quarterly and annual reporting against base service levels relating to track condition, track availability, temporary speed restrictions and below rail delays (Queensland Government 2001, p.23).

In advocating the development of integrated performance reports, the ANAO suggested reporting on physical condition, functionality, utilisation and financial performance (ANAO 1996, p.44). This appears to be reflected in the requirements set by Australian regulators for track owners. For example, in addition to performance reporting, the ARTC Undertaking provides for ARTC to publish on an annual basis, information relating to the following costs:[17]

- unit maintenance cost: expenditure on outsourced infrastructure maintenance and maintenance contract management
- unit train control cost: expenditure on train control and transit management
- unit operations cost: operations planning and management and a pro-rata allocation of ARTC corporate overheads.

In North America, the six US Class 1 railroads and the two major Canadian railways publish performance data for each railroad, on a weekly basis, on:[18]

- Total Cars On Line
- Average Train Speed
- Average Terminal Dwell Time
- Bill of Lading Timeliness.

4.7.4 Other issues in performance analysis

In concluding, mention should also be made of the possibility (real or perceived) that the results of modelling and estimation processes are biased if:

- the basis for each assumption is not reasonably justified or has not been established by research
- data inputs are assessed to be incomplete or out of date, and
- information asymmetry is seen to bias the estimation process.

Information asymmetry occurs where one party to a proposed transaction has relatively more information about matters relevant to the transaction and is able to exploit the information advantage to the detriment of the other party.[19] Identifying and benchmarking similar railways and corridors from operational, spatial, temporal and organisational perspectives can assist in reducing the problems created by information asymmetry. When benchmarking, care must also be taken in ensuring agreement on the definition of the scope of proposed infrastructure projects and in the selection of performance measurements and base years.

In addition, increased knowledge of current infrastructure improvement proposals should provide for enhanced understanding of the relationship between the proposed infrastructure improvement and likely transport outcomes. Inherent in this relationship is the interdependence between 'above rail' and 'below rail', wherein the combination of 'track quality' and 'net revenue' should be optimised. A selection of current, publicly available rail infrastructure proposals is provided at Annex 4.

Issues on data collection and analysis, including survey methods and the elimination of bias, are discussed in Part 3.

17 For 2002/03 unit cost calculations, see http://www.artc.com.au/access/Access_2.10.htm.

18 See http://www.railroadpm.org/ for current performance data by railroad and definitions of performance measures.

19 The theory of information asymmetry was espoused initially by George Akerlof and developed further by Mirrless and Vickrey in the early to mid-1990s and subsequently, in relation to markets, by Akerlof, Spence and Stiglitz. In 1970 George Akerlof wrote 'The Market for Lemons: Quality Uncertainty and the Market Mechanism, a study of the effects of information asymmetry on the second-hand market for used cars. In 1996 James Mirrless and William Vickrey were awarded the Bank of Sweden prize in Economic Sciences in Memory of Alfred Nobel for their contributions to the theory of incentives under asymmetric information. The 2001 Prize was awarded to George Akerlof, A. Michael Spence and Joseph Stiglitz for their analyses of markets with asymmetric information.

4.8 Bibliography

The bibliography is arranged in order of the Sections of part 4. The citations under each heading include the references given in the text and some other works that are of relevance and interest to the subject matter covered in the specific sections.

4.8.1 Needs for impact estimation

ARTC (2001b), *Interstate Rail Network Audit – Final Report.* Report prepared by Booz Allen Hamilton, Melbourne, at **http://www.artc.com.au/library/final_reportv14.pdf**. Australian Rail Track Corporation, Adelaide.

Banister, D and Berechman, Y (2001), Transport investment and the promotion of economic growth. *Journal of Transport Geography 9,* pp. 209–18.

Laird, P, Newman, P W G, Bachels, M A and Kenworthy, J R (2001), *Back on Track— Rethinking Transport Policy in Australia and New Zealand*. (UNSW Press: Sydney)

Nash, C (2000), Modelling performance: rail. In Hensher, D A and Button, K J (eds). *Handbook of Transport Modelling*. (Pergamon-Elsevier: Oxford), pp. 565–75.

4.8.2 Rail costs

ARTC (2001a), *Australian Rail Track Corporation Ltd Access Undertaking Explanatory Guide*, at http://www.artc.com.au/docs/news/pdf/access_undertaking.pdf, Australian Rail Track Corporation, Adelaide.

ARTC **(2003),** *Annual Report 2002/03*, Note 2 to Financial Statements. **http://www.artc.com.au/library/annual_report_2003_f.pdf.** Australian Rail Track Corporation, Adelaide.

Association of American Railroads (AAR) (2002a), *Analysis of Class 1 Railroads: 2000*.

Association of American Railroads (AAR) (2002b), *AAR Railroad Cost Indexes, March 2000.*

BIE (1993), *International performance indicators—rail freight update 1993*, Research Report 52. AGPS, Canberra.

BTE (1999), *Competitive neutrality between road and rail*. Working Paper 40, Bureau of Transport Economics, Canberra.

BTRE (2003), *Rail infrastructure pricing: principles and practice*. Report 109. Bureau of Transport and Regional Economics, Canberra.

Dennis, N (2001), *Auditing.* Paper presented to QR Railway Civil Engineering Course 2001. Unpublished mimeograph, p. 4 & 9.

EDI Rail (2003), Media statement dated 18 March 2003.

Laird, P (2000), The smooth running study. In *Queensland Transport 2003, Rail Studies.* Queensland Government, Brisbane.

Laird, P (2002), The straight track study. In *Queensland Transport 2003, Rail Studies*. Queensland Government, Brisbane.

Laird, P, Michell, M and Adorni-Braccesi, G (2002), Sydney–Canberra–Melbourne high speed train options. *Papers of the 25th Australasian Transport Research Forum*. October, Canberra. CD-ROM. (BTRE: Canberra).

Martin, M (2001), *Ballast specification.* Paper presented to QR Railway Civil Engineering Course 2001. Unpublished mimeograph, p. 2.

Northern Territory Department of Transport and Works (NTDoT&W) (1999), *Economic evaluation of Darwin – Alice Springs railway*. Report prepared by Booz, Allen Hamilton, Melbourne, at **http://www.aarc.com.au/aarc/news/miscel/booz-allen%20&%20hamilton%20report.pdf**

NSW Grain Infrastructure Advisory Committee (GIAC) (2004), *Report on road / rail options for grain logistics*. Final draft report January 2004, at **http://www.transport.nsw.gov.au/giac/GIAC-Report.pdf**

Profillidis, V A (2000), *Railway Engineering* (2nd ed). (Ashgate Publishing Co: Aldershot).

Queensland Competition Authority (QCA) (2000), *Usage related infrastructure maintenance costs in railways*. Working Paper, QCA, Brisbane.

REEBIE Transportation Management Consultants (2004), CO$TLINE® Rail Cost Analysis Model (RCAM). Reebie Transportation Management Consultants (Reebie Associates) was acquired by Global Insight's in 2005. The CostLine modal cost model tools are still available following the absorption of Reebie Associates by Global Insight in early 2005. These can be accessed from the Global Insight website at: http://www.globalinsight.com/gcpath/Costline.pdf.

4.8.3 Rail benefits

ARTC (2001), *Interstate Rail Network Audit—Audit of ATC Performance Targets* April, Appendix B. Available at **http://www.artc.com.au/library/appendix_b.pdf**

ATSB (2004), Rail safety. Australian Transport Safety Bureau, Canberra, at **http://www.atsb.gov.au/rail/index.cfm.**

BAH (2001a), *Summary report: interstate rail network audit.* Report to ARTC. Booz, Allen and Hamilton, Sydney. **http://www.artc.com.au/library/summary_reportv7.pdf**

BAH (2001b), *Appendix A—evaluation methodology: interstate rail network audit.* Report to ARTC. Booz, Allen and Hamilton, Sydney. **http://www.artc.com.au/library/appendix_a.pdf**

Banister, D and Berechman, Y (2001), Transport investment and the promotion of economic growth. *Journal of Transport Geography 9,* pp. 209–18.

BTE (2001), *The Blackspot Programme 1996-2002: an evaluation of the first three years*. Report 104. Bureau of Transport Economics, Canberra.

de Jong, G (2000), Value of freight travel time savings. In Hensher, D A and Button, K J (eds). *Handbook of Transport Modelling.* Pergamon-Elsevier: Oxford).

Ferreira, L J and Camenzuli, A (1996), Railway investment: appraisal and business planning issues. *International Journal of Transport Economics XXIV* (2), pp. 179–91.

Infras/IWW (2002), *External Costs of Transport—Accidents, Environmental and Congestion Costs in Western Europe*, Zurich/Karlsruhe.

Laird, P (2000), The smooth running study. In *Queensland Transport 2003, Rail Studies*. Queensland Government, Brisbane.

Laird, P (2002), The straight track study. In *Queensland Transport 2003, Rail Studies*. Queensland Government, Brisbane.

OXERA Environmental (2000), *The wider impacts of road and rail investment*, paper prepared for The Railway Forum, United Kingdom, at **http://www.railwayforum.com/getfile.php?id=19**.

Pratt, C. (2002), *Estimation and Valuation of Environmental and Social Externalities for the Transport Sector.* Paper presented at the Australasian Transport Research Forum (ARTF) Conference in Canberra, October.

4.8.4 Rail infrastructure improvement projects, programs and strategies

ANAO (1996), Asset management handbook. Australian National Audit Office, Canberra.

BAH (2001a), *Summary report: interstate rail network audit*. Report to ARTC. Booz Allen and Hamilton, Sydney. http://www.artc.com.au/library/summary_reportv7.pdf

Commonwealth Department of Transport and Regional Services (DoTARS) (2004), *Draft National Transport Investment Appraisal Guidelines.* Department of Transport and Regional Services, Canberra.

Ferreira, L J and Camenzuli, A (1996), Railway investment: appraisal and business planning issues. *International Journal of Transport Economics XXIV* (2), pp.179–91.

Hayashi, Y and Roy, J R (eds). (1996) *Transport, Land-Use and the Environment.* (Kluwer Academic: Dordrecht).

Laird, P (2000), The smooth running study. In *Queensland Transport 2003, Rail Studies.* Queensland Government, Brisbane.

Laird, P (2002), The straight track study. In *Queensland Transport 2003, Rail Studies.* Queensland Government, Brisbane.

Queensland Government (2001), *Rail Network Strategy for Queensland,* Queensland Transport, Brisbane.

4.8.5 Estimation and simulation techniques

ARTC (2000), *Signalling project boosts rail efficiency.* Latest News, 27 April 2000, at http://www.artc.com.au/library/news_270400.pdf

BAH (2001b), *Appendix A—evaluation methodology: interstate rail network audit*. Report to ARTC. Booz, Allen and Hamilton, Sydney. http:www.artc.com.au/library/appendix_a.pdf

Howlett, P G (2000), The optimal control of a train. *Annals of Operations Research 98*, pp. 65–87.

Howlett, P G and Pudney, P J (1995), *Energy-Efficient Train Control.* (Springer-Verlag: London).

Howlett, P G and Pudney, P J (2000), Energy-efficient driving strategies for long-haul trains. *Proceedings of CORE 2000 Conference on Railway Engineering*, 21–23 May, Adelaide.

iQR (2004), *Universal Traffic Control (UTS) system: products and services,* at http://www.iqr.com.au/default.aspx?sid=2q54.

Jones, M, Simpson, B and Alchin, S (2001), Strategic models for new railway planning – a New South Wales perspective. *Papers of the 24th Australasian Transport Research Forum 24,* Hobart, April, CD-ROM (Department of Infrastructure, Energy and Resources: Hobart).

Laird, P (2000), The smooth running study. In *Queensland Transport 2003, Rail Studies.* Queensland Government, Brisbane.

Laird, P (2002), The straight track study. In *Queensland Transport 2003, Rail Studies*. Queensland Government, Brisbane.

VICSIG (2004), *Glossary of terms used for safeworking (Victorian version),* at http://www.vicsig.net/index.php?page=glossary&type=safeworking.

Vuchic, V R (1981), *Urban Public Transportation: Systems and Technologies.* (Prentice-Hall: Englewood Cliffs NJ).

4.8.6 *Performance analysis and measurement*

ANAO (1996), Asset management handbook. Australian National Audit Office, Canberra.

BAH (2001a), *Summary report: interstate rail network audit*. Report to ARTC. Booz, Allen and Hamilton, Sydney. **http://www.artc.com.au/library/summary_reportv7.pdf**

BAH (2001b), *Appendix A—evaluation methodology: interstate rail network audit.* Report to ARTC. Booz, Allen and Hamilton, Sydney. **http://www.artc.com.au/library/appendix_a.pdf**

Oum T H, Waters, W G and Yu, C (1998), A survey of productivity and efficiency measurement in rail transport. *Journal of Transport Economics and Policy 33* (1), pp.9–42.

Queensland Government (2001), *Rail Network Strategy for Queensland,* Queensland Transport, Brisbane.

4.8.7 *Summary of external values used in recent rail studies*

ACIL (2001)*, Rail in sustainable transport.* Report to the Rail Group of the Standing Committee on Transport, May 2001 at **http://www.transport.qld.gov.au/qt/railport.nsf/index/rail_in_sustainable_transport**

ARTC (2001b), *Interstate Rail Network Audit – Final Report.* Report prepared by Booz, Allen Hamilton, Melbourne, at **http://www.artc.com.au/library/final_reportv14.pdf**. Australian Rail Track Corporation, Adelaide.

BTE (1999), *Competitive neutrality between road and rail.* Working Paper 40, Bureau of Transport Economics, Canberra.

BTRE (2002), *Greenhouse gas emissions from transport: Australian trends to 2020.* Report 107, Bureau of Transport and Regional Economics, Canberra.

De Jong, G (2000), Value of freight travel time savings. In Hensher, D A and Button, K J (eds). *Handbook of Transport Modelling.* Pergamon-Elsevier: Oxford).

Laird, P (2001), Land freight external costs in Queensland. In *Queensland Transport 2003, Rail Studies.* Queensland Government, Brisbane.

Nash, C (2000), Modelling performance: rail. In Hensher, D A and Button, K J (eds). *Handbook of Transport Modelling.* (Pergamon-Elsevier: Oxford), pp.565–75.

OXERA Environmental (2000), *The wider impacts of road and rail investment,* paper prepared for The Railway Forum, United Kingdom, at **http://www.railwayforum.com/getfile.php?id=19**

QR Network Access (2002), *Comparison of Greenhouse Gas Emissions by Australian Intermodal Rail and Road Transport: a report commissioned from Affleck Consulting Pty Ltd*, QR, Brisbane.

Annex 1 Representative parameter values

Table A1.1: A summary of values of rail cost parameters

SERIAL (a)	PARAMETER (b)	EST. VALUE[20] (c)	MEASURE (d)	REMARKS (e)
1	Track maintenance	$10,100	per track km	Maintenance costs will vary significantly depending on asset condition, usage, terrain, soil type, climate and weather.
2	Train control	$292	per '000 train kms	Train control costs are influenced by the type of control; however, dis-aggregation by system type is not feasible.
3	Locomotive capital cost	$4.5m to $6m	–	Estimate $1.68m/'000 hp, depending on the nature of the contract with the manufacturer.
4	Rollingstock capital cost	$150,000	per wagon	Wagons may be configured in 'five packs'.
5	Locomotive ½ life re-fit cost	$1.5m to $2m	–	If not included in purchase contract.
6	Wagon ½ life re-fit cost	$15,000 to $22,500	–	If not included in purchase contract.
7	Locomotive economic life	20 to 25 years	–	
8	Wagon economic life	25 to 30 years	–	
9	Locomotive maintenance	$0.80 to $1.50	per km	or $30 to $55 per hour
10	Wagon maintenance	$45 to $50	per '000 kms	
11	Fuel consumption (diesel)	3 to 5.4 litres	per '000 GTK	Older locomotives have higher fuel usage
12	Terminal lift price (empty TEU)	$15	per TEU	Not including storage and power costs
13	Terminal lift price (full TEU)	$20	per TEU	Not including storage and power costs
14	Locomotive crew cost (two person)	$100 to $130	per train crew hr	
15	Load damage	$2	per train per day	
16	'Below Rail' asset economic life	–	–	See Table 4.1 in body of report

20 Note that the dollar values are drawn from ARTC (2001).

Annex 2 Summary of externality values used in recent rail studies

This annex contains tables relating to the estimated costs of:

A1. Congestion

A2. Accidents

Each table contains a summary of monetary values for the designated externality. The ranges of values have been extracted from known publicly available reports.

Table A2.1: Range of values for congestion

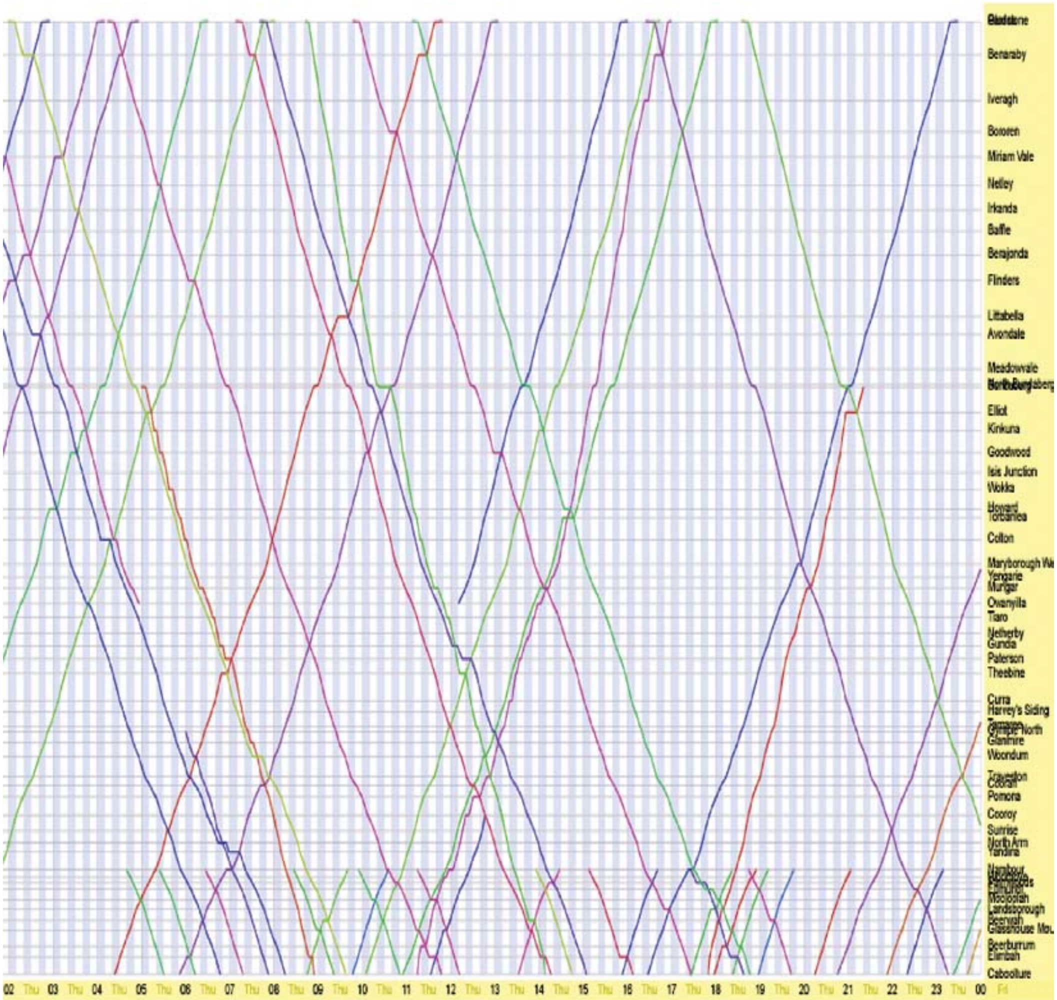
SERIAL	DESCRIPTION	ROAD VALUE	ROAD MEASURE	RAIL VALUE	RAIL MEASURE	SOURCE
[a]	[b]	[c]	[d]	[e]	[f]	[g]
1	European congestion costs (freight transport)	12	ECU /'000 t km	23	ECU /'000 t km	ACIL 2001, *Rail in Sustainable Transport*, report to the Rail Group of the Standing Committee on Transport, May 2001, Table 30 citing European Conference of Ministers of Transport (ECMT) 1998, *Efficient Transport for Europe: Policies for Internalisation of External Costs*, OECD Publications Service, Paris.
2	European congestion costs (passenger transport)	4	ECU /'000 pax km (car only)	8	ECU /'000 pax km	As above
3	United Kingdom value of passenger travel time	14	£ /pers/hr	18	£ /pers/hr	OXERA Environmental 2000, *The Wider Impacts of Road and Rail Investment*, paper prepared for The Railway Forum, United Kingdom, Table 2.2.
4	United Kingdom value of freight travel time	36 to 48	1999 $US per shipment per hour	0.09 to 1.29	1999 $US per tonne per hour	de Jong G. 2000, *Value of Freight Travel Time Savings*, in Hensher D. and Button K.J. (eds), *Handbook of Transport Modelling*, Elsevier Science (Pergamon), Oxford, Tables 1 & 2, pp.558–559.
5	Australian congestion cost	0.03	cents/ntk	–		BTE 1999, *Competitive Neutrality between Road and Rail*, Working Paper 40, Table III.1. (data from Columns G and H, sources and derivations of data detailed in notes to Table).
6	Congestion costs used to calculate benefit of diverting tonnage from road to rail (metropolitan)	0.09	cents/ntk	–		ARTC 2001b, *Interstate Rail Network Audit—Final Report*, report prepared by Booz Allen Hamilton, Melbourne, Appendix A, Table 5.3, page 24.
7	Estimated land freight congestion costs in Queensland (metropolitan)	0.09	cents/ntk	–		Laird P. 2001, *Land Freight External Costs in Queensland, in Queensland Transport 2003*, Rail Studies, Queensland Government, Brisbane, Table 5.1, page 20.

Table A2.2: Range of values for accidents

SERIAL	DESCRIPTION	ROAD		RAIL		SOURCE
		VALUE	MEASURE	VALUE	MEASURE	
[a]	[b]	[c]	[d]	[e]	[f]	[g]
1	European accident costs (freight transport)	21.0000	ECU/'000 t km	0.7500	ECU/'000 t km	ACIL 2001, *Rail in Sustainable Transport*, report to the Rail Group of the Standing Committee on Transport, May 2001, Table 27, citing European Conference of Ministers of Transport (ECMT) 1998, *Efficient Transport for Europe: Policies for Internalisation of External Costs*, OECD Publications Service, Paris.
2	European accident costs (passenger transport)	41.0000	ECU/'000 pax km (car+bus)	27.0000	ECU/'000 pax km (rail pax)	As above.
3	Australian accident cost	0.3200	cents/ntk	0.0300	cents/ntk	BTE 1999, *Competitive Neutrality between Road and Rail*, Working Paper 40, Table III.1. (data from Columns G and H, sources and derivations of data detailed in notes to table).
4	Assumed accident cost for trucks operating in the Darwin to Alice Springs corridor	0.1830	cents/ntk	-		Northern Territory Department of Transport and Works 1999, *Economic Evaluation of Darwin–Alice Springs Railway*, report prepared by Booz, Allen Hamilton, Melbourne, Section 2.8.
5	Australian accident cost (extrapolated from BTRE publications—baseline data includes passenger and freight transport)	10.1080	$/'000nt	0.3663	$/'000nt	[i] BTRE 2000, *Road Crash Costs in Australia*, Report 102, BTRE, Canberra. [ii] BTRE 2003a, *Rail Accident Costs in Australia*, Report 108, BTRE, Canberra. [iii] BTRE 2003b, Australian Transport Stats, DoTARS, Canberra.
6	As above	0.1131	$/t km (million)	0.0014	$/t km (million)	As above
7	Accident costs used to calculate benefit of diverting tonnage from road to rail	0.3200	cents/ntk	0.0300	cents/ntk	ARTC 2001, *Interstate Rail Network Audit—Final Report*, report prepared by Booz Allen Hamilton, Melbourne, Appendix A, Table 5.3, page 24.
8	Estimated land freight accident costs in Queensland	0.0700	cents/ntk	0.0024	cents/ntk	Laird P. 2001, *Land Freight External Costs in Queensland, in Queensland Transport 2003*, Rail Studies, Queensland Government, Brisbane, Table 5.1, page 20.

Annex 3 Sample train path diagram

This train path diagram represents all train movements on the Gladstone–Caboolture line in Queensland. This line carries a mixture of train traffic, including local passenger services, general freight and coal trains. Train density increases on the line as it approaches its southern end (Caboolture).



Annex 4 Selection of current rail investment proposals

The following tables list the projects that have been proposed by the Australian Rail Track Corporation (ARTC), the Railway Technical Society of Australasia (RTSA), Queensland Rail (QR), the Australasian Railway Association (ARA) and Pacific National (PN).

Table A4.1: Potential AusLink Projects identified by the Australian Rail Track Corporation (ARTC)

POTENTIAL PROJECT AS IDENTIFIED BY ARTC	BENEFIT AS STATED BY ARTC	SOURCE
Advanced train management and communications system (entire DIRN)	Enhanced network capacity (35%). Soft (technology) investment to deliver same benefits at a lower cost than physical investment. Enhanced safety. Single communications framework.	David Marchant, CEO, ARTC Presentation at AusRAILPLUS2003 (November) Conference, Sydney
Double stacking Melbourne–Sydney	Improve volume and capacity and lower unit rail cost. Increases intermodal flexibility at and between major ports. Reduce road volumes.	
Double stacking Sydney–Parkes (to Perth)	Improved container handling from Sydney port to regions and west to Perth.	
North–south rail straightening program	Remove significant network deviations between Melbourne and Sydney and Newcastle and Brisbane. Improve rail transit times and average speeds. Modelled on 1970s/80s road straightening.	
Northern Sydney Freight Access Route (North Sydney–Gosford)	Overcome freight train curfew during peak hours. Enhanced northern Sydney freight capacity. Create more capacity for general freight and urban passengers. Improve service reliability.	
Sydney bypass via Hunter Valley (double stack)	Improve Melbourne–Brisbane volume capability. Double stacking Sydney–Perth. Worthwhile alternative if Sydney– Newcastle was capacity constrained to freight trains.	
Inland freight route Melbourne– Brisbane	Melbourne–Brisbane rail corridor with efficient transit and cycle times. Allows double stacking from central Qld to Perth.	
Inland freight route Melbourne–Toowoomba (Stage 1)	Melbourne–Brisbane rail corridor with efficient transit and cycle times. Allows double stacking from central Qld to Perth. Still requires road conversion at Toowoomba.	

Note: The report on the ARTC *Interstate Rail Network Audit* (2001) provides further project details. Also, the ARTC *Annual Report for 2002–03* (page 25) indicates that a $872m forward investment program has been developed with the support of the Australian Government, for improving the track to be subject to the lease with NSW.

Table A4.2: 'Fix The Rails' and 'AusLink Plus' projects identified by the Railway Technical Society of Australasia (RTSA).

POTENTIAL PROJECT AS IDENTIFIED BY RTSA	BENEFIT AS STATED BY RTSA	SOURCE
Realignment of all poorly aligned mainline track between Campbelltown and Junee, such as in the Cullerin Ranges (near Gunning)	–	Fix the Rails—NSW, RTSA brochure dated January 2000
Basic improvements between Maitland and the Border Loop	–	
Upgrading of signalling and safe working systems between: › Casino (NSW) and Greenbank (Qld) › Harden and Wallendbeen › Exeter and Medway	–	
Realignment of the excessively curved mainline track on the eastern slopes of Adelaide Hills	Slowest section on DIRN. Reduce transit time, improve fuel efficiency and allow heavier trains to operate	Fix the Rails— SA/WA, RTSA brochure dated September 1999
Gauge standardisation of remaining non-urban broad gauge lines in SA	–	
Replacement of light weight rail on the Nullarbor	Allow for faster and heavier trains	
Double stacking: › Adelaide–Melbourne › Melbourne–Albury	Extension of existing Perth–Adelaide capability	Fix the Rails—Vic, RTSA brochure dated June 1999
Gauge standardisation of remaining broad gauge lines on the north-east corridor to Albury	–	
Upgrading of railway signalling	In places 19th century technology remains in use	
Three major rail deviations on Sydney–Melbourne corridor, namely: › 'Hoare deviation' between Bowning and near Frampton [ARTC Track Audit - S2 project] › 'Wentworth route' between Menangle and Mittagong › Centennial deviation between Goulburn and Yass.	Numerous reports have detailed 'steam age' alignment that needs straightening if rail freight is to be efficient and competitive with line haul road freight. The worst 20 per cent of the Sydney–Melbourne track needs to be rebuilt to modern engineering standards.	RTSA response dated 14 January 2004 to AusLink Green Paper
Bypass Bethungra Spiral	To also benefit Melbourne–Parkes trains	

Table A4.3: Future infrastructure improvements identified by Queensland Rail (QR)

POTENTIAL PROJECT AS IDENTIFIED BY QR	BENEFIT AS STATED BY QR	SOURCE
Conversion of nominated passing loops on Mt Isa system from right hand running to mainline and loop configuration	–	QR Mt Isa System Information Pack Issue 1, August 2001
Installation of Dragging Equipment Detection on Mt Isa system	–	
Installation of Wheel Impact Load Detector on Isa system	–	
Deviations	–	QR North Coast Line (North) System Information Pack Issue 1, August 2003
Trial of Dragging Equipment Detection	–	
Communications upgrade	–	
Resignalling Stuart	Funded by Queensland Government as a project through Transport Services Contract (Rail Infrastructure)	
UTC/DTC	–	
Timber bridge elimination Mackay–Townsville	Possibly funded by Queensland Government as an asset strategy through Transport Services Contract (Rail Infrastructure)	
Deviations Nambour–Rockhampton	–	QR North Coast Line (South) System Information Pack Nambour–Gladstone, Rocklands–Rockhampton Issue 1, July 2003
Rerailing Caboolture–Gladstone	–	
Security fencing of corridor–Gladstone	–	
Communications upgrade	–	
Installation of Remote Control Signalling between Acacia Ridge and the Border	–	QR Standard Gauge / Dual Gauge System Information Pack Issue 1, April 2002

Table A4.4: Australasian Railway Association (ARA) investment recommendations

PROPOSED INVESTMENT IDENTIFIED BY ARA	BENEFIT AS STATED BY ARA	SOURCE
Immediate rail 'Hot Spot' investment program be implemented to upgrade the east coast north–south rail link based on the ARTC Track Audit		ARA response dated January 2003 to AusLink Green Paper
Access to ports and terminals through urban areas	Most problematic issue facing rail freight	http://www.ara.net.au/policy/directions.php
Standardised and fully inter-operable digital communications technologies	Improved safety, increased track utilisation, reduced congestion, options for better rollingstock maintenance warning systems	

Table A4.5: Detailed rail infrastructure plan developed by Pacific National (PN)

PN PROPOSAL	RATIONALE AS STATED BY PN	SOURCE
PN has developed a detailed rail infrastructure plan and looks forward to proposing projects for inclusion in AusLink at the appropriate time		PN response dated February 2003 to AusLink Green Paper
Sydney dedicated freight line	Pressure on infrastructure created by public transport demand have driven need for dedicated freight line	
Funding should be identified and provided on a corridor basis as opposed to a project by project basis	Valuable flexibility is achieved by giving greater weight to the end user's requirements rather than to specific projects. Pacific National commends the ARTC rail track audit for its focus on outcomes rather than inputs. Within a corridor, the performance of individual projects is affected by the performance of adjacent infrastructure and other assets along the route. What are needed are effective end-to-end solutions. These can only be found by planning on an entire corridor basis. As we will explain below, funding decisions will also be more effective if taken on this basis. Part of funding should be allocated to special purpose programs or strategies, such as the Black Spot Road Programme, which focus on regional or local transport issues.	

APPENDIX

A Abbreviations

AADT	Average annual daily traffic
AAR	Association of American Railroads
ABS	Australian Bureau of Statistics
AGO	Australian Greenhouse Office
ANAO	Australian National Audit Office
ARA	Australasian Railway Association
ARTC	Australian Rail Track Corporation
ATSB	Australian Transport Safety Bureau
ATC	Australian Transport Council
BAH	Booz Allen Hamilton
BC	Base Case
BCA	Benefit–cost analysis
BCR	Benefit–cost ratio
BIE	Bureau of Industry Economics
BTCE	Bureau of Transport and Communications Economics
BTE	Bureau of Transport Economics
BTR	Bureau of Tourism Research
BTRE	Bureau of Transport and Regional Economics
CAPM	Capital Asset Pricing Model
COAG	Council of Australian Governments
CRC	Cooperative Research Centre
CTC	Centralised Train Control
DIRN	Defined Interstate Rail Network
DORC	Deprived Optimised Replacement Cost
ES	Electric Staff
FTTS	Freight travel time savings

FYRR	First-year rate of return
GDP	Gross Domestic Product
GGAP	Greenhouse Gas Abatement Program
GTK	Gross tonne kilometre
IBCR	Incremental benefit–cost ratio
IRR	Internal rate of return
IT	Information technology
LCV	Light commercial vehicle
MACO	Management administrative and corporate overhead (costs)
NIMBY	Not in my backyard
NBIR	Net benefit investment ratio
NDI	Noise depreciation index
NPV	Net present value
NSW GIAC	New South Wales Grain Infrastructure Advisory Committee
NT DoT&W	Northern Territory Department of Transport and Water
NTK	Net tonne-kilometre
OD	Origin–destination
PAT	Preferred arrival time
PCE	Post-completion evaluation
PCU	Passenger car unit
PSC	Public sector comparator
PTTS	Passenger travel time savings
QCA	Queensland Competition Authority
QR	Queensland Rail
RIAM	Road Infrastructure Assessment Model
RSS	Residual sum of squares
RTA	Roads and Traffic Authority
SCOT	Standing Committee on Transport
SDE	Schedule delay early
SDL	Schedule delay late
SMT	Strategic Merit Test
SMVU	Survey of Motor Vehicle Use
TTS	Travel time savings
UK	United Kingdom
VTTS	Value of travel time savings
WACC	Weighted average cost of capital
WBP	World best practice
WTP	Willingness-to-pay

APPENDIX

B Glossary

Adjusted benefit–cost analysis	The adjusted BCA technique is a hybrid of the BCA and multi-criteria analysis techniques. It produces an alternative ranking of initiatives with objectives given different weights from the weights implicit in standard BCAs. Adjusted BCA retains the dollar measuring rod of standard BCA. The relative significance of particular benefits and costs is altered by adjusting them using subjectively determined weights.
Appraisal	Process of determining impacts and overall merit of a proposed initiative, including the presentation of relevant information for consideration by the decision-maker. Undertaken in Phase 5.
Area	Defined geographic space and all the transport routes within it. Incorporates the pathways that enable the movement of people and freight between the diverse and multi-directional set of origins and destinations within the area. Most relevant in urban settings.
Assessment	Generic term referring to quantitative and qualitative analysis of data to produce information to aid decision-making.
Average Annual Daily Traffic (AADT)	Total number of vehicles passing a point on a road in a year divided by 365 (or 366 for a leap year).
Base Case	A BCA is always a comparison between two alternative states of the world. The Base Case is the state of the world in the absence of the initiative being implemented. The Project Case is the state of the world with the initiative being implemented.
Benefit–cost analysis (BCA)	Analysis of the benefits and costs to society of a proposed initiative. Aims to value benefits and costs in monetary terms and provide a summary indication of the net benefit. Mainly undertaken in Phase 5 of the Framework.
Benefit–cost ratio (BCR)	Ratio of the present value of economic benefits to the present value of economic costs of a proposed initiative. Indicator of the economic merit of a proposed initiative presented at the completion of benefit–cost analysis. Commonly used to aid comparison of initiatives competing for limited funds.
Business Case	A document that brings together the results of all the assessments and analyses undertaken of a proposal for an initiative. It is the formal means of presenting information about a proposal to aid decision-making. It includes all information needed to support a decision to proceed with the proposal and to secure necessary approvals from the relevant government agency.

Challenge	Reason for action that results from a gap between actual and desirable outcomes. In this document, used as a generic term that covers related terms such as problem, issue, deficiency, opportunity and need.
Congestion pricing	The policy of charging drivers a fee that varies inversely with the level of traffic on a congested roadway. The aim of congestion pricing is to ensure that, when roadway space is a scarce resource, it is allocated efficiently between competing users. It is also known as variable pricing and, in the US, as value pricing.
Consumers' surplus	The surplus of consumers' willingness-to-pay over and above what they actually pay for a given quantity of a good or service. It is measured as the willingness-to-pay area under the demand curve above the price paid.
Corridor	The parallel or competing modal routes between two locations (e.g. road and rail routes between two capital cities). A corridor is multi-modal where more than one mode operates, and uni-modal where only a single mode operates (e.g. in many rural areas).
Default values	Standard unit costs that can be applied across the board to obtain an estimate of externality costs, if more appropriate values for a particular situation are not available. Employing a default value is usually preferable to the alternative of giving it a zero value.
Depreciation	The amount that an asset reduces in value over one year, due to wear and tear or environmental factors. Adapted from the *Macquarie Dictionary.*
Discount rate	The interest rate at which future values are discounted to the present and vice versa.
Discounting	The process of converting money values that occur in different years to a common year.
Diverted traffic	Freight or passengers that switch from one mode or route to another as the result of an initiative.
Dose-response relationship	The relationship between dose and response (effect) wherein all possible degrees of response between minimum detectable response and a maximum response are producible by varying the dose, e.g. how exposure to pollution gives rise to physical effects. Dose-response assumes that a response to a toxic agent can be measured, and the effects increase in some kind of a relationship with the dose.
Economic efficiency	Any change for which the gainers could compensate the losers out of their gains and still have some gain left over is said to be an improvement in economic efficiency. Maximum economic efficiency is said to be obtained when no further changes of this type are possible.
Elasticity	A mathematical measure used in economics to describe the strength of a causal relationship between two variables. An elasticity value can be interpreted as the percentage change in the dependent variable in response to a one per cent change in the independent variable.
Environmental Impact Statement	A detailed written statement analysing the environmental impacts of a proposed action, adverse effects of the initiative that cannot be avoided, alternative courses of action, short-term uses of the environment versus the maintenance and enhancement of long-term productivity, and any irreversible and irretrievable commitment of resources.
Evaluation	Specific process of reviewing the outcomes and performance of an initiative after it has been implemented. Undertaken in Phase 8.

Externality	An effect that one party has on another that is not transmitted through market transactions. An example is noise pollution from vehicles: those operating the vehicles disturb other parties such as nearby residents, but a market transaction between these parties is absent.
Existing traffic	Traffic that uses the infrastructure in question in both the Base and Project Cases (in contrast with diverted and generated traffic). The quantity is existing traffic is, by definition, the same in the Base and Project Cases.
Financial analysis	The evaluation of the benefits and costs, measured in cash-flow terms, to a single entity.
First-year rate of return (FYRR)	Benefits minus operating costs in the first full year of operation of an initiative, divided by the present value of the investment costs, expressed as a percentage. The first-year rate of return is used to determine the optimum timing of initiatives.
Generalised cost	The sum of money price and user cost, with any additional costs to complete the door-to-door journey valued at money prices. Synonymous with private generalised cost.
Generated traffic	Freight or passengers that are induced by an initiative; that is, they would not exist but for the initiative.
Incremental BCR (IBCR)	Ratio of the present value of increase in benefit to the increase in investment cost that results from switching from one option to the adjacent, more expensive option. The incremental BCA is used to choose between different options for a particular initiative, having different levels of investment cost.
Infrastructure	Civil engineering structures that have been built to facilitate the movement of people and/or goods for various social and business reasons.
Infrastructure operating costs	The costs of providing the infrastructure after the initiative has commenced operation, e.g. maintenance, administration, operating costs of a facility.
Initiative	Any action to address a transport challenge. It could consist of an infrastructure or non-infrastructure intervention. The term ‘project’ is often used for such actions but it is limited by a perceived association with infrastructure.
Investment costs	The costs of providing the infrastructure before the initiative has commenced operation, e.g. planning and design, site surveying, site preparation, investigation, data collection and analysis, legal costs, administrative costs, land acquisition, construction costs, consequential works, construction externalities.
Intelligent transport systems	Integrated application of modern computer and communications technologies to transport systems to improve transport safety, use of infrastructure, transport operations and the environment.
Internal rate of return (IRR)	The discount rate that makes the net present value equal to zero.
Jurisdiction	Australian Government, state or territory government, local government, or a combination.
Life of initiative	The number of years over which the benefits and costs of an initiative are assessed following completion of construction.
Link	Homogeneous segment of a route. Includes an inter-modal facility.
Maintenance	Incremental work to restore infrastructure to an earlier condition or to slow the rate of deterioration. Distinct from construction and upgrading.

Money price	The money price paid to use a service.
Multi–criteria analysis	A loose collection of tools to assist decision-making where the aim is to promote a number of different objectives or criteria.
Multi-modal	Has several meanings. Can refer to passenger or freight movements that use more than one transport mode (e.g. road and rail). A 'multi-modal focus' means an approach to addressing transport challenges that considers the full range of potential solutions across all modes.
Mutually exclusive	In the BCA context, the term is used to refer to options where choice to adopt one option precludes adoption of all the other options.
Network	Collection of routes that provide inter-connected pathways between multiple locations for similar traffics. Can be multi-modal (typically comprising several uni-modal networks) or uni-modal.
Net present value (NPV)	The discounted present value of a stream of benefits and costs over time. The term 'net' signifies that costs have been included in the stream as negative values.
Nominal prices	A value or price at a given time. Nominal prices rise with inflation.
Objective	Statement of a desired outcome that has not yet been attained.
Opportunity cost	The value forgone by society from using a resource in its next best alternative use. Synonymous with resource cost.
Option	Alternative possible solution to a challenge.
Options analysis	Identification and assessment of alternative initiatives/solutions and variants of initiatives/solutions that promote the same set of objectives or address the same set of problems.
Parameters	Qualitative values applied consistently in appraisals. They are usually unit costs of impacts.
Perceived price	The subset of private generalised cost that is actually perceived by the user. For example, car drivers may perceive time but not vehicle operating costs.
Post-completion evaluation (PCE)	A review of a completed set of actions to determine whether the desired or forecast ends have been realised, and to explain the reasons for the outcomes. The aim is to discover lessons for the future. Undertaken in Phase 8.
Private cost	Cost incurred by an individual transport user or service provider. Private costs are valued at money prices, where applicable and may include user costs but exclude external costs imposed on others.
Program	Suite of appraised initiatives to be delivered within a specified time frame and sequence.
Real prices	Prices that have been adjusted to remove effects of inflation. They apply for a particular Base Year, e.g. 2004 dollars.
Reliability	The probability that a system will perform its intended function under stated conditions for a stated period of time. Adapted from ISO 15686-3. As a quality attribute of a transport service, on-time performance.
Resource cost	The value forgone by society from using a resource in its next best alternative use. Synonymous with opportunity cost.
Risk	A state in which the number of possible future events exceeds the number of events that will actually occur, and some measure of probability can be attached to them (Bannock et al 2003, p. 338).
Road-user costs	Costs of operating vehicles on roads, including time costs. Crash costs may or may not be included.

Route	Physical pathway connecting two locations for a particular mode. In land transport, consists of a continuous length of infrastructure (road, rail line). In shipping and aviation, delineated by operating or regulatory or administrative practices (shipping lane, air route).
Shadow price	A social cost used in a BCA that differs in magnitude from the corresponding private cost.
Strategic Merit Test (SMT)	Largely qualitative series of questions that provides a first-order determination of the 'strategic merit or fit' of an identified initiative. Identifies proposals that should proceed to the next stage of appraisal, proposals that require further scoping, and proposals that should be abandoned because they lack strategic fit. Also includes checks to ensure that the initiative has been properly formulated and is feasible.
Strategic planning	High-level planning involving fundamental direction-setting decisions. Narrows down the types of options that will be pursued. Involves consideration of present and future environments. Asks questions such as: 'Are we doing the right thing?' 'What are the most important issues to respond to?' and 'How should we respond?' Balances many competing considerations including value judgements, subjective assessments and political considerations. Involves iteration, stakeholder consultation and analysis.
Social generalised cost	The full cost to society valued at resource cost, including user costs and external costs. Any additional costs to complete the door-to-door trip are valued at resource cost.
Traffic	As used throughout this volume, transport users in general regardless of mode and unit of measurement, e.g. passengers, cars, trains.
Transport system	For a particular jurisdiction (or a multi-jurisdictional setting), comprises the following elements: › relevant transport networks—sets of routes that provide inter-connected pathways between multiple locations for similar traffics › transport user sub-system—people, goods and vehicles/wagons/etc. using the network › regulatory and management sub-system—regulatory regime and systems for managing the traffic that uses the network (including access arrangements, registration and licensing, traffic management centres and intelligent transport systems) › transport operating environment—e.g. land- use development patterns that generate traffic on the transport network › physical environment—e.g. geographic features, climate, air quality, and social environment e.g. accessibility, amenity, liveability, and › social environment—e.g. accessibility, amenity, liveability.
Vehicle operating cost	The costs of operating a vehicle, including fuel, oil, tyres and repair and maintenance costs.
User costs	Costs incurred by a transport user in addition to the money price—waiting time, time in transit, unreliability, damage to freight, passenger discomfort, additional costs to complete the door-to-door journey. Quality attributes such as time and reliability need to be expressed in dollar terms based on user valuations.
Willingness-to-pay (WTP)	The maximum amount a consumer is willing to pay for a given quantity of a particular good or service (rather than go without it). Total value that consumers place on a given quantity of a good or service. It is measured as the total area under the demand curve up to given quantity.

Index

S

T

U

V

W

Australian Transport Council